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.:	**D1.1.2**
Deliverable title:	**Description of algorithms**
Planned delivery date:	
Actual delivery date:	
Deliverable leader:	PP3 (CMCC)
Contributing partners:	PP6 (ATOS), PP1 (UPatras)
Dissemination Level:	IN = Internal

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

Document information and history

Deliverable description (from AF)
The D.1.1.2 'Description of algorithms' is a document containing the description of the algorithms used in order to compute the selected drought indicators SPEI, NDVI and CWSI.

Version N.	Date	Author [Person and Organisation]	Reviewer [Person and Organisation]	Notes
V.01	22/01/2025	M. Balzarolo	1st Internal review Internal review	
		V. Gentile		
		L. Duran		
		D. Odunayo		
V.02	19/03/2025	A. Argiriou	2nd Internal review	
		M. Samouris		

Table of Contents

1 Executive Summary

This document contains the description of the algorithms used to compute the selected drought indicators SPEI, CWSI and NDVI, whose choice is the outcome of a collaborative work of involved scientific partners and stakeholders, carried out in the frame of WP1 and for the Period 2 from M7 to M12. Specifically, all partners of Activity 1.4, and participants of workshop 3, in particular the pilot sites, agree on the need to identify SPEI indicator as the Minimum Viable Product (MVP), because is the most appropriate to quantify the contribution of both precipitation deficits and high temperatures to drought severity events in arid or semi-arid regions, such as the involved countries in the Mediterranean area.

1.1 Role of deliverable

The D.1.1.2 'Description of algorithms' is a document containing the description of the algorithms used in order to compute the selected drought indicators SPEI (Standardised Precipitation Evapotranspiration Index), CWSI (Crop Water Stress Index) and NDVI (Normalized Difference Vegetation Index). These indicators, chosen in concert with the scientific partners and involved stakeholders are useful for the development of the Vulnerability Assessment Tool (VAT), whose final aim is to support and orientate policy makers and stakeholders in the management and mitigation of drought events occurrence.

1.2 Relationship to other GERM OF LIFE deliverables

The deliverable D.1.1.2. 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 and which will be described in deliverable D.1.3.1 at M12.

1.3 Structure of the document

This deliverable is organised in the following sections:

2 Drought indicators

In the framework of GoL, the need of understanding the correlation and impact of drought on ecosystems with a focus on vegetation growth and productivity is mandatory to identify possible mitigation strategies useful for the designing of the Vulnerability Assessment Tool. Therefore SPEI, CWSI and NDVI have been selected for their significance and correlation to drought and especially to agro-meteorological drought and its impact on socio economic drought.

2.1 SPEI

The steps to calculate the SPEI are the following:

1. Data import from Copernicus Data Store

   A code has been created to import relevant variables. The selected dataset is ERA5 Land because of its higher spatial resolution.

Variable	Data set	Timestep	Comment
2m_temperature	derived-era5-land-daily-statistics	day	Must be imported with the three different daily statistics (mean, min and max)
total_precipitation	reanalysis-era5-land	hour	hourly data necessary because cumulative variables not available in era5 land daily

2. Extraction from NetCDF format and concatenation of variables at daily time step
3. Calculate PET using Hargreaves Samani as primary formulation:

$$ET_0 = K.R_a.(T_{mean} + 17.8).\sqrt{T_{max} - T_{min}}$$

Where ET_0 = Potential Evapotranspiration (mm/day)

K = Coefficient (0.0023 for daily calculations)

R_a = Extraterrestrial radiation (MJ/m²:day)

T_{mean} = Mean daily temperature (°C)

T_{max} = Mean daily temperature (°C)

T_{min} = Mean daily temperature (°C)

The extraterrestrial radiation, R_a, for each day of the year and for different latitudes is estimated from the solar constant, the solar declination and the time of the year by:

$$R_a = \frac{24(60)}{\pi} G_{SC} d_r [\omega_s \sin(\varphi)\sin(\delta) + \cos(\varphi)\cos(\delta)\sin(\omega_s)]$$

where:

R_a = extraterrestrial radiation [MJ/m^2/day],

G_{SC} = solar constant = 0.0820 MJ/m^2/min,

d_r = inverse relative distance Earth-Sun,

ω_s = sunset hour angle [rad],
φ = latitude [rad],
δ = solar declination [rad].

The latitude, φ, expressed in radians is positive for the northern hemisphere and negative for the southern hemisphere. The conversion from decimal degrees to radians is given by:

$$\varphi = \frac{\pi}{180}\left(\mathrm{Lat}_{\mathrm{deg}} + \frac{\mathrm{Lat}_{\mathrm{min}}}{60}\right)$$

The inverse relative distance Earth-Sun, d_r, and the solar declination δ, are given by:

$$d_r = 1 + 0.033 \cos\left(\frac{2\pi J}{365}\right)$$

$$\delta = 0.409 \sin\left(\frac{2\pi J}{365} - 1.39\right)$$

where

J = number of the day in the year between 1 (1 January) and 365 or 366 (31 December).

The sunset hour angle, w_S, is given by:

w_S = arccos (-tan(φ) tan(δ))

Note : for this first step Hargreaves Samani has been prioritised as the most data frugal and adapted to meteorological forecast to be used in production.

As a second step, Thornthwaite formulation can be added to be used itself or with Hargreaves Samani (using the median of both equation as recommended in Santini et al 2023[1]) .

Penman-Monteith equation can also be implemented as a second step but requires additional variables and hence a specific model to forecast.

Variable collected for PET alternative formulations and other relevant indicators (CWSI):

Variable	Data set	Time step	Comment
2m_temperature	derived-era5-land-daily-statistics	day	Must be imported with the three different daily statistics (mean,, min and max)

[1] Santini, M.; Noce, S.; Mancini, M.; Caporaso, L. A Global Multiscale SPEI Dataset under an Ensemble Approach. Data 2023, 8, 36. https://doi.org/10.3390/data8020036

total_precipitation	reanalysis-era5-land	hour	hourly data necessary because cumulative variables not available in era5 land daily
"2m_temperature", "instantaneous_surface_sensible_heat_flux", "surface_net_solar_radiation", "surface_sensible_heat_flux", "surface_thermal_radiation_downwards", "evaporation", "potential_evaporation", "soil_temperature_level_1", "soil_temperature_level_2", "soil_temperature_level_3", "soil_temperature_level_4", "soil_type", "volumetric_soil_water_layer_1", "volumetric_soil_water_layer_2", "volumetric_soil_water_layer_3", "volumetric_soil_water_layer_4", "10m_u_component_of_wind", "10m_v_component_of_wind", "2m_dewpoint_temperature",	derived-era5-land-daily-statistics		

"surface_pressure", "top_net_thermal_radiation", "uv_visible_albedo_for_direct_radiation"			

4- Calculate SPEI-1 and SPEI-3

The next step is to calculate the difference between the precipitation *P* and PET for the month *i* is calculated:

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

which provides a simple measure of the water surplus or deficit for the analyzed month (while taking into account global warming processes).

The calculated D_i values are aggregated at different time scales. The difference $D^k_{i,j}$ in a given month *j* and year *i* depends on the chosen time scale *k*. For example, the accumulated difference for one month in a particular year *i* with a 12-month time scale is calculated using:

$$X^k_{i,j} = \sum_{l=13-k+j}^{12} D_{i-1,l} + \sum_{l=1}^{j} D_{i,l}, \quad \text{if} \quad j < k \quad \text{and}$$

$$X^k_{i,j} = \sum_{l=j-k+1}^{j} D_{i,l}, \quad \text{if} \quad j \geq k,$$

where $D_{i,l}$ is the *P* − PET difference in the first month of year *i*, in millimeters. The following step is to fit the Deficit series with a function to standardize the series and obtain the SPEI. The most commonly used is the log-logistic distribution[2] .The probability density function of a three-parameter log-logistic distributed variable is expressed as:

$$f(x) = \frac{\beta}{\alpha}\left(\frac{x-\gamma}{\alpha}\right)^{\beta-1}\left[1+\left(\frac{x-\gamma}{\alpha}\right)^{\beta}\right]^{-2},$$

[2] Vicente-Serrano, S. M., Beguería, S., & López-Moreno, J. I. (2010), a multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index. *Journal of climate*, *23*(7), 1696-1718.

where α, β, and γ are scale, shape and origin parameters, respectively, for D values in the range ($\gamma > D < \infty$).

Parameters of the log-logistic distribution can be obtained following different procedures. Among them, the L-moment procedure is the most robust and easy approach (Ahmad et al. 1988). When L moments are calculated, the parameters of the Pearson III distribution can be obtained following Singh et al. (1993):

$$\beta = \frac{2w_1 - w_0}{6w_1 - w_0 - 6w_2},$$

$$\alpha = \frac{(w_0 - 2w_1)\beta}{\Gamma(1 + 1/\beta)\Gamma(1 - 1/\beta)}, \quad \text{and}$$

$$\gamma = w_0 - \alpha\Gamma\left(\frac{1+1}{\beta}\right)\Gamma\left(\frac{1-1}{\beta}\right),$$

where $\Gamma(\beta)$ is the gamma function of β.

The probability distribution function of the D series, according to the log-logistic distribution, is given by

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

The $F(x)$ values for the D series at different time scales adapt very well to the empirical $F(x)$ values at the different observatories, independently of the climate characteristics and the time scale of the analysis. Figure 9 shows an example of the results for the 3- and 12-month series of Albuquerque, São Paulo and Helsinki, but similar observations were made for the other observatories and time scales. This demonstrates the suitability of the log-logistic distribution to model $F(x)$ values from the D series in any region of the world.

With $F(x)$ the SPEI can easily be obtained as the standardized values of $F(x)$. For example, following the classical approximation of Abramowitz and Stegun (1965),

$$\mathrm{SPEI} = W - \frac{C_0 + C_1 W + C_2 W^2}{1 + d_1 W + d_2 W^2 + d_3 W^3},$$

where

$$W = \sqrt{-2\ln(P)} \quad \text{for} \quad P \leq 0.5$$

and P is the probability of exceeding a determined D value, $P = 1 - F(x)$. If $P > 0.5$, then P is replaced by $1 - P$ and the sign of the resultant SPEI is reversed. The constants are C_0 = 2.515517, C_1 = 0.802853, C_2 = 0.010328, d_1 = 1.432788, d_2 = 0.189269, and d_3 = 0.001308. The average value of SPEI is 0, and the standard deviation is 1. The SPEI is a standardized variable, and it can therefore be compared with other SPEI values over time and space. An SPEI of 0 indicates a value corresponding to 50% of the cumulative probability of *D*, according to a log-logistic distribution.

2.1.1 Comparison of different models for PET

Different formulas on PET calculation were compared in order to get an insight of the deviations among them so that we can come to a conclusion on the formula that will be used to calculate potential evapotranspiration, needed for SPEI. The different formulas tested are:

	Formula	Timescale
FAO-56 Penman-Monteith	$\frac{0.408\Delta(R_n-G)+\gamma\frac{900}{(T+273)}u_2(e_s-e_a)}{\Delta+\gamma(1+0.34u_2)}$	Daily PET
Hargreaves Samani	$0.0023(T_{mean}+17.8)(T_{max}-T_{min})^{0.5}\ 0.408R_a$	Daily PET
Hargreaves Samani modified	$0.0013(T_{mean}+17)\lvert T_{max}-T_{min}-0.0123P\rvert^{0.76}\ 0.408R_a$	Daily PET
Thornthwaite	$16(\frac{L}{12})(\frac{N}{30})(\frac{10Td}{I})^a$	Monthly PET

In the above formulas, daily input data were used in the Penman-Monteith (Daily G soil heat flux in Penman-Monteith has been approximated equal to zero as in Allen et al., 1988 https://www.fao.org/4/X0490E/x0490e00.htm), Hargreaves and Hargreaves modified equations, to calculate daily PET values. The modified Hargreaves formulation is the one presented by Droogers et al, 2001[3]. In the analysis these daily values were then converted to monthly PET values, by summing the daily values of each month, in order to be compared with the monthly PET values provided by Thornthwaite formula.

[3] Droogers, P., & Allen, R. G. (2002). Estimating reference evapotranspiration under inaccurate data conditions. *Irrigation and drainage systems*, *16*, 33-45.

FAO-56 Penman-Monteith

Δ = Saturation vapor pressure curve (kPa/°C): $\frac{4098[0.6108exp(\frac{17.27Tmean}{Tmean+237.3})]}{(Tmean+237.3)^2}$ (Tmean: daily mean temp.)

R_n = Daily Net Radiation (MJ/m^2)

T = Daily mean temperature (°C)

γ = Psychrometric constant (kPa/°C): 0.000665*101.3$[\frac{293-0.0065z}{293}]^{5.26}$ (z: altitude in m)

u_2 = Wind speed at 2m (m/s)

e_s = saturated vapor pressure (kPa): $\frac{e(Tmax)+e(Tmin)}{2}$ ($e_{Tmax\ (Tmin)}$ = 0.6108exp($\frac{17.27Tmax(Tmin)}{Tmax(Tmin)+237.3}$))

e_a = Actual vapor pressure (kPa): $\frac{e(Tmin)(RHmax/100)+e(Tmax)(RHmin/100)}{2}$ (RHmax, RHmin: daily max, min)

Hargreaves Samani & Hargreaves Samani modified

$T_{mean,\ max,\ min}$ = Daily mean, max, min temperature (°C)

R_a = Daily extraterrestrial radiation (MJ/m^2) (defined in 2.1 SPEI)

P: Daily precipitation (mm)

Thornthwaite

L = Sunlight hours: $\frac{24}{\pi}$*$\cos^{-1}$(-tanφ * tanδ) (φ, δ defined in 2.1 SPEI)

N = Number of days in the month

T_d = Average daily temperature of the month being calculated (°C)

I = Heat index (°C): $\sum_{i=1}^{12} (\frac{Tmi}{5})^{1.514}$ (T_{mi}: mean temperature of each month of the year)

a = ($6.75\ 10^{-7}$)*I^3 – ($7.71\ 10^{-5}$)*I^2 + ($1.792\ 10^{-2}$)*I + 0.49239

Formulas comparison

The comparison is applied on the monthly cumulative PET values derived from the daily ones, and the impact on SPEI values is examined as well. For this comparison, historical data from Copernicus ERA5 were used (1960-2022) and interpolated to the target location of Ioannina, a region with similar characteristics to the Hellenic pilot site.

The FAO-56 Penman-Monteith is assumed as the reference method since it's often mentioned as the most accurate one, even though it's the most challenging to implement, since it requires much more input data than the other formulas.

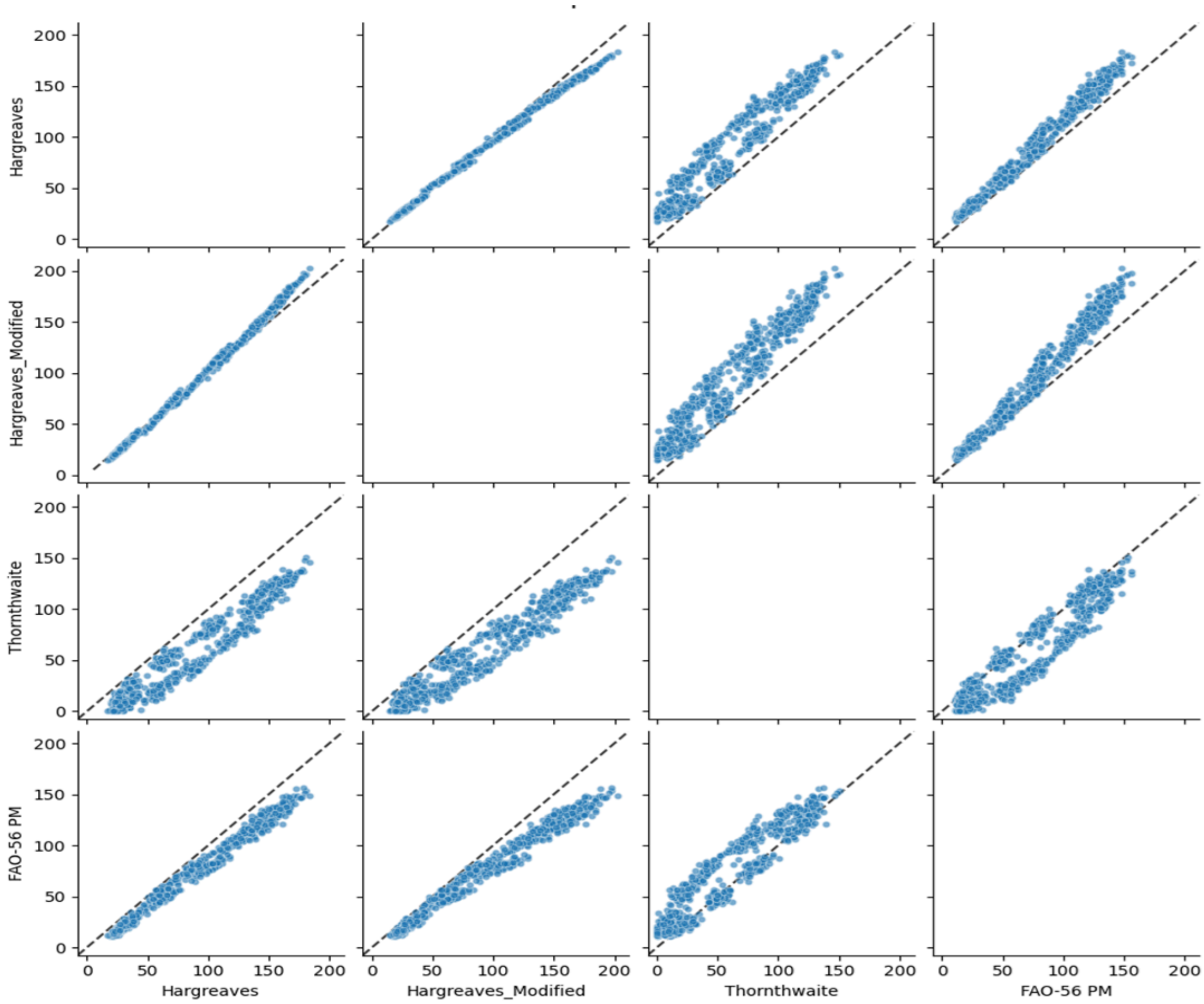

***Figure 1** Pairwise scatterplots of monthly PET values in mm.*

Focusing on the results of the pairwise scatterplots we see that Thornthwaite underestimates potential evapotranspiration compared to Penman-Monteith, while Hargreaves and modified Hargreaves formulas overestimate potential evapotranspiration compared to Penman-Monteith.

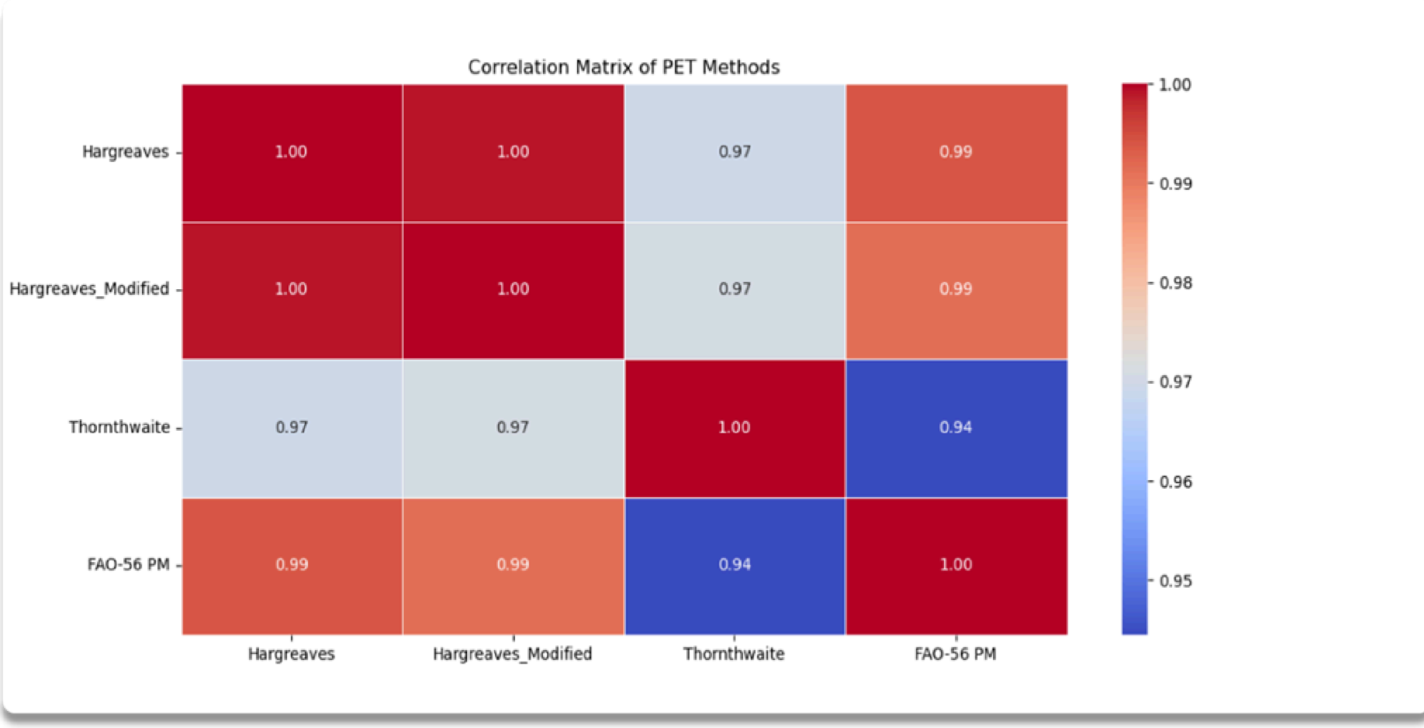

Figure 2 *Correlation matrix of monthly PET values among the different formulas*

The correlation matrix indicates a satisfying correlation of monthly PET values among the different methods. Thornthwaite shows a slightly lower correlation, probably because it relies only on temperature, but still the correlation coefficient is very good.

To quantify the overestimation/underestimation of monthly PET compared to Penman-Monteith, the residual values were calculated by subtracting PET calculated by the alternative formulas to PET calculated by Penman-Monteith:

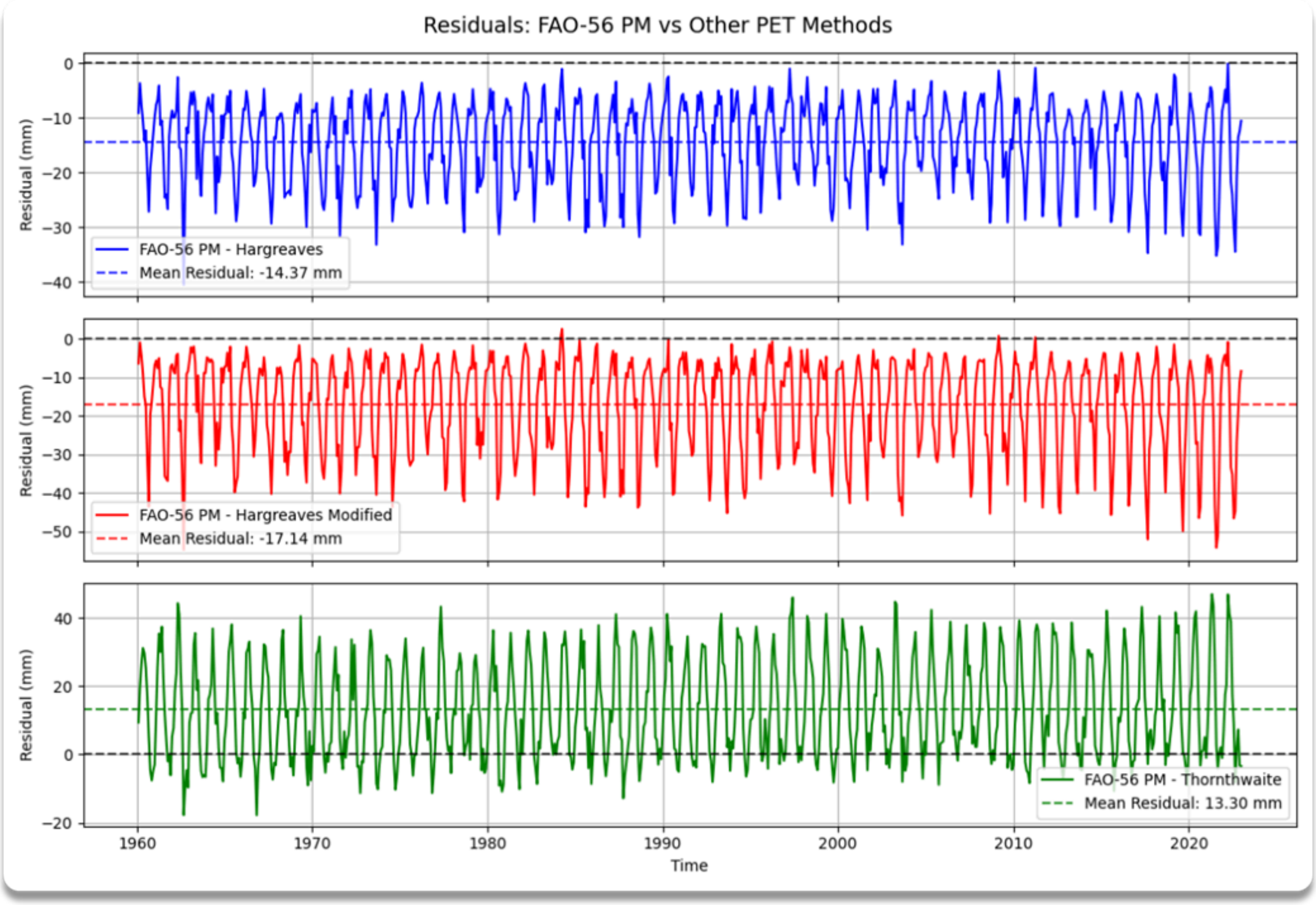

***Figure 3** Residuals time series. Each method's monthly PET values are subtracted to the Penman-Monteith's monthly PET value to obtain the residuals.*

From the timeseries of the residuals values we better understand the degree of overestimation/underestimation of the formulas compared to Penman-Monteith. On average, monthly PET is overestimated by 14.37 mm using Hargreaves, 17.14 mm using modified Hargreaves and is underestimated by 13.30 mm using Thornthwaite, compared to Penman-Monteith. Through the corresponding histograms we can have a view of the frequency of occurrence of each overestimation/underestimation value and the range of these:

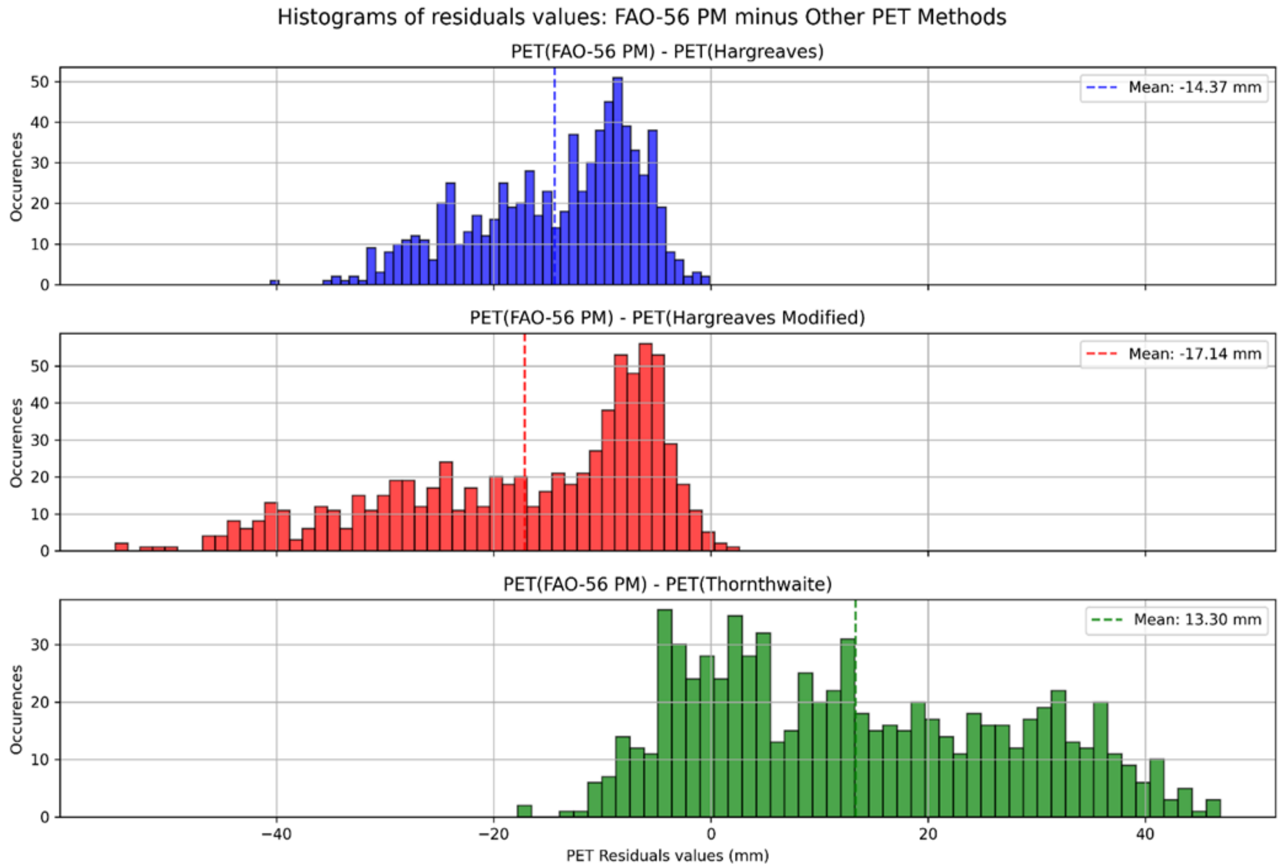

Figure 4 *Histograms of residuals values between Penman-Monteith and the other formulas*

To put this comparison into a percentage perspective, the percentage of bias was calculated through the formula:

$$\text{Pbias} = \left| \frac{\sum_{i=1}^{n} PET(\#) - PET(Penman-Monteith)}{\sum_{i=1}^{n} PET(Penman-Monteith)} \right| * 100\%$$

where # is Hargreaves, Hargreaves modified and Thornthwaite formulas.

	Pbias (%)
Hargreaves	20.7
Hargreaves modified	24.7
Thornthwaite	(-)19.2

Given the p-bias values we can support that Hargreaves and Hargreaves modified overestimate PET by 20.7% and 24.7% respectively, while the underestimation of PET by Thornthwaite is in the range of 19.2%.

Impact on SPEI values

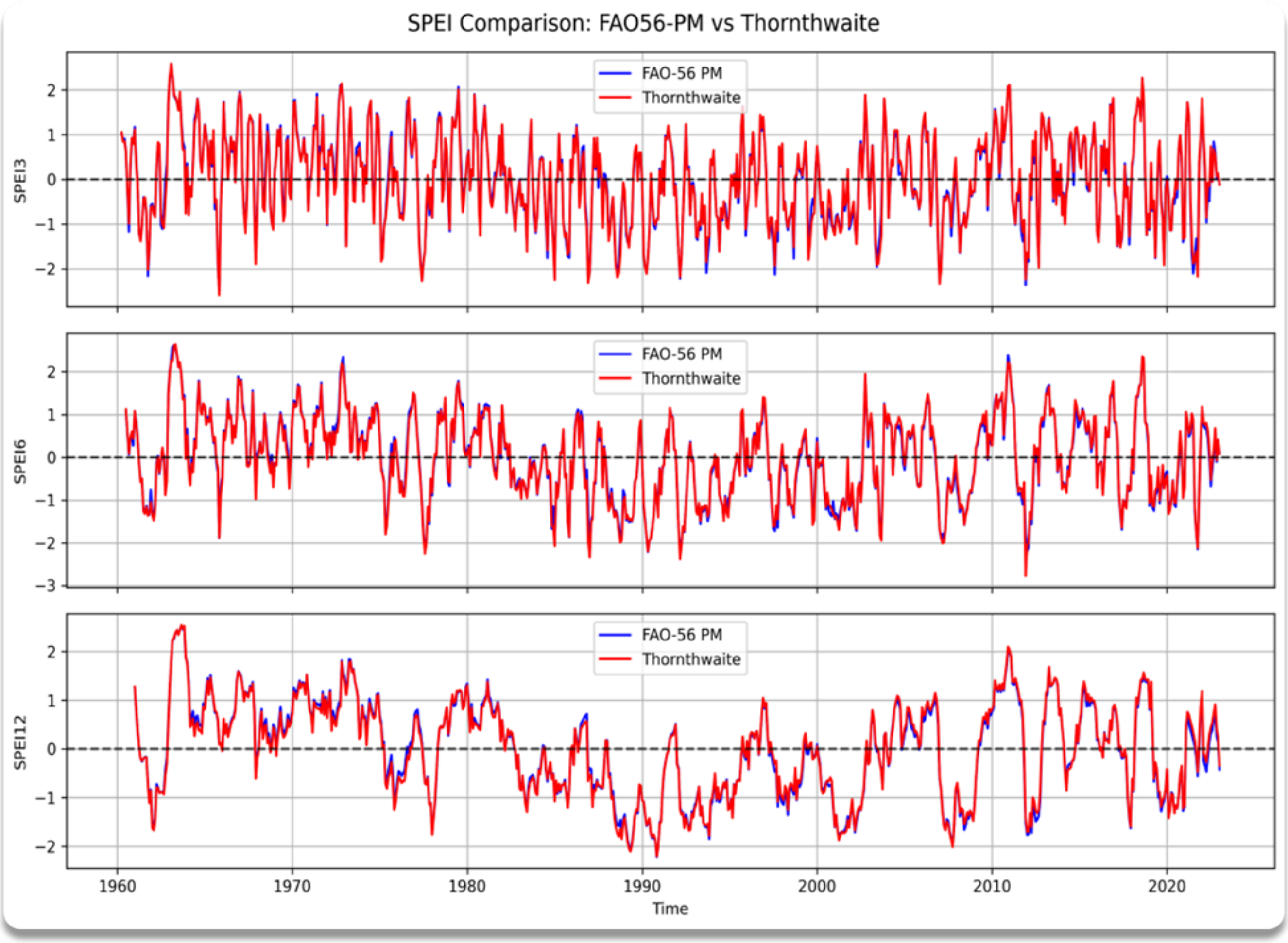

Figure 5 *SPEI-3, SPEI-6 and SPEI-12 timeseries under 2 scenarios: 1) PET calculated by Penman-Monteith and 2) PET calculated by Thornthwaite.*

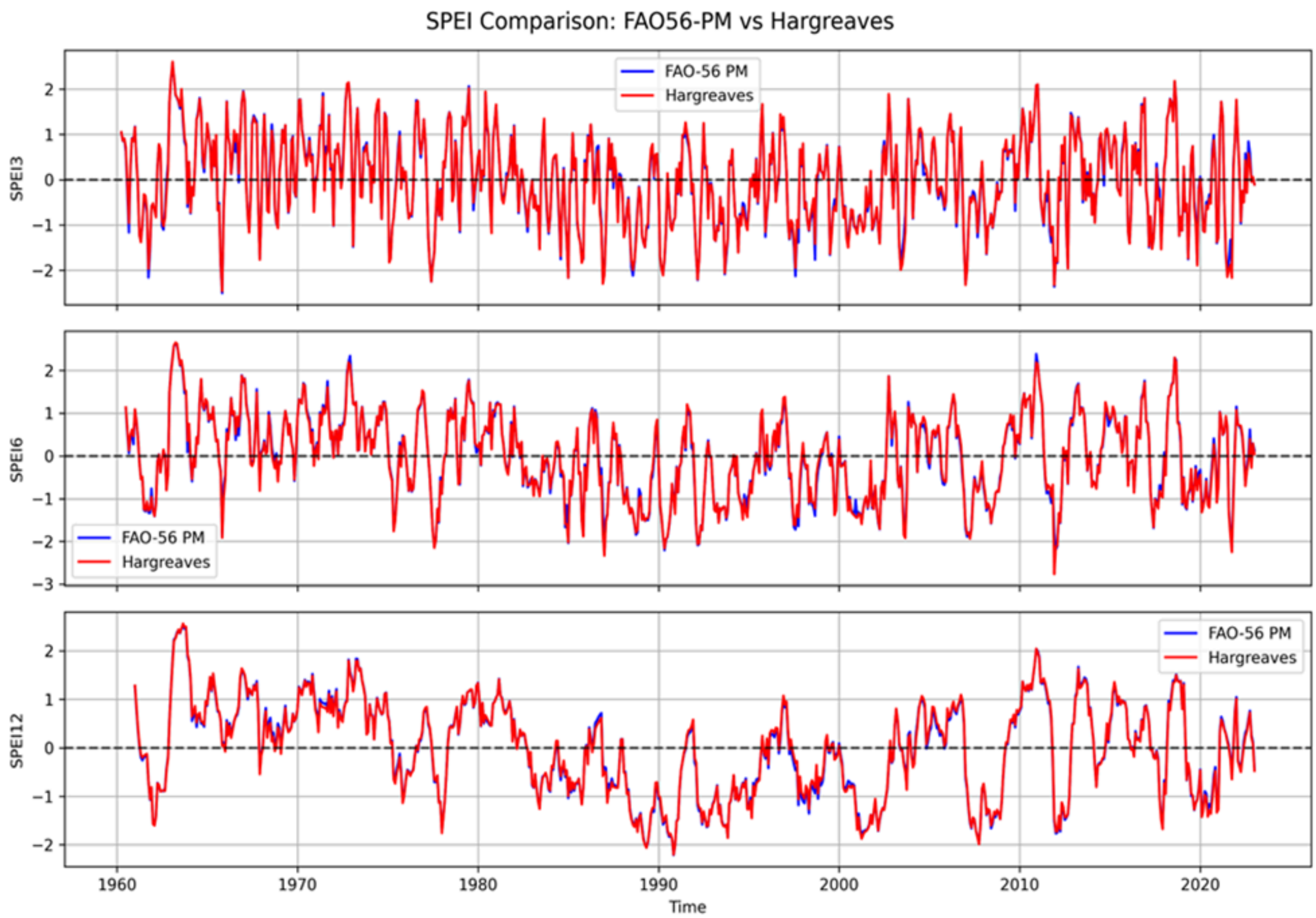

Figure 6 *SPEI-3, SPEI-6 and SPEI-12 timeseries under 2 scenarios: 1) PET calculated by Penman-Monteith and 2) PET calculated by Hargreaves.*

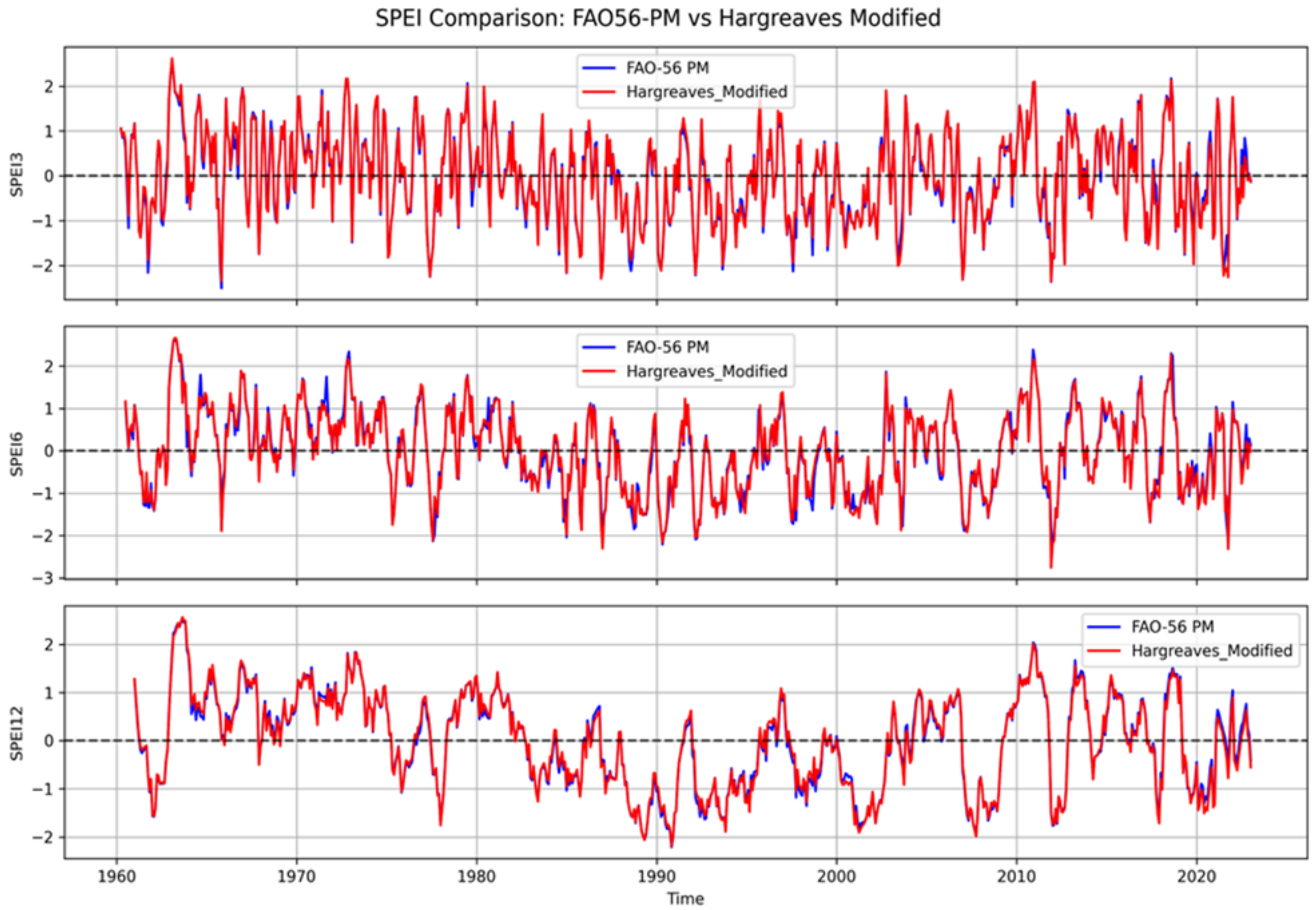

***Figure 6** SPEI-3, SPEI-6 and SPEI-12 timeseries under 2 scenarios: 1) PET calculated by Penman-Monteith and 2) PET calculated by Hargreaves modified.*

The SPEI time series indicate that the effect on SPEI values is relatively small and the results are satisfying among the different PET formulas. There are some values that show some deviation, but the overall correlation is good, as we can also see through the scatterplots of the SPEI values:

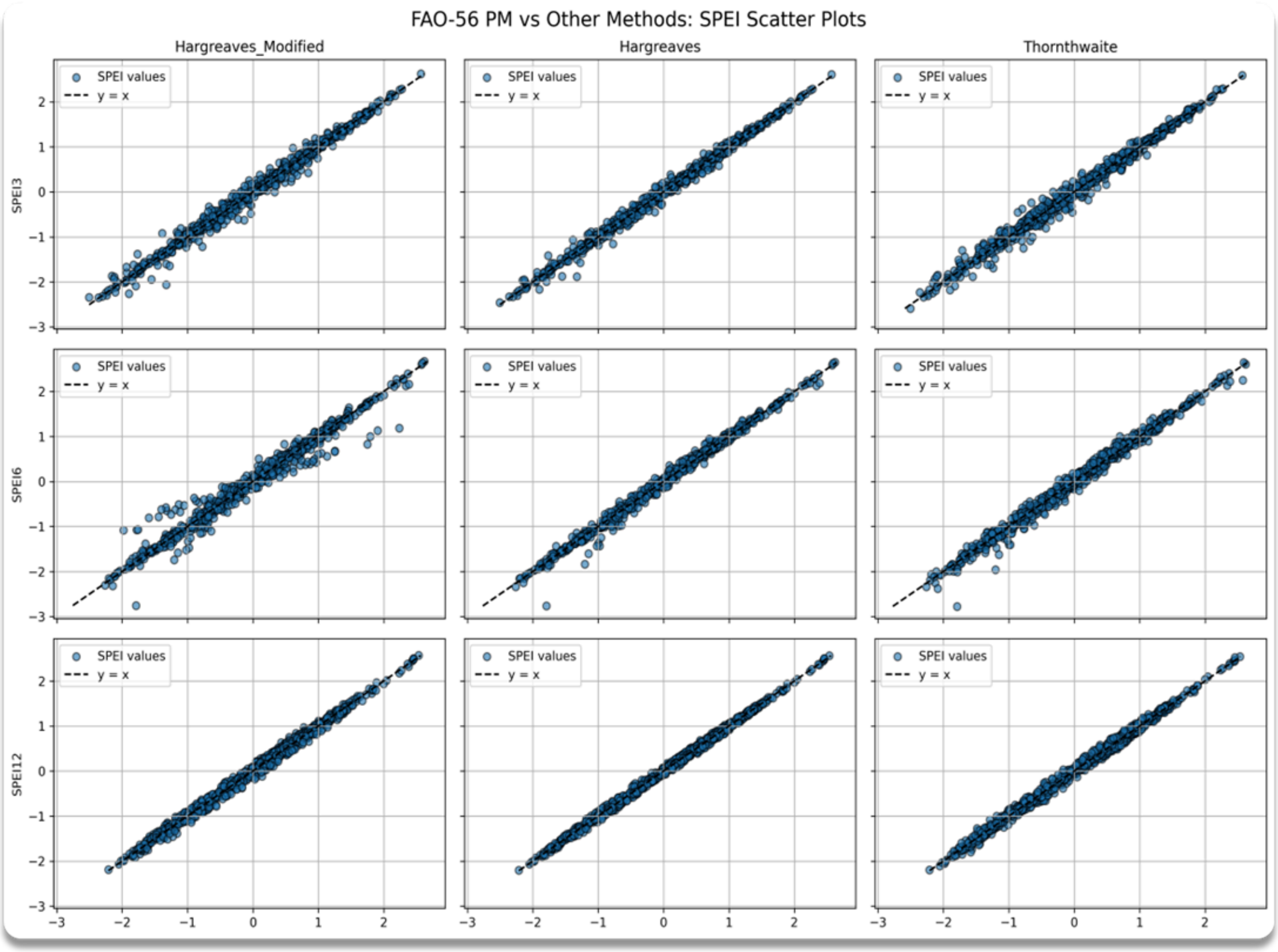

Figure 7 *Scatterplots of SPEI values by using different PET formulas. SPEI having used Penman-Monteith is in the x-axis while SPEI having used the other formulas is in the y-axis of the scatterplots.*

Summary

- ➔ Observed pattern of monthly PET values: Thornthwaite < Penman-Monteith < Hargreaves / Hargreaves modified
- ➔ High correlation among the different formulas
- ➔ The effect on SPEI values is relatively small and there is high agreement among the different formulas to be used. So:
- ➔ No need to work with Penman-Monteith which requires many input data and even assumptions on the formula

- ➔ Considering that in the frame of the project we need to calculate a drought vulnerability index, it's more desirable to overestimate PET instead of underestimating it, so the use of Hargreaves Samani formula could be prioritized

2.2 CWSI

Surface energy balance is central to understanding of surface biogeochemical processes and can be written as:

$$R_n = \lambda ET + H + G$$ [4]

where R_n is the net radiation, H is the sensible heat flux, λET is the latent heat flux in which λ is the latent heat of evaporation of water and ET is the rate of evaporation of water, G is the soil heat flux. When a plant goes from a no-stress to a stressed water condition, the ratio between potential evapotranspiration, i.e. the evapotranspiration of a crop with adequate amount of water, and the actual one changes, ranging from 1 to 0. The Crop Water Stress Index (CWSI), is calculated according to the following equation:[5]

$$CWSI = 1 - \frac{E}{E_p}$$

where E is the actual evapotranspiration and E_p is the potential evapotranspiration. This index is indicative of the plant's stress level and ranges from 0 to 1. We can regard the potential evapotranspiration as the difference between the net radiation and the soil heat flux [6], therefore we can rewrite CWSI as:

$$CWSI = 1 - \frac{\lambda ET}{R_n - G}$$

[4] Barry, Roger G. *Mountain weather and climate*. Routledge, 2013.
[5] Jackson, Ray D., et al. "Canopy temperature as a crop water stress indicator." *Water resources research* 17.4 (1981): 1133-1138.
[6] Renhua, Zhang, et al. "The potential information in the temperature difference between shadow and sunlit of surfaces and a new way of retrieving the soil moisture." *Science in China Series D: Earth Sciences* 44 (2001): 112-123.

In Figure 1 is reported trend of CWSI computed over a flat vineyard terrain. Physical data were acquired at Lison station, that is one of the ICOS[7] network stations. We can observe the natural periodic trend of CWSI over time. Minimum values occur during summer seasons when actual evapotranspiration is obviously far lower than potential evapotranspiration and plants are in a stressed water condition.

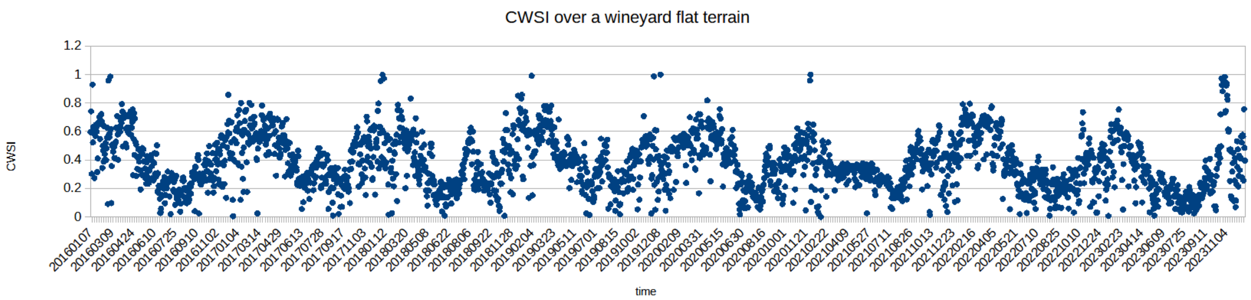

Figure 1. Trend of CWSI computed over a flat vineyard terrain

2.3 NDVI

The normalized difference vegetation index (NDVI) is widely-used for quantifying the vigour level of vegetation. It is calculated according to the following equation:[8]

$$NDVI = \frac{NIR - RED}{NIR * RED}$$

where NIR and RED are the near-infrared and the red visible radiation coming from vegetation respectively. The NDVI index ranges from -1 to 1. It assumes a high value in presence of growing, healthy vegetation and values close to zero in absence of vegetation or no green leaves. NDVI values less than 0 and close to -1 usually correspond to radiance values coming from the water body.

2.4 Input data, data source and range of valuability

[7] https://www.icos-cp.eu/observations/station-network

[8] Rouse, J. W., Haas, R. H., Schell, J. A., & Deering, D. W. (1974). Monitoring vegetation systems in the Great Plains with ERTS. NASA Spec. Publ, 351(1), 309.

In table 1 the input data necessary for the computation of each indicator, input data source and the range of valuability are summarized for each selected indicator.

Table 1. Input data, data source and range of valuability

	CWSI	SPEI	NDVI
Input data	Net radiation (R_n), Soil heat flux (G), Latent heat flux (λET)	Meteorological data	NIR radiation red visible radiation
Source of input data	Global meteorological datasets	Global meteorological datasets	Remotely sensed or in situ radiometric data
Range	[0;1] from not water stressed to stressed condition	Extremely wet (EW) 2.00 and above Severely wet (VW) 1.50 to 1.99 Moderately wet (MW) 1.00 to 1.49 Near normal (NN) −0.99 to 0.99 Moderately dry (MD) −1.00 to −1.49 Severely dry (SD) −1.50 to −1.99 Extremely dry (ED) −2.00 and less[9]	[-1;1] from water body to healthy vegetation

[9] Vicente-Serrano, S. M., Beguería, S., LópezMoreno, J. I. (2010). A Multiscalar Drought Index Sensitive to Global Warming: The Standardized Precipitation Evapotranspiration Index. Journal of Climate, 23(7), 1696–1718. https://doi.org/10.1175/2009JCLI2909.1

3 Retrieving and Processing Sentinel-2 Data from Google Earth Engine: main steps

1 Authenticate and Initialize GEE: Access Google Earth Engine using Python.
2 Apply Cloud Masking: Use the QA60 band to remove clouds and cirrus contamination.
3 Retrieve Sentinel-2 Data: Filter based on date range, cloud percentage, and spectral bands.
4 Compute Vegetation NDVI: Computed from the NIR (B8) and Red (B4) bands.
5 Extract Data for a Specific Location: Use a point geometry and extract pixel values within a buffered region.
6 Store and Process Extracted Data: Convert the extracted data into a structured Pandas DataFrame for further analysis and visualization.

3.1 Step 1: Authenticate and Initialize Google Earth Engine

Before accessing Sentinel-2 data, Google Earth Engine (GEE) must be authenticated and initialized within the Python environment.

```python
import ee
# Authenticate and initialize Google Earth Engine
ee.Authenticate()
ee.Initialize()
```

3.2 Step 2: Define Cloud Masking Function

To ensure high-quality data, a cloud masking function is applied. The QA60 band of Sentinel-2 imagery is used to identify and remove cloud-contaminated pixels.

```python
Python
def mask_s2_clouds(image):
    qa = image.select('QA60')  # Select the QA60 band for cloud
masking
    cloud_bit_mask = 1 << 10  # Cloud mask
    cirrus_bit_mask = 1 << 11  # Cirrus mask
    mask = (
        qa.bitwiseAnd(cloud_bit_mask).eq(0)
        .And(qa.bitwiseAnd(cirrus_bit_mask).eq(0))
    )
    return image.updateMask(mask).divide(10000)  # Apply the mask
and scale reflectance values
```

3.3 Step 3: Define Function to Extract Sentinel-2 Data

This function retrieves Sentinel-2 surface reflectance data for a specified location and time range. It applies cloud masking and selects relevant spectral bands.

```python
Python
def extract_s2(sitecode, lon, lat, start_date, end_date):
    lonlat = ee.Geometry.Point(lon, lat)

    sen2 = (
        ee.ImageCollection("COPERNICUS/S2_SR_HARMONIZED")
        .filterDate(start_date, end_date)
        .filter(ee.Filter.lt("CLOUDY_PIXEL_PERCENTAGE", 10))
        .map(mask_s2_clouds)
        .select(['B1', 'B2', 'B3', 'B4', 'B5', 'B6', 'B7', 'B8',
'B8A', 'B9', 'B11', 'B12'])
    )

    def ndvi_index(image):
```

```
        NDVI_param = image.normalizedDifference(['B8',
'B4']).rename('NDVI')
        return image.addBands([NDVI_param])
    coll = sen2.map(ndvi_index)
    # Extract data from GEE to a Pandas-compatible format
    try:
        values = coll.getRegion(lonlat, scale=10).getInfo()
    except Exception as e:
        print(f"Error extracting data for {sitecode} from
{start_date} to {end_date}: {e}")
        return None
    return values
```

3.4 Step 4: Processing Extracted Data

Once Sentinel-2 data is extracted, it is processed into a Pandas DataFrame for further analysis.

```python
import pandas as pd
def process_s2_data(df, start_date, end_date):
    grouped = df.groupby('sitecode')
    df_data = []
    for sitecode, group in grouped:
        lon, lat = group.iloc[0][['longitude', 'latitude']]
        extracted_data = extract_s2(sitecode, lon, lat, start_date,
end_date)
        if extracted_data is None:
            continue
        header = extracted_data[0]
        data = extracted_data[1:]
        for row in data:
            if len(row) >= 15:
```

```
                df_row = [
                    row[0], sitecode, row[1], row[2], row[4],
row[5], row[6], row[7], row[8],
                    row[9], row[10], row[11], row[12], row[13],
row[14], row[15], row[16]
                ]
                df_data.append(df_row)
    df_sen2 = pd.DataFrame(df_data, columns=[
        'Date', 'sitecode', 'lon', 'lat', 'B1', 'B2', 'B3', 'B4',
'B5', 'B6', 'B7', 'B8', 'B8A', 'B9', 'B11', 'B12', 'NDVI'])
    return df_sen2
```

3.5 Summary of the Workflow

1. Authenticate and initialize GEE.
2. Apply cloud masking using QA60 band.
3. Retrieve Sentinel-2 data by filtering for date range, cloud percentage, and relevant spectral bands.
4. Compute vegetation indices such as NDVI and STR for further analysis.
5. Extract data at a specific site and convert it into a structured format.
6. Store and process the extracted data into a Pandas DataFrame for further analysis and visualization.

3.6 NDVI Computation for a Circular Area (1 km Radius) of the Greek Site

The following Python script utilizes Google Earth Engine (GEE) to compute the Normalized Difference Vegetation Index (NDVI) for a circular area with a 1 km radius around the Greek study site. The script processes Sentinel-2 imagery on a monthly basis, applies cloud masking, computes NDVI, and downloads the processed images as GeoTIFF files.

```
import ee
import os
import requests
from datetime import datetime, timedelta

# Initialize the Earth Engine API
ee.Initialize()

# Cloud masking function
def mask_s2_clouds(image):
    qa = image.select('QA60')
    cloud_bit_mask = 1 << 10
    cirrus_bit_mask = 1 << 11
    mask = (
        qa.bitwiseAnd(cloud_bit_mask).eq(0)
        .And(qa.bitwiseAnd(cirrus_bit_mask).eq(0))
    )
    return image.updateMask(mask).divide(10000)

# NDVI computation function
def ndvi_index(image):
    NDVI_param = image.normalizedDifference(['B8', 'B4']).rename('NDVI')
    return image.addBands([NDVI_param]).set("system:time_start",
image.get("system:time_start"))

# Study site for Greece pilot
lonlat = ee.Geometry.Point(22.080389, 38.170750)
buffered_bbox_greece = lonlat.buffer(1000).bounds()

# Define start and end dates
start_date = datetime(2024, 3, 1)
end_date = datetime(2025, 3, 1)

# Define output directory
directory_path = "/users_home/cmcc/oa33824/work_dir/pilot/basemap/greece"
os.makedirs(directory_path, exist_ok=True)

# Process data in monthly chunks
current_date = start_date
while current_date <= end_date:
    next_date = current_date + timedelta(days=30)  # Process one month at
a time
```

```
    print(f"Processing: {current_date.strftime('%Y-%m-%d')} to
{next_date.strftime('%Y-%m-%d')}")

    try:
        # Generate the monthly basemap (RGB composite)
        s2_basemap = (
            ee.ImageCollection("COPERNICUS/S2_SR_HARMONIZED")
            .filterDate(current_date.strftime('%Y-%m-%d'),
next_date.strftime('%Y-%m-%d'))
            .filter(ee.Filter.lt("CLOUDY_PIXEL_PERCENTAGE", 10))
            .map(mask_s2_clouds)
            .select(["B2", "B3", "B4"])  # Blue, Green, Red
            .median()
            .clip(buffered_bbox_greece)  # Make sure to clip after
mosaicking
        )

        # Get download URL
        url = s2_basemap.getDownloadURL({
            "scale": 10,
            "region": buffered_bbox_greece,  # Ensure consistency
            "format": "GeoTIFF",
            "crs": "EPSG:4326"
        })

        # Download the image
        response = requests.get(url, stream=True)
        file_path =
f"{directory_path}/basemap_{current_date.strftime('%Y-%m')}.tif"

        with open(file_path, "wb") as file:
            file.write(response.content)

        print(f"Downloaded basemap for {current_date.strftime('%Y-%m')}")

    except Exception as e:
        print(f"Error downloading basemap for
{current_date.strftime('%Y-%m')}: {e}")

    # Move to the next month
    current_date = next_date
```