Interreg
Euro-MED

Co-funded by
the European Union

Germ of Life

Interreg Euro-MED Project GERM OF LIFE

"Digital Drought Risk Management enabling the drought mitigation and adaptation strategies for the restoration of the ecosystem equilibrium in Mediterranean European Countries".

Test Project (Thematic Project)
Mission: NATURAL HERITAGE

Duration: 33 months from 01/01/2024
Coordinator: UNIVERSITY OF PATRAS

Deliverable ID.:	**D1.4.1**	
Deliverable title:	**Drought Risk Monitoring and prediction service – Technical documentation**	
Planned delivery date:	31.12.2025	
Actual delivery date:	31.12.2025	
Deliverable leader:	PP6 (ATOS)	
Contributing partners:	PP3 (CMCC), PP1 (UPatras)	
Dissemination Level:	x	EX = External

This project has received funding from the European Interreg Euro-MED programme under Subsidy Contract (V1.mars.2022)- Project n° Euro-MED0200878

This deliverable reflects only the authors' view and the Commission is not responsible for any use that may be made of the information it contains.

Document information and history

Deliverable description (from AF)
The D.1.4.1. The deliverable of this task covers the pipeline for data processing and calculation of drought indicators (source code) to be implemented in the proposed platform (dashboard), along with the exploratory ML models to forecast the selected droughts indicators and soil moisture (source code). Note: this document is the technical documentation associated to the source code, that is available on the project git repository.

Version N.	Date	Author [Person and Organisation]	Reviewer [Person and Organisation]	Notes
V.01	22/12/2025	L. Duran	P. Saranti	

Table of contents

1 Executive Summary

The Drought Risk Monitoring and Prediction Service, developed under the Interreg Euro-MED Germ of Life (GoL) project, delivers a digital system to monitor and forecast drought risks in Mediterranean Europe. This service integrates advanced data processing, machine learning, and key drought indicators—SPEI, CWSI, and NDVI—to support climate adaptation and ecosystem resilience.

The project addresses the Mediterranean region's vulnerability to drought by standardizing drought indices and providing decision-support tools for stakeholders. It combines Earth observations, climate data, and predictive models to enable early warning and proactive risk management. The technical approach includes calculating the Standardized Precipitation Evapotranspiration Index (SPEI) using historical and real-time meteorological data, and forecasting vegetation health through Normalized Difference Vegetation Index (NDVI) derived from Sentinel-2 satellite imagery. The system also incorporates ground station data and ECMWF SEAS5 forecasts to generate actionable insights.

Data processing pipelines for SPEI and NDVI have been implemented, with SPEI calculated using the Hargreaves-Samani equation and NDVI extracted via Google Earth Engine. Forecasts are generated using ensemble simulations to account for uncertainty. Pilot tests in Spain, Greece, Portugal, and Italy demonstrate strong performance, particularly in Spain (Nash-Sutcliffe Efficiency of 0.875) and Portugal (0.905), while further adjustments are needed for the Greek site.

The project's architecture relies on a centralized database (Dotsoft) and a virtual machine (Atos) hosting algorithms and models. The deliverable provides technical documentation for these processes, ensuring transparency and reproducibility. Moving forward, the focus will be on generalizing models across pilot sites, engaging stakeholders, and supporting policy integration for drought resilience in Mediterranean Europe.

1.1 Role of deliverable

The D.1. 4.1 deliverable supports the key output of Activity WP1.4 : pipeline for data processing and calculation of drought indicators (source code) to be implemented in the proposed platform (dashboard), along with the exploratory

ML models to forecast the selected droughts indicators and soil moisture (source code).

This document is the technical documentation supporting the code for the algorithms and models, explaining the choices made and providing equations and model parametrisation.

1.2 Relationship to other GERM OF LIFE deliverables

The deliverable relies on deliverable D.1.1.2. which describes the algorithms used to derive or compute the selected drought indicators SPEI, CWSI and NDVI at the Pilot Test Areas (PTAs). These indicators were chosen on the basis of the preliminary study reported in the deliverable D.1.1.1 and further considerations on the availability of remotely sensed images and derived dataset as well as the usefulness and importance of them in the designing of VAT at the PTAs planned in the Activity 1.3 described in deliverable D.1.3.1 at M12.

1.3 Structure of the document

This deliverable is organised in the following sections:

1. Project description (purpose, architecture, data

2 Project description

The Germ of Life project is an Interreg Euro-MED initiative that develops a digital, preventive drought risk management system for Mediterranean Europe, combining advanced data, models and decision-support tools to support climate change adaptation and ecosystem resilience. It targets drought-prone territories in six Euro-Mediterranean countries and aims to operationalize proactive mitigation and adaptation strategies, with a strong focus on nature-based solutions and technological innovation.

2.1 Project purpose and scope

Germ of Life addresses the Mediterranean region's high exposure to multifaceted droughts and related impacts on agriculture, ecosystems and societies, by promoting standardized drought indices and integrated vulnerability assessments as a basis for risk-informed decision making. The project supports climate change adaptation by monitoring ecosystem vulnerability and resilience, explicitly considering ecosystem-based and nature-based approaches in water- and land management strategies.

Technical approach

The project designs and tests a preventive drought risk management approach built on newly available Earth observations, climate and environmental data, coupled with drought prediction models for territorial monitoring and early warning of emerging drought conditions. It develops a digital decision-support environment, including tools for vulnerability assessment and an innovation procurement platform for technological and nature-based solutions, to guide stakeholders in selecting and implementing effective mitigation and adaptation options.

Governance, partnership and implementation

Germ of Life is co-financed by the Interreg Euro-MED Programme, with a total budget of about 3 million euros, and is led by the University of Patras in collaboration with nine other partners from Greece, Italy, Spain, France, Portugal and Cyprus. The consortium brings together regional authorities, research centres and private companies to co-design, pilot and evaluate solutions across representative Mediterranean territories, fostering cross-border knowledge exchange and harmonized drought risk policies.

Expected outcomes

By integrating big data analytics, machine learning and multi-source indicators of meteorological, hydrological and agro-ecological drought, Germ of Life aims to improve the precision and automation of soil–vegetation–water risk assessment. The project is expected to deliver operational tools, pilot-tested in real-world contexts, that enhance early warning, support ecosystem restoration and strengthen the capacity of policy-makers and practitioners to design robust drought adaptation and mitigation strategies in Mediterranean Europe.

General architecture

The Germ of Life project is partitioned between different servors: - The first one is hosting the database of the project (managed by Dotsoft) - The second one is hosting the algorithms and machine learning models to compute and forecast the droughts indicators (managed by Atos). Infotrend have created the servors and are administrating them.

Data sources

The different sources of data are the following: - ERA5Land dataset on Copernicus Datastore. The data has been dowloaded since 1950 up to now, and is downloaded directly on Germ of life servor (Atos) and it is used to: 1. make past SPEI time series available 1. allow the statistical fit to store SPEI parameters to be used for current and forecasted values of SPEI 1. have the relevant variables to train NDVI model. - Near real time data is coming directly from ground stations (pilot sites of Germ of Life projects) from on site instruments. These data for the 4 pilot sites of the project are stored within the Germ of Life database of the project (Dotsoft) and then imported to the algorithms and models (Atos) through an API (API documentation). - Forecast of meteorological data from ECMWF : SEAS5 data set. This data is under licence. It is delivered through the national provider of this forecast, the National Hellenic Weather service. The data is sent directly to Germ of Life servor (Atos). - Historical NDVI has been retrieved from Sentinel 2 data from Google Earth Engine.

Standardized Precipitation Evapotranspiration Index

SPEI definition

The Standardized Precipitation Evapotranspiration Index (SPEI) is an important drought indicator that integrates both precipitation and temperature data to assess drought conditions, particularly in regions where both factors significantly contribute to drought severity, such as arid or semi-arid areas. It is designed to measure the balance between precipitation and potential evapotranspiration (PET), making it especially useful for evaluating agricultural drought and its impacts on ecosystems and socio-economic conditions. It is a standardized variable, meaning that SPEI values can be compared across different times and locations, with an average value of 0 and a standard deviation of 1.

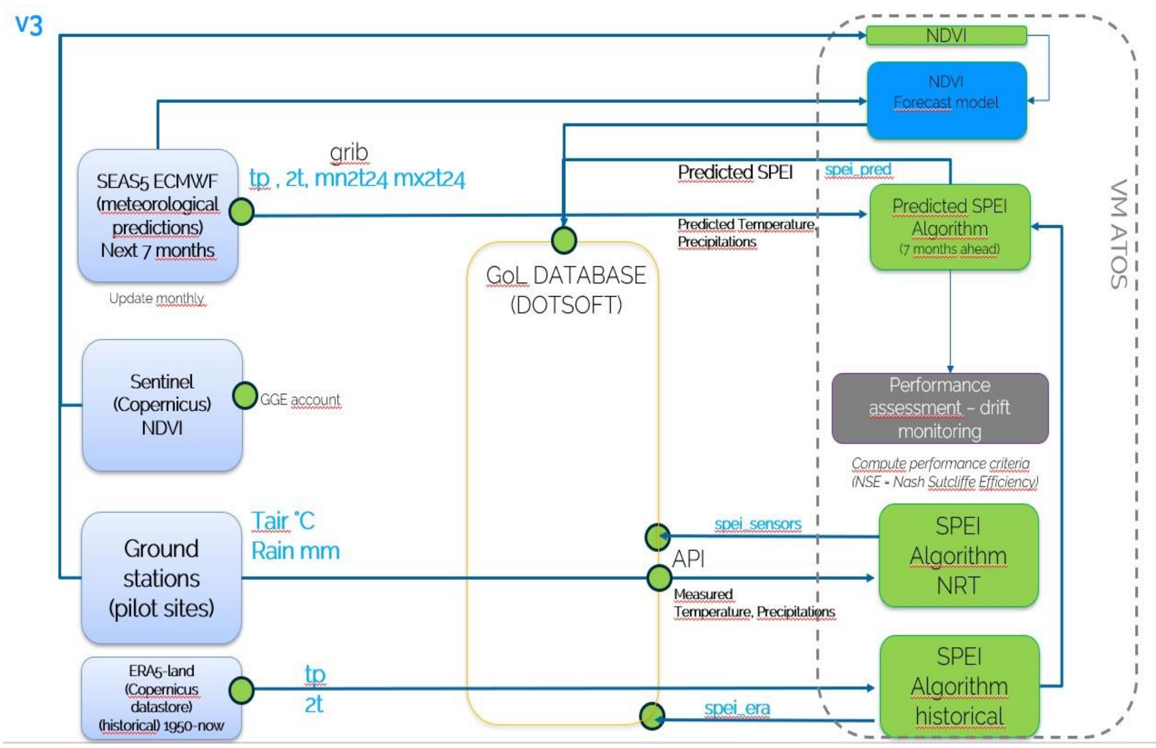


Figure 1: alt text

SPEI algorithm

1. Evapotranspiration

The computation of SPEI requires the following meteorological time series, typically at a monthly temporal resolution: * Precipitation PP (in mm) * Potential evapotranspiration PET (in mm), computed using mean, maximum and minimum daily temperature. This data is imported from Copernicus Data Store (ERA5 Land).

The Hargreaves-Samani formula has a version that has been calibrated to iprove accuracy under different climatic conditions by adjusting the empirical coefficient. This equation has been selected in this project (to account for specific mediterranean conditions).It is given by:

$$ET_0 = k_{rs}(T_{mean} + 17.8)(T_{max} - T_{min})^{0.5} R_a$$

Where:

- T_{mean}: mean daily temperature (°C)
- T_{max}: maximum daily temperature (°C)
- T_{min}: minimum daily temperature (°C)
- R_a: extraterrestrial radiation $MJ\ m^{-2}$
- k_{rs}: empirical coefficient (typically 0.0023) Monthly PET:

$$PET_{monthly} = \sum_{\text{days in month}} ET_0$$

References:

- Hargreaves, G. H., & Samani, Z. A. (1985). *Reference crop evapotranspiration from temperature*. Applied Engineering in Agriculture, 1(2), 96–99.
- Sperna Weiland, F. C., Tisseuil, C., Dürr, H. H., Vrac, M., & van Beek, L. P. H. (2012). *Selecting the optimal method to calculate daily global reference potential evaporation from CFSR reanalysis data for application in a hydrological model study*. Hydrology and Earth System Sciences, 16(3), 983–1000. https://doi.org/10.5194/hess-16-983-2012

2. Climatic Water Balance

The next step to compute SPEI is to compute the water balance:

$D_i = P_i - \mathrm{PET}_i$

3. Accumulation at Time Scale *k*

To compute SPEI at a specific time scale kk (e.g., 1, 3, 6, or 12 months), the water balance series DD is aggregated over kk consecutive time steps. For each time step tt, the accumulated water balance Dk(t) is given by:

$D_k(t) = \sum_{j=0}^{k-1} D(t-j)$

This results in a time series of kk-month accumulated water balance values, which forms the basis for the subsequent statistical analysis.

4. Log-Logistic Distribution CDF

A probability distribution is fitted to the accumulated water balance series Dk for each calendar month (to account for seasonal variability). The standard SPEI methodology employs a three-parameter log-logistic distribution with parameters:

- Scale parameter α
- Shape parameter β
- Location (origin) parameter γ

The distribution parameters are estimated using the method of Probability Weighted Moments (PWMs), which is robust for fitting the log-logistic distribution to hydroclimatic data. For each value xx in the accumulated series DkDk, the non-exceedance probability F(x) is computed from the fitted log-logistic distribution:

[$F(x) = \frac{1}{1+\left(\frac{x-\gamma}{\alpha}\right)^{-\beta}}, \quad x > \gamma$]

Exceedance probability is then given by: $P = 1 - F(x)$

5. SPEI Standardization

The exceedance probability PP is transformed into the standardized SPEI value using the inverse of the standard normal distribution. A common rational

approximation is: $z = \sqrt{-2lnP}$

Constants: c_0 = 2.515517, c_1 = 0.802853, c_2 = 0.010328 d_1 = 1.432788, d_2 = 0.189269, d_3 = 0.001308

If $p \leq 0.5$: [SPEI = $-\left(z + \frac{c_0+c_1 z+c_2 z^2}{1+d_1 z+d_2 z^2+d_3 z^3}\right)$]

If $p > 0.5$: Use $1 - p$ and reverse sign.

This transformation yields SPEI values with a mean close to 0 and a standard deviation close to 1, allowing for consistent interpretation across locations and time scales. Refrence for SPEI process : reference spei website ##6. Interpretation of SPEI Values

The resulting SPEI time series can be interpreted as follows:

- SPEI > 0: wet conditions relative to the long-term average
- SPEI < 0: dry conditions relative to the long-term average Common drought severity classes are defined as:

- SPEI > 1.0: very wet
- 0.5 < SPEI ≤ 1.0: moderately wet
- −0.5 ≤ SPEI ≤ 0.5: near normal
- −1.0 ≤ SPEI < −0.5: moderately dry
- −1.5 ≤ SPEI < −1.0: severely dry
- SPEI < -1.5 : extremely dry

Data flow

Calculation: SPEI is calculated using the following steps:

* Data Import: Relevant data, such as temperature and precipitation, is importedfrom sources like the Copernicus Data Store (ERA5 Land).

* Extraction: Data is extracted from formats like NetCDF and concatenated toa daily time step.
* Potential Evapotranspiration (PET): PET is calculated using the Hargreaves-Samani equation modified.

Historical and near real time flow:

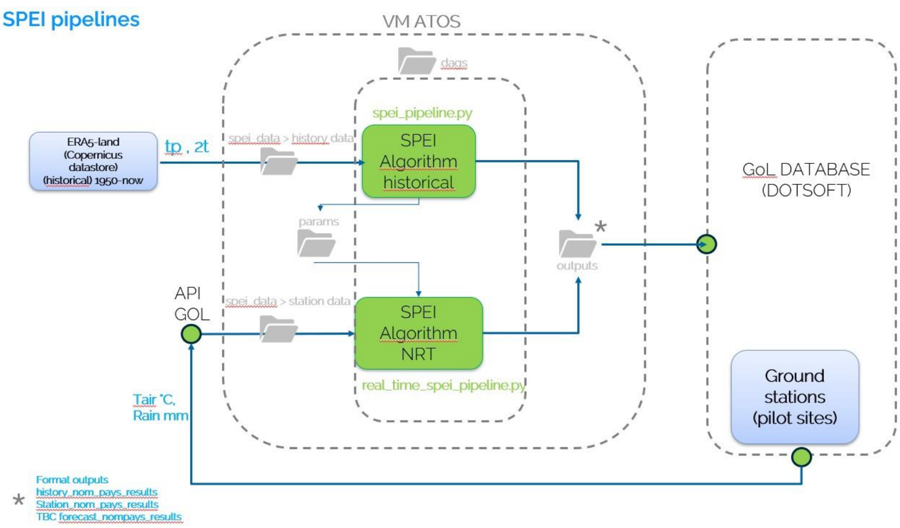


Figure 2: Historical SPEI architecture

Forecasting flow:

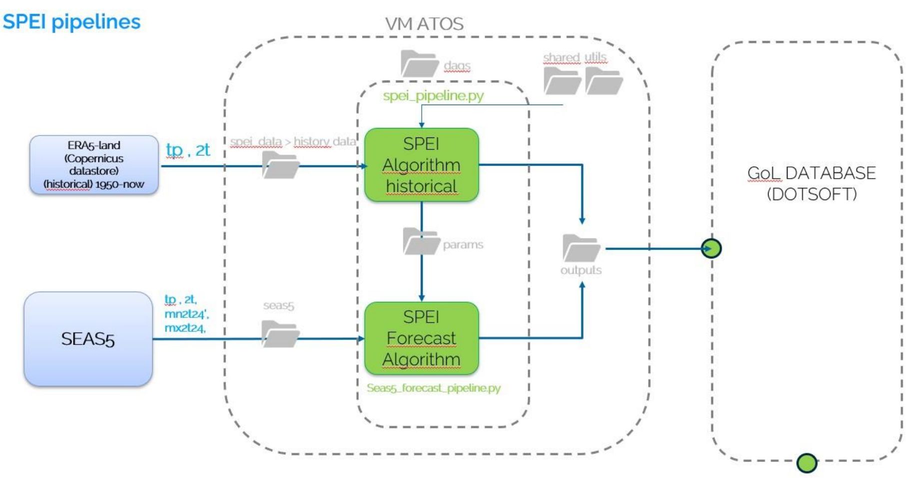


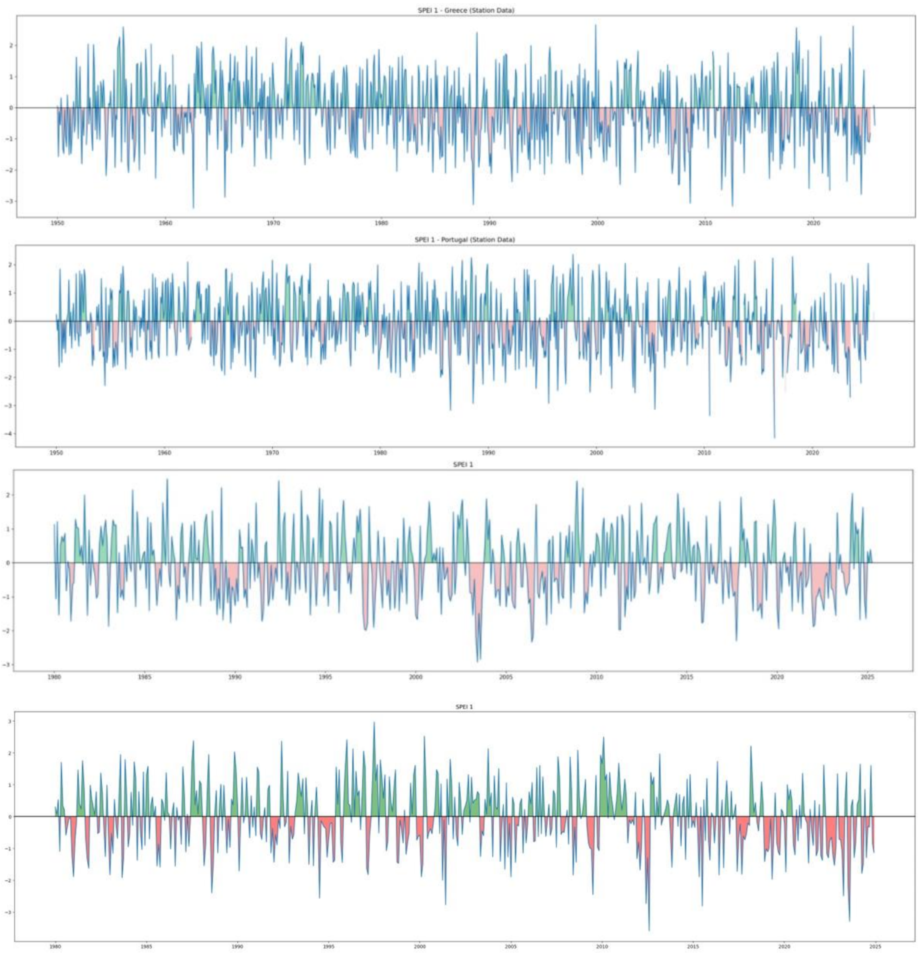
Figure 3: Forecasting SPEI architecture

Parametrization within GoL project and results

Historical SPEI calculation

- the fitting is done on time period 1960-1991 to match reference work in climatology science and fater considering the effect of a more recent fitting period (that would cause an undersestimating of the severity of droughts compared to past events).
- the fitted parameters are computed within the historical spei pipeline and stored in the params folder.

Resulting historical SPEI time series for the four pilots sites:



“Near real time” monthly calculation

The parameters stored on the historical fit are used to compute the monthly SPEI the first of each month for the previous month. The pipeline recomputes the latest months from the station data and adding the latest month SPEI. Only the latest month is then added to the Germ of Life database (Dotsoft).

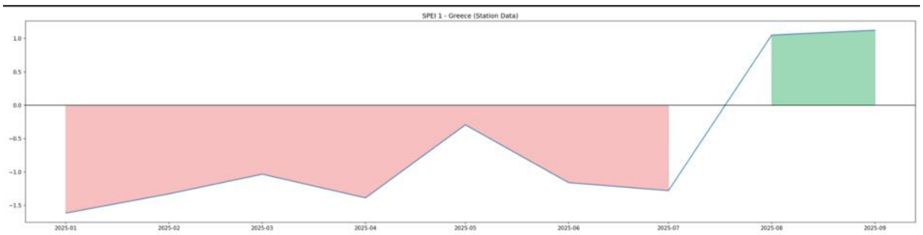


Figure 4: SPEI computed on Greek pilot with automated pipeline

Forecast

Similarly, the parameters of the statistical fit are used to generate the SPEI forecast. The inputs varaibles are from the SEAS5 dataset:

* daily average of temperature at 2 m 2t
* maximum daily temperature at two meters mx2t24
* minimum daily temperature at two meters mn2t24
* total daily precipitations tp

The SEAS5 dataset is available for the 7 months ahead, from the individual forecast runs V-v-b: Single level 24-hourly dataset(SEAS5 V-v-b) and V-v-a: Single level - 6-hourly (SEAS5 V-v-a).

The forecast SPEI is then forecasted for each of the 51 members of the individual simulations. The average of all resulting SPEI is then stored in the outputs folder, ready to be written in Germ of life database for visualisation. These individal members are also used to characterize the part of uncertainty due to the meteorological forecast.

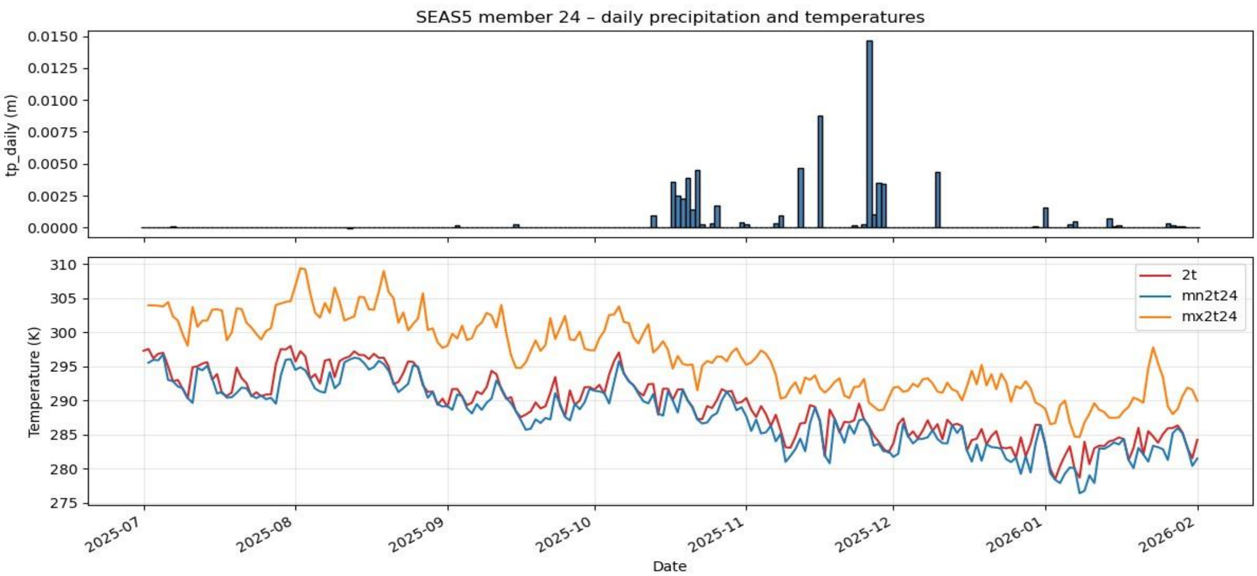


Figure 5: Example of SEAS5 simulation from the dataset V-v-b

Process of import : Sentinel-2 NDVI Pipeline (Google Earth Engine)

This module is designed to extract and process NDVI (Normalized Difference Vegetation Index) time series from Sentinel-2 Surface Reflectance (S2_SR_HARMONIZED) imagery using Google Earth Engine (GEE).

It supports two main use cases:

- Point-based NDVI time series extraction (from longitude / latitude coordinates)
- Spatial NDVI map download in GeoTIFF format around a given point.

The pipeline includes:

- Cloud and cirrus masking
- Cloud coverage filtering
- Vegetation pixel selection
- NDVI computation
- Temporal aggregation

In mask_s2_clouds Function : we use masks for: Clouds + Cirrus + Nonvegetated pixels

Then normalizes Sentinel-2 reflectance values.

Principle:

- Uses the QA60 band:
  - * Bit 10 : clouds
  - * Bit 11 : cirrus
- Uses the SCL (Scene Classification Layer) band:
  - – Class 4 = Vegetation
- Applied Processing :
  - – Creation of a clear-sky mask (no clouds, no cirrus)
  - – Vegetation pixel filtering
  - – Mask application
  - – Reflectance scaling (/ 10000)

- Computes NDVI for a Sentinel-2 image : NDVI Formula :

$$NDVI = \frac{B8 - B4}{B8 + B4}$$

* B8: Near Infrared (NIR)

* B4: Red

- Expected Input : DataFrame with columns: longitude | latitude
- Processing Steps
- Loop over each point
- Call extract_sentinel2
- Clean extracted data
- Extract acquisition date from Image ID
- NDVI averaging by:
  - Date
  - Longitude
  - Latitude

NDVI Extraction Using Neighboring Pixels (Robust Point Sampling)

- Purpose : Increase the probability of extracting a valid NDVI value when the central pixel is masked (clouds, cirrus, SCL filtering) by sampling its 8 neighboring pixels. This approach improves data availability and robustness for point-based NDVI time series, especially in cloudy regions or at field boundaries.
- Methodology

The process for each acquisition date is the following:

* Identify the central pixel corresponding to the target point

* Extract NDVI values for the 8 surrounding pixels (Moore neighborhood)

* Apply the same cloud and vegetation masks
* Retain valid (non-masked) NDVI values
* Aggregate values using a statistical rule median
* Neighborhood Definition: At 20 m resolution, the neighborhood corresponds to:Central pixel ±1 pixel offset in X and Y directions * Resulting grid:

  [p1 | p2 | p3]

  [p4 | pc | p5]

  [p6 | p7 | p8]

· Output : Final DataFrame: Dates | longitude | latitude | NDVI The NDVI data is stored in the ndvi_data folder.

NDVI Forecast model

Data sources

Historical data

- The NDVI data has been retrieved through Google Earth Engine as described above.
- The variables selected for training the model were selected within the set of variable available in the SEAS5 dataset from ECMWF based on relevance and initial testing:

The according variables were downloaded from Copernicus Datastore ERA5Land dataset for the historical time period corresponding to the availability of the NDVI time series (1980-2025 after initial 2016-2025).

variable	id	name	unit
swvl1	39	Volumetric soil water layer 1	$m^3\ m^{-3}$
swvl2	40	Volumetric soil water layer 2	$m^3\ m^{-3}$
swvl3	41	Volumetric soil water layer 3	$m^3\ m^{-3}$
swvl4	42	Volumetric soil water layer 4	$m^3\ m^{-3}$
stl1	139	Soil temperature level 1	K
10u	165	10 metre U wind component	$m\ s^{-1}$
10v	166	10 metre V wind component	$m\ s^{-1}$
2t	167	2 metre temperature	K
2d	168	2 metre dewpoint temperature	K
stl2	170	Soil temperature level 2	K
stl3	183	Soil temperature level 3	K
stl4	236	Soil temperature level 4	K
msshfl	172146	Mean surface sensible heat flux	$W\ m^{-2}$
mslhfl	172147	Mean surface latent heat flux	$W\ m^{-2}$
erate	172182	Evaporation	m of water s^{-1}
tp	228	Total precipitation	m

Input Forecast data

The variables selected for each site as relevant to forecast NDVI are retrieved from SEAS ECMWF dataset directly on the virtual machine of the project through the national provider (National Hellenic Weather Service). Th edata set used is individual forecast runs V-v-b: Single level 24-hourly dataset(SEAS5 V-v-b) and V-v-a: Single level - 6-hourly (SEAS5 V-v-a).

Data flow

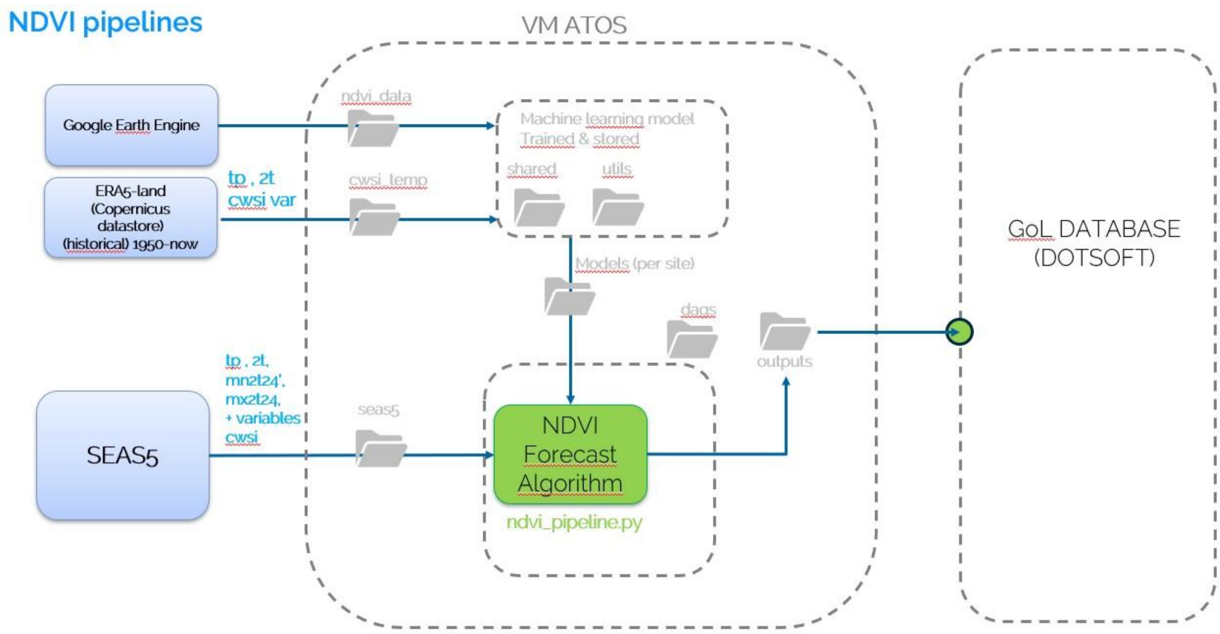


Figure 6: NDVI pipeline architecture

Training process

A baseline algorithm has been testing with a satisfying performance : polynomial regression with tailored features engineering. Artifical neural networks or deep learning did nota chieve significantly higher performances and are more energy consuming, so the choice was made to keep a frugal algorithm whenever possible. For this stage of the project, a model per pilot site has been developed. Generalisation models are to be tested in final stage of the project or following projects.

Pre processing

The NDVI was smoothed following a methodology inspired by Savistky-Golay filter method with an added correction for side effect. The adapted method is a convolution to reduce the time series. The middle time is t, 5 points before and five point after are included in a moving average with weights more important to the center point following f(x) = 1- cos(x), with a step of 1 for the moving average. Side effects of the extreme borders are dealt with padding (duplication added at the begnning and end of time series). The convolution is dealing with missing values (NaN are 1i (complex number), then imaginary numbers are divided by a mask + ratio).

functions used are : interpolation, convolve, kernelacos.

Feature enginnering (choice of inputs)

Convolution (lags) applied on the total precipitation are derived from the reference paper Zhou et al., 2025 (https://doi.org/10.1016/j.jag.2025.104498). It was also iteratively tested for the Spanish site with a similar order of magnitude of 72 days. functions used : convol function using numpy (convol).

For soil temperature variables, the tests included adding all variables separately and combined in different ways (averaged). Heat fluxes dipslayed episodes of droughts correlated to one of each others. Surface latent heat flux appeared as an important contributor, different combination of heat fluxes were tested, espacially ratio of them.

The seasonnality was extracted using ARIMA without trend.

After the initial exploration, the following variables and synthetic features were the ones enabling the higher performance for the Spanish site:

- Season = seasonnality
- t2m_max = maximal daily temperature at 2 m
- tp60 = total precipitation time series with a convolution by 60 days (determined by cross-correlation analysis and confirmed by literature review)

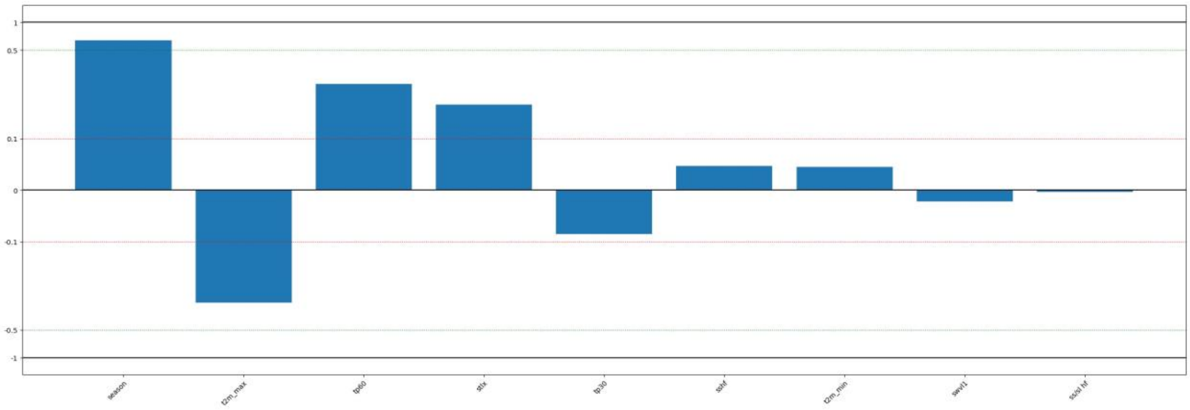


Figure 7: Contributing variables, Spanish pilot

- stlx = average of soil temperature at different depths (stl1 [0-7cm], stl2 [7-28cm], stl3[28-100cm] from ERA5land)
- sshf = surface sensible heat flux
- t2m_min = minimal daily temperature at 2m
- swvl1 = volumetric soil water layer 1
- ss/slhf = ratio between surface sensible heat flux and surface latent heat flux

relevant fucntions: > class PolyN for the convolution model Additionnal comments:
* Random = seed
* 100 values used for initial training

Criteria of performance is the Nash Sutcliffe efficiency (1 = perfect fit, à= model is as good as average, negative value= model is worse than the average of the time series) and the Root Mean Square Error (RMSE). The relevant script is stored in Utils folder. > Utils> class math for criteria of performance

Results

Spanish pilot

Performance : Nash Sutcliffe Efficiency = 0.875 ; RMSE = 0.104.

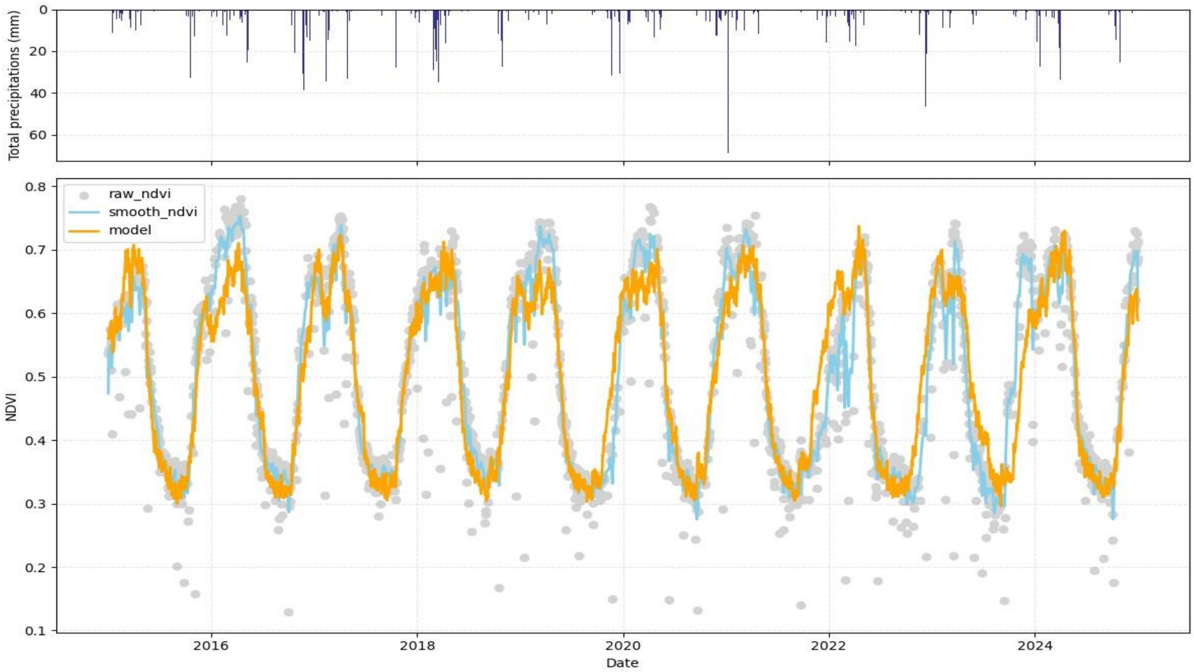
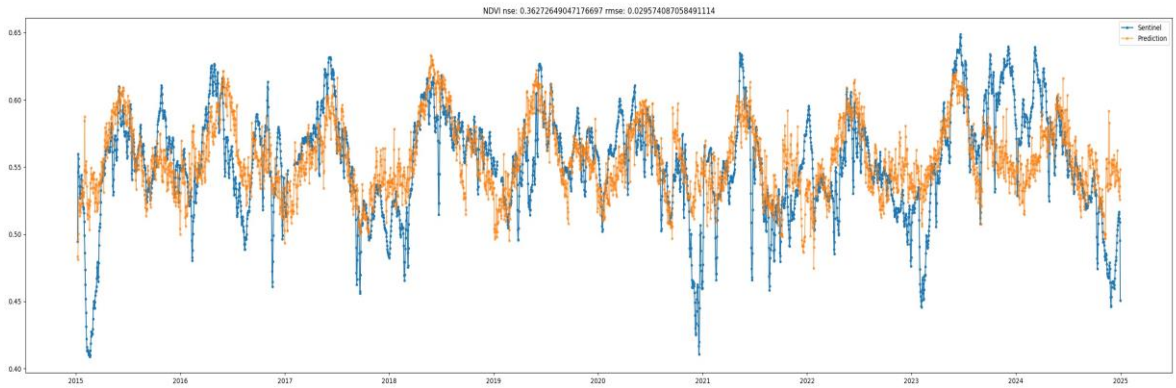


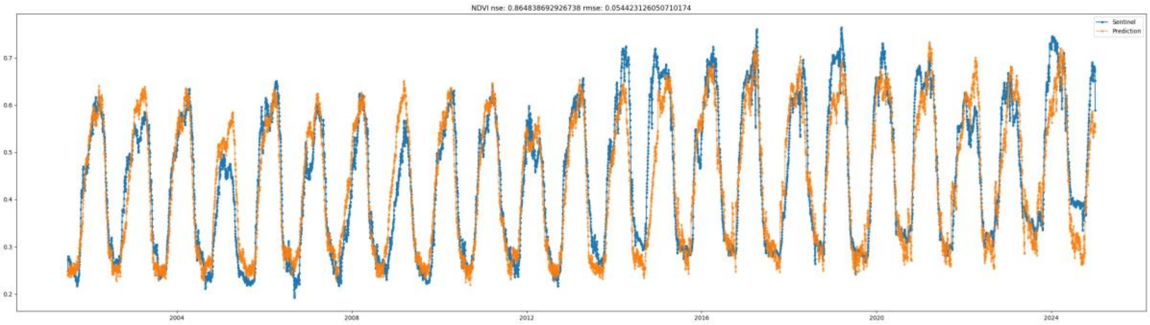
Figure 8: Spanish pilot NDVI model

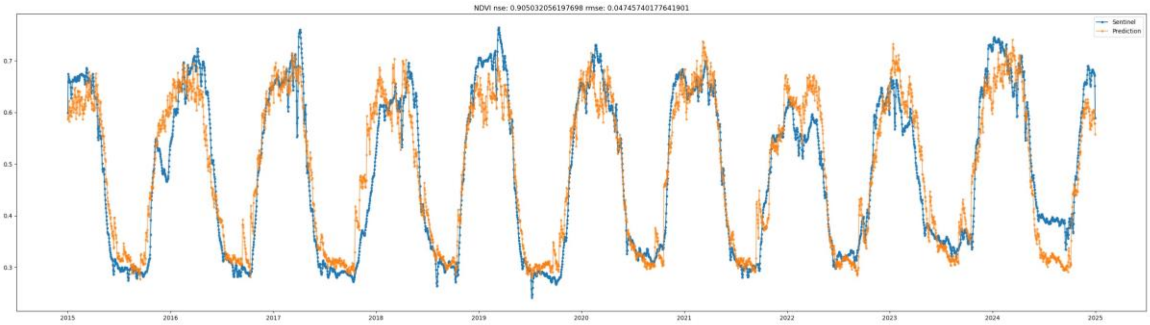
Greek pilot (initial results)



The behaviour of NDVI is less seasonally marked and the model requires further adjustments, possibly exploring additional variables as inputs or different feature engineering. Performance. NSE = 0.362, RMSE = 0.029.

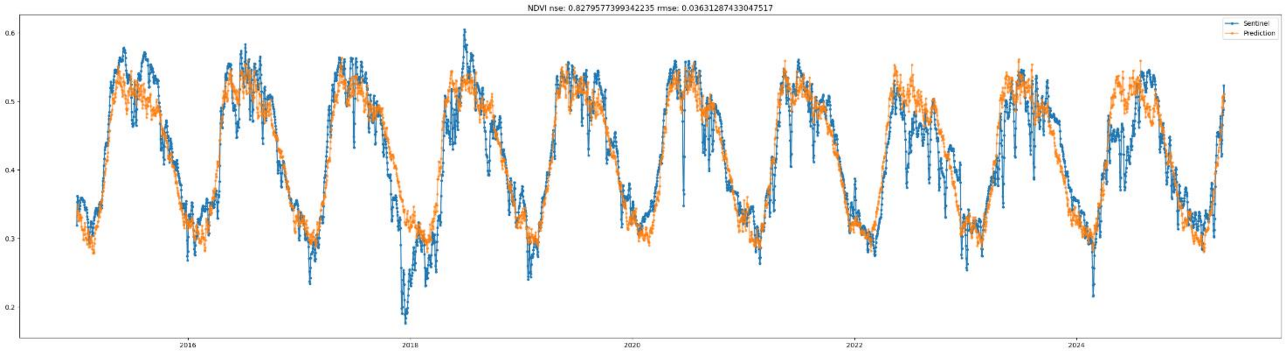
Portuguese pilot





Performance on 2015-2025: NSE = 0.905, RMSE= 0.047 Performance on period 2000-2025: NSE = 0.85, RMSE=0.054

Italian pilot



Performance on 2016-2025: NSE= 0.8279, RMSE = 0.0363.

Further analysis of error analysis and quantification of uncertainty is to be carried during Activity 2.3 of Work Package 2, and will determine possible fine tuning, or bias correction to be applied if required.