remote sensing

Supplementary material of: Synergistic Use of Multispectral Data and Crop Growth Modelling for Spatial and Temporal Evapotranspiration Estimations

Supplement to section 2: Materials and Methods

2.1. SAFY_swb and the Ensemble Kalman Filter Data Assimilation Method

The assimilation scheme proposed in this paper is summarized in **Error! Reference source not found.**.

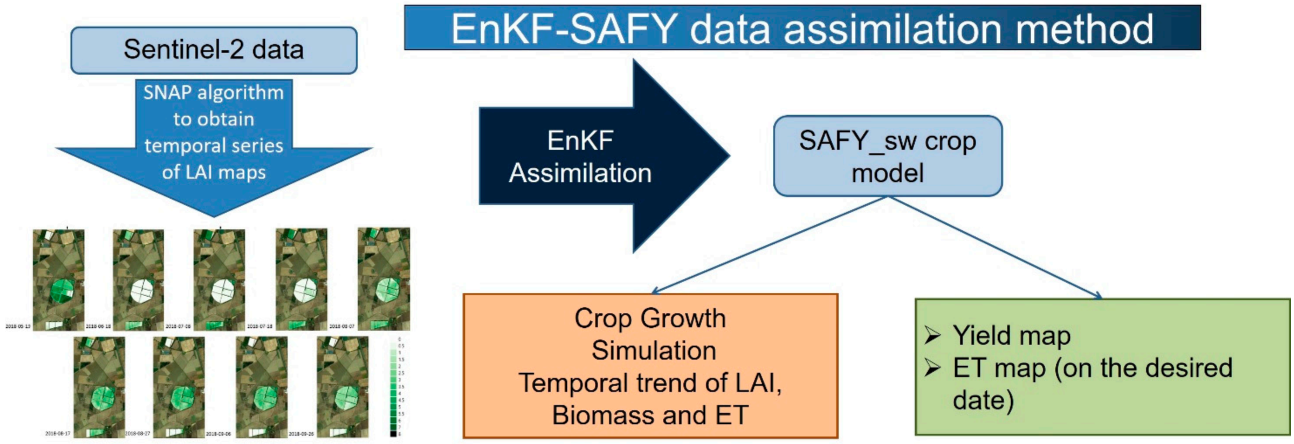

Figure 1. Simplified diagram of the EnKF-SAFY_swb assimilation algorithm.

2.1.1. SAFY-swb

Simple Algorithm For Yield (SAFY) [1] is a crop growth model mainly driven by the photosynthetically active solar radiation absorbed by plants (APAR). It is an application of the Monteith's concept [2], which describes the daily growth of dry aboveground biomass (DAM) as a function of the incoming global radiation (R_g) and Leaf Area Index (LAI) , as shown in Equations (1) and (2).

$$\Delta DAM(i) = R_g(i) \cdot R2P \cdot \varepsilon_I(i-1) \cdot ELUE \cdot F_T \quad (1)$$

Where R2P is the global to incoming photosynthetically active radiation (PAR), ε_I is the light interception efficiency (expressed in Equation (2)), ELUE (expressed in [g·MJ−1]) is the light use efficiency, and F_T is a temperature stress function [21].

The light interception efficiency is represented as follow:

$$\varepsilon_I(i-1) = 1 - e^{-k_{ex} \cdot LAI(i-1)} \quad (2)$$

Where k is the light interception coefficient and LAI is the Leaf Area Index.

F_T is expressed as follow:

$$F_T = \begin{cases} 1-\left(\dfrac{T_{opt}-T_a}{T_{opt}-T_{min}}\right)^2 \ if\ T_{min} < T_a < T_{opt} \\ 1-\left(\dfrac{T_a-T_{opt}}{T_{maxt}-T_{opt}}\right)^2 \ if\ T_{max} > T_a > T_{opt} \\ 0\ if\ T_a < T_{min}\ or\ T_a > T_{max} \end{cases} \tag{3}$$

Where: T_a = air temperature, T_{opt}=optimal temperature for crop functioning, T_{min} and T_{max}=extreme temperature values for crop functioning.

The dynamics of leaf area index (LAI) is simulated from the balance between leaf extent during growth ($\Delta LAI+, eq. 4$) and leaf disappearance during senescence (($\Delta LAI-, eq. 6$)). These two phenological phases are identified based on a degree-day approach from accumulated air temperature (thermal time ΣT_a). During growth, the aerial phytomass production is distributed into leaf and non-leaf mass according to the partition function P_L, then the increase of leaf mass is converted in in- crease of leaf area (($\Delta LAI +$,) according to the value of the specific leaf area (SLA). This leads:

$$\Delta LAI+= \Delta DAM \cdot P_L(\sum T_a) \cdot SLA \tag{4}$$

Where P_L is an empirical function of two parameters, P_{La} and P_{Lb} [21], and it is expressed as follow:

$$P_L\left(\sum T_a\right) = 1 - P_{La} \cdot e^{P_{Lb} \cdot \Sigma T_a} \tag{5}$$

The senescence of leaves starts when accumulated air temperature has reached a given threshold (STT). It increases with thermal time at a rate determined by the Rs parameter. It ends when LAI has returned to a value lower than the initial one, indicating total senescence. This leads to:

$$\Delta LAI-= LAI \cdot \sum T_a \cdot S_{TT}/\mathrm{R_S} \tag{6}$$

A detailed description of SAFY algorithms is provided by [1].

Previous studies [3–7] have proved the efficiency of this model in estimating the production of different crops, especially when implemented with estimated LAI measurements obtained from EO data. This is due to the simplicity of the model, a feature that makes it particularly suitable for use in synergy with EO data.

SAFY has been modified by [4]. The authors introduced a soil water balance to simulate root water uptake and crop water stress dynamics. In that version the soil profile is represented as a five layers soil, with a depth for each soil (from top to bottom) respectively of: 10 cm, 25 cm, 50 cm, 100 cm and 150 cm. For each soil layer the water balance is calculated for each day of the crop cycle. Following the changes made by [4], Equation (1) becomes:

$$\Delta DAM(i) = R_g(i) \cdot R2P \cdot \varepsilon_I(i-1) \cdot ELUE \cdot F_T \cdot WST(i) \tag{7}$$

where $WST(i)$ is a daily function expressed as the ratio of actual and potential plant transpiration, which proportionally affects (decreases) the daily biomass accumulation. WST represents the daily water stress to which the crop is exposed, and it is a function of the water balance, in turn a function of precipitation, runoff, drainage, and evapotranspiration.

Evapotranspiration (ET) is an important component of the water balance and it is composed of soil evaporation and plant transpiration. In the SAFY version proposed by [4] ET contribution to the soil water balance is considered separately. The Potential ET (PET) is initially calculated using the Priestley-Taylor, while potential plant transpiration (PTrsp) and potential soil evaporation (PESoil) are calculated as functions of LAI and PET. The actual soil evaporation (AESoil) is calculated as a function of soil water balance, and finally actual plant transpiration is calculated as a function of potential root water uptake.

A more detailed explanation of the changes made to SAFY that introduces the water balance, is provided by [4]. The changes made by [4]ensure that the daily biomass increase (Equation (7)) is strictly linked to two biophysical variables that can be measured by EO data: LAI and ET.

In the present study, some small changes have been made to the version of SAFY presented by [4]. In this case, the quantity of water supplied through irrigation was included in the calculation of the water balance. Additionally, the evaluation of potential evapotranspiration was calculated using the Penman-Monteith instead of the Priestley-Taylor equation used by [4]. Actually, in the current implementation, the reference evapotranspiration (ET_0) was calculated using ET_0 Calculator, a software developed by FAO [9], from which PET was obtained using an appropriate specific crop coefficient, as suggested by [9]. Thus ET_0 becomes one of the input meteorological variables of SAFY, PET is calculated only according to the selected crop, thus avoiding the repetition of the calculation at each single run of the model. Hereafter we will refer to the SAFY model used in this study (based on the version proposed by Kang and modified as previously described) as SAFY_swb.

2.1.2. Ensemble Kalman Filter Method Assimilation

The DA scheme used in this work for SAFY_swb, hereafter called EnKF-SAFY_swb, is based on the method developed by [13].

The EnKF algorithm developed for SAFY in this work is based on the theory of [3] ,which considers the observations as random variables, therefore adding random perturbations to the observed values. It is possible to divide the assimilation algorithm into the following steps:

1. Generation of an ensemble of n = 200 vectors, each one containing the values of the i parameters P_i, in this case i = 1, 2, . . . , 9. To each element, corresponding to a nominal parameter value, an error value was added, randomly drawn from a truncated normal distribution N(0, 1), with lower and upper limits.
2. Simulations with the SAFY model in order to obtain a value of LAI for each element of the ensemble at the date when the first satellite image was acquired. An error ε is added to the simulated LAI values. This error ε was randomly generated from a normal distribution N(0, Q), in which the standard deviation Q was arbitrarily chosen as 20% of the LAI value, to take into account the uncertainty of the model. In this way a matrix φ_{t1} was defined:

$$\varphi_{t1} = \begin{bmatrix} L^1 L^2 \dots L^{100} \\ P_1^1 P_1^2 \dots P_1^{100} \\ \dots \dots \dots \dots \\ P_5^1 P_5^2 \dots P_5^{100} \end{bmatrix} \quad (8)$$

where each column is an element of the ensemble and represents a random configuration of the model using parameters P_i and the corresponding LAI value obtained L^n at the time t_1 of the first satellite acquisition.

3. Generation of the vector M_{t1}, where each element is composed by the LAI observed at time t_1, i.e., retrieved from remote sensing data, plus an error τ_{t1}^j drawn from N[0, var(τ_{t1}^j)], where var(τ_{t1}^j) is a variance expressing the measurement error, which in our case was inferred from the comparison of the LAI values estimated from remote sensing, with the ground measurements.
4. Computation of the variance-covariance matrix of φ_{t1} for the 100 ensemble elements.
5. Calculation of the Kalman gain using the variance-covariance matrix:

$$K_{t1} = \frac{\sum_{t_1} R^T}{R \sum R^T + \mathrm{var}(\tau_{t1})} \quad (9)$$

Where: R = (1,0,…,0) with number of elements equal to the number of ensemble elements.

6. Update φ_{t1} of as follows:

$$\varphi_{t1,K}^{j} = \varphi_{t1}^{j} + K_{t1}(M_{t1}^{j} - R\varphi_{t1}^{j}) \quad (10)$$

Where j is the j[th] column of φ_{t1}.

7. Replacement of LAI and parameters values (the elements of φ_{t1}) with those calculated at step 6 (the elements of $\varphi_{t1,K}^{j}$).
8. Repetition from step 3 for each satellite observation date. When the last observation has been assimilated, SAFY runs to the end of the crop growth cycle and outputs the yield.

The biophysical variable assimilated is the LAI. In a first evaluation step, LAI was calculated from synthetic data to evaluate the assimilation efficiency of the method (see section 2.2). Subsequently it was estimated from EO data acquired by MSI (the multispectral instrument on board of Sentinel-2) and processed using the SNAP module for LAI assessment [44] (more details in section 2.3.2).

The number of elements for each ensemble was set to 200. As reported in literature [3,10], 100 elements are a good compromise between the minimization of the random component typical of EnKF and the computational cost of the algorithm. Having improved the computational cost of the algorithm, it was possible to double the number of ensembles to further minimize the random component.

Two other important parameters of EnKF scheme are the error on the model simulations (ε) and the error on the measurements τ [11], the first was adjusted as suggested by [3], the second was calibrated according to the validation of the LAI estimates retrieved from Sentinel-2 data using the SNAP S2Toolbox module made by [12].

2.2. Assimilation Efficiency Assessment

In order to evaluate the assimilation efficiency of EnKF-SAFY_swb, a procedure based on synthetic data, proposed by [13], was applied. This allowed to rigorously quantify the advantage of LAI assimilation (using the EnKF method) in SAFY_swb, compared to the simple use (called open loop) of the model. Furthermore, using synthetic data, it was possible to generate different scenarios not only for environmental and meteorological conditions, but also for the number of observations to be assimilated and different errors on the measurements. In this way the assimilation efficiency has been evaluated for several LAI errors and numbers of observations to be assimilated. This configuration proposed by [13], defined as general case, is presented in more detail in section 2.2.1. Once the ranges of measurement error and the number of observations for which it is convenient to use EnKF-SAFY_swb was established, the scheme proposed by [13] was repeated for a specific case (section 2.2.2). The specific case is similar, in terms of number of observations and LAI error, to the real case subsequently studied, but with a dataset composed by several heterogeneous scenarios (section 2.3).

The idea of the synthetic data procedure is to simulate the crop cycle in a hypothetic scenario using the crop model (in this case SAFY_swb) and assuming the combination of inputs and the consequent outputs represent “true” values. Then the model is run again, using a standard calibration, either using or not the assimilation of an “observed” variable. The latter is obtained using the “true” value plus an error). Finally, the results obtained using or not the assimilation are compared, in order to obtain an index that quantifies the assimilation efficiency.

The procedure is composed by 5 phases (Figure 2). In phase 1 the input dataset, composed by observed meteorological data, independent parameters, and environmental-dependant parameters, is set.

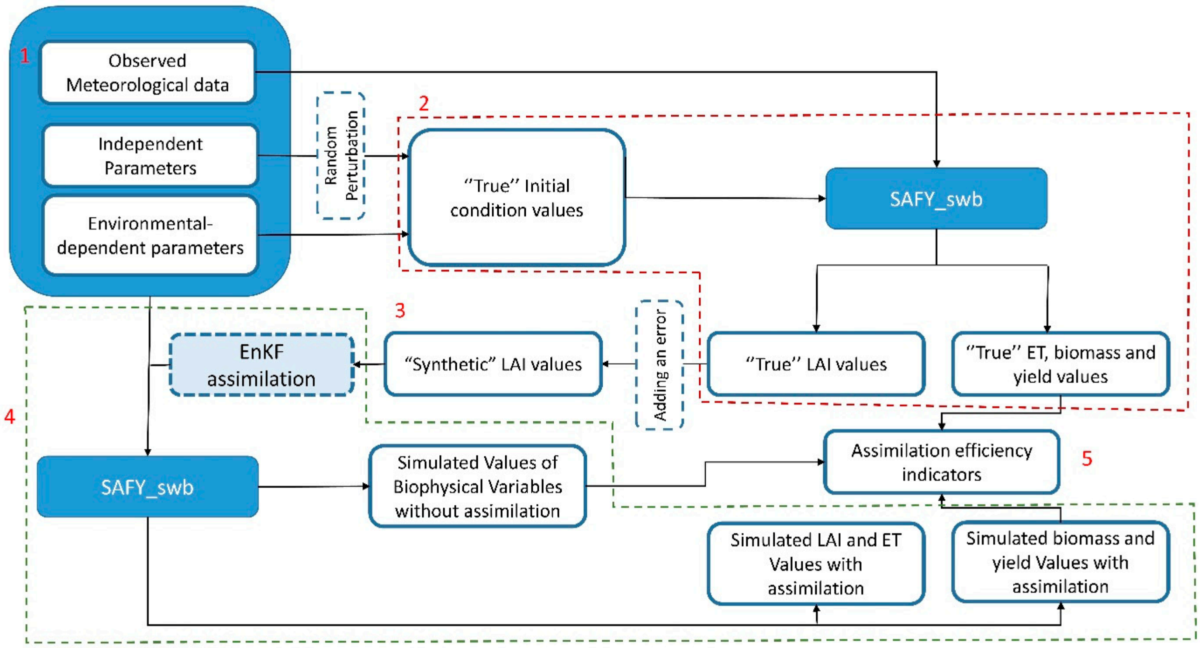

Figure 2. Scheme for the analysis of the assimilation efficiency of the EnKF-SAFY_swb method.

The input dataset was set up in this study starting from five real scenarios. These scenarios are described in Table 1. Daily climatic information, soil texture and type of crop are the characteristics considered to define the five initial scenarios through a series of parameters. The five real case scenarios were selected for covering a large climatic, soil and crop conditions, in order to create a robust synthetic dataset for diverse conditions. Temperatures, precipitations and solar radiation described the climatic characteristics. They were collected from literature [3,4,14] and are not varied to generate the synthetic dataset.

Table 1. Overview of the five initial scenarios used to generate the synthetic dataset. Since soil characteristics are not available, the parameters referring to them have been set on typical values from literature [48,49].

ID	Geographic Location	Climatic Characteristics	Soil Characteristics	Crop
Scenario 1	Lat: 45.20 Lon: 10.84	Precipitations per year are around 233 mm. They are mainly in late spring and summer. Temperatures between -6°C in winter and 37 in summer, average annual temperature of 17°C.	Silt	Maize
Scenario 2	Lat: 41.87 Lon:-93.08	Precipitations per year are around 639 mm. They are mainly in late spring and summer. Temperatures between -16°C in winter and 25 in summer, average annual temperature of 17°C.	Loam	Maize
Scenario 3	Lat: 41.89 Lon: 12.19	Precipitations per year are around 761 mm. Rainfall is distributed throughout the 4 seasons, especially in autumn. Heavy but sporadic rainfall in summer. Temperatures between 5°C in winter and 32 in summer, average annual temperature of 13°C.	Sandy clay loam	Winter wheat

Scenario 4	Lat: 43.02 Lon: 11.68	Precipitations per year are around 674 mm. Rainfall is distributed throughout the year, particularly abundant in autumn. Temperatures between -3°C in winter and 30 in summer, average annual temperature of 11°C.	Loam	Winter wheat
Scenario 5	Lat: 40.18 Lon: 116.34	Precipitations per year are around 306 mm, almost totally during summer. Cold and dry winter with minimum temperatures of -12°C, wet and warm summer with maximum temperatures of 35°C. Average annual temperature of 13°C.	Clay loam	Winter wheat

In this study, we have defined as "independent parameters" those which describe a scenario and which do not uniquely depend on a single characteristic, but on a combination of characteristics. The list of independent parameters is shown in Table 2. The random variation of these parameters generated several fields for each scenario that made up the synthetic dataset.

Table 2. List of independent parameters. Parameters of SAFY_swb not uniquely connected to crop, management, or soil.

ID	Parameter Name	Unit
Pfen_PrtA	Partition to leaf function	-
Pfen_PrtB	Partition to leaf function: parameter 2	-
Pfen_SenA	Sum of temperature for senescence	°C
Pfen_SenB	Rate of senescence	°C/day
Pfen_MrgD	Day of emergence	Day (in day of year)
Pgro_Lue	Effective Light-use efficiency	g /MJ
Pgro_R2P	Global to PAR incident radiation ratio	-
Pgro_Kex	Extinction of Radiation in Canopy	-
Pgro_Sla	Specific Leaf-Area	m^2/g
Pgro_P2G	Partition Coefficient to Grain	1/°C
Pgro_Ms0	Initial dry above-ground global variation	g/m^2

During phase 1, the independent parameters are varied with a random perturbation. This perturbation is characterized by an error randomly chosen in a range of $\pm 15\%$ of parameter value, using a Gaussian random function. This operation is repeated n times, where n represents the number of fields for each scenario. The final number of simulations generated will be equal to $5 \cdot n$ (where 5 represents the number of initial scenarios and n the number of fields for each scenario). We therefore define the set of independent and environmental-dependent parameters of each field as "true" initial condition.

In phase 2 (upper part of Figure 1), the SAFY_swb model is run and uses as input the "true" initial condition values and the meteorological data. The obtained model outputs are considered the "true" values of the biophysical quantities we want to observe (LAI, ET, biomass, and yield). The set of "true" initial condition values and "true" biophysical variable values together build up the synthetic dataset and will be used as a reference to determine the assimilation efficiency of the EnKF method for SAFY_swb.

Phase 3 is characterised by the simulation of the observed LAI, i.e., an estimate of LAI obtained from EO data is simulated by adding an error. In the general case (section 2.2.1), a list of typical errors is added, increasing the number of simulations at $5 \cdot n \cdot n_{LAIerr}$, where n_{LAIerr} is the number of error values considered. In this phase, also the number of assimilated observations is introduced. The number of assimilated observations (n_a) is established a priori. It can vary as in the general case or be fixed as in the specific case (section 2.2.2). In case it varies, the number of variation (i.e., the number of cases with different number of assimilated observations days) increases the number of final simulations, turning out to be: $5 \cdot n \cdot n_{LAIerr} \cdot n_{asc}$ (where n_{asc} is the number of cases for different numbers of assimilations considered). The need to distinguish two distinct cases arises to solve a computational efficiency problem, so the first set of simulations, generated in the "General Case", was used to quantify the influence of the number of assimilated observations and the error on the assimilated variable on the efficiency of the method. The second set of simulations ("Specific Case") was done to fix the number of observations and LAI measurements error to values similar to those occurring in the real case study and it was used to evaluate the assimilation efficiency of the model for this specific condition.

During phase 4 the SAFY_swb model was run twice for each field, with and without assimilation, in order to obtain the biophysical variables for both cases and to compare them with the "true" values. The independent parameters to run the model are initially set in both cases using a generic calibration retrieved from the literature [1,3,14–17]. In phase 5 the simulated values of biophysical variables, in presence of assimilation and simulated values of biophysical variables without the assimilation are compared with the "true" values following the method proposed by [13] for calculating the Assimilation Efficiency Index (AE), as detailed here below.

First of all, the relative mean absolute errors (RMAE) were calculated, both for the method with assimilation ($RMAE_{WA}$) and for the method without assimilation ($RMAE_{NO}$), as indicated in the Equations (11) and (12), respectively:

$$RMAE_{WA} = \frac{1}{N_f} \sum_{i=1}^{N_f} \frac{|V_T - V_{SWA}|}{|V_T|} \tag{11}$$

$$RMAE_{NO} = \frac{1}{N_f} \sum_{i=1}^{N_f} \frac{|V_T - V_{SNO}|}{|V_T|} \tag{12}$$

where:

N_f: number of fields (i.e., number of hypothetic scenarios);

V_T: "true" value of Reference Biophysical variable (e.g., yield);

V_{SWA}: simulated value of Reference Biophysical variable calculated using the assimilation method;

V_{SNO}: simulated value of Reference Biophysical variable calculated running the crop model without the assimilation method.

Subsequently the AE was calculated following Equation (13):

$$AE = 100 \cdot (1 - \frac{RMAE_{WA}}{RMAE_{NO}}) \tag{13}$$

For completeness, the assimilated variable (the LAI) was also compared with the true values. A number of eight observations (arbitrarily chosen value), taken at a constant interval in the time interval between the onset of senescence and harvest, were considered and the RMSE was calculated. In a similar way the RMSE was also calculated for biomass and ET.

2.2.1. General Case

We have defined the General Case as the application of the methodology proposed by [38] which aims to evaluate how the variation in the number of observations assimilated and the measurement error of the variable to be assimilated (in this case the LAI) affect the assimilation efficiency of EnKF -SAFY_swb.

The LAI to be assimilated (LAI_M) is obtained by adding an error proportional to the same value to the "true" value of LAI, as shown in Equation (14).

$$LAI_M(d) = LAI_T(d) \pm \varepsilon_{LAI} \cdot LAI_T(d) \tag{14}$$

where:

d: day of assimilation;
LAI_T: "true" value of LAI;
ε_{LAI}: coefficient of LAI error on the measurements;
ε_{LAI} considered in the general case are: 0.05, 0.1, 0.15, 0.2, 0.25, 0.3.

In this case, also the number of assimilated observations was varied. A set of 5, 10, 15 and 20 observations (and corresponding assimilations) were considered. In this way, 600 simulations were finally carried out. For the general case, the yield was used as reference variable to calculate $RMAE_{WA}$, $RMAE_{NO}$ and, consequently, the assimilation efficiency index AE (Equations (11)–(14)).

2.2.2. Specific Case

The specific case was analysed to obtain the assimilation efficiency for conditions very similar to the real case of study used to validate the EnKF-SAFY_swb method. The number of assimilated observations is fixed to 6, because this is the minimum number of Sentinel-2 images available for the area of interest during the crop growth cycle. The error of LAI measurement is fixed to 20%, as indicated in literature [44] for LAI derived from Sentinel-2 data using the SNAP software. Since the number of observations and the error on LAI are fixed, the number of simulations is equal to the number of fields, so it was possible to increase the number of fields in order to obtain a wider variety of potential scenarios. In this case, the number of fields has been set to 1000 for each initial scenario, for a total of 5000 simulations.

2.3. Grosseto Case Study (Central Italy)

All data acquired in situ used for the validation of this study were collected during the 2018 SurfSense campaign funded by ESA [54]. The field campaign was part of a larger activity in the context of future EO programmes promoted by Copernicus, the European Union's Earth Observation Programme [6]. Specifically, the campaign described by [54] aims to demonstrate the validity of the Copernicus Candidate mission High Spatio-Temporal Resolution Land Surface Temperature Monitoring (LSTM). The LSTM mission aims to address water, agriculture and food security issues by monitoring the variability in LST (and hence evapotranspiration) at the European field scale enabling more robust estimates of field-scale water productivity [54]. LST is closely linked to ET, the reference variable chosen in this study to demonstrate the effectiveness of the EnKF-SAFY_swb method. Although the field campaign was not specifically designed for the presented study, the recorded data are very useful for the presented analysis.

2.3.1. Study Area and In Situ Data

The study area is in a farm (Le Rogaie) located in Central Italy, in the province of Grosseto, in the coastal zone of Southern Tuscany (Figure 3). It is characterized by a Mediterranean climate, with very mild and wet winters and very hot summers. During winter, minimum temperatures during the night under 0°C are common as well maximum temperatures over 30°C during the days of summer. Detailed climatic data of 2018 are shown in Figure 4. Consorzio Lamma, Laboratory of Meteorology and Environmental Modelling have provided temperatures, precipitation and daily solar radiation data. The meteo station was located at 42.769 N and 11.016 E.

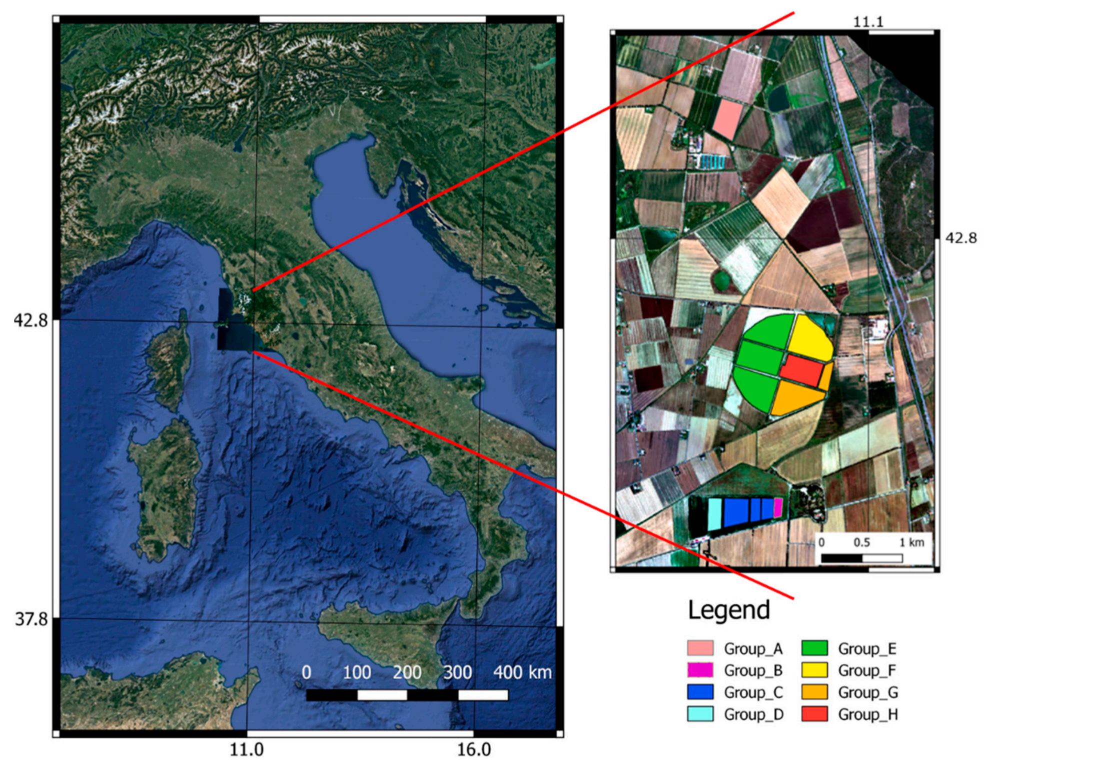

Figure 3. Study area: Le Rogaie, Grosseto, Tuscany (Italy). The base map used for the image on the left was downloaded from the google catalogue made available for the QGIS plugin QuickMapService. The base map used for the image on the right is true colour composite based on HyPlant data recorded on 20 July 2018 during SurfSense 2018 [54].

Since data of daily solar radiation was not available for some days, the RadEst model [18] was applied to estimate the parameter for the missing days and fill the gaps. The reference evapotranspiration has been calculated using the FAO software ET0 Calculator [19].

These fields were classified into 8 groups (from A to H) according to similar characteristics such as sowing date, management of irrigation and average trend of the LAI. The fields belonging to the groups A (in the northern part), B, C and D (in the southern part), of a total area of 33 ha were managed using fixed sprinkler irrigation systems during the maize growth period, while the fields belonging to the groups D, E, F, G, H located within a circular area of 72 ha with a diameter of about one km were irrigated by a rotating pivot system, which is normally operated 24 h a day in the period June-August.

Details of the soil properties, obtained from an intensive sampling campaign carried out during summer 2018 in the study area are reported in Table 3.

Table 3. Summary of soil property statistics of the study site (Grosseto, Central Italy).

Soil property	Min	Max	Mean	St.dev.
pH	7.1	8.2	7.5	0.2
Electrical Conductivity (mS cm^{-1})	158.0	2970.0	1472.5	548.5
Sand (%)	7.7	34.9	18.2	6.1
Silt (%)	38.7	67.3	55.0	6.2
Clay (%)	12.3	43.9	26.8	6.8
Carbonates (g kg^{-1})	73.0	148.0	122.8	11.3

Total nitrogen (g kg-1)	0.1	0.2	0.2	0.0
Organic carbon (g kg-1)	0.9	1.7	1.4	0.2
C/N	6.6	8.9	7.5	0.5

A full irrigation cycle is completed within four days. A malfunction of the system did not allow proper irrigation between 16th and 22nd of July.

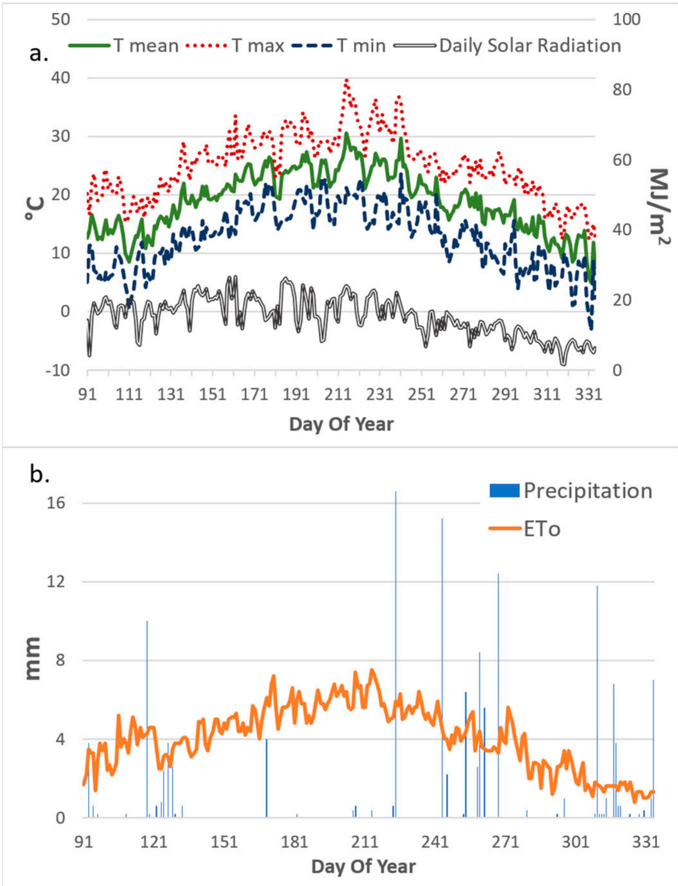

Figure 4. Study area: Le Rogaie, Grosseto, Tuscany (Italy). The base map used for the image on the left was downloaded from the google catalogue made available for the QGIS plugin QuickMapService. The base map used for the image on the right is true colour composite based on HyPlant data recorded on 20 July 2018 during SurfSense 2018 [54].

Turbulent flux data were acquired using an Eddy-covariance (EC) system installed inside the pivot area (11.07073, 42.83203) in accordance to the EuroFlux methodology [20] .The EC system was operational between the 143rd and 241st day of the year 2018 (May 23rd till August 29th). The fluxes (including the latent heat flux data in W/m^{-2} used in this study) were calculated) on half hourly intervals using the ECpack software [21].

2.3.2. Airborne Measurements

The airborne data used in this study were gathered as part of the Surfsense2018 ESA campaign [22]. The sensor used was the Thermal Airborne Spectrographic Imager (TASI-600), a push broom hyperspectral thermal sensor system specifically designed for airborne data acquisition. It is sensitive to wavelengths in the long wave infrared (LWIR) and measures the intensity of emitted radiance from the imaged target across 32 spectral bands in the range of 8 to 11.5 μm. The data were acquired on the 199th and 201st day of the year (July 18th and 20th), between 12:00 and 16:00 (local time), following the flight plan shown in Figure 5. The data acquired with the TASI were used to estimate the Surface Temperature (T_s).

Figure 5. Airborne flight plan used to record airborne data with the TASI-600 and HyPlant DUAL sensor. Airborne data were recorded during the 2018 Surfsense campaign [54]. The base map was elaborated from Sentinel-2A data acquired on 8/7/2018.

Simultaneously, optically reflective data were acquired using the HyPlant DUAL airborne imaging spectrometer. HyPlant DUAL consists of two push-broom hyperspectral line scanners, which provide contiguous spectral information from 370 nm to 2500 nm with 3 nm spectral resolution in the VIS/NIR and 10 nm spectral resolution in the SWIR spectral range.

Both the data acquired with the TASI and HyPlant were processed during the Surfsense2018 campaign [54] to obtain the estimations of the instantaneous Latent Heat Flux (LE). Those estimation were carried out using the Simplified Surface Energy Balance Index (S-SEBI) model described in [32,33]:

$$LE = \Lambda(R_n - G) \tag{15}$$

where Λ is the instantaneous evaporative fraction (adimensional), R_n is the instantaneous net radiation flux and G is the instantaneous soil heat flux (both expressed in $\frac{W}{m^2}$). According to [32], Λ was calculated as:

$$\Lambda = \frac{T_H - T_S}{T_H - T_{LE}} \tag{16}$$

where $T_s$is the land surface temperature and T_H and T_{LE} are the temperatures corresponding to dry and wet surface conditions for a given albedo value and can be obtained from the scatterplot between surface temperature and albedo according to [57].

R_n is expressed as follow:

$$Rn = (1 - \alpha)\, Rg + \varepsilon\, L^{\downarrow} - \varepsilon\, \sigma\, T_S^4 \tag{17}$$

where α is the albedo, Rg ($\frac{W}{m^2}$) is the incoming shortwave radiation, ε is the surface emissivity, $L^{\downarrow}$ ($\frac{W}{m^2}$) the incoming longwave radiation, σ ($\frac{W}{m^2 \cdot K^4}$) is the Stefan-Boltzmann constant and T_s is the surface temperature measured in °K.

G was calculated as suggested by [58]:

$$G = [0.0038\,(T_S - 273.15\,) + 0.0074\,\alpha^2\,(1 - 0.98\,\mathrm{NDVI}^4)]\mathrm{Rn} \quad (18)$$

All the details about the airborne acquired data and their processing to obtain the instantaneous LE are explained in detail in [23].

2.3.3. From Instantaneous LE to Daily Actual ET

In order to compare the daily evapotranspiration simulated by EnKF-SAFY_swb with the data acquired in situ using Eddy Covariance (EC) and through airborne combining TASI and HyPlant is necessary to convert the instantaneous LE ($\frac{W}{m^2}$) to daily actual Evapotranspiration (mm). For the EC data the daily latent heat flux (LE_d) was calculated by simply integrating the half hourly latent heat flux (LE_i) during the period of daylight using as time interval dt = 30 min. For the estimates of LE_i obtained from the airborne data it was necessary to use a model that related LE_d to LE_i.

With a view to developing a methodology that can be adapted in the future to the use of daily land surface temperature data acquired via satellite (for example LSTM), it was decided to use a single estimate of LE_i to derive the daily ET. The methodology used to obtain this conversion is the one proposed by [23] and [24]. They affirm that the ratio between solar total daily irradiance and instantaneous solar irradiance is equal to the ratio of LE_d (expressed in $\frac{MJ}{m^2 \cdot day}$) and instantaneous latent heat flux LE_i (expressed in $\frac{W}{m^2}$). The relationship between LE_d and LE_i proposed by [59,60] is shown in Equation (19):

$$LE_d = LE_i \frac{2N}{\pi \sin(\frac{\pi t}{N})} \quad (19)$$

Where N is the daylight period (calculated as a function of latitude or knowing the sunrise and sunset times of a given day of the year) and t the unit time of reference.

To obtain the daily ET expressed in mm, LE_d was divided by the latent heat of evaporation of water.

2.3.4. Satellite Data

A set of nine Level 2 cloud free Sentinel 2 images covering the study area were selected for the analysis. The level 2 images were processed with the SNAP module S2toolbox [25]to estimate the LAI (Table 4).

Table 4. List of Sentinel-2 Level 2 images used for this study and their day of acquisition (column 2). In the column called Validation, the images classified as "no" were used for the assimilation, while the images classified with capital letters were used to validate the results of the group of fields corresponding to the latter.

ID	Day of Acquisition	Validation
S2A_MSIL2A_20180419T101031	139	no
S2A_MSIL2A_20180618T101021	169	no
S2A_MSIL2A_20180708T101031	188	no
S2A_MSIL2A_20180718T101031	199	no
S2A_MSIL2A_20180807T101021	219	no
S2A_MSIL2A_20180817T101021	229	no
S2A_MSIL2A_20180827T101021	239	A-C-D
S2A_MSIL2A_20180906T101021	252	no
S2A_MSIL2A_20180926T101021	269	B-E-F-G-H

The trend of average LAI for each group of fields (defined as described in section 2.3.1) and the corresponding standard deviation were calculated and are shown in Figure 6. For each trend all the images were used for the assimilation, except the last image of the trend, which was used to validate the simulations generated with the EnKF-SAFY_swb.

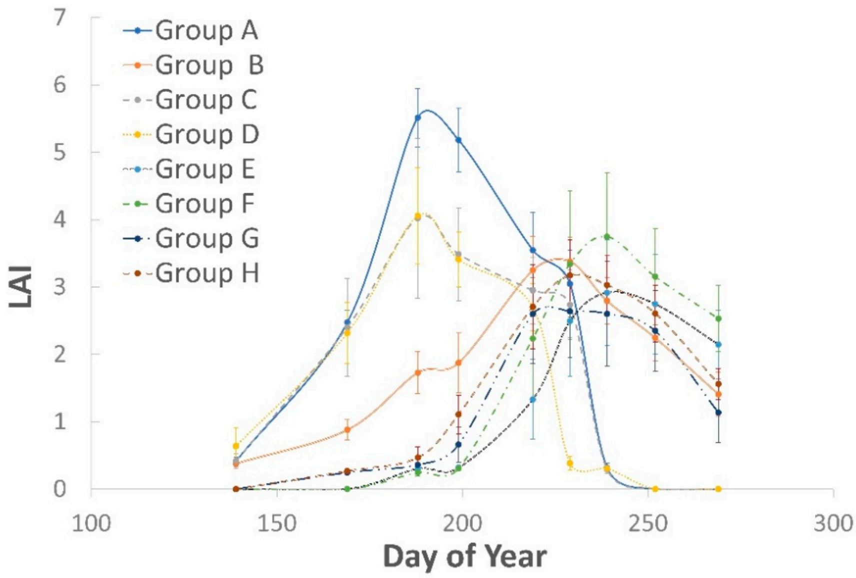

Figure 6. Average LAI Trend estimated from Sentinel-2 data. The error bars represent the standard deviation for each group of fields.

Each group of fields (Table 5) has the same sowing date, harvest date and number of assimilations. For each group of fields, the LAI is monitored during the crop cycle by the number of images assimilated (specified in Table 4) and the last image of the cycle was used for the validation (the acquisition date of the images used for validation for each group of fields is specified in Table 4).

Table 5. List of groups of fields. Each group has same sowing date, harvest date and number of images assimilated.

Group	Sowing date	Harvest date	Number of LAI observations
A	17 April 2018	24 August 2018	6
B	27 April 2018	03 October 2018	8
C	17 April 2018	30 August 2018	6
D	12 April 2018	22 August 2018	6
E	10 June 2018	27 October 2018	6
F	10 June 2018	27 August 2018	6
G	22 May 2018	22 October 2018	7
H	22 May 2018	22 October 2018	7

2.3.5. EnKF-SAFY_swb Method Pre-Processing

As mentioned in section 2.1 the majority of parameters of the model are kept constant during the run of EnKF-SAFY_swb (Table 6 and Table 7), while 5 parameters are varied during the model runs (Table 8).

Table 6. List of model fixed parameters and their calibration values.

ID	Description	Value	Unit	Reference
Pfen_PrtA	Partition-to-leaf function: parameter 1	0.279	-	[14]
Pfen_PrtB	Partition-to-leaf function: parameter 2	0.0022	-	[14]
Pfen_SenA	Sum of temperature for senescence	1100	°C	[14]
Pfen_SenB	Rate of senescence	5463	$°C/_{day}$	[14]
MxRDP	Maximum GDD required to reach maturity (°C)	1660	°C	[50]
Pgro_Ms0	Emergence Dry Mass Value	2.5	$^{g}/_{m^2}$	[50]
Maize.MxDay	Maximum Days from emergence to physiological maturity	175	days	[50]
Maize.RunWin	Window size for running average temperature	18	°C	[50]
Maize.RunAvg	Running average daily mean temperature before planting	12	°C	[50]
Maize.RunMin	Running average daily min temperature before planting	8	°C	[50]
Maize.ErlPlant_doy	Earliest planting date	91	Day of Year	[50]
Maize.EmGDD	GDD required from planting to emergence	80	°C	[50]
Maize.Tmin	Minimum Temperature for Plant Development (°C	10	°C	[50]
Maize.Topt	Optimal Temperature for Plant Development (°C)	30	°C	[50]
Maize.Tmax	Maximum Temperature for Plant Development	40	°C	[50]
RootRatio	root weight to length ratio	9800	cm/g	[50]
RtGrtRate	root depth growth rate	0.16	$\frac{cm}{deg \cdot day}$	[50]
MxRDP	maximum root depth	120	cm	[50]
Rnff	runoff factor	0.2	-	[52]
SALB	Soil Albedo	0.18	-	[53]
DrnCoeff	profile drainage coefficient	6.99	$^{cm}/_{day}$	[54]
MxRWU	maximum root water uptake rate	0.0035	$\frac{cm^3[water]}{cm[root] \cdot day}$	[54]

Table 7. Soil parameters. Calibrated by crossing [54] and [61]. They were kept constant during EnKF-SAFY_swb run.

ID	Unit	Layer 1	Layer 2	Layer 3	Layer 4	Layer 5
Layer Depth	cm	10	15	25	50	50
Field Capacity (FC)	cm^3/cm^3	0.46	0.46	0.46	0.45	0.45
Wilting Point (WP)	cm^3/cm^3	0.30	0.30	0.30	0.29	0.29
Air Dry	cm^3/cm^3	0.23	0.22	0.2	0.22	0.22
Saturation (Sat)	cm^3/cm^3	0.54	0.54	0.54	0.54	0.54

Table 8. List of parameters that vary in EnKF-SAFY_swb.

ID	Min	Max	Mean	Standard Deviation	Description
Pgro_Lue	2.88	3.52	3.2	0.32	Effective Light use efficiency
Pgro_R2P	0.423	0.517	0.47	0.047	Global to PAR incident radiation ratio (Climatic Efficiency)
Pgro_Kex	0.45	0.55	0.5	0.05	Light interception coefficient
Pgro_Sla	0.018	0.022	0.02	0.002	Specific Leaf-Area
Pgro_P2G	0.00585	0.00715	0.0065	0.00065	Partition coefficient To Grain

The parameters of the model kept constant during the run of SAFY_swb are shown in Table 6 and Table 7. They are calibrated using information obtained in situ, from airborne data and EO data [22] and information retrieved by literature [1,14–17]. The calibration is intentionally generic, because the idea is to test the efficiency of the EnKF-SAFY_swb methodology in the absence of a specific calibration for each field which is generally a great difficulty in using crop growth models.

The parameters to be changed during the execution of EnKF-SAFY_swb were chosen based on the sensitivity analysis made by[26]. To simplify the computational process of the algorithm only the five parameters that most affect the model were selected to be varied.

As suggested by literature [3,10,27,28], the size N of the ensemble used in the EnKF DA method (N corresponds to the number of simulations for each pixel) has to be big enough to ensure the convergence of the model [11]. For this study N was set to 200. The error of the model is considered as a percentage of the mean simulated LAI and is set to 20 % [3], as suggested by the results obtained evaluating the assimilation efficiency of the developed methodology (sections 2.2 and 3.1). The error of the measurement was set at 20 % as described by [22].

In Table 9 the number of pixel and the number of point for each group of fields are shown.

Table 9. Number of pixel and number of points used for assimialtion for each group of fields.

ID	Number of Pixels	Number of Point Used for the Assimilation
Group A	675	4050
Group B	101	808
Group C	988	5928
Group D	359	2154
Group E	3029	18174
Group F	1196	7176
Group G	1215	8505
Group H	799	5593