HELLENIC REPUBLIC MINISTRY OF DEVELOPMENT GENERAL SECRETARIAT FOR RESEARCH AND INNOVATION **HELLENIC FOUNDATION FOR RESEARCH AND INNOVATION**	

Greece 2.0 NATIONAL RECOVERY AND RESILIENCE PLAN
"BASIC RESEARCH FINANCING" (Horizontal support for all Sciences)
ID 16618 – Subproject 1 (MIS: 5163923)

WP5

D5.1 Report on Data Analysis techniques

Spatially Explicit Digital Twin of the Greek Agro-Hydro-System

ID 14815

Plan Details

Report on data analysis techniques
Deliverable number: D5.1
Creation Date: 10/04/2025
Last modification date: 19/04/2025
Dissemination Level: Public

1. Introduction

This deliverable presents the data analysis methods adopted and applied within the framework of Work Package 5 (WP5) of the DT-Agro project. WP5 focuses on processing, evaluation and adaptation of effective data analysis techniques. The analyses focus on the parameters used in the DT-Agro model, which include meteorological, soil data and land surface parameters that support its development and validation. The aim is to ensure that these data are efficiently utilized to support the development of the DT-Agro and to enable the analysis of agricultural, hydrological, and environmental processes.

Task 5.1 is addressed in this deliverable, which covers the investigation and implementation of data analysis techniques for the utilization of the extensive Earth Observation (EO) datasets used and produced by DT-Agro (M12-M18). The work presented here establishes the foundation for subsequent analyses in Task 5.2 (M18-M24) and for the design of digital, spatially explicit services under Task 5.3 (M18-M24). Tasks 5.2 and 5.3 will be analyzed in deliverables D5.2 and D5.3.

Several EO-based and many other datasets were evaluated for their suitability in supporting the DT-Agro framework. These included multiple satellite-derived products which offer different spatial and temporal resolutions. The work reported in this deliverable includes the processing and harmonization of all datasets, the evaluation of meteorological data, the analysis of topographic data (DEM), and the implementation of the Curve Number (CN) estimation method, integrating land cover, soil, and imperviousness data.

2. Description of Work & Results

All the data integrated into DT-Agro are from a wide range of EO and ground-based sources. Key datasets, which will be analyzed onwards, include meteorological reanalysis data (AgERA5), national meteorological station records, soil data from Greek and international repositories, the EU Digital Elevation Model (EU-DEM), and Copernicus Land Monitoring Service (CLMS) products such as land cover, imperviousness, surface soil moisture (SSM), and Normalized Difference Vegetation Index (NDVI). All spatial datasets were harmonized to a 100 m grid to ensure full compatibility with the spatial resolution defined for the DT-Agro model.

This resolution represents an optimal balance between national-scale applicability and computational feasibility.

Based on the abovementioned data, all the parameters required for the model are calculated dynamically in spatially distributed form. This workflow involves coordinate system unification (Greek Grid, EPSG:2100), clipping to the area of Greece, resampling to common spatial resolution, temporal aggregation and format modifications (such as from .nc to .csv, or .nc to GeoTIFF). The results of the analyses will be presented in the next deliverables of this WP. Challenges that arose during the data processing will be analyzed below.

2.1 DEM Analysis

Regarding topographic information, several Digital Elevation Models (DEMs) were reviewed in early stages of the model, which included datasets offering various resolutions. Following this evaluation, the Digital Elevation Model over Europe (EU-DEM) produced by the European Space Agency & European Union (2019) was selected as the elevation data source for DT-Agro.

The hydrological model configuration relied on the DEM over Europe (EU-DEM) which offers a resolution of 30m, uniform coverage and open accessibility, making it an optimal choice for hydrological and terrain analyses within the DT-Agro. The DEM was resampled to 100 m to ensure full consistency with the spatial framework of the DT-Agro. The digital twin uses this DEM to dynamically simulate water flow under varying rainfall and climate conditions. Products derived from the DEM will also interact with updated datasets to reflect real-time changes in hydrological conditions.

The DEM is being processed and analyzed to derive key topographic and hydrological parameters required by the DT-Agro. To ensure accuracy and consistency across the study area, the DEM was spatially aligned with all other datasets in the same coordinate system (Greek Geodetic Reference System 1987 - GGRS87).

All fundamental hydrological terrain derivatives required by the model were then produced. The preprocessing workflow included filling to generate a hydrologically corrected DEM, followed by the computation of flow direction and flow accumulation to delineate drainage networks and contributing areas. Additional topographic layers were derived from the EU-DEM, including flow length, slope, and cell-to-cell travel time, which characterize the spatial distribution of water movement across the landscape. Velocity-related rasters were also produced, encompassing inland (hillslope) flow velocity, overland/channel flow velocity, and the resulting total flow velocity grid. Together, these hydrologically relevant raster datasets form the foundation for runoff routing, infiltration estimation, and water balance calculations within DT-Agro, ensuring that all spatially distributed hydrological processes are initialized with consistent and physically meaningful topographic information.

The DEM will be dynamically integrated with updated meteorological and land cover data, allowing the Digital Twin to visualize and model hydrological responses to changing climatic and environmental conditions.

2.2 Meteorological Data Analysis

2.2.1 Observed meteorological data

Daily data derived from the Hellenic National Meteorological Service (HNMS) for a period from 1971 to 2024 are utilized to analyze long-term climatic trends, seasonal variability, and daily weather patterns that influence irrigation needs. The historical dataset provided by HNMS covers the period 1971-2004, containing daily measurements of rainfall, temperature, and other key variables for multiple meteorological stations across Greece. In addition, updated daily data for some of the 140 stations were provided from HNMS, resulting in an extended dataset spanning from 1 January 2005 to the most recent records available for each station. The coverage of the new data varies depending on the station: some stations include data up to 30 April 2024, while others have shorter series. These datasets include observed variables of temperature, precipitation, relative humidity, and wind speed. All station datasets were carefully merged into continuous time series to create the most complete possible record for each variable and location.

During data analysis, a challenge was the fragmented structure of the available datasets, as meteorological records from individual stations were distributed across multiple files, often separated by parameter or time period. To address this, a structured data management and preprocessing workflow was implemented. Individual files were merged, standardized, and reformatted into continuous station-based time series. Another issue encountered was the presence of missing or incomplete data for certain variables and time periods, including gaps spanning several days or, in some cases, entire years. To mitigate this, a combination of statistical and quality-control techniques was applied. Short gaps were filled using temporal interpolation or regression with nearby stations. These procedures ensured that only reliable data were used for calibration, evaluation, and model development.

Monthly meteorological data were collected specifically for the year 2023 from many (777 in total) meteorological stations across Greece, i.e. monthly mean temperature (degrees Celsius) and cumulative precipitation (mm). Despite the overall availability of the dataset, incomplete records at several stations resulted in missing data for certain months. Particular attention has been given to this dataset, since it was used for the evaluation of the AgERA5 dataset (a widely used reanalysis product with more information provided below).

Next step is to extend this evaluation to include data from 2024, which are currently being collected by the team.

2.2.2 AgERA5

After assessing multiple criteria, the project team decided to adopt the AgERA5 dataset as EO meteorological data source. AgERA5 is a climate reanalysis product developed under the C3S (Copernicus Climate Change Service, 2020) and derived from ERA5, offering daily global meteorological information specifically tailored for agricultural and agro-ecological studies (Soulis et al., 2025). It provides data from 1979 to the present on key variables such as precipitation, temperature, solar radiation, and wind speed at spatial resolution of 0.1° grid

(Boogaard et al., 2020). The service is based on the fifth generation of ECMWF atmospheric re-analyses of the global climate, better known as ERA5. AgERA5 'connects' users in the agricultural domain to the new ERA5 data set.

Using the coordinates of the points that correspond to 140 meteorological stations operated by the HNMS, data was obtained from the AgERA5 dataset, which is part of the Copernicus Climate Data Store (CDS). The AgERA5 grid cells that match with the area of Greece are defined by the bounding box 34.55°N to 41.92°N and 19.09°E to 30.15°E. Daily 24h mean temperature (air temperature at a height of 2 meters above the surface in Kelvin) and precipitation flux (total volume of liquid water (mm^3) precipitated over the period 00h-24h local time per unit of area (mm^2), per day in mm day^{-1}) data for the years 1979-2024 were downloaded for Greece using CDS's API and Python programming language, which allows automated access to large volumes of climate data. In addition to these variables, daily reference evapotranspiration (ET_o) (mm day^{-1}), also provided within the AgERA5 agrometeorological indicators, was downloaded for the same area and temporal period. This parameter is calculated using the Penman-Monteith method as described by the FAO56 guidelines, it represents the rate at which a well-watered reference crop loses water to the atmosphere through evapotranspiration and transpiration. Including ET_o in the analysis ensures that the DT-Agro can characterize atmospheric water demand consistently and can support crop water balance assessments. Together, these variables form the basis for describing local climatic conditions relevant to agricultural and hydrological applications.

The station coordinates of those 140 extraction points were expressed in standard geographic coordinates (WGS84 latitude and longitude). Each HNMS station has a unique identifier, which was preserved throughout the data extraction and processing pipeline to ensure that individual series could always be traced back to their original location.

Three main scripts were prepared, one used to extract the daily mean temperature, one for precipitation flux, and another one for reference evapotranspiration ET_o. Other scripts were needed to unzip downloaded files and transform NetCDF files to .csv. To comply with CDS request limits and to avoid oversized downloads, the retrieval was structured so that data were requested one year at a time for all stations. Each request returned a compressed file in .zip format, which contained NetCDF files holding the daily time series for the requested variable and year. The downloaded files in .zip format were named systematically to include the variable type, the year, and the station identifier (e.g.: 2m_temperature_1979_16600.0.zip, precipitation_flux_1979_16600.0.zip, reference_evapotranspiration_1979_16600.0.zip).

Once downloaded, the data were processed in Python using the xarray and pandas libraries. The .zip archives were first extracted. Then, the processed data were organized into folders, each folder containing data for one station and exported into .csv format with the same systematic file naming to ensure that the variable, year and station could always be identified. Additionally, the NetCDF files could be opened using QGIS software, although this

option was not used in the present project. All NetCDF handling, extraction and conversion tasks were performed exclusively through Python

To ensure quality and consistency, the workflow was tested initially on a single station, verifying both the units of the variables and the alignment of the time axis. Splitting the requests by year proved necessary not only to avoid “request too large” errors from the CDS system but also to make the downloading process more reliable. By following this procedure, a complete dataset was created that contains daily temperature and precipitation records for all 140 stations from 1979 through 2024. The final outputs are station-wise CSV files, each containing a continuous daily time series that is ready for use in further statistical analysis, hydrological modeling, or agricultural applications.

Evaluation of AgERA5 against meteorological observations

The year 2023 was selected for the evaluation, as it represents the most recent period with relatively complete station records across Greece. Out of all the data downloaded from the CDS, data from the year 2023 were used to validate AgERA5 against available station observations.

AgERA5 variables of daily 24h mean temperature and precipitation flux data were compared with ground-based station measurements after they were aggregated into monthly values, to identify potential biases, quantify uncertainties, and determine whether the dataset adequately represents the climatic conditions of Greece. To evaluate the performance of the AgERA5 dataset against observed meteorological station records across Greece for the year 2023, monthly mean temperature and cumulative monthly precipitation were compared. For the standardization of the units, some pre-processing was also needed, the parameter of 24h mean temperature was converted to degrees Celsius from Kelvin. For the analysis, the corresponding AgERA5 grid cell value was extracted for each of the 777 stations for each month of the year using nearest-neighbor interpolation, i.e. each station got the corresponding value from the nearest grid AgERA5 point.

The following statistical values are calculated:

- Bias, to assess systematic over- or underestimation (AgERA5 – Observed)
- Mean Absolute Error (MAE), to measure the average magnitude of error
- Root Mean Square Error (RMSE), to highlight typical error size with greater weight on large deviations
- Pearson Correlation Coefficient (r), to evaluate the ability of AgERA5 to reproduce the observed temporal variability

The statistical indicators mentioned, were estimated to assess accuracy and temporal consistency. The metrics were computed separately for temperature and precipitation, and results were summarized across stations using descriptive statistics and graphical visualizations (histograms, boxplots, and scatter plots).

The analysis provided valuable insights into the performance of AgERA5 under Greek climatic conditions and highlighted regional differences. Indicatively, results show good correspondence for air temperature with an example R^2 value of 0.7831 for January 2023 indicating a strong correlation between AgERA5 and observed data. For precipitation the agreement was notably weaker, with R^2 = 0.2905 for the same month.

Regarding the whole temperature dataset, mean bias was +0.45 °C, indicating that AgERA5 slightly overestimates observed temperatures on average. The MAE and RMSE were 2.06 °C and 2.44 °C, respectively. These values fall within the good to excellent range for temperature validation studies, suggesting that AgERA5 accurately reproduces both the spatial and temporal temperature variability across Greece. Despite the generally low errors, a few stations showed higher discrepancies (up to +18 °C), likely associated with local topographic effects or microclimatic conditions that are not fully captured by the coarser AgERA5 grid resolution (~10 km). Nevertheless, the small national mean bias and low overall error confirm that AgERA5 temperature data are acceptable for regional-scale modeling applications.

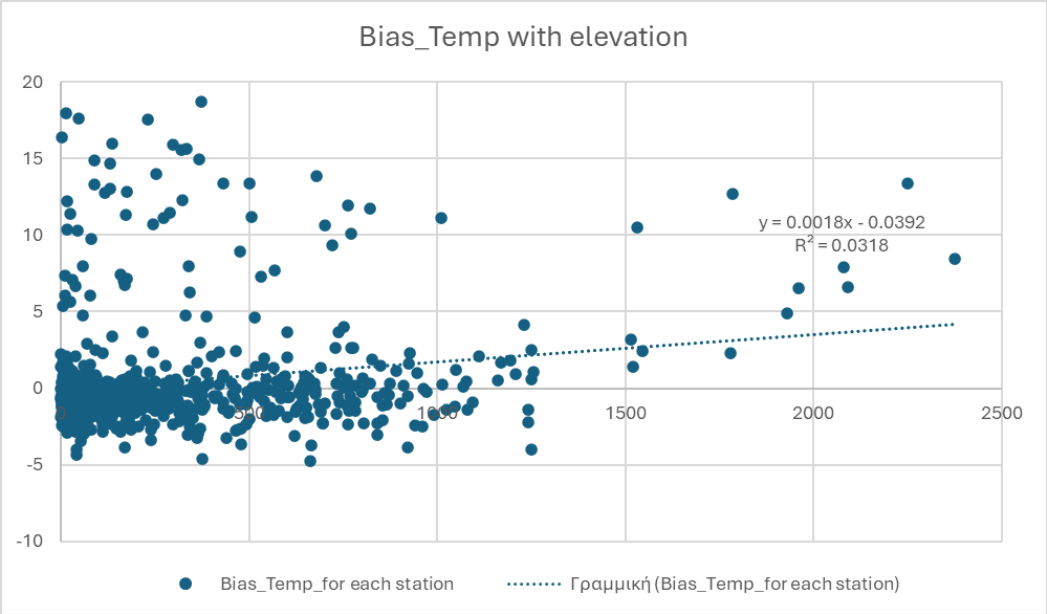

Figure 1: Relationship between temperature bias and elevation using the data of the 777 meteorological stations

The relationship between temperature bias and elevation (Figure 1) indicates that AgERA5 performs reasonably well at lower altitudes, where most stations show biases within ±5 °C. However, at higher elevations, the dataset tends to overestimate temperatures, with several stations showing increasingly positive biases above 1000 m and especially above 2000 m. This pattern reflects the smoothing of terrain in the reanalysis grid, which limits its ability to capture local mountain conditions. While AgERA5 provides reliable estimates for lowland areas, caution is required when applying it in high-altitude regions.

For precipitation, AgERA5 showed moderate accuracy, which is typical for reanalysis products in regions with complex terrain such as Greece. The mean bias was–-2.75 mm, indicating a minor underestimation of monthly totals. The mean MAE and RMSE were 34.94 mm and 52.01 mm, respectively, which reflect reasonable but variable agreement with the observed data. The relatively high standard deviations and extreme values (bias up to ±120 mm and RMSE up to 313 mm) suggest that local errors are significant at certain stations, particularly in mountainous and coastal areas where rainfall patterns are highly localized. Nonetheless, at the national level, AgERA5 captures the overall distribution and seasonality of precipitation. Results may be further improved through bias correction or local calibration using available station data.

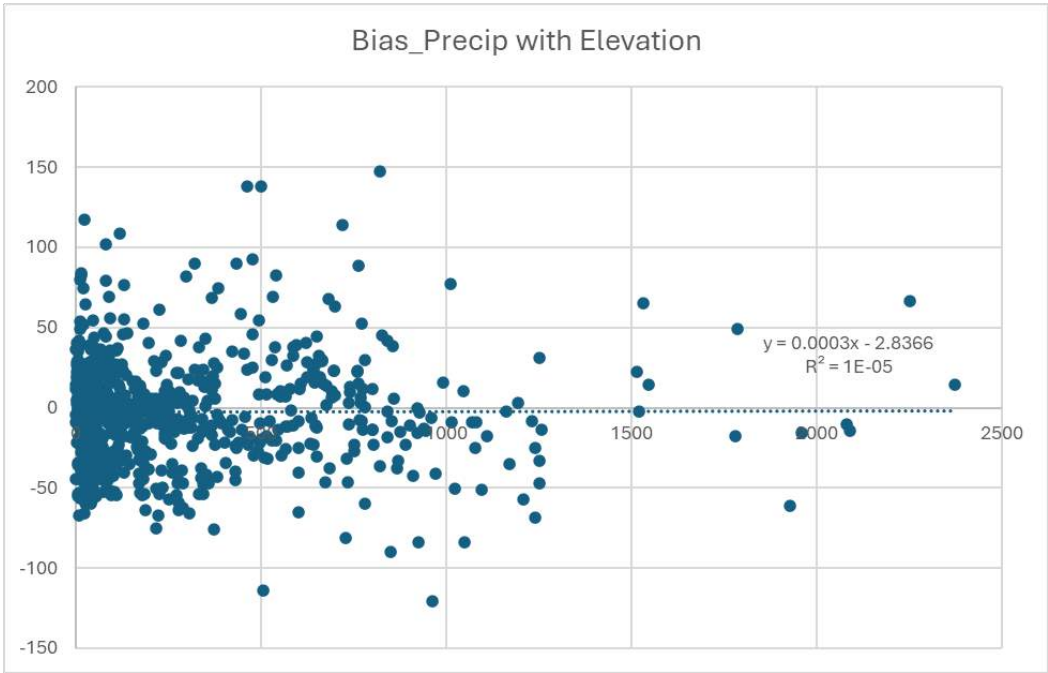

Figure 2: Relationship between precipitations bias and elevation using the data of the 777 meteorological stations

The analysis of precipitation bias as a function of elevation (Figure 2) showed an almost negligible relationship, with an R2 value on the order of 10^{-5}. This indicates that AgERA5 precipitation errors are not systematically influenced by station altitude. Instead, the observed discrepancies are likely related to other factors such as local microclimates, spatial resolution of the reanalysis grid, or the representation of convective and orographic rainfall processes.

Spatial evaluation of meteorological inputs based on station observations

As part of the data analysis techniques developed in WP5, a spatial evaluation was implemented to evaluate the reliability and representativeness of meteorological data used by DT-Agro. These techniques are intended to assess and document the quality of the input data and the methodological choices adopted for bias correction and spatial interpolation.

To characterize the spatial structure and seasonal variability of meteorological observations across Greece, monthly spatial diagnostics based on the coefficient of determination (R^2) were developed for air temperature (indicative results in Figures 3-6) and precipitation (indicative results in Figures 7-10) using ground-based meteorological station data from year 2023. The purpose of this analysis was to assess the internal spatial consistency of station observations and to support the development of gradient-based interpolation schemes for meteorological variables within DT-Agro.

Daily observations from the available meteorological station network were aggregated to monthly values for the year 2023. Separate analyses were performed for temperature and precipitation. Temperature typically exhibits smooth spatial gradients primarily controlled by elevation and regional-scale atmospheric processes, whereas precipitation displays stronger spatial heterogeneity due to orographic effects and local circulation patterns.

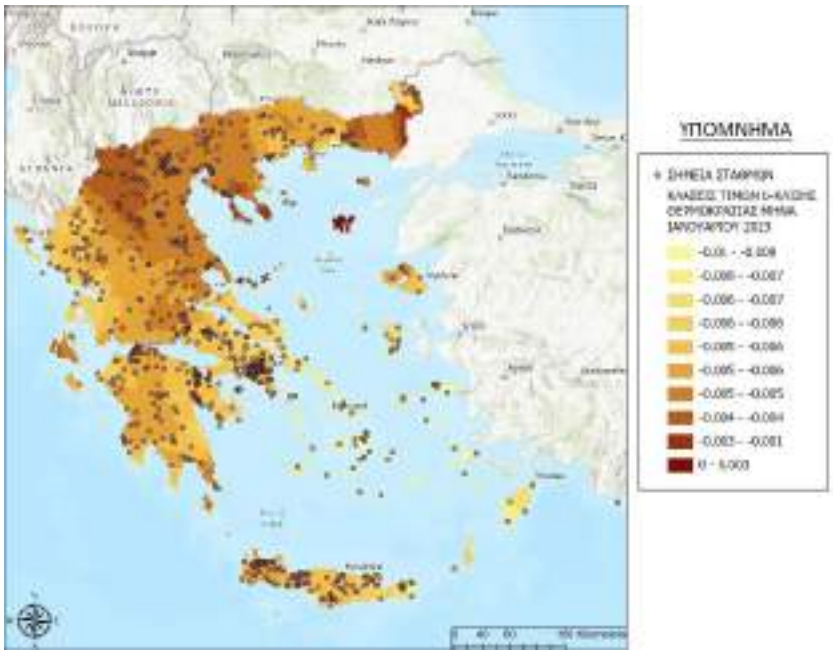

Figure 3: Monthly temperature gradient-class map for January 2023 derived from observations at 777 meteorological stations.

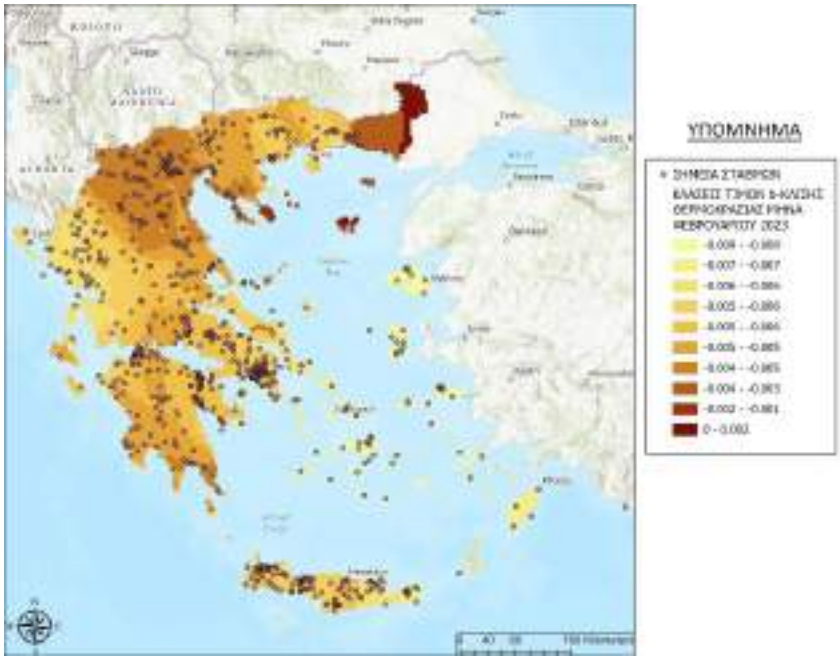

Figure 4: Monthly temperature gradient-class map for February 2023 derived from observations at 777 meteorological stations.

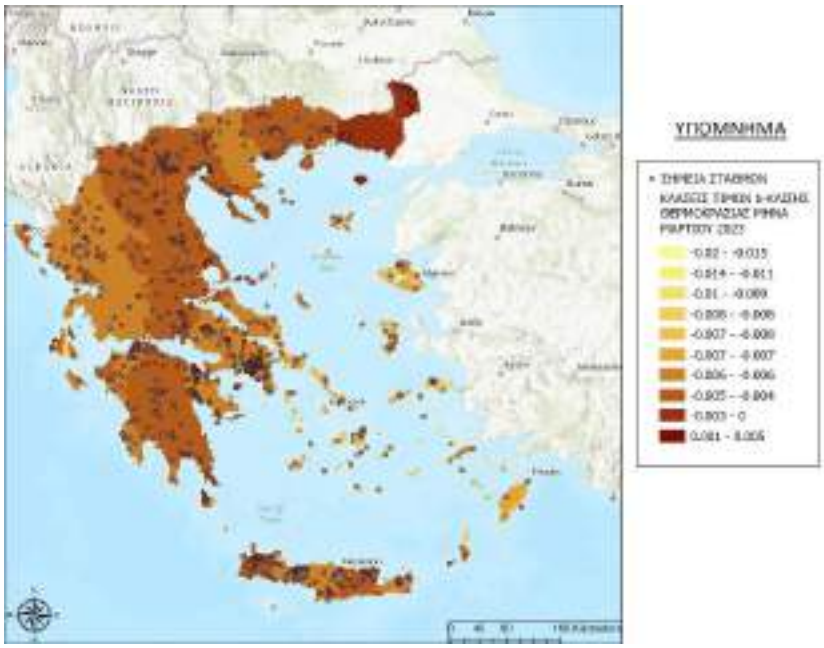

Figure 5: Monthly temperature gradient-class map for March 2023 derived from observations at 777 meteorological stations.

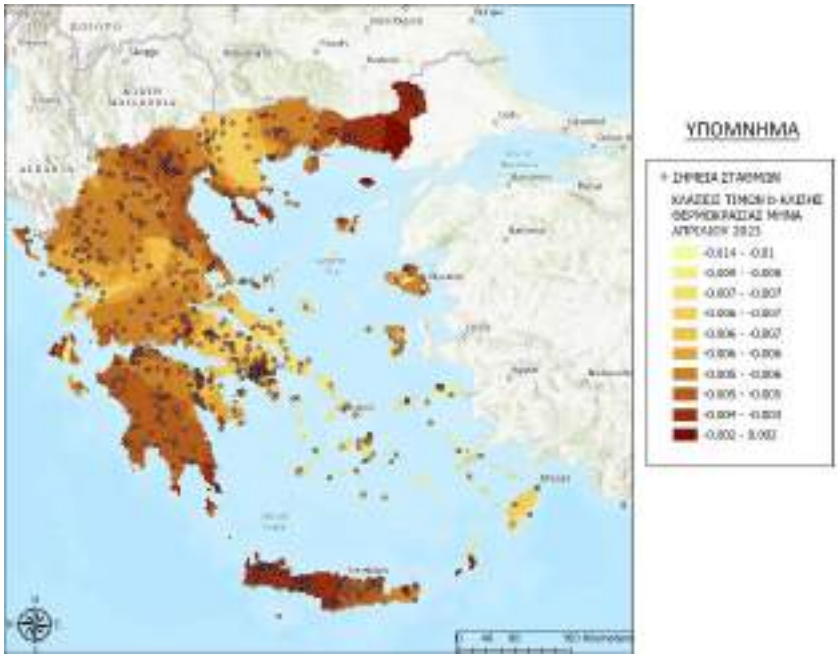

Figure 6: Monthly temperature gradient-class map for April 2023 derived from observations at 777 meteorological stations.

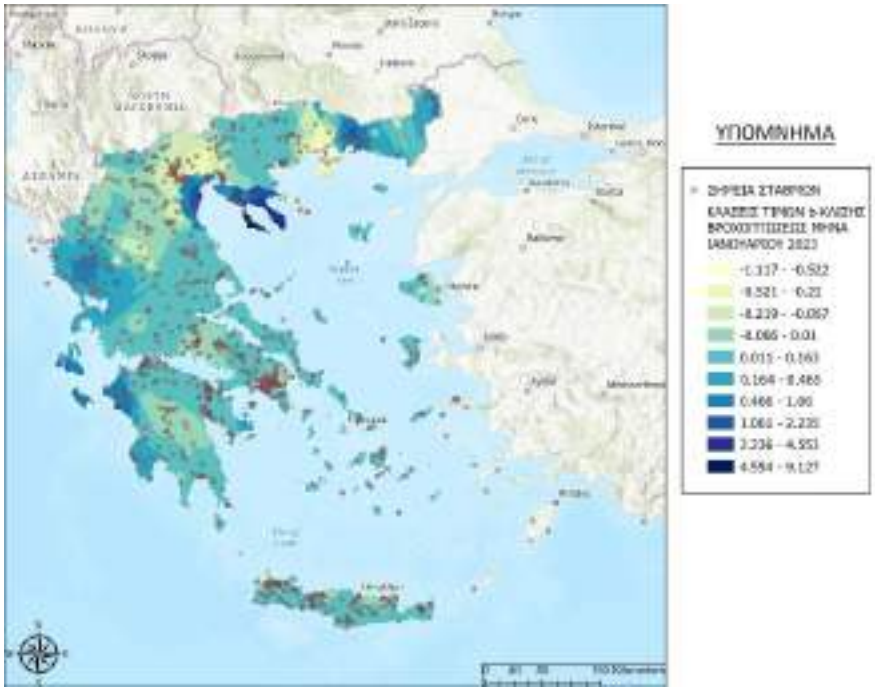

Figure 7: Monthly rainfall gradient-class map for January 2023 based on 777 meteorological stations.

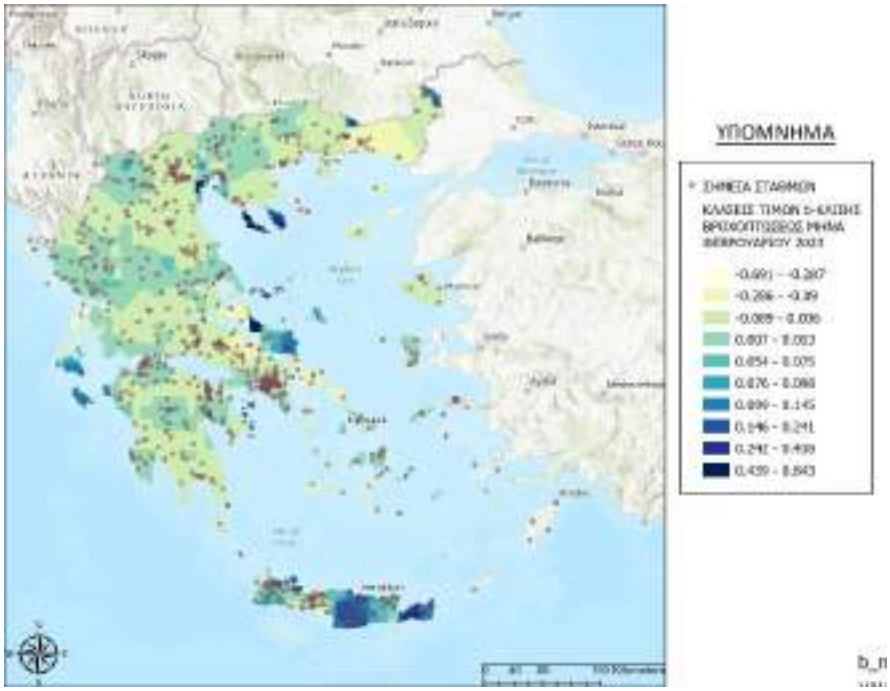

Figure 8: Monthly rainfall gradient-class map for February 2023 based on 777 meteorological stations.

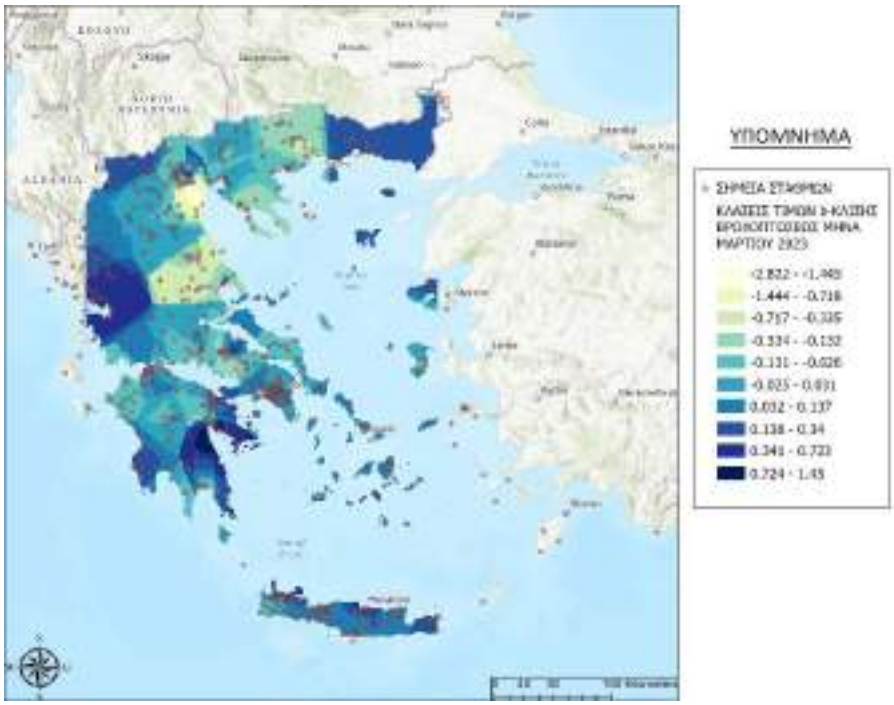

Figure 9: Monthly rainfall gradient-class map for March 2023 based on 777 meteorological stations.

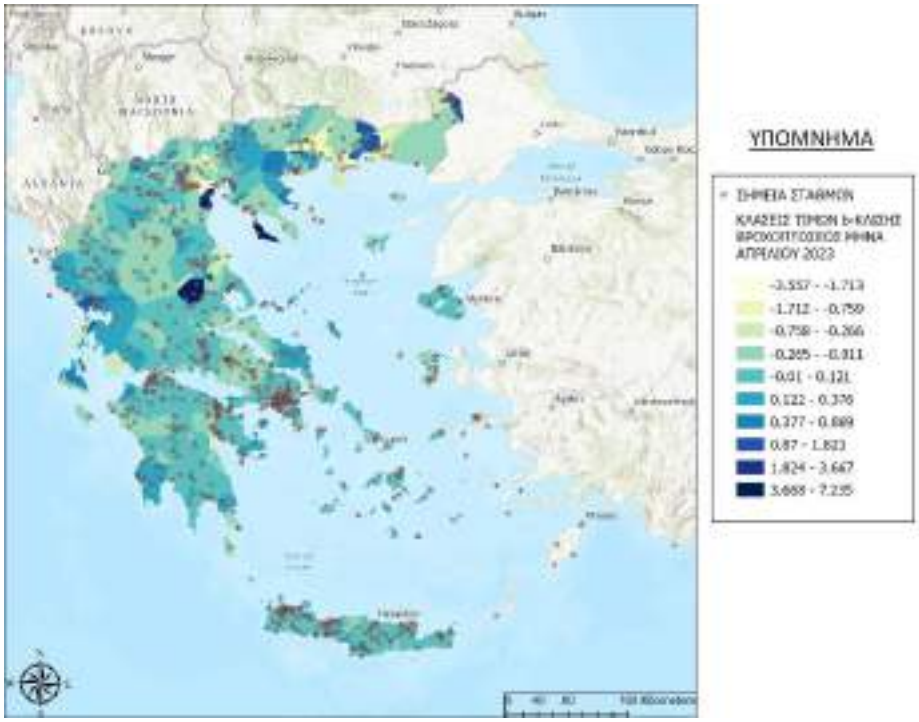

Figure 10: Monthly rainfall gradient-class map for April 2023 based on 777 meteorological stations.

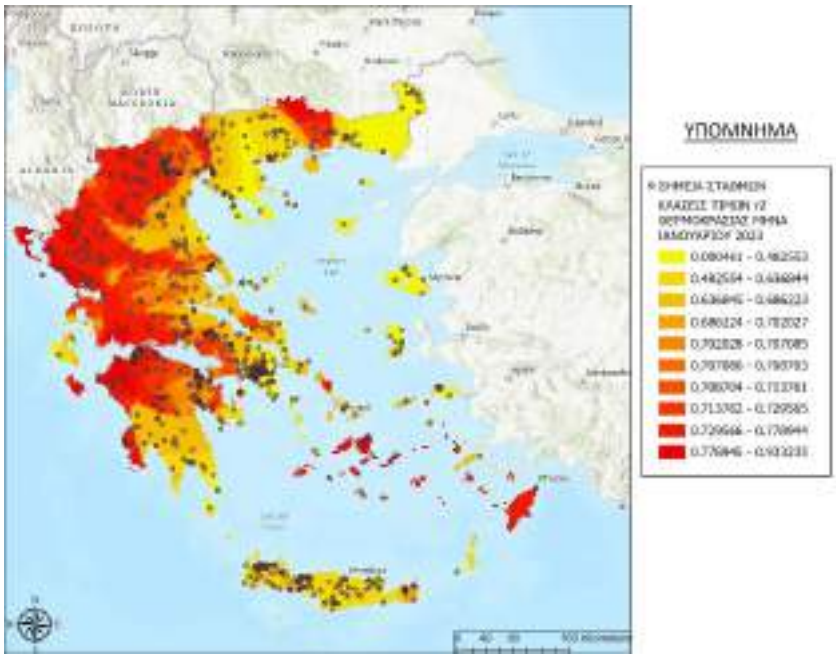

Figure 11: Monthly R^2 class map for air temperature in January 2023, derived from station observations. Classes summarize the spatial consistency of temperature variability for the selected month and support seasonal evaluation of interpolation performance.

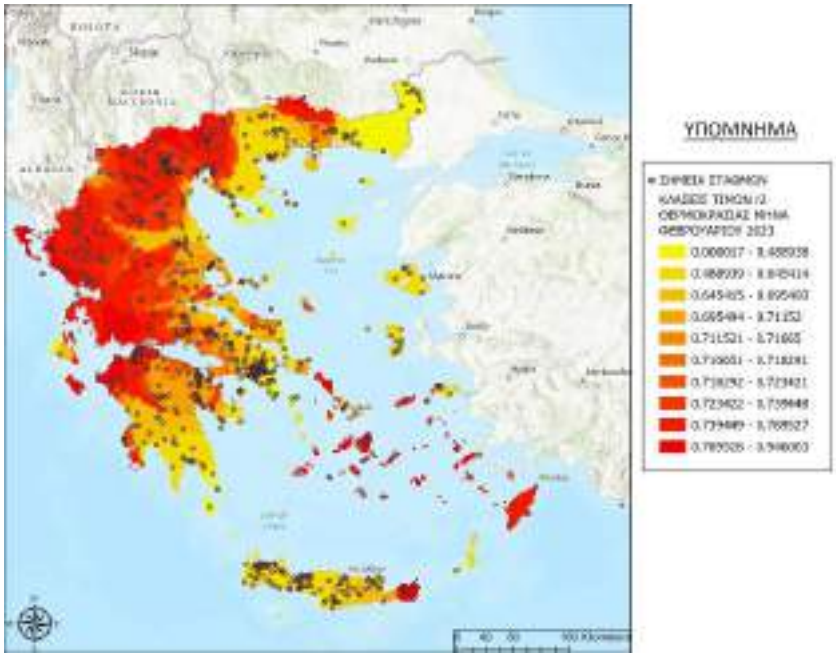

Figure 12: Monthly R^2 class map for air temperature in February 2023, derived from station observations. Classes summarize the spatial consistency of temperature variability for the selected month and support seasonal evaluation of interpolation performance.

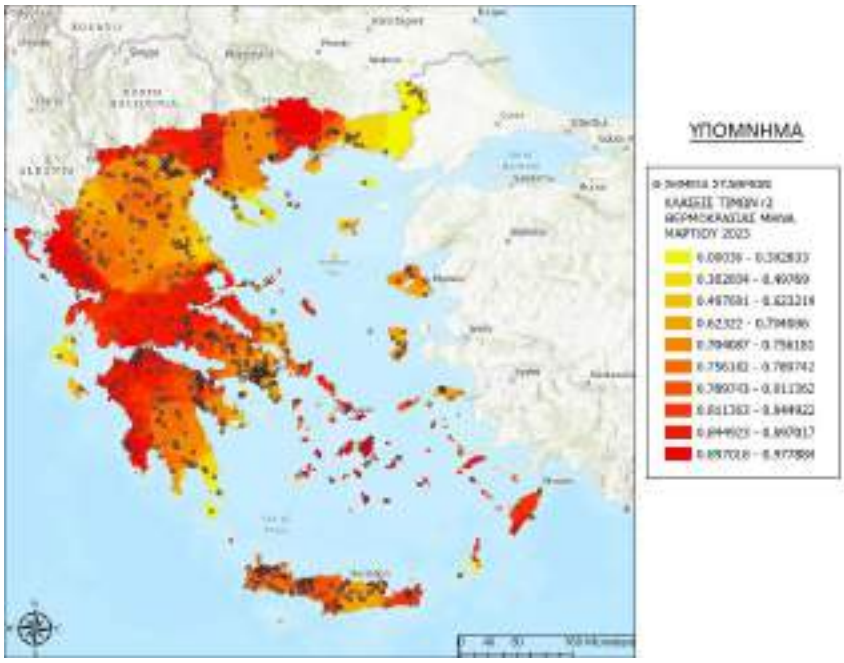

Figure 13: Monthly R^2 class map for air temperature in March 2023, derived from station observations. Classes summarize the spatial consistency of temperature variability for the selected month and support seasonal evaluation of interpolation performance.

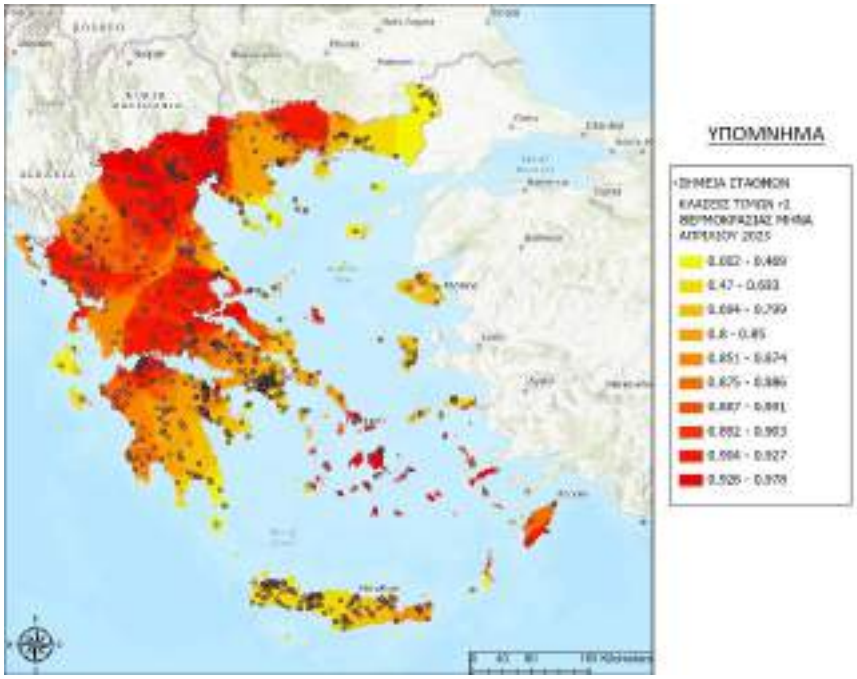

Figure 14: Monthly R^2 class map for air temperature in April 2023, derived from station observations. Classes summarize the spatial consistency of temperature variability for the selected month and support seasonal evaluation of interpolation performance.

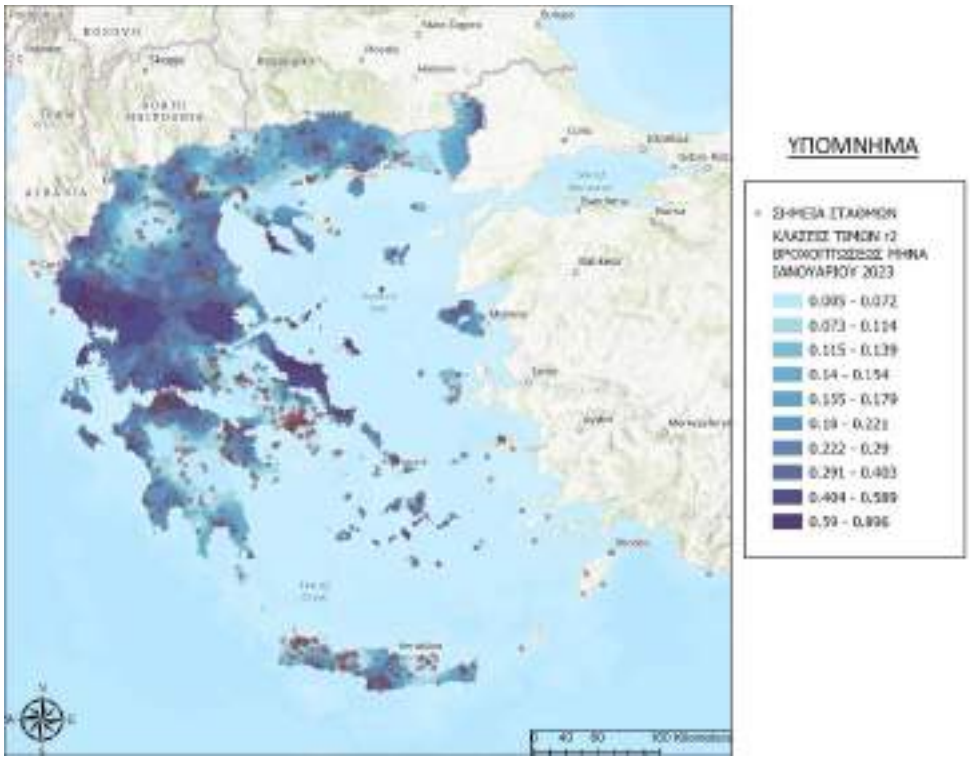

Figure 15: Monthly R^2 class map for precipitation in January 2023, based on station observations. The classification highlights seasonal differences in spatial coherence, reflecting the influence of synoptic versus local rainfall processes.

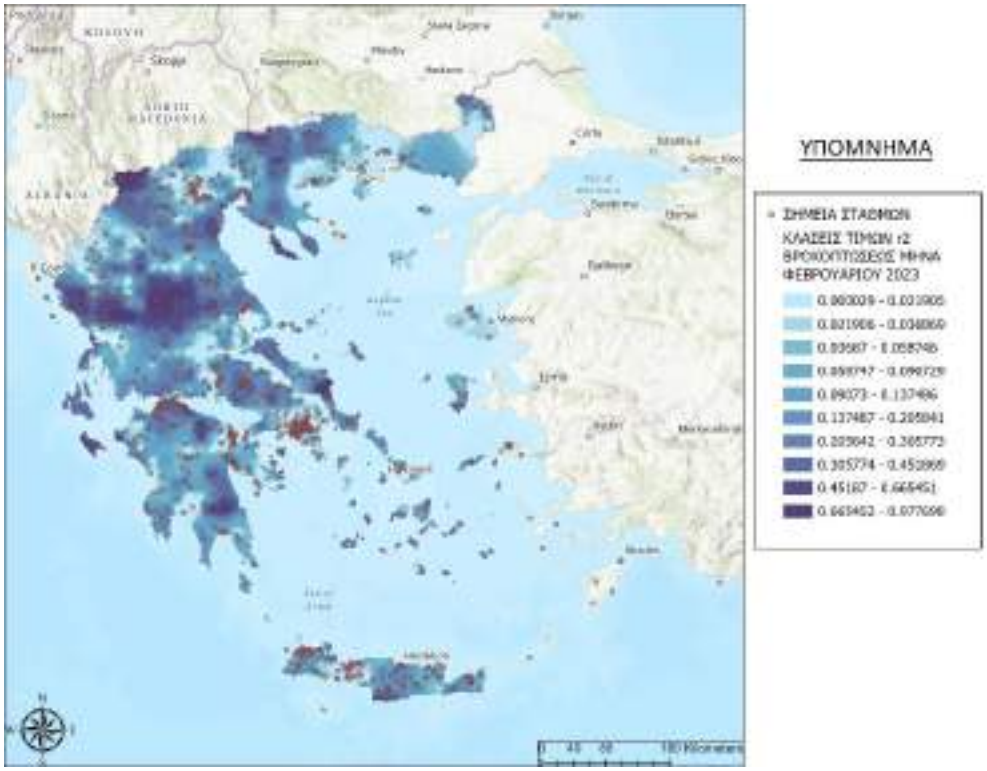

Figure 16: Monthly R^2 class map for precipitation in February 2023, based on station observations. The classification highlights seasonal differences in spatial coherence, reflecting the influence of synoptic versus local rainfall processes.

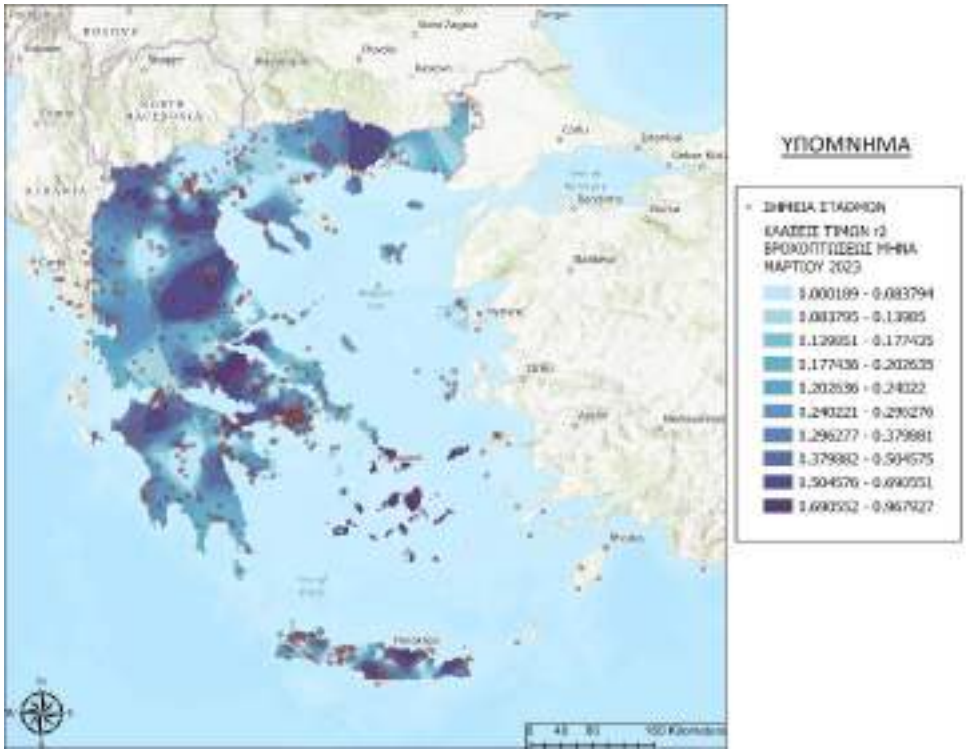

Figure 17: Monthly R^2 class map for precipitation in March 2023, based on station observations. The classification highlights seasonal differences in spatial coherence, reflecting the influence of synoptic versus local rainfall processes.

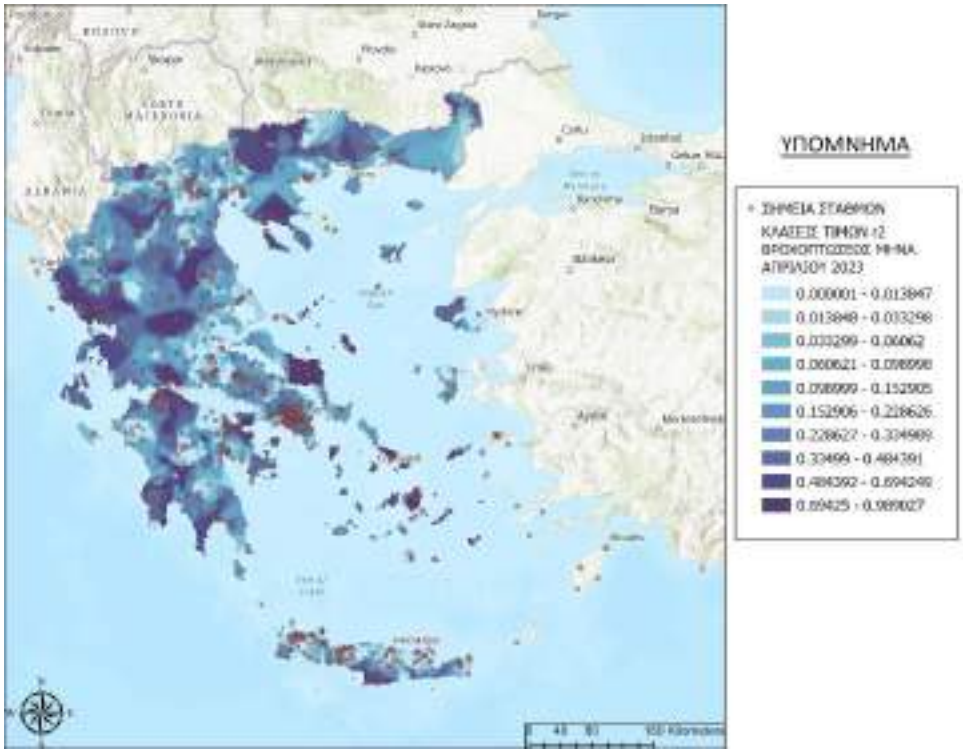

Figure 18: Monthly R^2 class map for precipitation in April 2023, based on station observations. The classification highlights seasonal differences in spatial coherence, reflecting the influence of synoptic versus local rainfall processes.

A moving-window approach was applied to quantify local spatial coherence in the station data. The window dimensions and moving step were first optimized by maximizing the determination coefficient (R^2) to ensure statistical robustness. Using these optimized parameters, we estimated the lapse rate and R^2 at each grid point of the study area. Subsequently, spatial interpolations were generated to create continuous maps of vertical gradients and their statistical reliability. This procedure was repeated systematically over the entire study area, resulting in spatially continuous monthly R^2 fields for both temperature and precipitation (Figure 19).

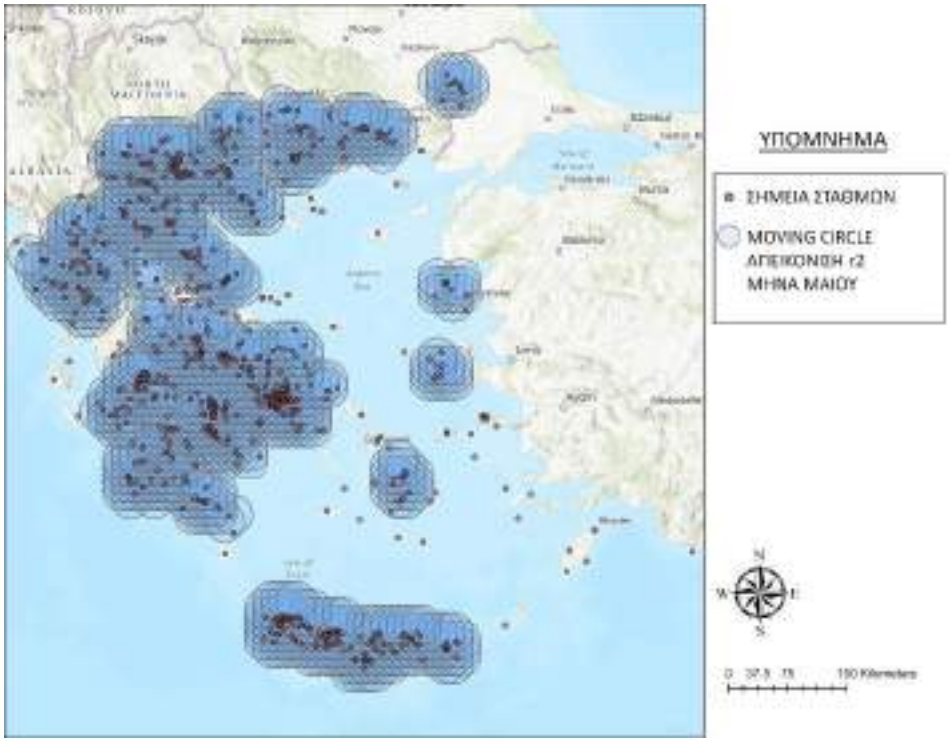

Figure 19: Spatial distribution of the local coefficient of determination (R^2) for May 2023, computed using a moving-circle approach on station observations. The map illustrates local spatial coherence of the variable and supports the assessment of neighborhood-scale variability.

The derived R^2 fields were subsequently classified into discrete classes to facilitate interpretation and comparison across space and time. These monthly R^2 classes highlight areas where station observations exhibit strong spatial coherence, as well as regions where local variability is high and spatial gradients are less well defined. Such information is particularly relevant in mountainous and coastal regions, where meteorological conditions can change sharply over short distances.

By repeating the analysis for each month, the methodology captures seasonal changes in the spatial consistency of meteorological observations. For precipitation, this allows the identification of periods dominated by widespread synoptic systems versus periods characterized by localized convective events. For temperature, the monthly R^2 diagnostics complement elevation-based gradient analyses by identifying months and regions where factors other than altitude, such as proximity to the sea or atmospheric stability, play a stronger role.

These monthly R^2 temperature and precipitation classes inform the selection and calibration of spatial interpolation approaches and provide a transparent assessment of the spatial representativeness of the station network.

2.2.3 Soil Data

Soil data were combined from various sources to cover the whole area of Greece so that CN values as well as the required hydraulic properties can be determined. The basic information used were data from the soil map of Greece (scale 1:30,000) (OPEKEPE, n.d.), which cover most of the cultivated areas. The study also made use of data sources like the European Soil Database (European Soil Database v2.0, scale 1:1,000,000) (Panagos et al., 2012) and the topsoil physical properties for Europe based on Land Use/Cover Area frame Survey (LUCAS) topsoil data (Panagos et al., 2012; Orgiazzi et al., 2018).

Efforts aim to improve global soil data so that they can be reliably applied within the model. These efforts are similar to the ones being undertaken for meteorological datasets, which also exhibit gaps in observational coverage. In both cases, the objective is to enhance the quality and completeness of global datasets to support robust modelling where local measurements are sparse or unavailable.

Soil data from three sources were compiled, which include the Greek Soil Map, the ISRIC SoilGrids global dataset at 250 m resolution and resources from the European Soil Data Centre (ESDAC). These datasets were evaluated and cross-compared to identify discrepancies in key parameters such as soil texture, bulk density, and organic carbon content. This comparative analysis formed the basis of a study, part of which was presented at EGU General Assembly 2025, Vienna, Austria, 27 Apr-2 May 2025 (Palli Gravani, S. et al., 2025). The full analysis conducted by the team, titled “Assessment of the accuracy of ISRIC and ESDAC soil texture data compared to the Soil Map of Greece: A statistical and spatial approach to identify sources of differences” is currently under review for publication.

The overall methodology was based on the processing, comparison, and evaluation of spatial and laboratory data related to soil texture. The process involved multiple stages, from data acquisition to statistical validation, ensuring robust evaluation and comparability across sources, aiming to assess the accuracy of international soil datasets for Greece.

Initially, raster datasets were collected from ISRIC SoilGrids, ESDAC, and the Greek national soil database. Each dataset was preprocessed by subsetting the topsoil layer (0–30 cm), clipping to the Greek territory, and projecting all rasters to the national coordinate system (EGSA87, EPSG:2100). Following preprocessing, values of key soil properties (sand, silt, and clay percentages) were extracted from the raster datasets for each sampling point. Differences between observed (Greek Soil Map) and predicted (ISRIC and ESDAC) values were then calculated to quantify errors for each soil property.

Subsequently, statistical analyses were performed to summarize and evaluate these differences, including univariate statistics and spatial variability assessments. Correlation between the values of soil properties were also calculated in the Soil Map of Greece and the corresponding predictions from the ISRIC and ESDAC database (Figure 20). The geographic distribution of absolute errors to identify regional patterns or clustering effects were also examined. To further interpret these results, soil texture classes were determined for each

sampling point using the USDA soil texture classification system. Predicted texture classes from the international datasets were then compared with those derived from the Greek Soil Map (Figure 21). The comparison employed categorical accuracy metrics, including producer's accuracy, user's accuracy, and overall accuracy, to assess the level of agreement between datasets.

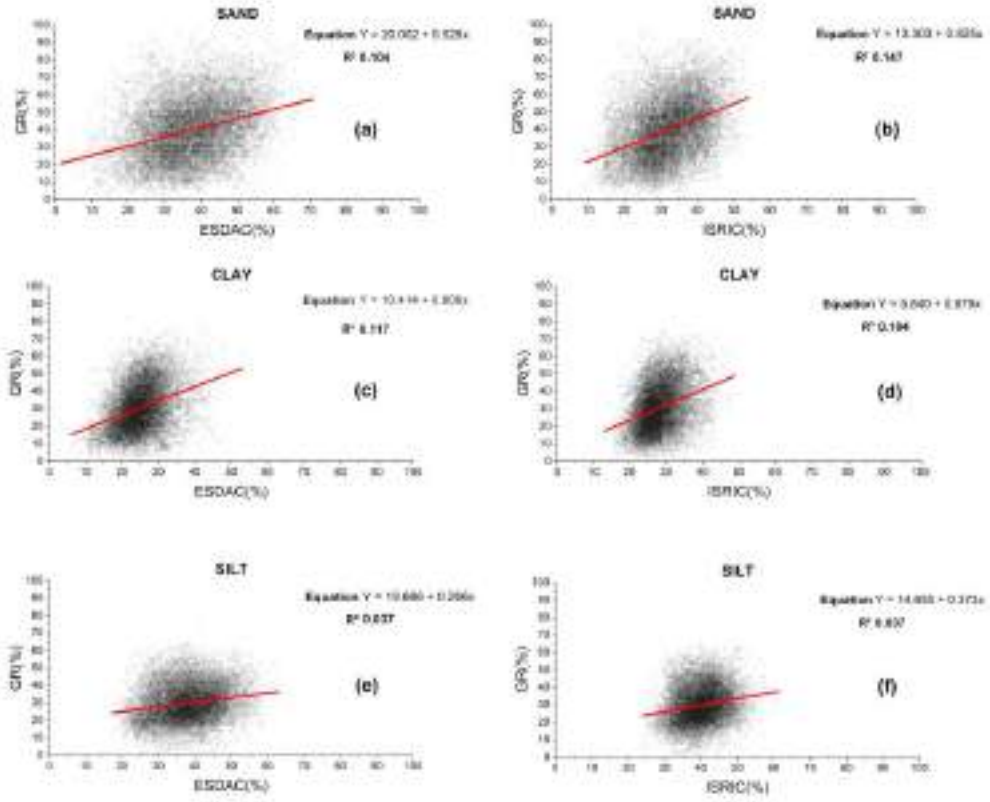

Figure 20: Scatterplots illustrating the correlation between the values of soil properties calculated in the Soil Map of Greece and the corresponding predictions from the ISRIC and ESDAC database. (a) Correlation of ESDAC sand content (%), (b) ISRIC sand content (%), (c) ESDAC clay content (%), (d) ISRIC clay content (%), (e) ESDAC silt content (%) and (f) ISRIC silt content (%).

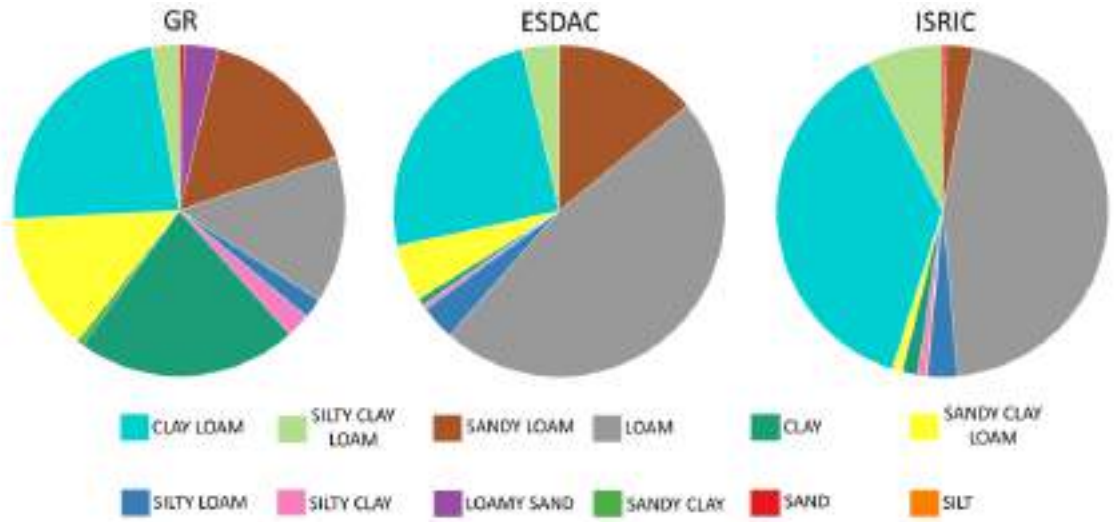

Figure 21: Distribution of soil texture classes based on the values of sand, clay, and silt fractions for each of the three datasets (GR, ESDAC, ISRIC).

To ensure the robustness of the findings, an alternative raster value extraction method was also tested, allowing for a comparison of residual errors and verification of the unbiasedness of the extraction process. Finally, validation of results was performed using supplementary Greek soil datasets, repeating the main analysis steps (data extraction, error calculation, and statistical evaluation) to confirm consistency and reliability across diverse sources.

Results showed significant discrepancies, demonstrating that their direct use for national-level assessments in a geologically diverse country like Greece can lead to inaccuracies. More specifically, ISRIC and ESDAC datasets failed to capture the full range of soil texture variability in Greece. The statistical comparison revealed very weak correlations and high errors between the global predictions and the measured ground-truth data. For all texture fractions, R^2 were low (e.g., ~0.15-0.19 for sand, ~0.12 for clay, and <0.04 for silt). The RMSE values were substantial, reaching up to 18.6% for sand, figures that are considerably higher than those reported in the original cross-validation studies for these global products. The analysis also showed that the statistical comparison revealed very weak correlations and high errors between the global predictions and the measured ground-truth data. For all texture fractions, R^2 were low (e.g., ~0.15-0.19 for sand, ~0.12 for clay, and <0.04 for silt). The RMSE values were substantial, reaching up to 18.6% for sand, figures that are considerably higher than those reported in the original cross-validation studies for these global products. The prediction errors are not entirely random but are spatially clustered in distinct hot and cold spots. Hot spots of high error were identified in specific regions, such as sandy coastal and island areas and clay-rich plains in central Greece. These discrepancies are strongly linked to parent material, as the global models, which utilize generalized geological maps, failed to accurately predict textures in soils derived from materials like dunes or clay deposits.

2.2.4 Copernicus Land Monitoring Service (CLMS) - Land Use/Land Cover Data

Land use and land cover (LULC) data from the CLMS were utilized in DT-Agro to characterize the spatial distribution of agricultural, urban, and forested areas across Greece. These datasets provide harmonized, high-resolution information on land surface conditions, which is essential for the hydrological and agro-environmental modeling processes of the DT-Agro.

The primary CLMS datasets used include the CORINE Land Cover (CLC) and the CLC Backbone, offering detailed land cover classifications that enable the identification of cropland, forest, grassland, and urban areas. These layers were complemented by the Imperviousness Density dataset, which represents the proportion of sealed surfaces (e.g., roads, buildings, pavements) within each grid cell. Imperviousness data are particularly critical for estimating surface runoff potential, as they directly affect infiltration capacity and hydrological response.

All LULC rasters were processed using ArcGIS Pro and Python to ensure spatial consistency and alignment with the other datasets. The steps include all Land Use and Land Cover (LULC) datasets, and more specifically the CORINE Land Cover (1990-2018) and CLCplus Backbone (2018, 2021, 2023) products, which were obtained from the European Union's Copernicus Land Monitoring Service (EEA/CLMS, 2020; 2023). To ensure full compatibility with other geospatial datasets used within the DT-Agro, all layers underwent systematic preprocessing and integration, as described below.

- **CORINE Land Cover (CLC)**

The CORINE Land Cover dataset, produced by the European Environment Agency (EEA), offers harmonized LULC information at a 100 m spatial resolution for Europe, with a 25-hectare minimum mapping unit. It has been released for multiple reference years (1990, 2000, 2006, 2012, 2018) (CORINE Land Cover 1990, 2000, 2006, 2012, 2018). CORINE layers (1990-2018) are used for historical analysis and land-use change scenario development

For DT-Agro, all available CORINE datasets were collected to provide a long-term temporal baseline of land cover evolution in Greece. This supports the analysis of land-use trends relevant to agricultural expansion, urbanization, deforestation, and other landscape changes that influence agro-hydrological conditions. Each CORINE layer was reprojected to EGSA87 (EPSG:2100), consistent with other spatial datasets used in the Digital Twin. The CORINE hierarchical classification system (Level 1-3, 44 classes) was preserved but aggregated when needed into broader categories relevant for hydrological modeling (e.g., urban, cropland, forest, grassland, bare land, water). The original CORINE classification (Levels 1-3) was retained in the database to allow flexible use across different analyses.

Each CORINE dataset was subjected to a standardized preprocessing workflow to ensure consistency with the DT-Agro spatial framework and compatibility with other datasets used in the Digital Twin. Initially, all CLC layers were clipped to the national extent of Greece and reprojected to the Greek Grid coordinate reference system (EPSG:2100). Although the

native CLC resolution is 100 m, consistency checks were performed to ensure exact alignment with the DT-Agro modeling grid.

For hydrological modeling purposes, CORINE classes were further aggregated into broader functional categories relevant to runoff and infiltration processes, such as urban and artificial surfaces, cropland, forest, grassland, bare or sparsely vegetated land, and water bodies. These aggregated classes were used to support the parameterization of the SCS-CN method, evapotranspiration calculations, and soil erosion modeling, while the detailed Level-3 information remains available for more refined analyses when required.

Overall, the preprocessed CORINE Land Cover datasets provide a consistent, spatially explicit representation of historical land use patterns in Greece and form a key input for long-term agro-hydrological analysis and scenario development within DT-Agro.

- **CLCplus Backbone**

To complement the CORINE dataset, the CLCplus Backbone was also employed, providing higher accuracy and more frequent updates. The CLC Backbone product is derived from Sentinel-2 time series analysis and offers a backbone layer for the future evolution of CORINE and national LULC mapping initiatives. The dataset provides consistent coverage for 2018, 2021, and 2023, at 10 m spatial resolution, which was resampled to 100 m for integration with the national model data. The 2023 CLCplus Backbone was used as the primary LULC layer for current model parameterization, while earlier years (2018, 2021) serve as references for detecting short-term land cover changes. A notable difference among the available years concerns the format and structure of the distributed data The 2023 CLCplus Backbone dataset was downloaded as multiple raster tiles from the WEkEO Copernicus platform (https://wekeo.copernicus.eu/) using an Area of Interest (AOI) covering the entire territory of Greece, while the 2018 and 2021 editions were distributed as single pan-European rasters.

All CLMS raster products were reprojected to Greek Grid (EPSG:2100) and resampled to a spatial resolution of 100 m to ensure compatibility with other DT-Agro datasets (meteorological, soil, DEM, etc.). The combination of LULC and imperviousness layers was used to support the Curve Number (CN) estimation methodology developed by (Soulis & Valiantzas, 2012). This integration allows Digital Twin to represent varying land surface characteristics dynamically and to simulate how land use changes, such as urban expansion or agricultural intensification affect runoff generation and soil moisture distribution.

For the 2023 dataset, all tiles covering the Greek territory were downloaded and merged into a seamless mosaic. The merging process was executed using an ArcPy script, which iterated through the tile directory and combined all rasters into a single dataset, ensuring complete national coverage while avoiding manual processing. Then, the 10 m raster was resampled to 100 m using the nearest-neighbor method to preserve categorical values, reprojected to the Greek Grid (EGSA87/EPSG:2100), and finally clipped to the exact national boundaries.

For the 2018 and 2021 datasets, preprocessing followed a simplified workflow. Each raster was reprojected to Greek Grid (EPSG:2100), resampled from 10 m to 100 m, and clipped to the Greek national boundary to ensure full spatial alignment with the 2023 product. The integration of these harmonized land-cover datasets with the imperviousness density layers from CLMS provides essential inputs for the CN estimation methodology based on Soulis & Valiantzas (2012). This combined framework enables the Digital Twin to represent spatial variability in land-surface characteristics more accurately and to simulate how land-use changes affect hydrological responses, runoff generation, and soil-moisture distribution across Greece.

- **NDVI**

The Normalized Difference Vegetation Index (NDVI) is a key indicator of vegetation health and biomass, derived from satellite imagery. For DT-Agro, NDVI provides critical insight into crop conditions, land cover monitoring, and agricultural planning. This report outlines the preprocessing steps applied to the Copernicus global NDVI dataset to generate a dataset suitable for analysis over Greece.

The source data consists of global NDVI GeoTIFFs provided by the CLMS. The dataset contains integer digital numbers (DN), representing NDVI values at 300 m spatial resolution with a 10-day temporal frequency. No-data values are typically encoded with the maximum integer (e.g., 255). The dataset that was downloaded is organized in multiple subfolders, each containing one or more raster tiles.

A preprocessing workflow was followed to convert to Float NDVI, to spatially clip to Greece and handle multiple files and subfolders. More specifically, raw integer values were converted to floating-point NDVI using the linear transformation: $NDVI_{float}=(DN\times0.004)-0.08$. After this transformation NDVI values range from -0.08 (bare soil) to 0.92 (dense vegetation). Original no-data pixels (DN ≥ 255) were masked and replaced with NaN to facilitate subsequent analysis. Then, the global NDVI values that were downloaded were clipped to the Greek territory using the coordinates NW (41.9202, 19.09755) and SE (34.55076, 29.54712). Clipping was implemented with rasterio.mask.mask() in Python, ensuring that output rasters retain spatial referencing (EPSG:4326). The preprocessing workflow supports, implemented in Python 3.13, batch processing for hundreds of files while maintaining consistent CRS and spatial resolution. The following libraries were used rasterio: reading, clipping, and writing GeoTIFFs. shapely: defining bounding box geometries. geopandas: optional, for handling shapefiles. numpy: numerical operations and masking. tqdm: progress tracking for large batch processing. To ensure compatibility with all other datasets all NDVI rasters were resampled from 300 m to 100 m spatial resolution.

These NDVI values are used to estimate crop coefficient K_c which represent the ratio of crop evapotranspiration (ET_c) to reference evapotranspiration (ET_o). K_c is critical for irrigation planning, water management, and crop growth monitoring. Several empirical relationships between NDVI and Kc have been proposed in the literature (e.g., Blanta et al., 2011;

Rozenstein et al., 2018). For the DT-Agro project, after reviewing existing models and testing multiple formulations, the relationship proposed by Montgomery et al. (2015) was selected as the most suitable. Indicatively, for cotton cultivation under conditions similar to those in Greece the following relationship is proposed:

$$K_c = 1.37 * NDVI – 0.086 \quad (1)$$

This equation provides realistic Kc values that closely match those recommended by FAO for the early stages of cotton crop development, while also capturing the temporal variability observed in satellite NDVI data. The use of the NDVI-K_c relationship allows for pixel-wise estimation of crop coefficients across Greece at 100 m resolution, enabling spatially explicit mapping of crop water requirements. This information supports DT-Agro's irrigation planning tools and contributes to the assessment of water use efficiency in cotton production systems.

Next, all the parameters needed directly and indirectly in the water balance equation (ETp, ETc) were calculated. Finally, the water balance model was applied with the only difference being that in this case the new values of actual evapotranspiration ET_c were used.

2.3 Hydrological Model

2.3.1 Runoff generation

Direct runoff is calculated using the Soil Conservation Service Curve Number (SCS–CN) method, which is widely applied for runoff estimation in agricultural watersheds. The method accounts for the combined effects of land use and land cover, soil properties and antecedent moisture conditions through a single parameter, the Curve Number (CN).

The initial abstraction ratio is set to the standard value of 0.2 so that CN remains the only calibration parameter of the method. To account for the effect of soil moisture conditions on runoff response, CN values are dynamically adjusted daily between dry and wet condition limits, depending on the simulated soil water content of the reference soil layer. This approach allows the runoff response to evolve continuously as soil moisture conditions change.

Accurate rainfall inputs are critical for this approach. For this reason, AgERA5 precipitation data were bias-corrected using observations from the HNMS prior to their use in runoff calculations. Bias correction is particularly important for SCS-CN applications, as small systematic errors in rainfall magnitude or frequency can lead to disproportionate errors in runoff estimates. Following bias correction, rainfall fields were spatially interpolated to generate continuous precipitation surfaces consistent with the model grid.

2.3.2 Soil hydraulic properties

Soil hydraulic conductivity at saturation (Ks), expressed in mm day^{-1}, is a fundamental parameter controlling infiltration and vertical water movement within the soil profile. In DT-Agro, Ks is spatially parameterized based on soil texture information, following pedotransfer approaches that relate hydraulic conductivity to sand, silt and clay fractions. This allows the representation of spatial variability in infiltration capacity across different soil types.

Soil moisture at saturation (θ_s) and soil moisture at field capacity (θ_{fc}) are also key variables of the soil water balance. The moisture at saturation represents the maximum water-holding capacity of the soil when all pores are filled with water, while field capacity defines the soil moisture content after gravitational drainage has ceased. Both parameters are spatially distributed and used to constrain soil water storage, percolation and plant-available water.

The shape factor b from the Brooks and Corey soil water retention model is used to describe the nonlinear relationship between soil water content and soil water potential. This parameter governs how soil moisture changes with pressure head and strongly influences unsaturated flow, deep percolation and soil moisture redistribution. The Brooks and Corey formulation provides a robust and computationally efficient framework for representing unsaturated hydraulic behavior within the Digital Twin.

2.3.3 Evapotranspiration

Evapotranspiration in DT-Agro is represented in reference, potential and crop evapotranspiration. Reference evapotranspiration (ET_o) is obtained from the AgERA5 dataset (mm day^{-1}) and describes the atmospheric demand for water. Potential evapotranspiration (ET_p) represents the maximum evapotranspiration under non-limiting soil moisture conditions. Crop evapotranspiration (ET_c) is calculated by adjusting ET_o using crop coefficients (K_c), which in DT-Agro are derived from NDVI. This allows crop water demand to vary in space and time according to vegetation conditions observed from EO data. Actual evapotranspiration is then limited by soil moisture availability, linking atmospheric demand with soil water conditions.

Crop evapotranspiration (ET_c) is calculated by adjusting ET_o using crop coefficients (K_c), which account for vegetation type, canopy development and phenological stage. In DT-Agro, K_c values are dynamically derived from NDVI, allowing temporal variability in vegetation condition to be captured directly from Earth Observation data. This approach enables the model to represent spatial and seasonal differences in crop water demand without relying solely on fixed, tabulated coefficients.

Actual evapotranspiration is subsequently constrained by soil moisture availability, linking atmospheric demand to soil hydraulic conditions and closing the soil water balance. The combined representation of ET_o, ET_p and NDVI-derived ET_c ensures a consistent and physically meaningful estimation of evapotranspiration across different land-cover and crop types.

2.3.4 Rainfall Data

Rainfall is a primary variable of the hydrological balance and is introduced into DT-Agro using daily precipitation data from the AgERA5 dataset. Prior to its use in the model, the AgERA5 precipitation data were subjected to a bias correction procedure based on ground-based observations provided by the HNMS. This step was implemented to reduce systematic deviations between reanalysis-derived precipitation and observed rainfall.

Bias correction is particularly critical for runoff estimation, as the SCS-Curve Number (SCS-CN) approach exhibits high sensitivity to precipitation inputs and initial abstraction thresholds. Following bias correction, precipitation inputs were spatially interpolated. The interpolation procedure allows point-based data to be transformed into spatially distributed rainfall surfaces, preserving spatial gradients and enabling their direct integration into the hydrological model. The interpolated and corrected rainfall data were then used as model inputs, interacting with soil hydraulic properties to control infiltration, surface runoff generation and soil water recharge, forming a key component of the water balance calculations.

2.3.5 Curve Number (CN) estimation

This is an important component of the data analysis, which considers the part of impervious surfaces on infiltration and surface runoff. Runoff estimation is based on the SCS-CN method and is extended in DT-Agro through explicit consideration of impervious surfaces.

The estimation of CN values for the model was performed using the simplified version of the Two-CN method (Soulis & Valiantzas, 2012), which incorporates the spatial distribution of impervious surfaces in each grid cell, where CN_a is set equal to 100, and CN_b characterizes the remaining area (pervious). In addition, the pervious CN is dynamically linked to simulated soil moisture to represent continuous antecedent moisture effects rather than discrete AMC classes. This method preserves compatibility with SCS-CN documentation and enables realistic representation of runoff sensitivity to even small impervious patches. This methodology integrates soil, land use/land cover, and imperviousness density information from CLMS to dynamically generate a spatially distributed CN for hydrological simulations within the digital twin.

Data used include Land use/Land cover from CORINE and CLC Backbone, Soil data from various sources (soil map of Greece (scale 1:30,000) (OPEKEPE, n.d.), European Soil Database (European Soil Database v2.0, scale 1:1,000,000) (Panagos et al., 2012) and the topsoil physical properties for Europe based on Land Use/Cover Area frame Survey (LUCAS) topsoil data (Panagos et al., 2012; Orgiazzi et al., 2018)), and Imperviousness Density from CLMS. Surface Soil Moisture (SSM) data from CLMS can be used to adjust CN values according to antecedent wetness conditions within DT-Agro.

With the simplified Two-CN method each grid cell can be represented by two fractions: an impervious part with CN_a = 100, corresponding to S_a = 0, and a pervious part with CNb estimated from land use and HSG combinations under AMC II conditions. Cells with imperviousness equal to or greater than 90% were considered fully impervious, and therefore directly assigned CN = 100.

All input raster datasets were clipped to the national extent of Greece, projected to the Greek Grid (EGSA87/EPSG:2100), and resampled to the common spatial resolution of 100 m to ensure consistency across datasets. CN values were initially assigned based on land use-soil group combinations and imperviousness density, following the simplified Two-CN

formulation. To facilitate calibration and improve computational efficiency, CN values were rounded and stored as integer values. This way, the calibration and adjustment of the parameter become computationally more efficient.

Overall, this CN estimation methodology integrates land cover, soil properties, imperviousness density, and soil moisture dynamics into a unified and spatially distributed framework. It enables DT-Agro to represent runoff processes realistically across a wide range of landscape conditions, from intensively cultivated agricultural areas to mixed rural–urban environments, supporting robust hydrological analysis within the Digital Twin.

3. Conclusions and Summary of Methodologies developed and Link to Upcoming Deliverables

The activities carried out under WP5 up to Month 18 have established the backbone for the DT-Agro. Through the integration and evaluation of multi-source environmental datasets, including meteorological, soil, land cover, and topographic information, the project has developed workflows for data processing, quality control, and model input generation.

These techniques not only enable the accurate representation of the Greek agro-hydrological system but also support the Digital Twin's capacity for dynamic updating, scenario simulation, and data-driven decision support. In the next phase (M18–M24), the focus will shift toward applying these methodologies for further analysis, incorporating climate variability and adaptation strategies, and initiating the design of spatially explicit digital services for agricultural and environmental management.

This deliverable (D5.1) has established a complete and reproducible set of data analysis methodologies supporting the development and operation of the DT-Agro. These methodologies address the acquisition, preprocessing, harmonization, and integration of Earth Observation (EO), meteorological, soil, land-cover, and topographic datasets, ensuring that all inputs required by the agro-hydrological model are spatially consistent, quality-controlled, and ready for analyses.

Comprehensive workflows were developed for processing Copernicus CLMS land use and land cover datasets, including all CORINE Land Cover (1990-2018) layers and the CLCplus Backbone datasets for 2018, 2021, and 2023. The 2023 CLCplus Backbone dataset, downloaded as tiled rasters from the WEkEO platform, required automated merging using ArcPy, followed by reprojection to EGSA87, resampling to 100 m, and extraction by the national boundary. The 2018 and 2021 products were harmonized through equivalent reprojection, resampling, and clipping procedures. Together with the CLMS imperviousness layers, these harmonized LULC datasets form the basis of the CN estimation methodology, which applies an adapted implementation of the simplified Two-CN method (Soulis & Valiantzas, 2012) to generate spatially distributed runoff parameters.

The meteorological data workflow was formalized through the adoption of the AgERA5 reanalysis dataset, downloaded through the CDS API, processed with Python, and converted from NetCDF to CSV. A full evaluation of AgERA5 temperature and precipitation for 2023 was carried out using bias, MAE, RMSE, and correlation coefficients, establishing a methodological foundation for both model validation and future operational updates of the Digital Twin.

For soil data, methodologies were developed to assess and compare soil properties from ISRIC SoilGrids, ESDAC datasets, and the Greek Soil Map. This included raster value extraction, statistical error quantification, USDA-based soil texture classification, spatial variability analysis, validation with independent Greek datasets, and targeted evaluation based on parent materials. These workflows ensure that soil variables used in the model are scientifically validated and suitable for spatially explicit hydrological applications.

Topographic processing workflows were also defined using the EU-DEM (30 m), which was resampled to 100 m and used to derive all hydrologically relevant terrain layers, including flow direction, flow accumulation, flow length, slope, travel time, and inland/overland flow velocities. These datasets constitute the terrain foundation for runoff routing, flow simulation, and water balance calculations within the Digital Twin.

Collectively, the methodologies developed in D5.1 establish a standardized, automated, and scalable data framework for DT-Agro. They directly support the next phases of the project: D5.2, which will apply these harmonized datasets to analyze historical and current agro-hydrological conditions, evaluate Digital Twin outputs, and assess climate and land-use scenarios. D5.3 will build upon these processed datasets and analytical results to design and test spatially explicit digital information services for end users.

By completing these methodological foundations, D5.1 fulfills the objectives of Task 5.1 and provides the technical groundwork upon which the analytical (D5.2) and service-oriented (D5.3) components of WP5 will be developed.

References

Blanta, A., Nicolas, D. R., Maliara, A., & Spyropoulos, N. (n.d.). Monitoring cotton crop evapotranspiration based on satellite data.

Boogaard, H., Schubert, J., De Wit, A., Lazebnik, J., Hutjes, R., Van der Grijn, G., (2020): Agrometeorological indicators from 1979 to present derived from reanalysis. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). DOI: 10.24381/cds.6c68c9bb (Accessed on 02-02-2025)

Copernicus Climate Change Service (2020): Agrometeorological indicators from 1979 to present derived from reanalysis. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). DOI: 10.24381/cds.6c68c9bb (Accessed on 02-02-2025)

CLCplus Backbone 2018 (raster 10 m) – version 2023_10. European Union's Copernicus Land Monitoring Service information,

https://sdi.eea.europa.eu/catalogue/copernicus/api/records/d32f1ee3-30c4-4052-9436-59a43b4c04dc. DOI: https://doi.org/10.2909/cd534ebf-f553-42f0-9ac1-62c1dc36d32c (Accessed on 10.03.2025).

CLCplus Backbone 2021 (raster 10 m) – version 2023_10. European Union's Copernicus Land Monitoring Service information, https://sdi.eea.europa.eu/catalogue/copernicus/api/records/23a5347f-65f0-478d-8b9d-70483fcfd699. DOI: https://doi.org/10.2909/71fc9d1b-479f-4da1-aa66-662a2fff2cf7 (Accessed on 10.03.2025).

CLCplus Backbone 2023 (raster 10 m) – version 2023_10. European Union's Copernicus Land Monitoring Service information, https://sdi.eea.europa.eu/catalogue/copernicus/api/records/d62ade7a-9c9c-405e-b88e-6ce0410d6a4f. DOI: https://doi.org/10.2909/b0bd43c6-1fa1-4d88-9c45-98b13a95d0b2 (Accessed on 10.03.2025).

CORINE Land Cover 1990 (raster 100 m) – version 2020_20u1. European Union's Copernicus Land Monitoring Service information, https://sdi.eea.europa.eu/catalogue/copernicus/api/records/c89324ef-7729-4477-9f1b-623f5f88eaa1?language=all DOI: https://doi.org/10.2909/c89324ef-7729-4477-9f1b-623f5f88eaa1 (Accessed on 10.03.2025).

CORINE Land Cover 2000 (raster 100 m) – version 2020_20u1. European Union's Copernicus Land Monitoring Service information, https://sdi.eea.europa.eu/catalogue/copernicus/api/records/ddacbd5e-068f-4e52-a596-d606e8de7f40?language=all DOI: https://doi.org/10.2909/ddacbd5e-068f-4e52-a596-d606e8de7f40 (Accessed on 10.03.2025).

CORINE Land Cover 2006 (raster 100 m) – version 2020_20u1. European Union's Copernicus Land Monitoring Service information, https://sdi.eea.europa.eu/catalogue/copernicus/api/records/08560441-2fd5-4eb9-bf4c-9ef16725726a?language=all DOI: https://doi.org/10.2909/08560441-2fd5-4eb9-bf4c-9ef16725726a (Accessed on 10.03.2025).

CORINE Land Cover 2012 (raster 100 m) – version 2020_20u1. European Union's Copernicus Land Monitoring Service information, https://sdi.eea.europa.eu/catalogue/copernicus/api/records/a84ae124-c5c5-4577-8e10-511bfe55cc0d?language=all DOI: https://doi.org/10.2909/a84ae124-c5c5-4577-8e10-511bfe55cc0d (Accessed on 10.03.2025).

CORINE Land Cover 2018 (raster 100 m) – version 2020_20u1. European Union's Copernicus Land Monitoring Service information, https://sdi.eea.europa.eu/catalogue/copernicus/api/records/71c95a07-e296-44fc-b22b-415f42acfdf0. DOI: https://doi.org/10.2909/960998c1-1870-4e82-8051-6485205ebbac (Accessed on 10.03.2025).

European Space Agency & European Union. (2019). *Copernicus DEM – Global and European Digital Elevation Model* (Version 2019_1) [Data set]. Copernicus Data Space Ecosystem. https://doi.org/10.5270/ESA-c5d3d65

OPEKEPE Greek Payment and Control Agency for Guidance and Guarantee Community Aid. Available online: https://www.opekepe.gr/opekepe-organisation-gr/opekepe-e-

services-gr/pliroforiaka-systimata/gewpliroforiako-systimaedafologikwn-dedomenwn (accessed on 08.01.2025).

Orgiazzi, A.; Ballabio, C.; Panagos, P.; Jones, A.; Fernández-Ugalde, O. LUCAS Soil, the largest expandable soil dataset for Europe: A review. Eur. J. Soil Sci. 2018, 69, 140–153.

Palli Gravani, S., Gerontidis, S., Kopanelis, D., Kairis, O., Soulis, K., and Kalivas, D.: Evaluation of digital maps of top-soil properties compared to large-scale laboratory soil data and synergies towards a better European soils' delineation, EGU General Assembly 2025, Vienna, Austria, 27 Apr–2 May 2025, EGU25-15613, https://doi.org/10.5194/egusphere-egu25-15613, 2025.

Panagos, P.; van Liedekerke, M.; Jones, A.; Montanarella, L. European Soil Data Centre: Response to European policy support and public data requirements. Land Use Policy 2012, 29, 329–338.

Rozenstein, O., Haymann, N., Kaplan, G., & Tanny, J. (2018). Estimating cotton water consumption using a time series of Sentinel-2 imagery. Agricultural Water Management, 207, 44–52. https://doi.org/10.1016/j.agwat.2018.05.017

Soulis, K., Dosiadis, E., Nikitakis, E., Charalambopoulos, I., Kairis, O., Katsogiannou, A., Palli Gravani, S., & Kalivas, D. (2025). Assessing AGERA5 and MERRA-2 Global climate datasets for Small-Scale Agricultural Applications. Atmosphere, 16(3), 263. https://doi.org/10.3390/atmos16030263

Soulis, K. X., & Valiantzas, J. D. (2012). SCS-CN parameter determination using rainfall-runoff data in heterogeneous watersheds – the two-CN system approach. Hydrology and Earth System Sciences, 16(3), 1001–1015. https://doi.org/10.5194/hess-16-1001-2012