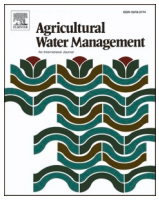
Agricultural Water Management 290 (2023) 108596

Contents lists available at ScienceDirect

Agricultural Water Management

journal homepage: www.elsevier.com/locate/agwat



Assessing the performance of the Gaussian Process Regression algorithm to fill gaps in the time-series of daily actual evapotranspiration of different crops in temperate and continental zones using ground and remotely sensed data

Dario De Caro [a], Matteo Ippolito [a,*], Marcella Cannarozzo [a], Giuseppe Provenzano [b], Giuseppe Ciraolo [a]

[a] Engineering Department, University of Palermo, Viale delle Scienze Ed. 8., 90128 Palermo, Italy
[b] Department of Agriculture, Food and Forest Sciences, University of Palermo, Viale delle Scienze Ed.4, 90128 Palermo, Italy

ARTICLE INFO

Handling Editor - Xiying Zhang

Keywords:
Daily actual evapotranspiration
Gap-filling
Gaussian Process Regression (GPR)
Agro-meteorological and remote sensed data
Sentinel-2
MODIS

ABSTRACT

The knowledge of crop evapotranspiration is crucial for several hydrological processes, including those related to the management of agricultural water sources. In particular, the estimations of actual evapotranspiration fluxes within fields are essential to managing irrigation strategies to save water and preserve water resources. Among the indirect methods to estimate actual evapotranspiration, ET_a, the eddy covariance (EC) method allows to acquire continuous measurement of latent heat flux (LE). However, the time series of EC measurements are sometimes characterized by a lack of data due to the sensors' malfunctions. At this aim, Machine Learning (ML) techniques could represent a powerful tool to fill possible gaps in the time series. In this paper, the ML technique was applied using the Gaussian Process Regression (GPR) algorithm to fill gaps in daily actual evapotranspiration. The technique was tested in six different plots, two in Italy, three in the United States of America, and one in Canada, with different crops and climatic conditions in order to consider the suitability of the ML model in various contexts. For each site, the climate variables were not the same, therefore, the performance of the method was investigated on the basis of the available information. Initially, a comparison of ground and reanalysis data, where both databases were available, and between two different satellite products, when both databases were available, have been conducted. Then, the GPR model was tested. The mean and the covariance functions were set by considering a database of climate variables, soil water status measurements, and remotely sensed vegetation indices. Then, five different combinations of variables were analyzed to verify the suitability of the ML approach when limited input data are available or when the weather variables are replaced with reanalysis data. Cross-validation was used to assess the performance of the procedure. The model performances were assessed based on the statistical indicators: Root Mean Square Error (RMSE), coefficient of determination (R^2), Mean Absolute Error (MAE), regression coefficient (b), and Nash-Sutcliffe efficiency coefficient (NSE). The quite high Nash Sutcliffe Efficiency (NSE) coefficient, and the root mean square error (RMSE) low values confirm the suitability of the proposed algorithm.

1. Introduction

The 2030 United Nations Agenda for Sustainable Development Goals (SDGs) supports action to bring good and healthy living conditions for the world population. Agriculture 4.0 is the last advancement in farming technology, based on four main pillars: increasing productivity, sustainable use of natural resources, resilience to climate change, and reducing food waste (Zhai et al., 2020). Regarding the sustainable use of natural resources, optimization of irrigation systems plays a key role because agriculture uses around 70% of the global freshwater consumption (United Nations, 2021). To improve water use in agriculture, political actions and institutional interventions should be finalized to regulate the pricing policies from one side and, from the other, farmers should improve water use efficiency for food production through

* Corresponding author.
E-mail address: matteo.ippolito@unipa.it (M. Ippolito).

https://doi.org/10.1016/j.agwat.2023.108596
Received 3 July 2023; Received in revised form 31 October 2023; Accepted 13 November 2023
Available online 18 November 2023

management optimization which depends, among the other factors, from the correct estimation of crop water requirement. According to the FAO-56 model (Allen et al., 1998), crop water requirements depend on actual crop evapotranspiration (ET_a), which is composed of two different terms: actual plant transpiration (T_a) and soil evaporation (E_a). For a given crop type and phenological stage, ET_a is governed by the atmospheric evaporative demand and soil water availability. Agro-hydrological models based on the soil water balance, SWB, (Van Dam et al., 1997; Marletto et al., 2007; Pereira et al., 2020) or surface energy balance models, SEB (Bastiaanssen et al., 1998; Awada et al., 2019) have been extensively applied to assess actual crop evapotranspiration. Indirect estimations of actual crop evapotranspiration, ET_a, can be obtained, according to the single crop coefficient approach (Allen et al., 2008), by:

$$ET_a = ETo\ K_c\ K_s \tag{1}$$

in which ETo is the crop reference evapotranspiration, K_c is the crop coefficient accounting for the crop phenological stage and, K_s is the water stress coefficient depending on soil water availability.

Even if several models based on the combination of the available climate variables have been proposed to estimate ETo, the Penman-Monteith equation has been largely applied when standard climatological records (air temperature, solar radiation, wind speed, and relative air humidity) are available (Allen et al., 1998). In some regions of the world where weather data are unavailable and/or poor, global atmospheric models are considered a valid alternative to compensate for the absence or improve the quality of data (Rodrigues and Braga, 2021). ERA5 and ERA5-Land (ERA5-L), provided by the European Centre for Medium-Range Weather Forecasts (ECMWF), are the last generation of the global reanalysis climate database. Specifically, in the ERA5-L the simulated land fields are controlled by a process of atmospheric forcing using air temperature (T_{air}), air humidity and air pressure. The values of these variables are corrected considering the altitude difference between the grid of the forcing and the higher resolution grid of ERA5-L (Muñoz Sabater, 2019). In a recent review presented by Muñoz Sabater et al. (2021), it is possible to notice the state-of-the-art associated with the use of ERA5-L for land and environmental applications. A climate reanalysis generates large datasets and gives a numerical description of the recent climate produced by combining models with observations. It contains estimates of atmospheric parameters for all locations on Earth and for a long time period. The ERA5 and ERA5-L databases are freely downloadable from the internet (https://www.copernicus.eu/en) and are characterized by a spatial resolution of 0.25° latitude by 0.25° longitude and 0.10° latitude and 0.10° longitude, respectively and a temporal coverage from 1950 up to five days before the real-time. In a recent study carried out in Sicily, Pelosi and Chirico (2021) indicated that the joint use of ERA5-Land reanalysis and CM-SAF satellite-based radiation data is suitable to assess crop reference evapotranspiration, ETo, if ground data were not available. For Continental Portugal, Paredes et al. (2018) used daily meteorological observations to assess the performance of the ERA-Interim weather variables when compared to ground observations and to evaluate the performance of computing daily ETo with reanalysis data. The values of crop coefficients K_c (1), for different growth stages and management are tabulated in the original FAO-56 manual or more recent publications accounting for the research carried out in the last two decades (Pereira et al., 2021; Rallo et al., 2021). To avoid the limitations associated with the tabular values, satellite images can represent a valid alternative to monitor the actual K_c remotely and in a precise way. Pôças et al. (2020) reported an exhaustive list of different K_c–Vegetation Indices (VIs) relationships, valid for a variety of crops, retrieved based on satellite images acquired by different sensors which provide information in the visible (VIS), near-infrared (NIR) and shortwave infrared (SWIR) regions.

The value of the stress coefficient K_s (1), can be determined based on the application of the water balance model, using different functions related to the soil water content (Allen et al., 1998). Therefore, the joint use of reanalysis and remote sensing data can represent a powerful database to retrieve information on climate variables and crop characteristics with spatial and temporal resolutions suitable for practical applications.

On the other hand, direct estimations of ET_a can be obtained, at a larger spatial scale, with micrometeorological techniques, such as the eddy covariance (EC) system. This technique, largely used in various agroforestry systems, allows to measure the vertical turbulent fluxes within the atmospheric boundary layers (Cammalleri et al., 2013; Anderson et al., 2017; Rozenstein et al., 2019; Corbari et al., 2020; French et al., 2020; Saitta et al., 2020). However, the results achievable depend on the quality and the continuity of the available energy fluxes time series. The occurrence of acquisition gaps of one or more sensors compromises the continuity of actual evapotranspiration records; anomalies are generally due to failure, improper installation of the sensors or incorrect calibration, as well as the poor maintenance of sensors or data logger. Moreover, the occurrence of outliers in time series can also compromise the output quality and could generate inconsistent results mainly when model simulations are applied.

Machine learning (ML) techniques have recently been proposed as suitable methods for time series gap-filling. ML algorithms have already been developed for the estimation of crop reference evapotranspiration (Jing et al., 2019; Kim et al., 2020; Krishnashetty et al., 2021), several authors have been investigating the suitability of ML models to predict directly ET_a since, in most cases, they disposed of long temporal series of data for training and validating the models (Granata et al., 2019; Walls et al., 2020; Talib et al., 2021; Mosre and Suárez, 2021). In particular, Rasmussen and Williams (2006) demonstrated that the ML model GPR (Gaussian Process Regression) can allow to retrieval of synthetic temporal series even if the availability of training data is limited. The GPR is a non-parametric model based on statistical Bayesian theory that can be used to solve supervised multivariate regression and classification issues such as high-dimensional, small-sample, and nonlinear problems (Rasmussen et al., 2006; Murphy, 2012).

Carter and Liang (2019) and Mosre and Suárez (2021) demonstrate that the quality of the results of ML algorithms application can be improved if the input variables include vegetation indices VIs as NDVI, NDWI as well as EVI.

The Normalized Difference Vegetation Index, NDVI (Rouse et al., 1974) is based on the reflectance in the regions of visible (VIS) and near-infrared (NIR), and the Normalized Difference Water Index, NDWI (Gao, 1996) consider the reflectance in the regions of NIR and shortwave (SWIR):

$$NDVI = \frac{\rho_{nir} - \rho_{red}}{\rho_{nir} + \rho_{red}} \tag{2}$$

$$NDWI = \frac{\rho_{nir} - \rho_{swir}}{\rho_{nir} + \rho_{swir}} \tag{3}$$

where ρ_{nir}, ρ_{red} and ρ_{swir} are the near-infrared, red and shortwave reflectance, respectively. According to Rouse et al. (1974), the authors retrieved the near-infrared and red reflectance values in band 5, from 0.6 to 0.7 μm, and in band 7, from 0.8 to 1.1 μm, respectively, using ERTS-1 MSS data. On the other hand, Gao (1996) used reflectance at 0.86 μm for near-infrared and at 1.24 μm for shortwave.

As shown by Pagano et al. (2023), the existing literature focused on the prediction of actual evapotranspiration ET_a, by means ML algorithms, is very limited. Moreover, many of the proposed models are site-specific and the replicability of the algorithm on different crops and climatic conditions is not yet addressed. Instead, in this paper, the performances of the proposed ML algorithm to predict actual evapotranspiration are validated on different crops, in different climatic conditions and considering different input data sources. Taking into account the reasons presented above, the objective of this study was to assess the suitability and the performance of the GPR algorithm to fill

gaps in the time-series of daily actual evapotranspiration (ET_a) of different crops acquired in temperate e continental zones. To this purpose, data, provided from different sources, were considered as input variables, like combinations of ground and remote sensing data, and reanalysis data when necessary. That's why a large part of the work consisted in verifying the suitability of the climate variables, comparing ground and reanalysis data in those experimental fields where both databases were available. Even data retrieved from two different satellites were compared when both databases were available.

2. Materials and methods

The performances of the GPR algorithm have been tested on six experimental fields: two in Italy, three in the United States of America, and one in Canada, with different crops and climatic conditions in order to consider the suitability of the ML model in various contexts. A comparison of ground and reanalysis data have been conducted where both databases were available and for one of the experimental filed two satellite databases were compared since contemporary Sentinel-2 and MODIS images were available. This last comparison has been conducted in order to verify the suitability of the MODIS product, which is characterized by a lower spatial resolution than Sentinel-2 product.

2.1. Experimental fields

The first experimental field is a Mediterranean citrus orchard (*Citrus reticulata Blanco,* cv. Mandarino Tardivo di Ciaculli) of about 0.4 ha, located near Palermo, Italy (38° 4' 53.4'' N, 13° 25' 8.2'' E). Fig. 1 shows the map of the field with the position of the sensors and tools installed on the ground. The experimental activities were carried out from March 2019 to September 2021. More details regarding the field and crop characteristics are reported in Ippolito et al. (2022).

The second experimental field is an olive orchard (cv. "Nocellara del Belice") extended about 6 ha (Fig. 2). The field, located in the South-West of Sicily, Italy (37°38'61'' N, 12°50'53'' E), about 5 km far from the town of Castelvetrano, is part of a larger irrigation district characterized by a flat landscape and rather homogeneous soil and crop types. The experimental activities were carried out over three years, from January 2009 to December 2011. More details regarding the field and crop characteristics are reported in Rallo and Provenzano (2013).

Both fields, citrus and olive orchard, are characterized by a *Hot summer Mediterranean climate (Csa)* (Kottek et al., 2006), according to the last version of the Köppen climatic classification, with rainfall concentrated in fall and winter and quite hot and dry summer.

The other four sites, with different crops and climatic classifications, are reported in Fig. 3. The US_ARM is situated near Billings, Oklahoma, USA (36° 36' 20.9''N, 97° 29' 19.59'' O). The main crop grown is hard red winter wheat ("common wheat," Triticum aestivum L., main var. Jaeger, Jagalene, and Fuller) with infrequent crop rotations (Fischer et al., 2007; Raz-Yaseef et al., 2015). The experimental activities were carried out from January 2005 to December 2010. The Köppen climatic classification (Kottek et al., 2006) is *Humid subtropical climate (Cfa)*.

The fourth site (US_Ne1) is located near Mead, Nebraska, USA (41° 09' 54.56''N, 96° 28' 35.50 O). This field is irrigated with a center pivot system. The site has ten years history of maize-soybean rotation under no-till (Suyker, 2022). The experimental activities were carried out from January 2001 to December 2020. The Köppen climatic classification (Kottek et al., 2006) is *Hot summer humid continental climate* (*Dfa)*.

The fifth site (US_CF1) is near Pullman, Washington, USA (46° 46' 53.53''N, 117° 4' 55.50'' O). The predominant crops are winter wheat, spring cereals (barley and wheat), and pulse crops (dry pea, lentil, and chickpea). The experimental activities were carried out from January 2018 to December 2020. According to the Köppen climatic classification (Kottek et al., 2006), the climate is *Warm, dry summer continental climate (Dsb)*.

The sixth and last site (CA_ER1) is near Guelph, Ontario, Canada (43° 38' 26.17''N, 80° 24' 44.35 O). This field is an agricultural site where corn is cultivated. The site is relatively flat and homogenous. The experimental activities were carried out from January 2016 to December 2019. The Köppen climatic classification (Kottek et al., 2006) is *Warm summer humid continental climate (Dfb)*.

Table 1 summarizes the crop types, the position, and the climatic zone of the experimental fields considered, as well as the analysis period.

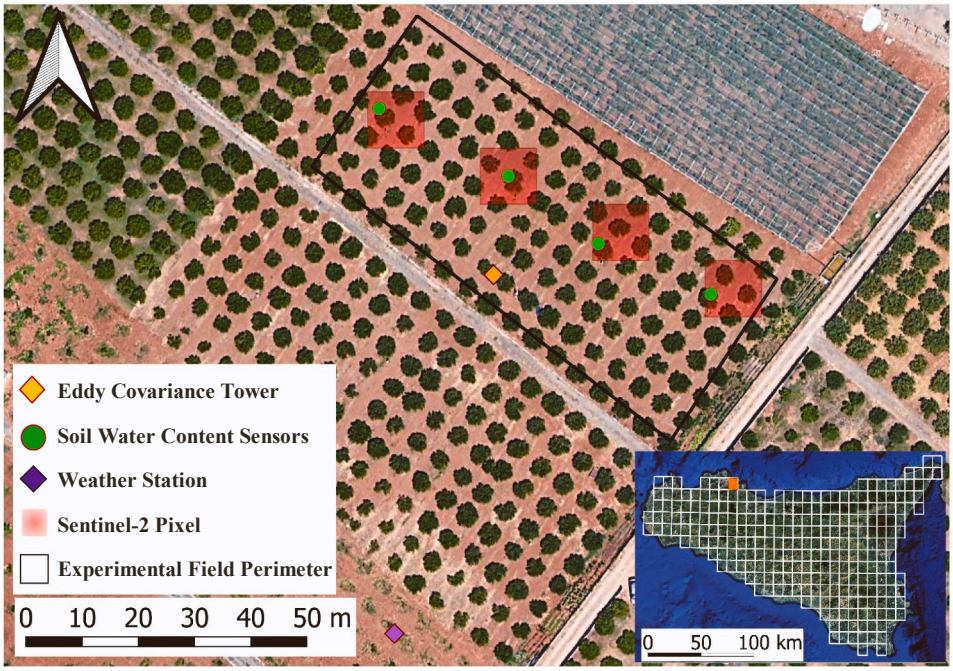


Fig. 1. Map of the citrus orchard with the position of Eddy Covariance tower (EC), soil water content sensors with the corresponding Sentinel-2 (L2A/L2B) pixels and WatchDog 2000 weather station (WD). The bottom right box shows Sicily island with the ERA5-Land grid (white) and the rectangle containing the experimental field (orange).

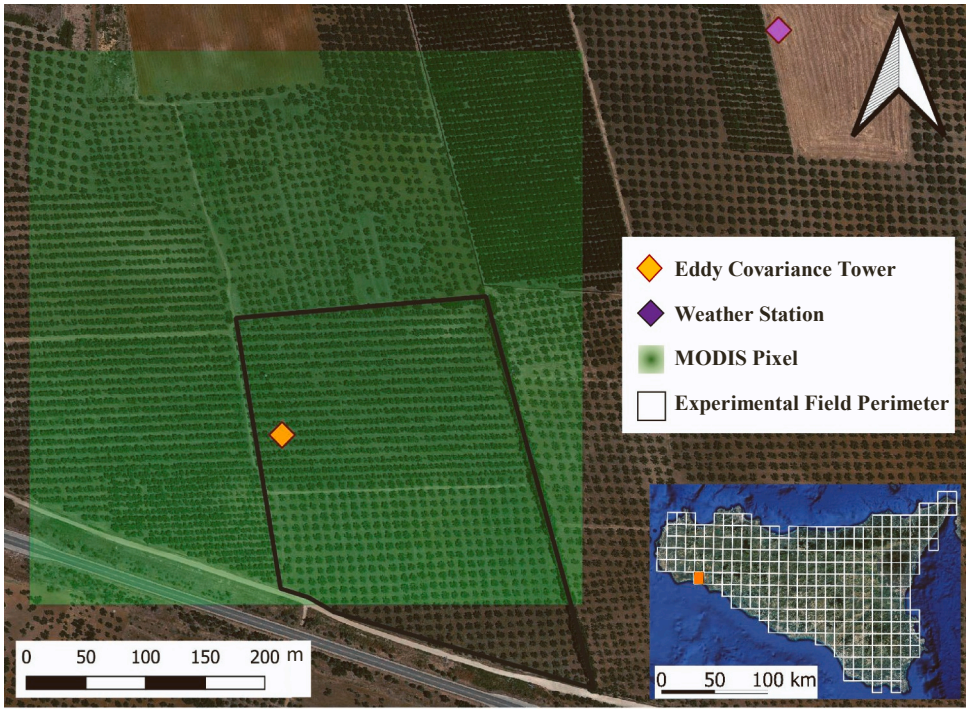


Fig. 2. Map of the olive orchard with the position of the EC tower and the SIAS weather station; the MODIS (MCD43A v006) pixel containing the experimental field is also shown. The bottom right box shows Sicily Island with the ERA5-Land grid (white) and the rectangle containing the experimental field (orange).

2.2. Climate data

The climatic data were obtained from different sources depending on the availability of ground measurements and/or reanalysis data.

2.2.1. Ground measurements

In the citrus orchard, a WatchDog 2000 (Spectrum Technologies, Inc., Aurora, IL, USA) standard weather station (WD) was installed near the field to collect, at half hourly steps, precipitation, P [mm], global solar radiation, R_s [W m^{-2}], air temperature, T [°C], air relative humidity, RH [%], wind speed [m s^{-1}] and direction [°] at 2 m height.

In the olive orchard, the same climate variables with hourly steps were measured by a standard weather station belonging to the Sicilian Agro-meteorological Information Service (SIAS) (www.sias.regione.sicilia.it/).

For the four American and Canadian fields, climate data from ground measurements were not available.

2.2.2. Reanalysis data

Additionally, for all the six fields, hourly reanalysis data of air temperature, T [°C], global solar radiation, R_s [W m^{-2}], dew-point temperature, T_{dew} [°C], and wind speed measured at 10 m above the ground, W_{s10} [m s^{-1}], were also downloaded from the ERA5-Land (ERA5-L) database (Muñoz Sabater, 2019), available in the climate change service portal provided by Copernicus, with the resolution of the spatial grid of 0.1° latitude and 0.1° longitude.

(https://cds.climate.co1pernicus.eu/cdsapp#!/home).

The hourly relative air humidity, RH [%], not available in the reanalysis database, was calculated as (Allen et al., 1998):

$$RH = 100\frac{e_a(T_{dew})}{e_s(T)} \quad (4)$$

where $e_a(T_{dew})$ and $e_s(T)$ are the actual and saturated vapor pressure corresponding at dew-point, T_{dew}, and actual air temperature, T, respectively.

Moreover, the hourly wind speed at 2 m above the ground, W_{s2} [m s^{-1}], was calculated based on the wind speed at the height z retrieved by the ERA5-L database, assuming valid the logarithmic wind speed profile (Allen et al., 1998):

$$W_{s2} = W_{sz}\frac{4.87}{\ln(67.8z - 5.42)} \quad (5)$$

where W_{s2} and W_{sz} are the wind speed at 2 m and z m above the soil surface. For ERA5-L, z is equal to 10 m.

Considering the aim of this study, all the weather variables at the half hourly and hourly time steps were then aggregated at the daily time step.

2.2.3. Quality assurance and quality control

The quality assurance procedures applied to the Italian ground climate measurements, provided by the WatchDog and SIAS stations, are reported in Fiebrich et al. (2010).

The quality of ERA5-Land data was evaluated by direct comparison to many in situ observations collected mainly for the period 2001–2018, as well as by comparison to additional model or satellite-based global reference datasets (Muñoz-Sabater et al., 2021).

2.3. Micrometeorological data

Micrometeorological data were retrieved from Eddy Covariance towers for the Italian orchards as well as from the AmeriFlux network for the Canadian and American orchards.

2.3.1. Italian fields

In both the Italian experimental fields, actual evapotranspiration, ET_a [mm], was estimated by an Eddy Covariance flux tower (EC). More details regarding the equipment of the EC towers in the citrus and olive orchard are reported in Ippolito et al. (2022) and in Cammalleri et al. (2013), respectively.

Sensible, H [W m^{-2}], and latent, LE [W m^{-2}], heat fluxes were evaluated as:

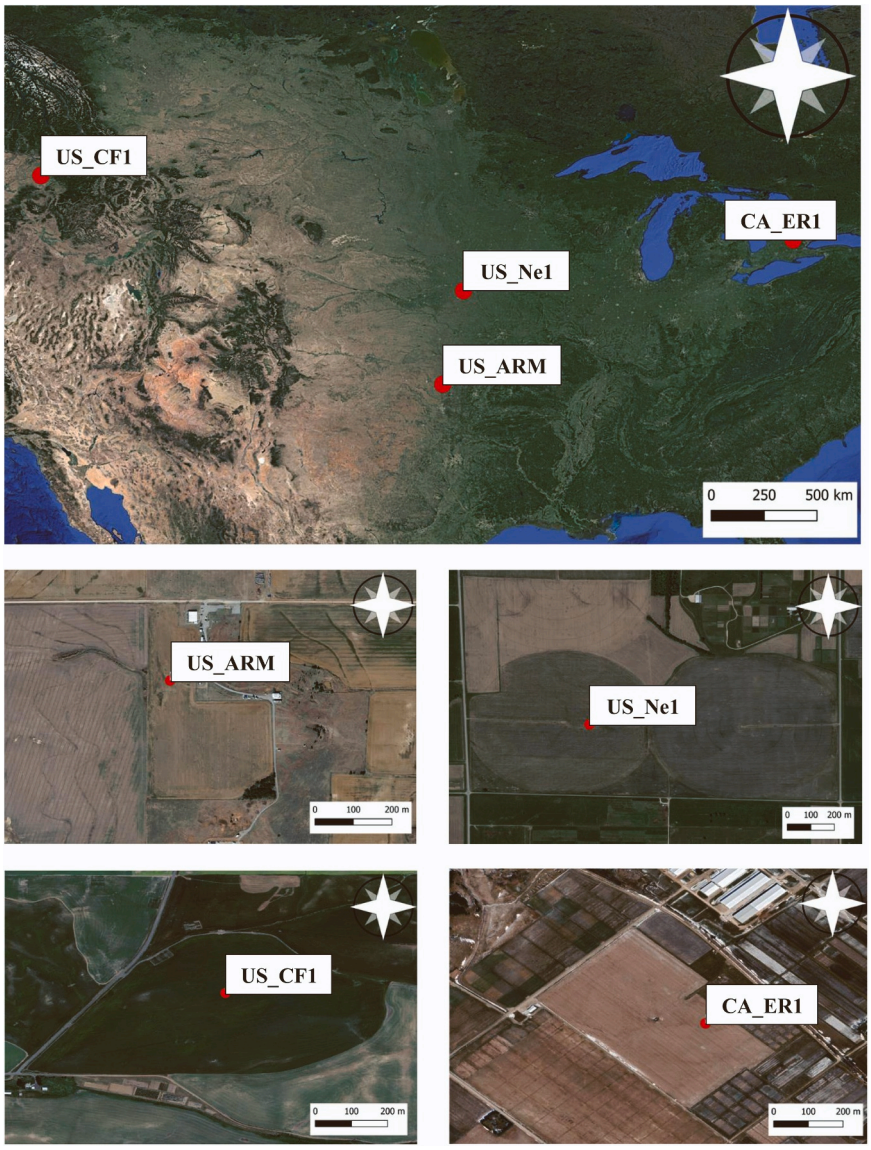


Fig. 3. Maps of the American and Canadian sites with the position of the EC towers (red points).

Table 1
Experimental fields.

Field	Crop	Latitude	Longitude	Climaticzone	Analysis period
Villabate	Citrus	38°4'53.4''N	38°4'53.4''E	Csa	2019–2021
Castelvetrano	Olives	37°38'61''N	12°50'53''E	Csa	2009–2011
US_ARM	Wheat	36°36'20.9''N	97°29'19.59''O	Cfa	2005–2010
US_Ne1	Maize-soybean	41°09'54.56''N	96°28'35.50''O	Dfa	2001–2020
US_CF1	Winter wheat	46°46'53.53''N	117°4'55.50''O	Dsb	2018–2020
CA_ER1	Corn	43°38'26.17''N	80°24'44.35''O	Dfb	2016–2019

$$H=\rho\, c_p\, \sigma_{WT} \quad (6)$$

$$LE=\lambda\, \sigma_{WQ} \quad (7)$$

where ρ [g/m^3] is the air density, c_p [J g^{-1} K^{-1}] is the air-specific heat capacity at constant pressure, σ_{WT} [m K s^{-1}], λ [J g^{-1}] is the latent heat of vaporization and σ_{WQ} [g m^{-2} s^{-1}] are the covariance between vertical wind speed and air temperature or water vapor density, respectively. All the fluxes were evaluated, using Eqs. (6) and (7), with a time step of 30 min using the software developed by Manca (2003). All the fluxes at the half hourly time steps were then aggregated at the daily time step.

2.3.2. American and Canadian fields

The values of latent heat flux (LE) for the American and Canadian sites were obtained from the AmeriFlux BASE data product (https://ameriflux.lbl.gov). Three sites' datasets had a half-hourly resolution, except the US_Ne1 site, which had an hourly resolution (Biraud et al., 2021; Suyker, 2022; Wagner-Riddle, 2021; Phillips and Huggins, 2022). Data shared under CC-BY-4.0 licence were downloaded. The LE at the half-hourly and hourly time steps were then aggregated at the daily time step.

2.3.3. Quality assurance and quality control

For the Italian field, the raw EC data were post-processed by means of the procedure implemented by Manca (2003), virtually analogous to the FLUXNET standard protocol (Mauder et al., 2008; Pastorello et al., 2014; 2020). Data de-trending was performed using a running mean, a coordinate rotation was applied to the sonic anemometer data to obtain a zero mean vertical and transversal wind speeds, and correction for spectral loss was performed. In addition, adjustments for high wind speeds on sonic temperature and Webb–Pearman–Leuning corrections for water vapor were applied (Moncrieff et al., 1997) before the final computation of half-hourly fluxes. The method suggested by Prueger et al. (2005), based on the Closure Ratio (CR), was used to assess the surface energy balance closure. When the energy storage in the soil is neglected, CR represents the slope of the regression line between the available energy given by the variance of the net radiation, Rn, and the soil heat flux, G, and the covariance of the turbulent heat fluxes (LE + H) only evaluated from the subset of half hourly data corresponding to Rn $\geq$ 100 W m^{-2}:

$$CR = \frac{LE + H}{R_n - G} \tag{8}$$

To guarantee the consistency of the daily data, only the days with 48 half-hourly measurements were considered.

AmeriFlux BASE data Quality Assurance (QA) and quality control (QC) follow the methodology adopted for processing the FLUXNET2015 dataset (Pastorello et al., 2014). The data for each site goes through QA and QC steps tailored to the generation of these derived data products. A few of these QA and QC steps are described in Pastorello et al. (2020). They provide a processing flowchart, which describes all the processing that is applied. Data QA and QC assesses units and sign conventions, timestamp alignments, trends, step changes, outliers based on site-specific historical ranges, multivariate comparisons, diurnal/-seasonal patterns, friction velocity filtering, and variable availability. Quality checks are done over single variables, multiple/combined variables, or more specialized tests.

2.4. Monitoring soil water contents

Only in the citrus orchard, the temporal dynamics of soil water content (SWC) were monitored with four "drill & drop" probes (Sentek Pty Ltd, Stepney, Australia), installed at a distance of 0.8 m from the trees' trunks and 0.3 m apart from one emitter. These sensors, based on the frequency-domain reflectometry (FDR) technique (Gaskin and Miller, 1996), allowed measuring apparent soil dielectric permittivity, whose values are strongly influenced by the monitoring SWC, at each 0.1 m depth, from the soil surface to 0.6 m depth, with a time step of about 30 min. FDR sensors have been extensively used to monitor SWC (Burgess et al., 2006; Sun et al., 2014; Singh et al., 2019; Campora et al., 2020).

The "drill & drop" sensor's response is represented by the scaled frequency, which includes the raw counts in air, water, and soil. The probes calibration equations for the different soils were obtained directly in the field. The probes were normalized, according to the recommendation of the manufacturer, to convert the raw counts, corresponding to the actual water content, into scaled frequency (SF). To determine SF, the frequencies in air and water were also acquired:

$$\mathrm{SF} = \frac{(F_a - F_s)}{(F_a - F_w)} \tag{9}$$

where F_a is the raw count in air, F_s is the raw count in the soil, and F_w is the raw count in water.

The manufacturer also provides a standard calibration power relationship that allows to retrieve SWC, expressed in cm^3 cm^{-3}, value starting from the SF values measured by the sensor:

$$\mathrm{SWC} = \alpha\ \mathrm{SF}^{\beta} \tag{10}$$

where α and β are the coefficients of the general equation proposed by the manufacturer, equal to 0.494 and 3.017, respectively.

2.5. Vegetation indices (VIs)

Sentinel-2 scenes and Moderate Resolution Imaging Spectroradiometer (MODIS) nadir reflectances time series were used to calculate the NDVI as well as the NDWI. Matlab® R2021b and QGIS version 3.4.3 were jointly used to process the images and evaluate both vegetation indices.

2.5.1. Sentinel-2

Since the Copernicus mission started operations in 2018 spring and the experiments started in 2019 winter, the spatio-temporal variability of vegetation indices for the citrus orchard was investigated only by using the Sentinel-2 multispectral images level 2 A.

Instead, since the experimental period in the olive orchard is prior to 2018, it was not possible to use the Sentinel-2 data. However, for this experimental field, it was decided to use the Sentinel-2 images, available for the years 2019–2021, to test the reliability of the MODIS data.

The images were downloaded and pre-processed using the R package toolbox "sen2r" (Ranghetti et al., 2020) In the period from 2019 to 2021, 191 and 65 available clear-sky scenes were selected for the citrus and the olive orchards, respectively; the near-infrared, red and shortwave reflectance were detected in the bands B08 ranging from 785 to 899 nm, B04 from 650 to 680 nm, and B11 from 1565 to 1655 nm of the Sentinel-2 electromagnetic spectrum. Based on Eqs. (2) and (3), maps of NDVI and NDWI with a spatial resolution of 10 m, were generated. To downscale the maps of NDWI at 10 m resolution, the value of a single pixel in the shortwave region (20 m resolution) was associated with the corresponding four values in the near-infrared region (10 m resolution). A continuous time series of daily VIs were obtained based on linear interpolations carried out between consecutive pairs of Sentinel-2 images acquired at two different dates (Pan et al., 2017).

2.5.2. Moderate Resolution Imaging Spectroradiometer (MODIS)

Moderate Resolution Imaging Spectroradiometer (MODIS) MCD43A3 Version 6 product (MCD43A v006) (Schaaf and Wang, 2015) was considered to obtain time-series of nadir reflectance at the different wavelengths for all the experimental fields, except for citrus orchard. The continuous time series of MODIS nadir reflectance were downloaded using the Google Earth Engine (GEE) platform.

The near-infrared, red and shortwave reflectance were detected respectively in band B2 ranging from 841 to 876 nm, in band B1 ranging from 620 to 670 nm and in band B6, from 1628 to 1652 nm of the MODIS electromagnetic spectrum.

2.5.3. Sentinel-2 vs MODIS

As previously mentioned, for the olive orchard, a comparison between Sentinel-2 and MODIS images was carried out in the period 2019–2021 to verify the suitability of MODIS images for estimate the vegetation indices NDVI e NDWI since this product is characterized by a spatial resolution very lower than the Sentinel-2 one. The results demonstrated that MODIS images can be considered suitable for the investigated areas since are quite homogeneous in terms of soil and crop types (Autovino et al., 2016).

2.6. Gaussian Process Regression (GPR) algorithm

To estimate the missing data in the ET_a time series, the Gaussian Process Regression (GPR) model was exploited for its power, among other machine learning algorithms, to make predictions relying on a few parameters. The model establishes a relation between the independent input variables, x_i, (climatic variables, soil water status, and VIs), and the dependent variable, y, represented by the actual crop evapotranspiration. Specifically, the regression model built by the GPR is:

$$y = f(x) + \varepsilon \sim N(m(x), k(x'x')) + \varepsilon \quad (11)$$

where x and y denote the input and output in the training dataset and f (x) is known as latent variable in the GPR model, and ε is the noisy observations of the true function expressed as a normal distribution characterized by a mean equal to zero and variance σ_n^2 estimated from the data. The Gaussian process $f(x)$ can be defined by its mean *m(x)* and covariance kernel *k(x,x')* functions, represented respectively by a vector and a matrix in the form:

$$m(x) = E[f(x)] \quad (12)$$

$$k(x,x') = E[(f(x) - m(x))(f(x') - m(x'))] \quad (13)$$

The mean (basis) function of the GPR model can be assumed constant, with a value set to zero, equal to the mean of the training dataset, or by using a linear function. The kernel (covariance) function represents a geometrical distance measure assuming that the more closely located inputs would be more correlated in terms of their function values. The covariance kernel function can be assumed as rational quadratic, squared exponential, exponential, or using a Matern kernel. More details on the GPR model can be found in Rasmussen and Williams (2006).

To identify the best *m(x)* and *k(x,x')* functions, the machine learning model based on the GPR was implemented by using a Matlab® script, which was used to test fifteen possible combinations of the mean and covariance kernel functions: three mean functions (Zero, Constant, and Linear) and five covariance kernel functions (Squared Exponential, Matern 5/2, Matern 3/2, Rational Quadratic, and Exponential).

The Squared Exponential kernel, also called Gaussian kernel or RBF kernel, is defined as (Rasmussen and Williams, 2006):

$$k(x,x') = \sigma_S^2 \exp\left(-\frac{1}{2}\frac{(x-x')^T(x-x')}{\lambda_S^2}\right) + \sigma_\varepsilon^2 X \quad (14)$$

where $(x-x')^T(x-x')$ can be regarded as the squared Euclidean distance between two eigenvectors and σ_S, λ_S and σ_ε are three hyperparameters. σ_S^2 controls the marginal variance of f(x) and is referred to as variance parameter, which is used to characterize the deviation of the fitting function from the signal mean value. When σ_S^2 is small, the fitting function deviates from the signal mean value slightly. When σ_S^2 is large, the fluctuation of the fitting function will become larger (Pan et al., 2021). λ_S is the relative characteristic length scale, which is used to describe the smoothness of the function. When λ_S is small, the dynamic response performance of the fitting function is better than when it is high, but it is accompanied by the risk of overshooting; when λ_S is large, the resultant function tends to be smooth. σ_ε^2 represent the measurement error. X is the indicator function.

The Matern function kernel is named by Stein (1999) after the work of Matern (1986). It can be written as:

$$k(x,x') = \sigma_M^2 \exp\left(-\frac{\sqrt{2\upsilon(x-x')^T(x-x')}}{\lambda_M}\right) \times \frac{\Gamma(p+1)}{\Gamma(2p+1)} \sum_{l=0}^{p} \frac{(p+l)!}{l!(p-l)!} \left(\frac{\sqrt{8\upsilon(x-x')^T(x-x')}}{\lambda_M}\right)^{p-l} + \sigma_\varepsilon^2 X \quad (15)$$

where υ is a chosen parameter equals to (p + ½), the definitions of σ_M and λ_M are like those of σ_S and λ_S, where the subscripts indicate the reference to the kernel function name. Γ is the Gamma function. The Matern kernel is characterized by the parameter υ. According to Rasmussen and Williams (2006), we set υ equal to 3/2 and 5/2.

The Rational Quadratic kernel can be treated as a scale mixture of Squared Exponential kernels with the different characteristic length-scales (Rasmussen and Williams, 2006):

$$k(x,x') = \sigma_R^2 \left(1 + \frac{(x-x')^T(x-x')}{2\alpha_R \lambda_R^2}\right)^{-\alpha_R} + \sigma_\varepsilon^2 X \quad (16)$$

The hyperparameters σ_R, λ_R and σ_ε are like those in the definition of Squared Exponential and Matern, where the subscripts indicate the reference to the kernel function name. Rational Quadratic has a positive-valued scale-mixture parameter α_R that can determine the relative weighting of large-scale and small-scale variances. Rational Quadratic kernel is equivalent to a scale mixture of Squared Exponential kernels with different relative characteristic length scales (Roberts et al., 2013). The rational quadratic kernel has a wide scope, which could help the prediction performance also when small input database is used and to improve the generalization ability and the dynamic response performance (Yang et al., 2022; Shi et al., 2022).

Finally, the mathematical formula of the Exponential kernel is as follow (Rasmussen and Williams, 2006):

$$k(x,x') = \sigma_f^2 \exp\left(-\frac{\sqrt{(x-x')^T(x-x')}}{\lambda_f}\right) + \sigma_\varepsilon^2 X \quad (17)$$

The hyperparameters σ_f, λ_f and σ_ε are defined previously, where the subscripts indicate the reference to the kernel function name.

This analysis was carried out by considering the complete dataset of the input variables acquired in the citrus orchard including:

i) the standard weather variables: global solar radiation (R_s), air temperature (T), relative air humidity (RH), and, wind speed measured at 2 m height (W_{s2});
ii) the Soil Water Content (SWC), expressive of the soil water status;
iii) two vegetation indices (VIs), such as NDVI and NDWI, accounting for the characteristics of the vegetation, with impact on the crop coefficient K_c (Lei and Yang, 2014; Pôças et al., 2020).

Once identified the best *m(x)* and *k(x,x')* functions, using the complete database, four different possible combinations of the input variables were considered to reduce the computational burden.

Table 2 summarizes the combinations of the investigated variables used as input, to run the model; where the first combination is used to find the best mean and kernel functions; instead, the other four to reduce the computational burden. The first combination contains the complete dataset that was also used to select the *m(x)* and *k(x,x')* functions. The second and the third combinations excluded, respectively, the VIs, and the SWCs, whose values are not always available. The fourth combination considers only the agrometeorological variables, whereas the last combination included only variables acquired from remote platforms and available online, such as the agrometeorological variables and the VIs.

To assess the model suitability using the best *m(x)* and *k(x,x')* functions previously retrieved, for the olive orchard, only the last three variables' combinations due to the unavailability of SWC measurements, were tested; instead, for the fields located in the USA and Canada only the 5th combination was evaluated. For these applications, the VIs were retrieved from the MODIS sensor.

2.7. N-fold cross-validation and statistical indicators

The best combination of *m(x)* and *k(x,x')* functions, as well as the performances associated with the other variables combinations, were evaluated using the N-fold cross-validation (Mosteller et al., 1968). Cross-validation is a statistical method in which the database is randomly divided into N different groups, each one containing the records corresponding to the natural number closest to N_t/N, in which N_t is the total number of records. Model validation follows an iterative

Table 2
Different combinations of input variables.

Variables' combination	Input	Tool	Symbol	Unit
1	Global Solar Radiation	WatchDog 2000	R_s	[W m^{-2} d^{-1}]
	Air Temperature		T	[°C]
	Relative Air Humidity		RH	[%]
	Wind Speed		W_{s2}	[m s^{-1}]
	Normalized Difference Vegetation Index	Sentinel-2	NDVI	[-]
	Normalized Difference Water Index		NDWI	[-]
	Soil Water Content	Drill and Drop	SWC	[cm^3 cm^{-3}]
2	Global Solar Radiation	WatchDog 2000	R_s	[W m^{-2} d^{-1}]
	Air Temperature		T	[°C]
	Relative Air Humidity		RH	[%]
	Wind Speed		W_{s2}	[m s^{-1}]
	Soil Water Content	Drill and Drop	SWC	[cm^3 cm^{-3}]
3	Global Solar Radiation	WatchDog 2000	R_s	[W m^{-2} d^{-1}]
	Air Temperature	or	T	[°C]
	Relative Air Humidity	SIAS	RH	[%]
	Wind Speed		W_{s2}	[m s^{-1}]
	Normalized Difference Vegetation Index	Sentinel-2	NDVI	[-]
	Normalized Difference Water Index	or	NDWI	[-]
		MODIS		
4	Global Solar Radiation	WatchDog 2000	R_s	[W m^{-2} d^{-1}]
	Air Temperature	or	T	[°C]
	Relative Air Humidity	SIAS	RH	[%]
	Wind Speed		W_{s2}	[m s^{-1}]
5	Global Solar Radiation	ERA5-Land	R_s	[W m^{-2} d^{-1}]
	Air Temperature		T	[°C]
	Relative Air Humidity		RH	[%]
	Wind Speed		W_{s2}	[m s^{-1}]
	Normalized Difference Vegetation Index	Sentinel-2	NDVI	[-]
	Normalized Difference Water Index	or	NDWI	[-]
		MODIS		

procedure in which a group is used as *test set*, while the other N-1 groups are used for training. The procedure is stopped after N iterations and therefore after using each group as *test set*. This iterative statistical analysis reduces the possibility of overfitting problems (Nguyen et al., 2021) occurring when the model fits well the training data but fails in the prediction phase due to the noise or random fluctuations in the training data (Namasudra et al., 2021). In this study, the value of N was set up as equal to five and therefore 80% of the entire database was used for training, while the remaining 20% was used for testing purposes.

The suitability of *m(x)* and *k(x,x')* functions, as well as the performance of the GPR model with the other four different variables combinations were assessed based on the following statistical indicators:

i) the Root Mean Square Error (RMSE, mm d^{-1}) whose target value is zero if there are no differences between simulated and observed values;
ii) the coefficient of determination (R^2, dimensionless) whose unitary target indicates that the variance of the observed values is totally explained by the model (Eisenhauer, 2003);
iii) the Mean Absolute Error (MAE, mm d^{-1}) representing the distance between the predicted and observed values, with a target of zero;
iv) the regression coefficient (b, dimensionless), whose target is one, representing the angular coefficient of the regression line between simulated and observed values forced to the origin;
v) the Nash-Sutcliffe efficiency coefficient (NSE, dimensionless), whose target values is one; a value of this indicator between 0.0 and 1.0 indicates an acceptable model performance, whereas a negative value indicates that the mean of observed values is a better predictor than the simulated ones and therefore the performance is unacceptable (Nash and Sutcliffe, 1970).

For each iteration, all the statistical indicators were evaluated. Finally, the model performances were analyzed as the mean of all the N performed iterations.

3. Input data analysis

Before the algorithm application, a comparison between ground and reanalysis data was conducted. This analysis, realized in the citrus orchard, where both databases were available, allowed to evaluate the suitability of reanalysis climate variables when ground data are not available.

Then, the series of actual evapotranspiration values for each experimental field were estimated and the consistencies of these series were analyzed to evaluate the percentage of missing data for each site.

Finally, for the olive orchard, as previously mentioned, a comparison between the NDVI and NDWI retrieved from Sentinel-2 and MODIS was evaluated in order to verify the suitability of MODIS images to estimate the VIs.

3.1. Ground vs reanalysis data

Fig. 4 shows the scatterplots between ground, acquired by the citrus orchard WD weather station, and reanalysis observations, retrieved by ERA5-L database, of global solar radiation, R_s (a), average air temperature, T_{avg} (b), relative air humidity, RH_{avg} (c), and wind speed measured at 2 m above the soil surface, W_{s2} (d). The colored bar indicates the day of the solar year (DOY). R_s values retrieved from the ERA5-L product were slightly overestimated (b = 1.03; RMSE = 32.68 W m^{-2}), being the relatively higher differences associated with the lower R_s values, mostly concentrated during winter (blue and red dots) (Fig. 4**a**). The average air temperature T_{avg} retrieved by the ERA5-L database resulted slightly underestimated if compared with the ground measurements (b = 0.93; RMSE = 2.25 °C). In this case, the underestimation was distributed along the year (Fig. 4**b**). In Fig. 4**c**, the point cloud is larger than the previous one, and the red and blue points, corresponding to the wet seasons, stay up the 1:1 line, while the yellow and orange points, stay under the1:1 line. Therefore, RH_{avg} evaluated based on the ERA5-L records is lightly overestimated in the wet season and underestimated in the dry season, since RMSE value is equal to

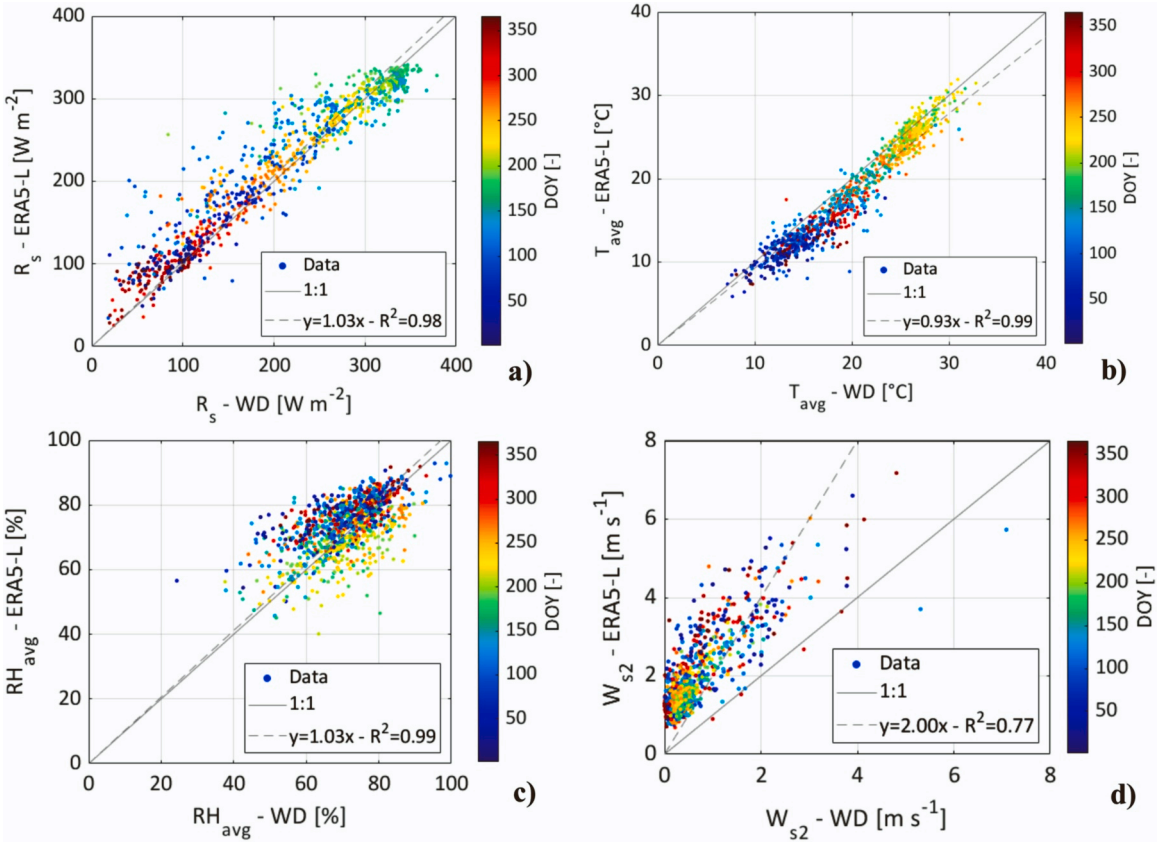


Fig. 4. a,d. Comparison between a) daily global solar radiation, R_s, b) average air temperature, T_{avg}, c) average air relative humidity, RH_{avg}, and d) wind speed measured at 2 m above the ground, W_{s2}, registered by the weather station installed in the citrus orchard and retrieved by the reanalysis data. The colored bar indicates the day of the year (DOY).

8.64%. Finally, the wind speed values at 2 m height, calculated from the ERA5-L reanalysis data referred at 10 m height by assuming a logarithmic wind profile, resulted generally higher than the ground measurements (Fig. 4**d**), with a coefficient of the regression line passing through the origin equal to b = 2.00 and RMSE value equal to 1.43 m s^{-1}.

Similar results for all the climate variables were obtained by Vanella et al. (2022) for the weather station installed near the olive orchard for which the comparisons with the ERA5-L data led to RMSE values respectively of 30.31 W m^{-2} for R_s, 1.37 °C for T_{avg}, 9.72% for RH_{avg} and 1.54 m s^{-1} for W_{s2}, air temperature. Vanella et al. (2022) considered acceptable these results because of ERA5-L provided daily ETo estimates with good accuracy. Therefore, in the examined area, reanalysis database such as the ERA5-L can be considered a suitable surrogate of the ground data, as evaluated by several authors for most weather stations installed in Sicily (Pelosi and Chirico, 2021; Vanella et al., 2022).

3.2. Micrometereological data

The scatterplots between the turbulent heat fluxes, H+LE, as a function of the available energy, R_n-G measured by the EC tower installed in the citrus orchard at the half hourly time step, for each of the three years, were already evaluated in Pagano et al. (2023). The slope of the regression line passing through the origin represents the closure ratio, CR, whose values resulted in satisfactory and equal to 0.98, 0.88 and 1.03 for 2019, 2020 and 2021, respectively.

For the olive orchard, the values of CR were estimated equal to 0.90 for 2009, 0.92 for 2010 in Cammalleri et al. (2013), and 1.02 for 2011 in Autovino et al. (2016).

For tree crops, Kustas et al. (1999) considered acceptable values of CR ranging between 0.80 and 0.90 even if values of CR equal to 1.08 and 1.03, were obtained by Er-Raki et al. (2009) in two citrus orchards in south Morocco characterized by a semi-arid Mediterranean climate. The CR estimates for the two fields conformed to the values found in the literature and are therefore considered reliable.

Table 3 summarizes the lack of data percentage for each field.

Table 3
Lack of data percentage.

Field	Days of observations	Days of measurements	Lack of data [%]
Villabate	933	501	46.3
Castelvetrano	1095	573	47.7
US_ARM	2191	1858	15.2
US_Ne1	7305	6512	10.9
US_CF1	1096	809	26.2
CA_ER1	1461	1160	20.6

3.3. Sentinel-2 vs MODIS

Fig. 5 shows the comparison between the NDVI and NDWI retrieved from Sentinel-2 (black line) and MODIS (red line) products for the olive orchard. The Sentinel-2 VIs, with the corresponding standard deviations, were obtained through zonal statistics considering all the Sentinel-2 pixels contained within the pixel MODIS. The major differences between the two VIs are concentrated during falls and winters (red and blue dots in the scatter plots on the right) probably due to the presence of ground weeds that grow after rain events (Ippolito et al., 2022), which can be detected with the Sentinel-2 images and not with the MODIS ones. The generally good agreements between the NDVI (b = 0.95: R^2 =0.99) and NDWI (b = 0.88; R^2 = 0.96) obtained from the two different platforms strongly depend on the homogeneity of soil and land characterizing the MODIS pixel (Autovino et al., 2016).

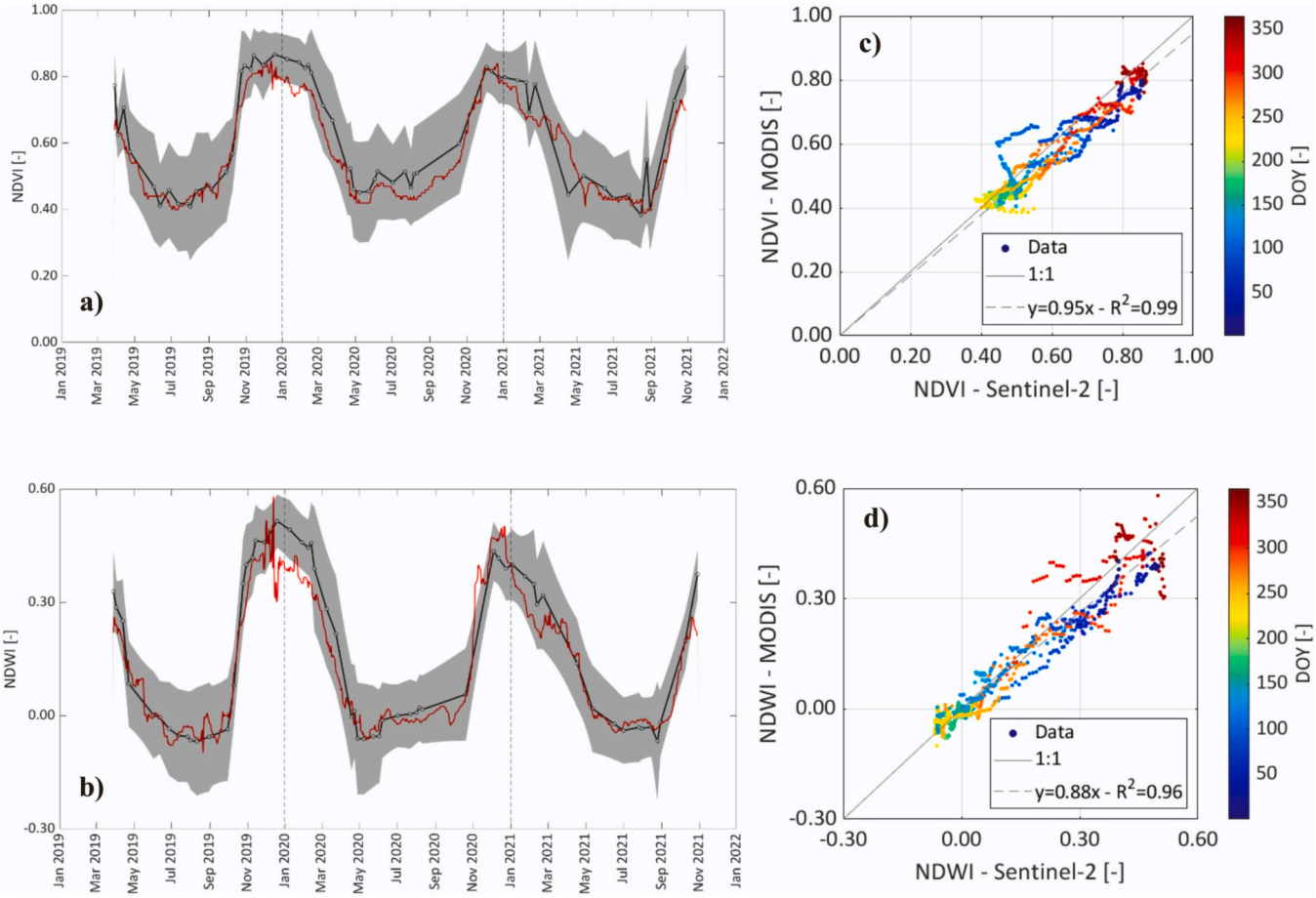


Fig. 5. a,d. Comparison between the temporal dynamics of NDVI (a) and NDWI (b) retrieved from Sentinel-2 satellites (black line) and MODIS (red line) products for the olive orchard. The grey region represents the values of the standard deviations, obtained through zonal statistics considering all the Sentinel-2 pixels contained within the pixel MODIS. The corresponding scatterplots of NDVI and NDWI obtained by Sentinel-2 versus the corresponding MODIS are also shown on the right (c, d).

Table 4
Statistical indicators for mean, *m(x)*, and kernel covariance, *k(x,x')*, functions obtained by considering the entire database which includes all the examined variables (combination 1).

		Covariance kernel functions k(x,x')				
	Mean function m(x)	Rational Quadratic	Squared Exponential	Matern 5/2	Matern 3/2	Exponential
RMSE	Zero	0.39	0.44	0.42	0.40	**0.38**
[mm d^{-1}]	Constant	0.40	0.45	0.43	0.42	0.39
	Linear	0.41	0.40	0.41	0.41	0.40
MAE	Zero	0.28	0.33	0.31	0.29	**0.28**
[mm d^{-1}]	Constant	0.29	0.32	0.31	0.30	0.28
	Linear	0.29	0.29	0.30	0.29	0.29
R^2	Zero	0.87	0.84	0.85	0.86	**0.88**
[-]	Constant	0.86	0.83	0.84	0.85	0.87
	Linear	0.86	0.87	0.86	0.86	0.87
b	Zero	0.98	0.97	0.97	0.98	**0.98**
[-]	Constant	0.97	0.97	0.97	0.98	0.98
	Linear	0.97	0.98	0.98	0.98	0.98
NSE	Zero	0.87	0.84	0.85	0.86	**0.88**
[-]	Constant	0.86	0.83	0.84	0.85	0.87
	Linear	0.86	0.87	0.86	0.86	0.87

4. Results and discussion

Once the input data as well as the actual evapotranspiration have been estimated, the GPR algorithm was tested.

The best mean, *m(x)*, and kernel covariance, *k(x,x')*, functions were assessed by the complete dataset related to the citrus orchard, which includes all the examined variable (see combination 1 in Table 2). Table 4 summarizes the statistical indicators associated with the fifteen combinations of the examined *m(x)* and *k(x,x')* functions, whereas Fig. 6 illustrates the scatterplots of the corresponding measured versus estimated actual evapotranspiration values when considering the complete database of the input variables and implementing the best and the worst *m(x)* and *k(x,x')* functions in the GPR model.

Despite the fairly limited differences in terms of the examined

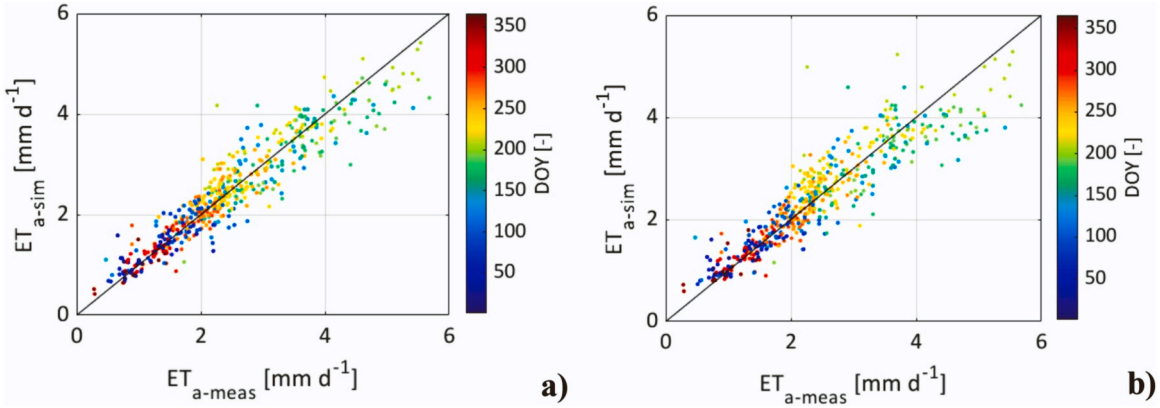


Fig. 6. a,b. Scatterplots of estimated versus measured ET_a obtained for the citrus orchard when considering the complete dataset (Combination 1) and assuming the best (a) and the worst (b) combination of the mean and kernel covariance functions. RMSE values for the best and the worst combination are equal to 0.38 mm d^{-1} and 0.45 mm d^{-1}, respectively.

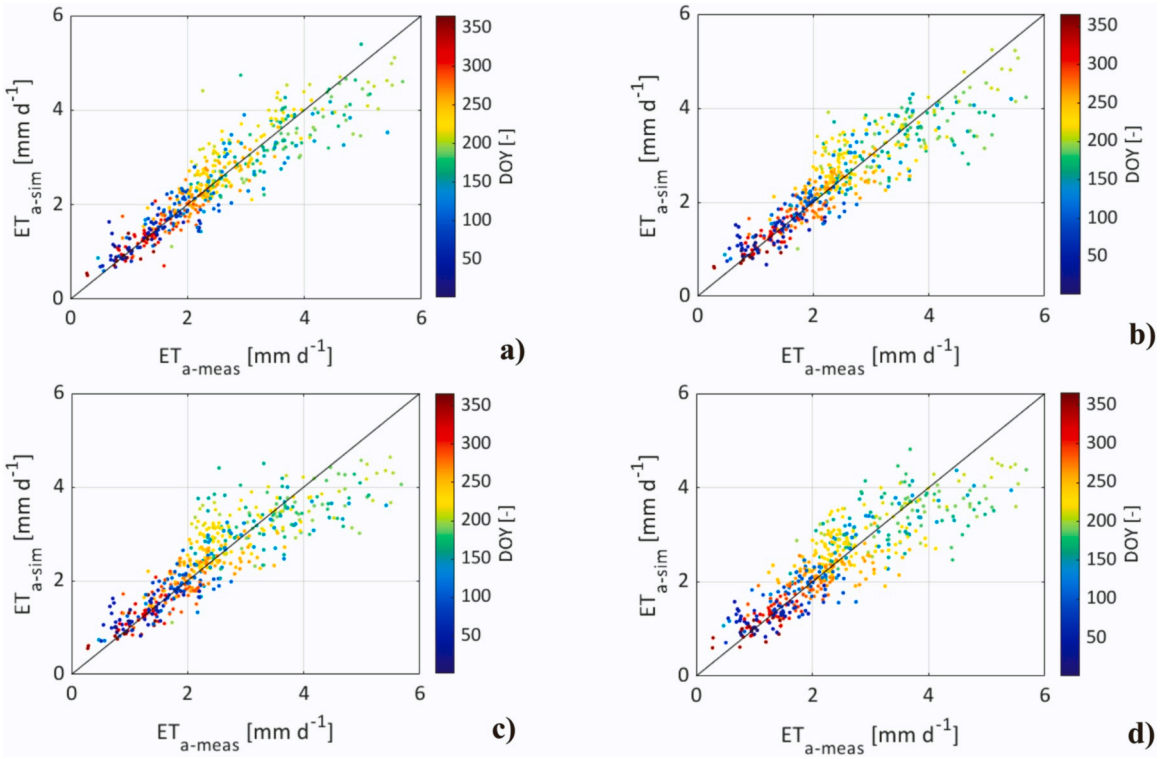


Fig. 7. a,d. Scatterplots of estimated versus measured ET_a obtained in the citrus orchard by implementing the GPR model with a zero mean function and an exponential kernel covariance function, and the four different combinations of the input variables (a) combination 2, b) combination 3, c) combination 4, d) combination 5).

Table 5
Statistical indicators associated with the GPR model implemented with the best *m(x)* and *k(x,x')* functions and the five combinations of the input variables, for the citrus orchard.

	RMSE [mm d^{-1}]	MAE [mm d^{-1}]	R^2 [-]	b [-]	NSE [-]
1	0.38	0.28	0.88	0.98	0.88
2	0.43	0.31	0.84	0.97	0.84
3	0.48	0.36	0.80	0.96	0.80
4	0.53	0.39	0.76	0.96	0.76
5	0.55	0.42	0.74	0.95	0.74

statistical indicators, the best result was obtained when assuming a zero mean function and an exponential kernel covariance function. Under these assumptions, the estimations of ET_a resulted characterized by RMSE = 0.38 mm d^{-1} and MAE = 0.28 mm d^{-1}. On the other hand, the worst result, characterized by RMSE = 0.45 mm d^{-1} and MAE = 0.32 mm d^{-1} was obtained when assuming a constant *m(x)* and a squared exponential *k(x,x')*. In general, however, regardless of the *m(x)* and *k(x,x')* functions, the model estimates quite well the daily actual evapotranspiration (NSE > 0.83).

Therefore, for the other variables' combinations and in all the other fields, the suitability of the ML model was tested using the best *m(x)*, and, *k(x,x')* functions, that is zero mean function, and an exponential kernel covariance functions, respectively.

For the other combinations of the input variables (see combinations

2÷5 in Table 2), Fig. 7 shows the scatterplots of measured versus estimated daily ET_a by implementing in the GPR model the best *m(x)* and *k(x,x')* functions; the estimated values are in better agreement with the measured ones in fall and winter days (blue and red dots). In spring and summer days (yellow, green, and cyan dots), the dispersion increases for values higher than about 2.5 mm d^{-1}. For the citrus orchard, Table 5 summarizes the statistical indicators associated with the GPR model implemented with the best *m(x)* and *k(x,x')* functions and the five input variables' combinations. As can be observed, whatever is the variable combination there is always an underestimation of actual evapotranspiration with slopes of the regression lines declining from 0.97 to 0.95 as the number of variables decreases. The RMSE and MAE values increase from 0.43 mm d^{-1} to 0.55 mm d^{-1} and from 0.31 mm d^{-1} to 0.42 mm d^{-1}, respectively. As well as, the R^2 and NSE values decrease from 0.84 to 0.74, for both the statistical indicators.

When comparing the second and third input variable combination, involving the measured soil water content SWC (combination 2) resulted in a better model performance than including the two vegetation indices (combination 3), as demonstrated by the lower RMSE and MAE and the higher R^2, b and NSE values. The possibility to include measurements of SWC among the input variables to improve the estimations of actual evapotranspiration was also indicated by Granata (2019). On the other hand, also the introduction of the VIs improved the model performances, as shown by Carter and Liang (2019) and Mosre and Suárez (2021). The worst performances were associated with combination 5, in which all the input variables are detected from information freely available online. However, even if compared to the others, the statistical parameters evidenced that the joint use of reanalysis and remote sensing data can still be considered acceptable for practical applications related to irrigation management, considering that all the required input data can be downloaded from the web, avoiding costly and time-consuming field measurements in areas where ground data are not available. According to Faramiñán et al. (2021) and Kang et al. (2019) the use of reanalysis and remoted sensed data could be a good alternative to retrieve ET_a with ML techniques.

Table 6
Statistical indicators associated with the GPR model implemented with the best *m(x)* and *k(x,x')* functions and the three combinations of the input variables, for the olive orchard.

	RMSE [mm d^{-1}]	MAE [mm d^{-1}]	R^2 [-]	b [-]	NSE [-]
3	0.50	0.38	0.67	0.94	0.67
4	0.51	0.39	0.65	0.93	0.65
5	0.50	0.38	0.66	0.94	0.66

Similar good results were obtained when considering the daily actual evapotranspiration data collected in the olive orchard. Even if the available database did not include measurements of SWC, it permitted a further validation of the proposed ML algorithm by considering only the three variable combinations including the weather data and VIs. Fig. 8 shows the scatterplots between measured and estimated ET_a by assuming a zero mean function and an exponential kernel covariance function and the three combinations of the input variables. The estimated values are in better agreement with the measured ones in fall and winter days (blue and red dots). The dispersion increases for values higher than about 3 mm d^{-1}. The statistical indicators associated with the three different variable combinations are summarized in Table 6. As can be observed, combinations 3 and 5 are characterized by quite similar results. This result is a consequence of the good performance of the ERA5-L in depicting the agrometeorological data measured on the ground nearby the olive orchard (Vanella et al., 2022). Slightly worse results were associated with combination 4 in which only measured weather data were used as input of the model.

The last analysis, which involves the American and Canadian sites, located in different climate zones, returns good performances. The approach was tested using only remote sensed data (combination 5).

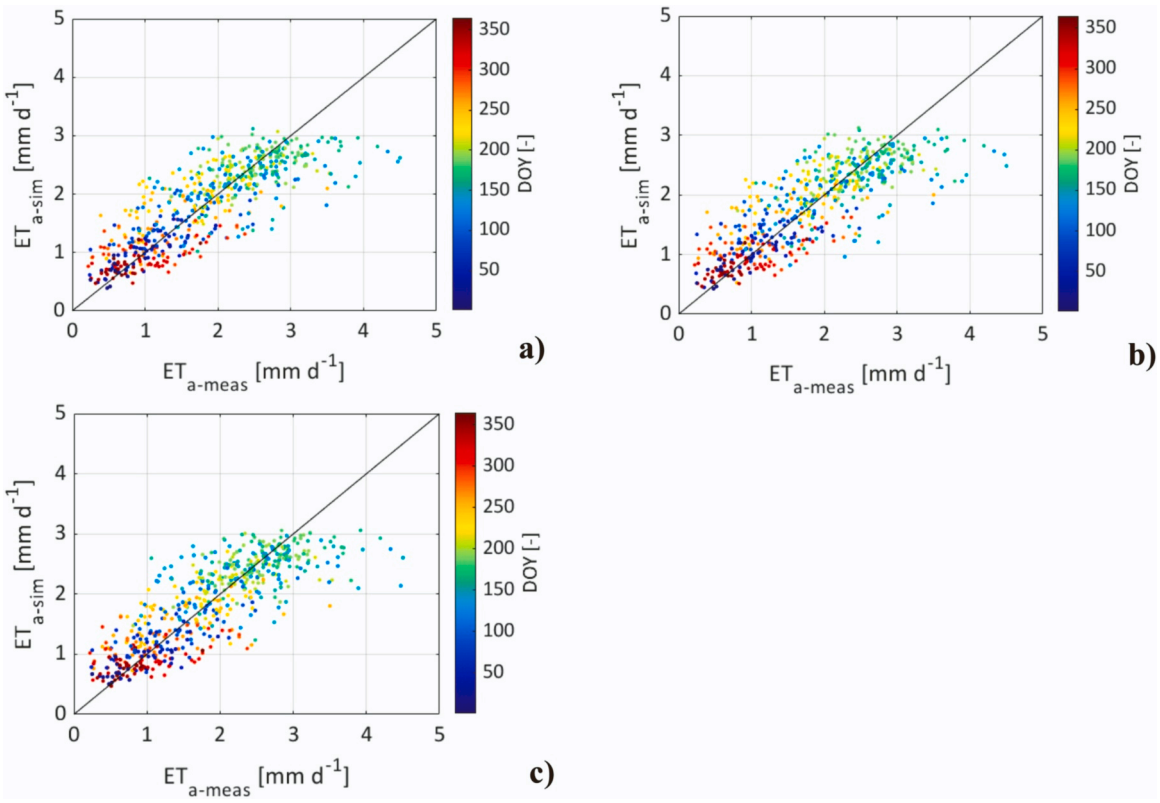


Fig. 8. a,c. Scatterplots of estimated versus measured ET_a obtained in the olive orchard by implementing the GPR model with a zero mean function, an exponential kernel covariance function, and the three different combinations of the input variables (a) combination 3, b) combination 4, c) combination 5).

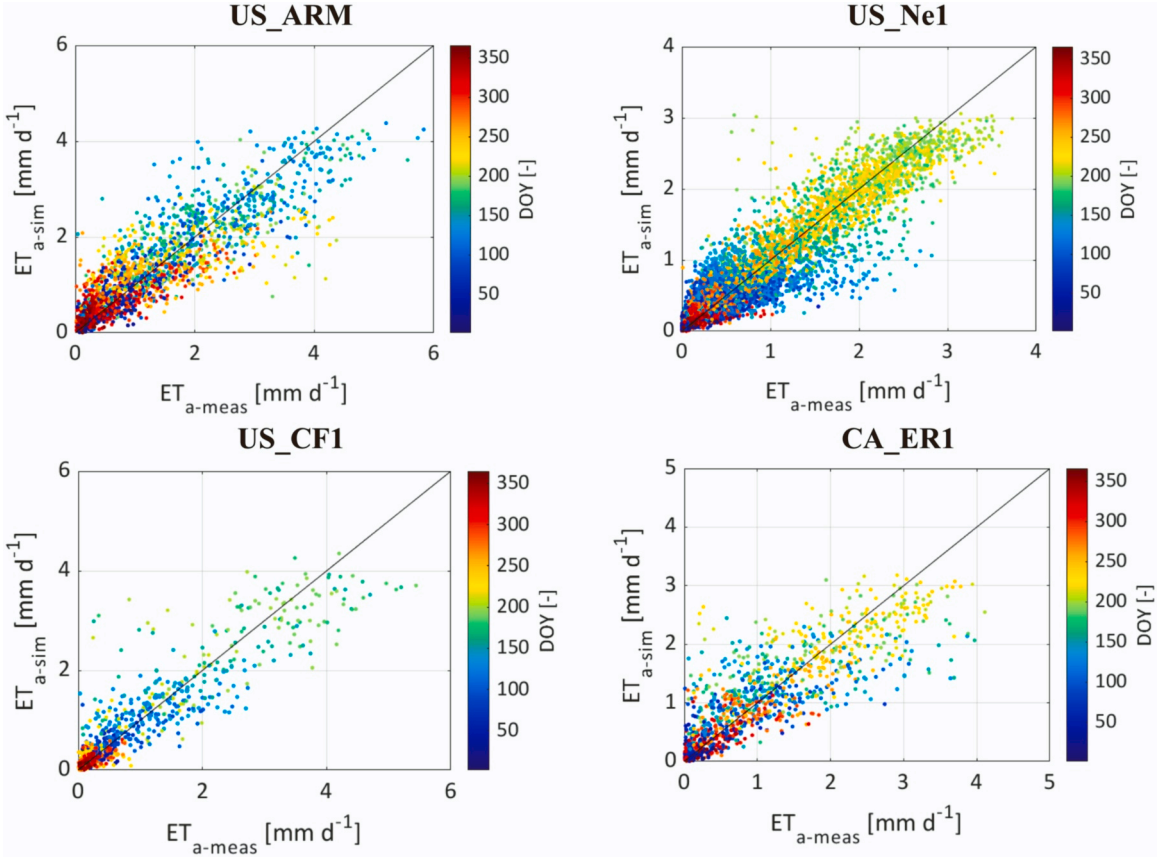


Fig. 9. Scatterplots of estimated versus measured ET_a obtained in the four sites by implementing the GPR model with a zero mean function, an exponential kernel covariance function, and the fifth combination of the input variables.

Table 7
Statistical indicators associated with the GPR model implemented with the best *m(x)* and *k(x,x')* functions and the fifth combination of the input variables, for the four sites in the United States of America and Canada.

		RMSE [mm d^{-1}]	MAE [mm d^{-1}]	R^2 [-]	b [-]	NSE [-]
5	**US_ARM**	0.57	0.41	0.74	0.90	0.74
	US_Ne1	0.30	0.21	0.87	0.93	0.87
	CA_ER1	0.52	0.36	0.70	0.86	0.70
	US_CF1	0.47	0.29	0.83	0.90	0.83

Fig. 9 shows the scatterplots between measured and estimated ET_a of daily actual evapotranspiration by assuming a zero mean function and an exponential kernel covariance function and the fifth combination of the input variables for each field. As can be noticed, the points are distributed differently according to the different climate classifications. As shown in Table 7, the RMSE values are in good agreement with the ones obtained in the other climatic zones. The values are always lower than 0.57 mm d^{-1}. Finally, the NSE values ranging from 0.70 to 0.87 confirm the model's possibility to estimate actual evapotranspiration values.

Finally, to have a clearer picture of the seasonal performances of the method, considering the best m(x) and k(x,x´), the NSE values were also evaluated by splitting the database in two. The division was done considering the cold seasons, fall and winter days, and the warm seasons, spring and summer days. The first included DOYs from 1 to 92 and from 275 to 366; the second from 93 to 274. The results, shown in Table 8, demonstrated that the model performances are better, in many cases, during the warm seasons than in the cold season. For two sites, US_ARM and US_CF1, the results are the best in the cold seasons but are also good in the others. For the warm seasons, these values are encouraging because these could permit good performance in the irrigation water management.

Table 8
Seasonal NSE values.

	NSE [-]	
	Fall/Winter	Spring/Summer
Villabate		
1	0.70	0.81
2	0.66	0.76
3	0.66	0.68
4	0.60	0.61
5	0.55	0.58
Castelvetrano		
3	0.39	0.45
4	0.33	0.43
5	0.35	0.45
US_ARM		
5	0.68	0.63
US_Ne1		
5	0.41	0.78
US_CF1		
5	0.60	0.55
CA_ER1		
5	0.65	0.79

5. Conclusion

In this study, the use of the Gaussian Process Regression (GPR) algorithm to fill gaps in the time series of daily actual evapotranspiration was investigated for six different sites located in different climates zone.

Initially, a detailed analysis of the available data, the methodologies, and tools adopted for the estimation of ET_a and the possible variables on which its value depends was carried out. In particular, a comparison between ground and reanalysis databases was made, which allowed to evaluate the suitability of reanalysis climate variables when ground data are not available. Two different satellites product, with different spatial resolutions, were compared, in order to analyze the possibility of use VIs, retrieved from images with a lower spatial resolution, when no other products are available. Five different combinations of variables were considered to verify the suitability of the ML approach when limited input data are available Then, the GPR model was tested. Initially, the best combination of mean, $m(x)$, and kernel covariance, $k(x,x')$, functions were identified considering all the input variables associated with the climate, soil water status as well as two vegetation indices expressive of plant vigor and surface water status in the citrus orchard where all the variables were available. Then, four additional variable combinations were also explored for the same orchard, to verify the performance of the estimation when limiting the number of input variables or considering only input variables freely downloadable from the web.

Further validation of the proposed ML algorithm with the three variable combinations not including SWC was carried out in the second experimental field, the olive orchard.

Finally, the approach was also applied using only remote sensed data for the experimental sites, located in different climate zones, in the United States of America and Canada.

The goodness of the proposed methodology is always confirmed by the quite high NSE index and by the low values of RMSE.

Therefore, the joint use of agrometeorological and remote sensing data with a GPR model can represent an opportunity to estimate missing data in the daily ET_a time series. Furthermore, in the zone where ground data are not available, the use of ERA5-Land agrometeorological archive and VIs retrieved from Sentinel-2 images or, in areas characterized by homogeneous soil and land use, from MODIS sensors can be considered a valid alternative to fill gaps in measured time series of daily actual evapotranspiration. Reliable values of actual evapotranspiration obtained using the proposed ML algorithm, allow for the estimation of accurate values of crop water requirements to improve water use in agriculture.

Future studies will assess the performance in ET_a estimations associated with the introduction, as an input variable, of high-resolution satellite-based soil moisture products.

Declaration of Competing Interest

The author declares that no conflict of interest exists.

Data Availability

Data will be made available on request.

Acknowledgements

The authors acknowledge the Ministero dell'Università e della Ricerca (MUR) and Università degli Studi di Palermo to co-finance the project "INtegrated Computer modeling and monitoring for Irrigation Planning in Italy, INCIPIT" (2017XWA834_006) in the frame of the PRIN 2017.

All the Authors organized experiments and edited the final version of the manuscript. Experimental setup, field data acquisition and processing were handled by DDC and MI. The authors wish to thank Massimo Mannino to ensure the maintenance of EC towers, Mr Emanuele Zarcone owner of the citrus orchard, and Mr Pierluigi Crescimanno owner of the olive orchard, where the experiments were carried out. The authors would like to express their profound gratitude to Prof. Giuseppe Provenzano (who sadly passed away in December 2022) for his significant contributions to the initial draft of this article, as well as for fostering extraordinary human and professional relationships.

References

Allen, R., Pereira, L., Raes, D., Smith, M., 1998. FAO Irrigation and drainage paper No. 56. Rome: Food and Agriculture Organization of the United Nations 56, 26–40.

Anderson, R.G., Alfieri, J.G., Tirado-Corbalá, R., Gartung, J., McKee, L.G., Prueger, J.H., Wang, D., Ayars, J.E., Kustas, W.P., 2017. Assessing FAO-56 dual crop coefficients using eddy covariance flux partitioning. Agric. Water Manag. 179, 92–102. https://doi.org/10.1016/j.agwat.2016.07.027.

Autovino, D., Minacapilli, M., Provenzano, G., 2016. Modelling bulk surface resistance by MODIS data and assessment of MOD16A2 evapotranspiration product in an irrigation district of Southern Italy. Agric. Water Manag. 167, 86–94. https://doi.org/10.1016/j.agwat.2016.01.006.

Awada, H., Ciraolo, G., Maltese, A., Provenzano, G., Hidalgo, M.A.M., Còrcoles, J.I., 2019. Assessing the performance of a large-scale irrigation system by estimations of actual evapotranspiration obtained by Landsat satellite images resampled with cubic convolution. Int. J. Appl. Earth Obs. Geoinf. 75, 96–105. https://doi.org/10.1016/j.jag.2018.10.016.

Bastiaanssen, W.G.M., Menenti, M., Feddes, R.A., Holtslag, A.A.M., 1998. A remote sensing surface energy balance algorithm for land SEBAL. 1. Formulation. J. Hydrol. 212–213, 198–212. https://doi.org/10.1016/S0022-1694(98)00253-4.

Biraud, S., Fischer, M., Chan, S., Torn, M., 2021. AmeriFlux BASE US-ARM ARM Southern Great Plains site- Lamont, Ver. 11–5, AmeriFlux AMP, (Dataset). https://doi.org/10.17190/AMF/1246027.

Burgess, P.J., Reinhard, B.R., Pasturel, P., 2006. Compatible measurements of volumetric soil water content using a neutron probe and Diviner 2000 after field calibration. Soil Use Manag 22, 401–404. https://doi.org/10.1111/j.1475-2743.2006.00060.x.

Cammalleri, C., Rallo, G., Agnese, C., Ciraolo, G., Minacapilli, M., Provenzano, G., 2013. Combined use of eddy covariance and sap flow techniques for partition of ET fluxes and water stress assessment in an irrigated olive orchard. Agric. Water Manag. 120, 89–97. https://doi.org/10.1016/j.agwat.2012.10.003.

Campora, M., Palla, A., Gnecco, I., Bovolenta, R., Passalacqua, R., 2020. The laboratory calibration of a soil moisture capacitance probe in sandy soils. Soil Water Res 15, 75–84. https://doi.org/10.17221/227/2018-SWR.

Carter, C., Liang, S., 2019. Evaluation of ten machine learning methods for estimating terrestrial evapotranspiration from remote sensing. Int. J. Appl. Earth Obs. Geoinf. 78, 86–92. https://doi.org/10.1016/j.jag.2019.01.020.

Corbari, C., Skokovic, D., Sobrino, J., Mancini, M., 2020. Evapotranspiration estimates at high spatial and temporal resolutions from an energy–water balance model and satellite data in the capitanata irrigation consortium. Remote Sens. 12 doi:10.3390/rs12244083.

Eisenhauer, J.G., 2003. Regression through the Origin. Teach. Stat. 25, 76–80. https://doi.org/10.1111/1467-9639.00136.

Er-Raki, S., Chehbouni, A., Guemouria, N., Ezzahar, J., Khabba, S., Boulet, G., Hanich, L., 2009. Citrus orchard evapotranspiration: comparison between eddy covariance measurements and the FAO-56 approach estimates. Plant Biosyst. - Int. J. Deal. Asp. Plant Biol. 1431 201–208. https://doi.org/10.1080/11263500802709897.

Faramiñán, A.M.G., Degano, M.F., Carmona, F., Rodriguez, P.O., 2021. Estimation of actual evapotranspiration using NASA-POWER data and Support Vector Machine. 2021 XIX Workshop on Information Processing and Control RPIC, pp. 1–5. doi: 10.1109/RPIC53795.2021.9648425.

Fiebrich, C.A., Morgan, C.R., McCombs, A.G., Hall, P.K., McPherson, R.A., 2010. Quality assurance procedures for mesoscale meteorological data. J. Atmos. Ocean. Technol. 27, 1565–1582. https://doi.org/10.1175/2010JTECHA1433.1.

Fischer, M.L., Billesbach, D.P., Berry, J.A., Riley, W.J., Torn, M.S., 2007. Spatiotemporal variations in growing season exchanges Of Co2, H2o, and sensible heat in agricultural fields of the southern great plains. Earth Interact. 11 (17), 1–21. https://doi.org/10.1175/EI231.1.

French, A.N., Hunsaker, D.J., Sanchez, C.A., Saber, M., Gonzalez, J.R., Anderson, R., 2020. Satellite-based NDVI crop coefficients and evapotranspiration with eddy covariance validation for multiple durum wheat fields in the US Southwest. Agric. Water Manag. 239, 106266 https://doi.org/10.1016/j.agwat.2020.106266.

Gao, B.C., 1996. NDWI - A normalized difference water index for remote sensing of vegetation liquid water from space. Remote Sens Environ. 58 (3), 257–266. https://doi.org/10.1016/S0034-4257(96)00067-3.

Gaskin, G.J., Miller, J.D., 1996. Measurement of soil water content using a simplified impedance measuring technique. J. Agric. Eng. Res. 63, 153–160.

Granata, F., 2019. Evapotranspiration evaluation models based on machine learning algorithms—a comparative study. Agric. Water Manag. 217, 303–315. https://doi.org/10.1016/j.agwat.2019.03.015.

Ippolito, M., De Caro, D., Ciraolo, G., Minacapilli, M., Provenzano, G., 2022. Estimating crop coefficient and actual evapotranspiration in citrus orchards with sporadic cover weeds based on ground and remote sensing data. Irrig. Sci. 1–18. https://doi.org/10.1007/s00271-022-00829-4.

Jing, W., Yassen, Z.M., Shahid, S., Saggi, M.K., Tao, H., Kisi, O., Salih, S.Q., Al-Ansari, N., Chau, K.W., 2019. Implementation of evolutionary computing models for reference evapotranspiration modelling: short review, assessment and possible future research directions. Eng. Appl. Comput. Fluid Mech. 13 (1), 811–823. https://doi.org/10.1080/19942060.2019.1645045.

Kang, M., Ichii, K., Kim, J., Indrawati, Y.M., Park, J., Moon, M., Lim, J.H., Chun, J.H., 2019. New gap-filling strategies for long-period flux data gaps using a data-driven approach. Atmosphere 10 (10). https://doi.org/10.3390/atmos10100568.

Kim, N., Kim, K., Lee, S., Cho, J., Lee, Y., 2020. Retrieval of daily reference evapotranspiration for croplands in south korea using machine learning with satellite images and numerical weather prediction data. Remote Sens. 1221 doi: 10.3390/rs12213642.

Kottek, M., Grieser, J., Beck, C., Rudolf, B., Rubel, F., 2006. World Map of the Köppen-Geiger climate classification updated. Meteorol. Z. 15, 259–263. https://doi.org/10.1127/0941-2948/2006/0130.

Krishnashetty, P.H., Balasangameshwara, J., Sreeman, S., Desai, S., Kantharaju, A.B., 2021. Cognitive computation models for estimation of reference evapotranspiration: A review. Cogn. Syst. Res. 70, 109–116. https://doi.org/10.1016/j.cogsys.2021.07.012.

Kustas, W.P., Zhan, X., Jackson, T.J., 1999. Mapping surface energy flux partitioning at large scales with optical and microwave remote sensing data from Washita '92. Water Resour. Res. 351 265–277.

Lei, H., Yang, D., 2014. Combining the crop coefficient of winter wheat and summer maize with a remotely sensed vegetation index for estimating evapotranspiration in the North China Plain. J. Hydrol. Eng. 191, 243–251 doi: 10.1061/(ASCE)HE.1943-5584.0000765.

Manca G., 2003. Analisi dei flussi di carbonio di una cronosequenza di cerro (Quercus cerris L.) dell'Italia centrale attraverso la tecnica della correlazione turbolenta. PhD Dissertation Thesis. Universita` degli Studi della Tuscia, Viterbo, 225 pp (in Italian).

Marletto, V., Ventura, F., Fontana, G., Tomei, F., 2007. Wheat growth simulation and yield prediction with seasonal forecasts and a numerical model. Agric. For. Meteorol. 1471 71–79. https://doi.org/10.1016/j.agrformet.2007.07.003.

Matern, B., 1986. Spatial variation: stochastic models and their application to some problems in forest surveys and other sampling investigations. Springer-Verlag,, Berlin.

Mauder, M., Foken, T., Clement, R., Elbers, J.A., Eugster, W., Grünwald, T., Heusinkveld, B., Kolle, O., 2008. Quality control of CarboEurope flux data – Part 2: Inter-comparison of eddy-covariance software. Biogeosciences 5, 451–462. https://doi.org/10.5194/bg-5–451-2008.

Moncrieff, J.B., Massheder, J.M., de Bruin, H.A.R., Elbers, T., Friborg, J., Heusinkveld, B., et al., 1997. A system to measure surface fluxes of momentum, sensible heat, water vapour and carbon dioxide. J. Hydrol. 188–189, 589–611.

Mosre, J., Suárez, F., 2021. Actual evapotranspiration estimates in arid cold regions using machine learning algorithms with in situ and remote sensing data. Water 136. https://doi.org/10.3390/w13060870.

Mosteller, F., Tukey, J.W., 1968. Data analysis, including statistics. In: Lindzey, G., Aronson, E. (Eds.), Handbook of Social Psychology, first edition 1954, Vol. 2. Addison–Wesley.

Muñoz Sabater, J., 2019. ERA5-Land hourly data from 1981 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). (Accessed on <10-Jan-2022>), 10.24381/cds.e2161bac.

Muñoz-Sabater, J., Dutra, E., Agusti-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D.G., Piles, M., Rodriguez-Fernández, N.J., Zsoter, E., Buontempo, C., Thépaut, J.N., 2021. ERA5-Land: a state-of-the-art global reanalysis dataset for land applications. Earth Syst., Sci. Data 13 (9), 4349–4383. https://doi.org/10.5194/essd-13-4349-2021.

Murphy, K., 2012. Machine learning: a probabilistic. Perspective Vol. 58.

Namasudra, S., Dhamodharavadhani, S., Rathipriya, R., 2021. Nonlinear neural network based forecasting model for predicting COVID-19 cases. Neural Process Lett. https://doi.org/10.1007/s11063-021-10495-w.

Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I — a discussion of principles. J. Hydrol. 103, 282–290. https://doi.org/10.1016/0022-1694(70)90255-6.

Nguyen, X.C., Nguyen, T.T.H., La, D.D., Kumar, G., Rene, E.R., Nguyen, D.D., Chang, S. W., Chung, W.J., Nguyen, X.H., Nguyen, V.K., 2021. Development of machine learning - based models to forecast solid waste generation in residential areas: a case study from Vietnam. Resour., Conserv. Recycl. 167, 105381 https://doi.org/10.1016/j.resconrec.2020.105381.

Pagano, A., Amato, F., Ippolito, M., De Caro, D., Croce, D., Motisi, A., Provenzano, G., Tinnirello, I., 2023. Machine learning models to predict daily actual evapotranspiration of citrus orchards under regulated deficit irrigation, 102133, ISSN 1574-9541 Ecol. Inform. Volume 76, 2023. https://doi.org/10.1016/j.ecoinf.2023.102133.

Pan, Y., Zeng, X.K., Xu, H., Sun, Y., Wang, D., Wu, J., 2021. Evaluation of Gaussian process regression kernel functions for improving groundwater prediction. J. Hydrol. 2021 603, 126960.

Pan, Z., Hu, Y., Cao, B., 2017. Construction of smooth daily remote sensing time series data: a higher spatiotemporal resolution perspective. Open Geospatial Data Softw. Stand. 2 https://doi.org/10.1186/s40965-017-0038-z.

Paredes, P., Martins, D.S., Pereira, L.S., Cadima, J., Pires, C., 2018. Accuracy of daily estimation of grass reference evapotranspiration using ERA-Interim reanalysis products with assessment of alternative bias correction schemes. Agric. Water Manag. 210, 340–353. https://doi.org/10.1016/j.agwat.2018.08.003.

Pastorello, G., Agarwal, D., Papale, P. et al., 2014. Observational data patterns for time series data quality assessment, paper presented at e-Science (e-Science), IEEE 10th International Conference on e-Science, Sao Paulo. DOI:10.1109/eScience.2014.45.

Pastorello, G., Trotta, C., Canfora, E., et al., 2020. The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data. Sci. Data 7, 225. https://doi.org/10.1038/s41597-020-0534-3.

Pelosi, A., Chirico, G.B., 2021. Regional assessment of daily reference evapotranspiration: Can ground observations be replaced by blending ERA5-Land meteorological reanalysis and CM-SAF satellite-based radiation data? Agric. Water Manag. 258, 107169 https://doi.org/10.1016/j.agwat.2021.107169.

Pereira, L.S., Paredes, P., Jovanovic, N., 2020. Soil water balance models for determining crop water and irrigation requirements and irrigation scheduling focusing on the FAO56 method and the dual Kc approach. Agric. Water Manag. 241, 106357 https://doi.org/10.1016/j.agwat.2020.106357.

Pereira, L.S., Paredes, P., Hunsaker, D.J., López-Urrea, R., Mohammadi Shad, Z., 2021. Standard single and basal crop coefficients for field crops. Updates and advances to the FAO56 crop water requirements method. Agric. Water Manag. 243, 106466 https://doi.org/10.1016/j.agwat.2020.106466.

Phillips, C.L., Huggins, D., 2022. AmeriFlux BASE US-CF1 CAF-LTAR Cook East, Ver. 3–5, AmeriFlux AMP, (Dataset). https://doi.org/10.17190/AMF/1543382.

Pôças, I., Calera, A., Campos, I., Cunha, M., 2020. Remote sensing for estimating and mapping single and basal crop coefficients: a review on spectral vegetation indices approaches. Agric. Water Manag. 233, 106081 https://doi.org/10.1016/j.agwat.2020.106081.

Prueger, J.H., Hatfield, J.L., Parkin, T.B., Kustas, W.P., Hipps, L.E., Neale, C.M.U., MacPherson, J.I., Eichinger, W.E., Cooper, D.I., 2005. Tower and aircraft eddy covariance measurements of water vapor, energy, and carbon dioxide fluxes during SMACEX. J. Hydrometeorol. 66, 954–960. https://doi.org/10.1175/JHM457.1.

Rallo, G., Provenzano, G., 2013. Modelling eco-physiological response of table olive trees (Olea europaea L.) to soil water deficit conditions. Agric. Water Manag. 120, 79–88. https://doi.org/10.1016/j.agwat.2012.10.005.

Rallo, G., Paço, T.A., Paredes, P., Puig-Sirera, À., Massai, R., Provenzano, G., Pereira, L. S., 2021. Updated single and dual crop coefficients for tree and vine fruit crops. Agric. Water Manag. 250, 106645 https://doi.org/10.1016/j.agwat.2020.106645.

Ranghetti, L., Boschetti, M., Nutini, F., Busetto, L., 2020. sen2r: An R toolbox for automatically downloading and preprocessing Sentinel-2 satellite data. Comput. Geosci. 139, 104473 https://doi.org/10.1016/j.cageo.2020.104473.

Rasmussen, C.E., Williams, C.K.I., 2006. Gaussian Processes for Machine Learning. The MIT Press. ISBN 026218253X. Massachusetts Institute of Technology. www.GaussianProcess.org/gpml.

Raz-Yaseef, N., Billesbach, D.P., Fischer, M.L., Biraud, S.C., Gunter, S.A., Bradford, J.A., Torn, M.S., 2015. Vulnerability of crops and native grasses to summer drying in the U.S. southern great plains, agriculture. Ecosyst. Environ. 213, 209–218. https://doi.org/10.1016/j.agee.2015.07.021.

Roberts, S.J., Osborne, M.A., Ebden, M., Reece, S., Gibson, N., Aigrain, S., 2013. Gaussian processes for time-series modelling. Philos. Trans. R. Soc. A. 371, 20110550 https://doi.org/10.1098/rsta.2011.0550.

Rodrigues, G.C., Braga, R.P., 2021. Evaluation of NASA-POWER reanalysis products to estimate daily weather variables in a hot summer mediterranean climate. Agronomy 116. https://doi.org/10.3390/agronomy11061207.

Rouse, J.W., Haas, R.H., Schell, J.A., Deering, D.W., 1974. Monitoring vegetation systems in the Great Plains with Erts, Third ERTS-1 Symposium NASA, NASA SP-351, Washington DC, 309–317.

Rozenstein, O., Haymann, N., Kaplan, G., Tanny, J., 2019. Validation of the cotton crop coefficient estimation model based on Sentinel-2 imagery and eddy covariance measurements. Agric. Water Manag. 223, 105715 https://doi.org/10.1016/j.agwat.2019.105715.

Saitta, D., Vanella, D., Ramírez Cuesta, J., Longo Minnolo, G., Ferlito, F., Consoli, S., 2020. Comparison of orange orchard evapotranspiration by eddy covariance, sap flow, and FAO-56 methods under different irrigation strategies, 05020002 J. Irrig. Drain. Eng. 146. https://doi.org/10.1061/(ASCE)IR.1943-4774.0001479.

Schaaf, C., Wang, Z., 2015. MCD43A4 MODIS/Terra+Aqua BRDF/albedo nadir BRDF adjusted ref daily L3 global - 500m V006 (Dataset). NASA EOSDIS Land Process. DAAC. https://doi.org/10.5067/MODIS/MCD43A4.006.

Shi, X., Jiang, D., Qian, W., Liang, Y., 2022. Application of the gaussian process regression method based on a combined kernel function in engine performance prediction. ACS Omega2022 7, 41732–41743. https://doi.org/10.1021/acsomega.2c05952.

Singh, A.K., Bhardwaj, A.K., Verma, C.L., Mishra, V.K., Singh, A.K., Arora, S., Sharma, N., Ojha, R.P., 2019. Soil moisture sensing techniques for scheduling irrigation. J. Soil Salin. Water Qual. 11, 68–76.

Stein, M.L., 1999. Interpoletion of Spatial Data. Springer-Verlag, New York.

Sun, Y., Zhou, H., Qin, Y., Schulze Lammers, P., Berg, A., Deng, H., Cai, X., Wang, D., Jones, S.B., 2014. Horizontal monitoring of soil water content using a novel automated and mobile electromagnetic access-tube sensor. J. Hydrol. 516, 50–55. https://doi.org/10.1016/j.jhydrol.2014.01.067.

Suyker, A., 2022. AmeriFlux BASE US-Ne1 Mead - irrigated continuous maize site (AmeriFlux AMP, (Dataset)). Ver 13–15. https://doi.org/10.17190/AMF/1246084.

Talib, A., Desai, A.R., Huang, J., Griffis, T.J., Reed, D.E., Chen, J., 2021. Evaluation of prediction and forecasting models for evapotranspiration of agricultural lands in the midwest us. J. Hydrol. 600, 126579.

United Nations, The United Nations World Water Development Report 2021: Valuing Water. UNESCO, Paris.

Van Dam, J.C. , Huygen, J. , Wesseling, J.G. , Feddes, R.A. , Kabat, P. , van Walsum, P.E. V , Groenendijk, P. , van Diepen, C.A. , 1997. Theory of SWAP version 2.0. Report 71. Technical Document 45, Wageningen, 167 pp.

Vanella, D., Longo-Minnolo, G., Belfiore, O.R., Ramírez-Cuesta, J.M., Pappalardo, S., Consoli, S., D'Urso, G., Chirico, G.B., Coppola, A., Comegna, A., Toscano, A., Quarta, R., Provenzano, G., Ippolito, M., Castagna, A., Gandolfi, C., 2022. Comparing the use of ERA5 reanalysis dataset and ground-based agrometeorological data under different climates and topography in Italy. J. Hydrol.: Reg. Stud. 42 https://doi.org/10.1016/j.ejrh.2022.101182.

Wagner-Riddle, C., 2021. AmeriFlux BASE CA-ER1 Elora Research Station, Ver. 3–5, AmeriFlux AMP, (Dataset). https://doi.org/10.17190/AMF/1579541.

Walls, S., Binns, A.D., Levison, J., MacRitchie, S., 2020. Prediction of actual evapotranspiration by artificial neural network models using data from a bowen ratio energy balance station. Neural Comput. Appl. 32 (17), 14001–14018.

Yang, S.K., Li, H.M., Gou, X.D., Bian, C., Shao, Q., 2022. Optimized Bayesian adaptive resonance theory mapping model using a rational quadratic kernel and Bayesian quadratic regularization. Appl. Intell. 2022 (52), 7777–7792.

Zhai, Z., Martínez, J.F., Beltran, V., Martínez, N.L., 2020. Decision support systems for agriculture 4.0: survey and challenges. Comput. Electron. Agric. 170, 105256 https://doi.org/10.1016/j.compag.2020.105256.