# SEBALIGEE v2: Global Evapotranspiration Estimation Replacing Hot/Cold Pixels with Machine Learning

Mario Mhawej[1], Xiang Gao[2], John Reilly[3], and Yaser Abunnasr[1]

[1]American University of Beirut
[2]Massachusetts Institute of Technology
[3]MIT

November 21, 2022

**Abstract**

An open source computer algorithm, the Surface Energy Balance Algorithm for Land-Improved (SEBALI), was designed to estimate actual evapotranspiration (ET) at a basin level. In this study, we build on later versions of SEBALI/SEBALIGEE to estimate ET at a 30-m resolution for any scale application using advanced machine learning approaches (SEBALIGEE v2). We evaluate the monthly ET estimated from the new algorithm across several fluxnet sites in US, China, Italy, Belgium, Germany, and France, yielding an Absolute Mean Error (AME) of 0.41 mm/day versus 0.48 mm/day in the original SEBALIGEE. Analyses of the ET in the US indicate that the annual wheat ET decreases significantly between 2013 and 2021 ($p < 0.05$), accompanied by a significant air temperature increase. Net solar radiation is found to be the most influencing factor on ET of corn and soybeans with $R^2$ values of ˜0.72.

**Hosted file**

essoar.10512468.1.docx available at https://authorea.com/users/527720/articles/596693-sebaligee-v2-global-evapotranspiration-estimation-replacing-hot-cold-pixels-with-machine-learning

# SEBALIGEE v2: Global Evapotranspiration Estimation Replacing Hot/Cold Pixels with Machine Learning

**Mario Mhawej[1], Xiang Gao[2], John M. Reilly[2], and Yaser Abunnasr [1, †]**

[1] Department of Landscape Design and Ecosystem Management, Faculty of Agricultural and Food Sciences, American University of Beirut, Bliss St., Beirut 2020-1100, Lebanon;

[2] Massachusetts Institute of Technology (MIT), Department of Earth, Atmospheric and PlanETry Sciences 77 Massachusetts Avenue, 54-918, Cambridge, MA 02139.

Corresponding author: Yaser Abunnasr (ya20@aub.edu.lb)

Key Points:

- An open source machine learning-based algorithm was designed to estimate actual evapotranspiration (ET) at a 30-m resolution for any-scale application.
- Among climate variables, we find net solar radiation to have the most influence on ET over agricultural lands in the CONUS.
- Over CONUS, annual wheat ET decreased significantly from 2013 to 2021, accompanied by a significant increase in air temperature.

Abstract

An open source computer algorithm, the Surface Energy Balance Algorithm for Land-Improved (SEBALI), was designed to estimate actual evapotranspiration (ET) at a basin level. In this study, we build on later versions of SEBALI/SEBALIGEE to estimate ET at a 30-m resolution for any scale application using advanced machine learning approaches (SEBALIGEE v2). We evaluate the monthly ET estimated from the new algorithm across several fluxnet sites in US, China, Italy, Belgium, Germany, and France, yielding an Absolute Mean Error (AME) of 0.41 mm/day versus 0.48 mm/day in the original SEBALIGEE. Analyses of the ET in the US indicate that the annual wheat ET decreases significantly between 2013 and 2021 ($p < 0.05$), accompanied by a significant air temperature increase. Net solar radiation is found to be the most influencing factor on ET of corn and soybeans with $R^2$ values of ~0.72.

## Plain Language Summary

Evapotranspiration, the combination of soil evaporation and plant transpiration, represents a major consumptive use of agricultural water. Accurate estimate of evapotranspiration plays an important role in improving water use efficiency and achieving sustainable water management in agriculture. We improve on existing algorithms for estimating global evapotranspiration from crops at a 30-m resolution for any-scale application through use of advanced machine learning approaches. We apply the algorithm to three major crops in the US from 2013

to 2021, and examine associations with various climate variables and drought conditions. The new version of the algorithm is an open-sourced and fully automated system that can be easily customized to estimate actual evapotranspiration, accounting for differences in management and other environmental conditions, in any region of interest across the globe.

## 1 Introduction

A sustainable water management plan is critical to meeting the goals of the UN 2030 Agenda on Sustainable Development Goals (Si, 2021). However, climate change poses a major threat to developing a resilient water management system that ensures crops' optimum health and productivity (Zhang et al. 2021). Nevertheless, the availability of abundant satellite observations and the advancement of artificial intelligence/machine learning technologies have provided unprecedented opportunity to tackle this challenge.

One essential component of water management systems calls for an accurate estimate of crop water demand, also known as the actual evapotranspiration (ET). ET is affected by many factors, including weather parameters and crop characteristics as well as management and environmental aspects. ET is typically expressed by relating to the reference evapotranspiration ($ET_0$) via the adjusted crop coefficient ($aK_c$, $ET = ET_0 * aK_c$) (Mhawej et al., 2021). $ET_0$ represents the evapotranspiration from a reference surface (eg., well-watered grass) and is only affected by weather parameters. $aK_c$ accounts for the crop evapotranspiration under non-standard field conditions with all kinds of stresses as well as sub-optimal crop management and environmental constraints. It represents the integrated effects of crop characteristics, leaf area, plant height, canopy cover, rate of crop development, irrigation method, soil and climate conditions, and management practices. $aK_c$ is specific for each crop, changes throughout the crop growth stages, and is usually determined experimentally (Mhawej et al., 2021). Knowledge of $aK_c$ for a given crop will provide crucial information about water use at each stage of the crop growing season (from sowing till harvest).

There are two common approaches to estimate ET: crop model and satellite remote sensing. Each of these techniques considers specific properties and has limitations. Crop models, such as AquaCrop (Steduto et al., 2009), CropSyst (Stockle et al., 1994), the Agricultural Production Systems sIMulator (APSIM) (Keating et al., 2003), are often applied on a parcel-scale, due to many site-specific parameters which need be calibrated to achieve a desired accuracy. This usually generates fragmented observations. Satellite remote sensing can readily provide continuous observations over large areas of various variables required to calculate ET at global coverage (Fadel et al., 2020). The widely used algorithms for estimating ET include the Surface Energy Balance Algorithm for Land (SEBAL) (Bastiaanssen et al., 1998), Mapping EvapoTranspiRation with Internalized Calibration (METRIC) (Allen et al., 2007), ET-Watch (Wu et al., 2008), the Atmosphere-Land EXchange Inverse (ALEXI) flux Disaggregation approach (DisALEXI) (Anderson et al., 2013), and operational Simplified Surface Energy Balance (SSEBop) (Senay et al., 2014). Our recently developed

SEBAL – Improved (SEBALI) (Mhawej et al., 2020a; 2020b; Allam et al., 2021) Google Earth Engine (SEBALI-GEE) (Mhawej and Faour, 2020; Mhawej et al., 2021) algorithm represents many improvements over the previously mentioned models (see section 3.1 for more details). The utilization of the GEE platform also enables prompt fetching and processing of all the needed inputs (Gorelick et al., 2017) and is therefore end-user friendly. However, SEBALI/SEBALIGEE and other algorithms mentioned above generally rely on the use of hot and cold pixels to construct a linear relationship between temperature differences (dTs) and their corresponding land surface temperatures (LSTs) in the sensible heat calculation (Abunnasr et al., 2022). This procedure may introduce biases when the relationship is extrapolated to other pixels for the dT computation (see section 3.2 for more details).

Here we propose an improved version of the SEBALI/SEBALIGEE model (SEBALIGEE version 2, hereinafter referred to as SEBALIGEE v2), also hosted over the GEE. SEBALIGEE v2 employs the Random Forest (RF) machine learning technique in place of the conventional hot/cold pixel approach, which enables more objective and accurate dT computation on any scale. In this study, we first evaluate the performance of SEBALIGEE v2 in estimating the ET at various Fluxnet sites across six different countries, including the United States, China, Italy, Belgium, Germany, and France. We then focus on three most commonly grown crops across the entire Contiguous United States (CONUS), namely, corn, soybeans and winter wheat, and deduce their $aK_c$ values from the SEBALIGEE v2-estimated ET and $ET_0$ between 2013 and 2021. We further assess the seasonal and annual variations of three crops' $aK_c$ and ET as well as how different climate variables and drought conditions are associated with their variations. The structure of the paper is as follows; in section 2, we describe the study area and various datasets used in this study. The method for developing the SEBALIGEE v2 is given in section 3. Section 4 presents the results. Summary and discussions are provided in section 5.

## 2 Study Area and Datasets

### 2.1 Study Area

Our study focuses on the Contiguous United States (CONUS) which is represented by 22 different climatic zones (Abunnasr and Mhawej, 2021). Agriculture is a major industry in the United States with nearly two million farmers (Page, 2018). It is mostly mechanized and concentrated in the Great Plains, Great Lakes and east of the Rocky Mountains. According to the United States Department of Agricultural National Agricultural Statistics Service (USDA NASS) report of 2015 (*https://www.nass.usda.gov/Charts_and_Maps/Crop_Progress_&_Condition/2015/index.php*), corn represents the highest production value of 52.3 billion USD, followed by soybeans of 40.3 billion USD and wheat of 11.9 billion USD. In 2021, these three crops corresponded to an area of 53, 46, and 12 million ha, respectively (Figure 1). Corn and soybeans are mainly located over the eastern half of the US called Corn Belt, whilst wheat plantations are largely found over the

western half, known as the Wheat Belt (Winkler et al., 2012). The western US is typically characterized by a cold semi-arid climate in the interior upper states and warm/hot semi-arid climate in the southwestern states. In the eastern US, the climate transitions from humid continental in northern areas into a humid temperate in the southern states.

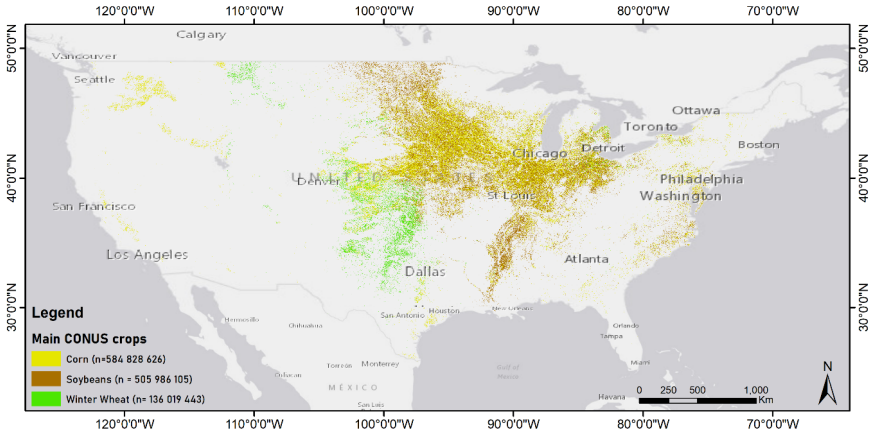


**Figure 1: Mapping main crops in the CONUS, 2021 ("n" corresponds to sample size of 30x30 pixels).**

2.2 Datasets

Various datasets have been compiled for the development of SEBALIGEE v2, including 1) the 30-m 16-day Level-2 Bottom of Atmosphere (BOA) Surface Reflectance (SR) Landsat-8 satellite (Vermote et al., 2016), which is used for the generation of Normalized Difference Vegetation Index (NDVI), LST, and albedo; 2) the monthly air temperature, dewpoint and wind speed from the fifth generation ECMWF atmospheric reanalysis of the global climate at 0.1° (~10-km, ERA5, Hersbach et al. 2020) used for the calculation of $ET_0$; 3) the 1-km Moderate Resolution Imaging Spectroradiometer (MODIS) daily snow cover version 6 (MOD10A1, Hall et al., 2016) used for cloud cover detection; 4) the 500-m annual MODIS Land Cover Type version 6 (MCD12Q1.006, Sulla-Menashe and Friedl, 2018) used for water body mapping; 5) the 90-m NASA Shuttle Radar Topographic Mission (SRTM) version 4 (Jarvis et al., 2008) used for the calculation of altitude and slope and $ET_0$; 6) the 30-m annual USDA NASS Cropland Data (Boryan et al., 2011) used for the corn, soybeans and winter wheat mapping; 7) air temperature at two different heights (i.e., 2-m and 10-m) from the Modern-Era Retrospective analysis for Research and Applications,

Version 2 at 0.625° × 0.5° (~ 50km, MERRA-2, Bosilovich et al., 2016) used for the sensible heat calculation. Two additional datasets are employed for the statistical analysis (section 4.2), including the monthly total precipitation from the Parameter-elevation Regressions on Independent Slopes Model AN81m at 4-km (PRISM, Daly et al., 2008) and various daily drought indices at 4-km derived from the Gridded Surface Meteorological (GRIDMET, Abatzoglou, 2012). The drought indices to be examined in our study include the widely-used Standardized Precipitation Evapotranspiration Index (SPEI) calculated at different time scales (Table S1 in Supplementary Materials). Given a wide range of spatial resolutions among various datasets (30-m ~ 50-km), our analyses will be performed at the 30-m (finest) grid scale. Any coarser dataset grid will overlay multiple 30-m grids with the same values across the grids. Although the ERA5 provides air temperature at 2-m and 10-m, we consider only those datasets that are directly available and accessible at the GEE platform. This will allow end-users to readily apply our algorithm for their applications. We expect that the use of climate variables from multiple data sets will not significantly change our results.

We use ET observations collected at multiple Eddy Covariance flux towers to evaluate the SEBALIGEE v2 performance (Table S2 and Figure S1 in Supplementary Materials). Most of these sites have been used in our previous studies related to the SEBALI/SEBALIGEE model. We focus on the sites which have the flux measurements overlapping the Landsat-8 data, specifically from September 2013 to December 2021. This results in a total of 11 sites across four climate regions (Hot-summer Mediterranean climate (Csa), Oceanic (Cfb), Monsoon-influenced warm-summer humid continental (Dwb), and hot-summer humid continental (Dfa)) in six countries. These datasets are available to download from the FLUXNET (https://fluxnet.fluxdata.org/) or AMERIFLUX (https://ameriflux.lbl.gov/) official data portal. Only the fully processed Eddy Covariance ET measurements in China were provided by the Chinese Academy of Science (CAS) from January 2016 to December 2017.

## 3 Development of the SEBALIGEE v2

### 3.1 SEBALI/SEBALIGEE model

SEBALI/SEBALIGEE is a satellite-based surface energy balance model developed in 2020 and commonly implemented over a basin scale. The basic principle behind SEBALI /SEBALIGEE is the widely known surface energy equation (Price, 1985) at the Earth's surface with the latent heat flux as a proxy for ET, derived as the residual of other energy terms (i.e., net radiation flux Rn, soil heat flux G and sensible heat flux H). SEBALI/SEBALIGEE integrates the strengths of existing models (i.e., SEBAL, METRIC, ET-Watch, and SSEBop) but also addresses gaps in their capabilities. Major improvements of SEBALI/SEBALIGEE over previous models include the reduction of required parameters (e.g., daily wind speed and relative humidity), the exclusion of the commonly-biased soil-related variables (e.g., wilting point and field capacity) as inputs, and a more sophisticated tree-like algorithm for automatically identifying hot and cold pixels

(Mhawej and Faour, 2020), water-based internal calibrations to achieve improved satellite-based estimates of sensible and latent heat fluxes, and the estimation of 30-m ET and $aK_c$ simultaneously. SEBALI/SEBALIGEE is a fully automated, open-source, user-friendly system by ingesting various readily available remote sensed datasets at the GEE platform, including vegetation indices, albedo, and LST, climate, and other ancillary datasets (Mhawej and Faour, 2020). Detailed information can be found in Mhawej et al. (2020a, 2020b; 2021) and Allam et al. (2021).

### 3.2 SEBALIGEE v2

SEBALIGEE v2 (Figure S2 in Supplementary Materials) essentially follows the same principle and procedure as SEBALI/SEBALIGEE in the calculation of ET, $ET_0$, and $aK_c$, except that RF machine learning method is implemented in place of the hot/cold pixels approach to resolve the air temperature differences for the sensible heat calculation and further reduce potential bias/uncertainty. In SEBALI/SEBALIGEE, hot and cold pixels are identified in two steps by accounting for the effects of both vegetation (agricultural areas only) and surface temperature. First, hot pixels are selected from bare or uncultivated lands with limited or lack of vegetation cover ($0 < NDVI <= 0.2$) while cold pixels are associated with a vegetation cover ($NDVI > 0.2$). Two types of pixels are further filtered based on the LST statistics (mean and standard deviation) of each pixel group. A linear relationship is constructed between dTs and LSTs of two groups and applied to the LSTs of other pixels across the whole satellite image for estimating a pixel-based spatial distribution of dT. This procedure may introduce several potential sources of uncertainty due to 1) the impacts of ever changing cloud/shadow conditions and/or land cover type through the time on the determination of hot/cold pixels and their respective LSTs and dTs; and 2) the extrapolation of the linear relationship to the pixels of the whole image.

The RF approach is a type of supervised learning algorithm that uses ensemble methods (bagging) to solve both regression and classification problems (Breiman 2001; Chen et al., 2020). In SEBALIGEE v2, RF is implemented on a monthly scale at the spatial resolution of MERRA-2 (50-km * 62.5-km) over the CONUS. We first obtain the median daily air temperature at 2-m and 10-m heights from MERRA-2 and calculate their difference at each grid for each month. Similarly, the median NDVI, LST and albedo values at 30-m are derived from one or two Landsat images of each month and then spatially aggregated to the resolution of MERRA-2. RF is then trained with Landsat NDVI, LST, and albedo as the independent variables and MERRA-2 air temperature difference as the dependent variable. The trained RF model is later applied to the NDVI, LST, and albedo at their native Landsat resolution to derive the spatial distribution of dT at 30m. The procedure is performed for each month from 2013 to 2021. In comparison with the hot/cold pixel approach, RF algorithm presents several unique strengths in the dT computation on a large scale: 1) it utilizes all the pixels within the extent of any region of interest and thus eliminates the

subjectivity and potential biases in identifying two types of extreme pixels; 2) the trained RF model is nonparametric and therefore the relationship between dependent (dTs) and independent variables is not necessarily linear; and 3) RF provides great flexibility to account for multiple independent variables in the classification model. These strengths are expected to lead to a more robust model with an improved performance. The RF is computationally very efficient. With 130 decision trees (the maximum allowable capacity at the GEE platform), it takes only a few seconds to produce a monthly dTs at the 30-m resolution over the CONUS. SEBALIGEE v2 (Figure S3 in Supplementary Materials) is easily portable and customizable to produce the relevant parameters specific to any region of interest, greatly increasing its applicability.

We have computed a total of 220 monthly ET estimates across 11 Eddy covariance tower sites of different climate zones in six countries (Table S2 in Supplementary Materials). A random selection of 70% of the samples is used for the calibration with the remaining 30% for validation. The model performance is evaluated with the monthly Root Mean Square Error (RMSE), Absolute Mean Error (AME) and the R-squared correlation between SEBALIGEE v2 monthly ET values and their flux towers' counterparts. The validated SEBALIGEE v2 is then applied for estimating the water requirements of three major crops, including corn, soybeans and winter wheat, in the CONUS between 2013 and 2021. Statistical analysis, including Mann-Kendall tests and step-wise regressions, were followed to understand how different climate variables affect the variations of the aKc for each crop and also identify the most influential factors.

## 4 Results and Discussion

Comparison of the SEBALIGEE v2 estimated ET values with the flux tower observations show good performance with a RMSE of 16.35 mm/month, an AME of 14.47 mm/month and a $R^2$ of 0.78 over the remaining 30% of total samples. SEBALIGEE exhibits a similar RMSE of 16.26 mm/month and an AME of 14.54 mm/month but a much lower $R^2$ of 0.66 with the six flux tower observations (excluding the new three tower sites in the US). Over the same flux towers considered in SEBALIGEE, SEBALIGEE v2 has an improved RSME (14.73 mm/month) and AME (12.22 mm/month) and a much higher $R^2$ (~0.83%). Our following analyses focus on the three crops in the CONUS.

### 4.1. Water requirements of corn, soybeans and winter wheat

Figure 2a shows the monthly climatology of ET for three crops from 2013 to 2021. These results are aggregated from the total number of parcels of each crop (sample size in Figure 1) based on the yearly crop plantation mask from USDA NASS cropland data layers. We see that the ET values of all three crops present similar distinct seasonal cycles with peaks occurring in late spring to the summer (May-Aug) and troughs between October and January. These features correspond well to the key stages of crop life cycle, such as onset of greenness, peak of the growing season, duration of the growing season, and initiation of the dry down period. Corn and soybean exhibit similar magnitudes of ET (36

~ 168 mm/month) and higher than winter wheat (33 ~ 122 mm/month), where the reproductive stages (i.e., emergence and nth leaf collar for corn, germination and seedling and rapid vegetative growth for soybeans, and heading, flowering and grain filling in winter wheat) are requiring the most water (Baum et al., 2019). Large standard deviations are observed in the ET estimates of all the crops, with ~ 32 mm/month for corn and soybean and ~ 43 mm/month for winter wheat, likely attributed to different parcels of the same crop exhibiting a wide range of planting times, crop health conditions, and management practices. Our ET curves are consistent with previous studies which suggested an optimum planting window between the last week of April and the first week of May for corn and soybeans (Nafziger, 1994; Norberg et al., 2010; Xu et al., 2019) and from mid-September to the first week of October (Nouri et al., 2017; Nasrallah et al., 2020) for winter wheat.

All three crops show a similar seasonal cycle of aKc with two distinct peaks (~ 1.2) in April and October, respectively. This may be associated with cultivation twice a year or in rotation with other crops in the same year. Due to the continuous pressure from more frequent occurrence of extreme events (e.g., droughts), multi-cultivation/rotation is a desired strategy for farmers to achieve increased annual crop yields. The lowest aKc values (~0.8) generally occur in winter and summer when crop growth is less active or a field is fallow. Corn and Soybean present overall higher aKc than winter wheat.

Interannual variations of ET and aKc are shown in Figure 2b for all three crops, with corn and soybean consistently presenting higher ET and aKc values than winter wheat. Corn shows the highest and lowest ET in the year 2014 and 2020, respectively. This is largely consistent with the interannual variations of SPEI. The incipient dry spell in 2013, as reflected by SPEI-360, triggered increased annual ET the year after in 2014 (Figure 2c). Inversely, the wet year of 2019 decreased the corn water consumption in 2020. The $R^2$ between SPEI-360 and one-year lag ET is around 0.6 for corn (Figure 2d).

The highest and lowest ET values occur in the year 2018 and 2015 for soybean as well as 2019 and 2013 for winter wheat, respectively. Similar relationships between SPEI-360 and one-year lag ET can be observed for winter wheat, with a $R^2$ value of 0.54. However, this does not hold for the soybean with a $R^2$ value of 0.12. Overall, soybeans and corn demonstrate higher aKc values (1.08 +/- 0.03 and 1.05 +/- 0.03, respectively) than winter wheat (0.94 +/- 0.04), suggesting corn and soybeans higher water requirements and stronger irrigation dependence. Note that these aKc values are aggregated from all the parcels. When the number of parcels over the CONUS is factored in, the small aKc differences observed among crops would have a large impact on water resources nationwide. Thus, more attention and effort are required to assess water management system particularly in the US, where irrigation water is often unsustainably pumped from groundwater, leading to a reduction in future production of corn, soybeans and winter wheat (Lopez et al., 2022).

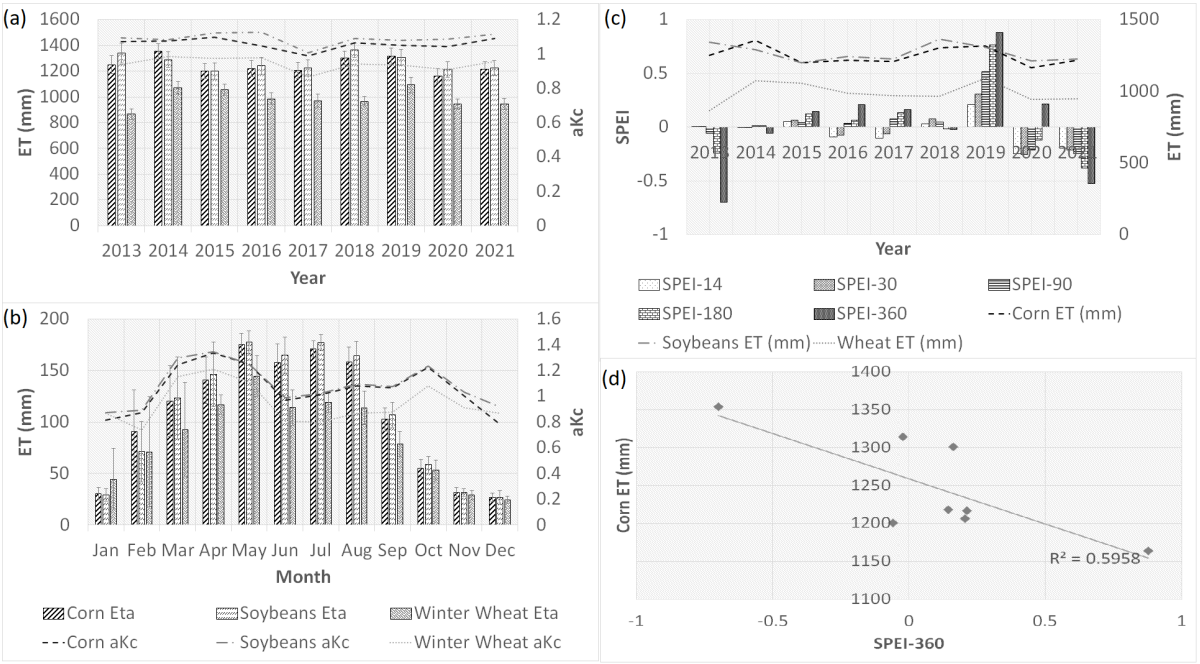


**Figure 2: Monthly climatology (a), annual mean (b) of ET and aKc and annual mean SPEI (c) for corn, soybeans and winter wheat across all parcels over the CONUS from 2013 to 2021; the error bar represents one standard deviation of all the years and parcels of each crop. (d) represents the R2 between SPEI-360 and one-year lag ET for corn.**

4.2 Statistical Analysis

4.2.1 Mann-Kendall Test

We further examine the annual trends of ET and aKc of three crops, climate variables (i.e., air temperature, dewpoint, wind speed, precipitation, surface net solar radiation) as well as the SPEI indices computed on multiple time scales (i.e., 14, 30, 90, 180 and 360 days) between 2013 and 2021. The Mann-Kendall test is used to check whether a set of data values is increasing or decreasing over time, and whether the trend in either direction is statistically significant (Mann 1945, Kendall 1975, Gilbert 1987). It does not assess the magnitude of change and cannot necessarily be extrapolated into the future. Still, the strength of applying the Mann-Kendall test is based on its ability to work on a small sample's size, usually larger than four (Gilbert 1987).

Our results indicate that only winter wheat ET presents a significant decreasing trend (-1.06 mm/year, significant at 5% level) over the last decade. Crop yields have been shown to continuously increase during the past decades (https://www.nass.usda.gov/). This finding is likely associated with the climate warming

induced increase in the air temperature (0.129°C, significant at 5% level) and resulting longer growing seasons (Rizzo et al., 2022). However, warming climate may also render the Corn Belt less sustainable to corn plantation and thus endanger further productivity gains (Rizzo et al., 2022). The exact reasons for the decrease in water demand is not certain, but could be attributed to the improved breeding and agronomic management (Grassini et al., 2013; Cooper et al., 2020). Another possible cause is that recent higher temperatures and drier conditions in the CONUS may lead to less ET—just no water to evaporate. Further research needs to be done to determine the root causes behind the winter wheat ET trend and better understand its impacts on productivity.

#### 4.2.2 Step-wise regression

There seems to be no strong association between aKc values of corn, soybeans and winter wheat and various climate variables as well as SPEI indices examined (Table S3 in Supplementary Materials). The highest $R^2$ values are obtained when five variables are used with the corresponding $R^2$ values of 0.35, 0.31 and 0.2 for corn, soybeans, and winter wheat, respectively. The weak association is likely related to the fact that aKc is a complex metric aggregated from many aspects, including crop characteristics, environmental conditions, and management practices. There is no simple way to incorporate all these different features into a single relationship (Allam et al., 2021). Our analyses indicate that the most influential factor is wind speed, followed by dewpoint temperature. Inclusion of additional variables may help improve the $R^2$ values marginally.

In contrast, ET appears to be well represented to various extents by climate variables and SPEI indices. As expected, the surface net solar radiation shows the most dominant impact on the crop ET, with $R^2$ values of 0.73, 0.71, and 0.54 for soybeans, corn, and winter wheat, respectively. In this context, agrivoltaic systems could serve as a viable option to lower the net solar radiation (Zainol Abidin et al., 2021) and decrease the crop water consumption for future farming. Wind speed appears to be the second most influential variable for corn and soybeans, but the SPEI-30 stands out for winter wheat. SPEI-14 appears to be the most dominant factor over corn and soybeans plantations in comparison to other SPEI time scales (i.e., 30, 90, 180 and 360 days). This information is useful to assist farmers and managers in determining the optimum time for cultivating these crops.

## 5 Conclusions

In this study, we develop a new version of the SEBALIGEE (v2) in which a machine learning RF algorithm is employed for the calculation of sensible heat flux in place of the widely used hot/cold pixel approach. This improvement allows SEBALIGEE's application to extend from a basin scale to any scale (e.g., parcel, basin, nation, continent, or globe). The performance of SEBALIGEE v2 in estimating the ET is evaluated with the observations from various Eddy Covariance flux towers across different climatic regions. Our results indicate satisfactory performance of the SEBALIGEE v2 in estimating ET as a key

proxy for crop water requirement, with a much lower AME (~ 0.48 mm/day) than those reported by some more complex models (e.g., 0.52 ~ 1.31 mm/day, Degano et al. 2018, Knipper et al. 2019, Allam et al. 2021, Asadi and Kamran, 2022). In comparison with the SEBALIGEE, v2 presents a similar bias but much higher $R^2$ (0.78 versus 0.66).

Further analyses focus on the SEBALIGEE v2 estimated ET and aKc values of three major crops in the CONUS from 2013 to 2021, including the seasonal and annual variations as well as the association with different climate variables and drought conditions. Our study considers a large study area of ~112 million ha per year with a huge number of parcels in each crop type, which indirectly takes into consideration the diverse agricultural management plans and micro-climates. Several findings were reported aiming towards increasing crop water productivity by scheduling irrigation and guiding future initiatives, at local, regional and national scales. This can provide useful insights into achieving an increased crop water productivity and a sustainable water usage. Furthemore, it perfectly fits within the Sustainable Development Goals (SDG) and the 2030 agenda proposed by the United Nation (UN).

The SEBALIGEE v2 is an open-source, fully automated, computationally efficient, readily-accessed, and easily-operated system for ET and aKc assessments. Non-expert end users can customize and implement it to any time period and study the region of their interest across the globe with minimal effort. The main outputs include processed monthly albedo, NDVI, LST, ET and aKc, and annual Land Cover. An option is also available for the statistical analysis of the relevant variables. The hosting GEE platform provides great accessibility to massive computational power and a large collection of geospatial databases, which facilitates the water productivity research that would otherwise be hindered by the scarcity of required resources.

Nevertheless, several aspects of the presented SEBALIGEE v2 motivate subsequent trials and development that include, 1) outputs at a higher spatial resolution by integrating 10-m sentinel-2 data (30-m in this study), 2) time laps studies that go back to the 1980's and 1990's by including other sensors such as Landsat 4, Landsat 5, Landsat 7, Advanced Spaceborne Thermal Emission and Reflectance Radiometer (ASTER), 3) outputs on weekly and daily scales, 4) validation of the SEBALIGEE v2 with more extensive tower observations, and 5) the seasonal responses of ET and aKc to the drought intensity and the impacts on crop yields. Improving the SEBALIGEE v2 with these aforementioned considerations will ultimately lead to persuasive and actionable insights for future water sustainable plans in the face of unavoidable and unpreventable global changes, population increase and resources' scarcity.

**Acknowledgments**

This research is funded by the Dean Office at the Faculty of Agricultural and Food Sciences, American University of Beirut.

**Open Research**

The evapotranspiration-related data used for ET assessments in the study are available from SEBALIGEE v2 [Software] via https://github.com/mariomhawej /SEBALIGEEv2 with Creative Commons Attribution. This code was produced from scractch under the Javascript language by the authors of this paper.

**References**


1. Abatzoglou, John T. "Development of gridded surface meteorological data for ecological applications and modelling." International Journal of Climatology 33, no. 1 (2013): 121-131.

2. Abdi, Abdulhakim M., Romain Carrié, William Sidemo-Holm, Zhanzhang Cai, Niklas Boke-Olén, Henrik G. Smith, Lars Eklundh, and Johan Ekroos. "Biodiversity decline with increasing crop productivity in agricultural fields revealed by satellite remote sensing." Ecological Indicators 130 (2021): 108098.

3. Abunnasr, Yaser, and Mario Mhawej. "Pervious area change as surrogate to diverse climatic variables trends in the CONUS: A county-scale assessment." Urban Climate 35 (2021): 100733.

4. Abunnasr, Yaser, Mario Mhawej, and Nektarios Chrysoulakis. "SEBU: A novel fully automated Google Earth Engine surface energy balance model for urban areas." Urban Climate 44 (2022): 101187.

5. Allam, Mona, Mario Mhawej, Qingyan Meng, Ghaleb Faour, Yaser Abunnasr, Ali Fadel, and Hu Xinli. "Monthly 10-m evapotranspiration rates retrieved by SEBALI with Sentinel-2 and MODIS LST data." Agricultural Water Management 243 (2021): 106432.

6. Allen, Richard G., Luis S. Pereira, Dirk Raes, and Martin Smith. "Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56." Fao, Rome 300, no. 9 (1998): D05109.

7. Allen, Richard G., Masahiro Tasumi, Anthony Morse, Ricardo Trezza, James L. Wright, Wim Bastiaanssen, William Kramber, Ignacio Lorite, and Clarence W. Robison. "Satellite-based energy balance for mapping evapotranspiration with internalized calibration (METRIC)—Applications." Journal of irrigation and drainage engineering 133, no. 4 (2007): 395-406.

8. Anderson, Martha C., William P. Kustas, Christopher R. Hain, C. Cammalleri, F. Gao, M. Yilmaz, I. Mladenova, J. Otkin, M. Schull, and R. Houborg. "Mapping surface fluxes and moisture conditions from field to global scales using ALEXI/DisALEXI." Remote sensing of energy fluxes and soil moisture content 207 (2013): 232.

9. Asadi, Mehdi, and Khalil Valizadeh Kamran. "Comparison of SEBAL, METRIC, and ALARM algorithms for estimating actual evapotranspiration of wheat crop." Theoretical and Applied Climatology (2022): 1-11.

10. Bastiaanssen, Wim GM, Massimo Menenti, R. A. Feddes, and A. A. M. Holtslag. "A remote sensing surface energy balance algorithm for land (SEBAL). 1. Formulation." Journal of hydrology 212 (1998): 198-212.

11. Baum, M. E., S. V. Archontoulis, and M. A. Licht. "Planting date, hybrid maturity, and weather effects on maize yield and crop stage." Agronomy Journal 111, no. 1 (2019): 303-313.

12. Behnassi, Mohamed, and Mahjoub El Haiba. "Implications of the Russia–Ukraine war for global food security." Nature Human Behaviour (2022): 1-2.

13. Béné, Christophe, Deborah Bakker, Mónica Juliana Chavarro, Brice Even, Jenny Melo, and Anne Sonneveld. "Global assessment of the impacts of COVID-19 on food security." Global Food Security 31 (2021): 100575.

14. Bing-Fang, W. U., Xiong Jun, Yan Na-na, Y. A. N. G. Lei-dong, and D. U. Xin. "ETWatch for monitoring regional evapotranspiration with remote sensing." 19, no. 5 (2008): 671-678.

15. Boryan, Claire, Zhengwei Yang, Rick Mueller, and Mike Craig. "Monitoring US agriculture: the US department of agriculture, national agricultural statistics service, cropland data layer program." Geocarto International 26, no. 5 (2011): 341-358.

16. Bosilovich, M. G., R. Lucchesi, and M. Suarez, 2016: MERRA-2: file specification. GMAO Office Note No. 9 (Version 1.1), p 73. http://gmao. gsfc.nasa.gov/pubs/office_notes

17. Breiman, Leo. "Random Forests. Statistics Department." University of California, Berkeley, CA 4720 (2001).

18. Chen, Cheng, Qiuwen Chen, Binni Qin, Shuhe Zhao, and Zheng Duan. "Comparison of different methods for spatial downscaling of GPM IMERG V06B satellite precipitation product over a typical arid to semi-arid area." Frontiers in Earth Science (2020): 525.

19. Cooper, Mark, Tom Tang, Carla Gho, Tim Hart, Graeme Hammer, and Carlos Messina. "Integrating genetic gain and gap analysis to predict improvements in crop productivity." Crop Science 60, no. 2 (2020): 582-604.

20. Daly, Christopher, Michael Halbleib, Joseph I. Smith, Wayne P. Gibson, Matthew K. Doggett, George H. Taylor, Jan Curtis, and Phillip P. Pasteris. "Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States." International Journal of Climatology: a Journal of the Royal Meteorological Society 28, no. 15 (2008): 2031-2064.

21. Degano, M. Florencia, Raúl E. Rivas, Juan M. Sánchez, Facundo Carmona, and Raquel Niclòs. "Assessment of the potential evapotranspiration MODIS product using ground measurements in the Pampas." In 2018 IEEE Biennial Congress of Argentina (ARGENCON), pp. 1-5. IEEE, 2018.

22. Fadel, Ali, Mario Mhawej, Ghaleb Faour, and Kamal Slim. "On the application of METRIC-GEE to estimate spatial and temporal evaporation rates in a mediterranean lake." Remote Sensing Applications: Society and Environment 20 (2020): 100431.

23. Gilbert, R.O. Statistical Methods for Environmental Pollution Monitoring, Wiley, NY. 1987.

24. Gorelick, Noel, Matt Hancher, Mike Dixon, Simon Ilyushchenko, David Thau, and Rebecca Moore. "Google Earth Engine: PlanETry-scale geospatial analysis for everyone." Remote sensing of Environment 202 (2017): 18-27.

25. Grassini, Patricio, James E. Specht, Matthijs Tollenaar, Ignacio Ciampitti, and Kenneth G. Cassman. "High-yield maize–soybean cropping systems in the US Corn Belt." In Crop physiology, pp. 17-41. Academic Press, 2015.

26. Hall, D. K., G. A. Riggs, and V. V. Salomonson. "MODIS/terra snow cover daily L3 global 500m grid, version 6." Boulder, CO: NASA National Snow and Ice Data Center Distributed Active Archive Center (2016).

27. Hersbach, Hans, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas et al. "The ERA5 global reanalysis." Quarterly Journal of the Royal Meteorological Society 146, no. 730 (2020): 1999-2049.

28. Jarvis, Andy, Hannes Isaak Reuter, Andrew Nelson, and Edward Guevara. "Hole-filled SRTM for the globe Version 4." available from the CGIAR-CSI SRTM 90m Database (http://srtm. csi. cgiar. org) 15, no. 25-54 (2008): 5.

29. Keating, Brian A., Peter S. Carberry, Graeme L. Hammer, Mervyn E. Probert, Michael J. Robertson, D. Holzworth, Neil I. Huth et al. "An overview of APSIM, a model designed for

farming systems simulation." European journal of agronomy 18, no. 3-4 (2003): 267-288.

30. Kendall, M.G. Rank Correlation Methods, 4th edition, Charles Griffin, London. 1975.

31. Knipper, Kyle R., William P. Kustas, Martha C. Anderson, Joseph G. Alfieri, John H. Prueger, Christopher R. Hain, Feng Gao et al. "Evapotranspiration estimates derived using thermal-based satellite remote sensing and data fusion for irrigation management in California vineyards." Irrigation Science 37, no. 3 (2019): 431-449.

32. Lopez, Jose R., Jonathan M. Winter, Joshua Elliott, Alex C. Ruane, Cheryl Porter, Gerrit Hoogenboom, Martha Anderson, and Christopher Hain. "Sustainable Use of Groundwater May Dramatically Reduce Irrigated Production of Maize, Soybean, and Wheat." Earth's Future 10, no. 1 (2022): e2021EF002018.

33. Mancuso, Giuseppe, Giulio Demetrio Perulli, Stevo Lavrnić, Brunella Morandi, and Attilio Toscano. "Sars-cov-2 from urban to rural water environment: Occurrence, persistence, fate, and influence on agriculture irrigation. A review." Water 13, no. 6 (2021): 764.

34. Mann, H.B. 1945. Non-parametric tests against trend, Econometrica no. 13 (1945):163-171.

35. Mhawej, Mario, Ali Nasrallah, Yaser Abunnasr, Ali Fadel, and Ghaleb Faour. "Better irrigation management using the satellite-based adjusted single crop coefficient (aKc) for over sixty crop types in California, USA." Agricultural Water Management 256 (2021): 107059.

36. Mhawej, Mario, and Ghaleb Faour. "Open-source Google Earth Engine 30-m evapotranspiration rates retrieval: The SEBALIGEE system." Environmental Modelling & Software 133 (2020): 104845.

37. Mhawej, Mario, Arnaud Caiserman, Ali Nasrallah, Ali Dawi, Roula Bachour, and Ghaleb Faour. "Automated evapotranspiration retrieval model with missing soil-related datasets: The proposal of SEBALI." Agricultural Water Management 229 (2020a): 105938.

38. Mhawej, Mario, Georgie Elias, Ali Nasrallah, and Ghaleb Faour. "Dynamic calibration for better SEBALI ET estimations: Validations and recommendations." Agricultural Water Management 230 (2020b): 105955.

39. Monteith, J. L. "Evaporation and environment. pp. 205-234. In GE Fogg Symposium of the Society for Experimental Biology." The State and Movement of Water in Living Organisms 19 (1965).

40. Nafziger, Emerson D. "Corn planting date and plant population." Journal of Production Agriculture 7, no. 1 (1994): 59-62.

41. Nasrallah, Ali, Hatem Belhouchette, N. Baghdadi, M. Mhawej, T. Darwish, S. Darwich, and G. Faour. "Performance of wheat-based cropping systems and economic risk of low relative productivity assessment in a sub-dry Mediterranean environment." European Journal of Agronomy 113 (2020): 125968.

42. Norberg, O. Steven, Clinton C. Shock, and Erik Bruno Goncalves Feibert. "Growing irrigated soybeans in the Pacific Northwest." (2010).

43. Nouri, Milad, Mehdi Homaee, Mohammad Bannayan, and Gerrit Hoogenboom. "Towards shifting planting date as an adaptation practice for rainfed wheat response to climate change." Agricultural water management 186 (2017): 108-119.

44. Page, Jessica R. "American Farmland Trust." Journal of Agricultural & Food Information 19, no. 4 (2018): 300-306.

45. Price, John C. "On the analysis of thermal infrared imagery: The limited utility of apparent thermal inertia." Remote sensing of Environment 18, no. 1 (1985): 59-73.

46. Rizzo, Gonzalo, Juan Pablo Monzon, Fatima A. Tenorio, Réka Howard, Kenneth G. Cassman, and Patricio Grassini. "Climate and agronomy, not genetics, underpin recent maize yield gains in favorable environments." Proceedings of the National Academy of Sciences 119, no. 4 (2022): e2113629119.

47. Senay, G. B., Prasanna H. Gowda, Stefanie Bohms, T. A. Howell, Mackenzie Friedrichs, T. H. Marek, and J. P. Verdin. "Evaluating the SSEBop approach for evapotranspiration mapping with landsat data using lysimetric observations in the semi-arid Texas High Plains." Hydrology and Earth System Sciences Discussions 11, no. 1 (2014): 723-756.

48. Si, Shuyang. "EMPIRICAL ANALYSES OF FOOD AND ENERGY ECONOMICS AND POLICY IN CHINA." (2021).

49. Steduto, Pasquale, Theodore C. Hsiao, Dirk Raes, and Elias Fereres. "AquaCrop—The FAO crop model to simulate yield response to water: I. Concepts and underlying principles." Agronomy Journal 101, no. 3 (2009): 426-437.

50. Stockle, Claudio O., Steve A. Martin, and Gaylon S. Campbell. "CropSyst, a cropping systems simulation model: water/nitrogen budgets and crop yield." Agricultural systems 46, no. 3 (1994): 335-359.

51. Sulla-Menashe, Damien, and Mark A. Friedl. "User guide to collection 6 MODIS land cover (MCD12Q1 and MCD12C1) product." USGS: Reston, VA, USA 1 (2018): 18.

52. Vermote, Eric, Chris Justice, Martin Claverie, and Belen Franch. "Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product." Remote Sensing of Environment 185 (2016): 46-56.S

53. Winkler, Julie A., Raymond W. Arritt, and Sara C. Pryor. "Climate projections for the Midwest: Availability, interprETtion and synthesis." US National Climate Assessment Midwest Technical Input Report 24 (2012).

54. Xu, Qin, Rakhal Sarker, Glenn Fox, and Daniel McKenney. "Effects of climatic and economic factors on corn and soybean yields in Ontario: a county level analysis." International Journal of Food and Agricultural Economics (IJFAEC) 7, no. 1128-2019-560 (2019): 1-17.

55. Zainol Abidin, Mohd Ashraf, Muhammad Nasiruddin Mahyuddin, and Muhammad Ammirrul Atiqi Mohd Zainuri. "Solar photovoltaic architecture and agronomic management in agrivoltaic system: A review." Sustainability 13, no. 14 (2021): 7846.

56. Zhang, Peng, Zhiling Guo, Sami Ullah, Georgia Melagraki, Antreas Afantitis, and Iseult Lynch. "Nanotechnology and artificial intelligence to enable sustainable and precision agriculture." Nature Plants 7, no. 7 (2021): 864-876.

57. Zhou, Baoyuan, Yang Yue, Xuefang Sun, Xinbing Wang, Zhimin Wang, Wei Ma, and Ming Zhao. "Maize grain yield and dry matter production responses to variations in weather conditions." Agronomy Journal 108, no. 1 (2016): 196-204.

*Journal of Geophysical Research Letters*

Supporting Information for

# SEBALIGEE v2: Global Evapotranspiration Estimation Replacing Hot/Cold Pixels with Machine Learning

Mario Mhawej[1], Xiang Gao[2], John M. Reilly[2], and Yaser Abunnasr[1, †]

1 Department of Landscape Design and Ecosystem Management, Faculty of Agricultural and Food Sciences, American University of Beirut, Bliss St., Beirut 2020-1100, Lebanon;
2 Massachusetts Institute of Technology (MIT), Department of Earth, Atmospheric and PlanETry Sciences 77 Massachusetts Avenue, 54-918, Cambridge, MA 02139.

## Contents of this file

**Table S1: SPEI Thresholds for Drought/Wet Classification**

| SPEI Value | Meaning |
|---|---|
| **2.0 or more** | extremely wet |
| **1.6 to 1.99** | very wet |
| **1.3 to 1.59** | moderately wet |
| **0.8 to 1.29** | slightly wet |
| **0.5 to 0.79** | incipient wet spell |
| **-0.49 to 0.49** | near normal |
| **-0.79 to -0.5** | incipient dry spell |
| **-1.29 to -0.8** | mild drought |
| **-1.59 to -1.3** | moderate drought |
| **-1.99 to -1.6** | severe drought |
| **-2.0 or less** | extreme drought |

## Table S2: Description of various Eddy Covariance Flux towers sites used in our study

| Country | Flux tower name | Location (longitude, latitude) | Data DOI | Dates |
|---|---|---|---|---|
| **Belgium** | BE_LON | 4.746E, 50.551N | 10.18140/FLX/1440129 | Jan 2014 – Nov 2014 |
| **China** | - | 115.788E, 40.349N | - | Jan 2016 – Dec 2017 |
| **France** | FR_GRI | 1.951E, 48.844N | 10.18140/FLX/1440162 | Sep 2013 – Oct 2014 |
| **Germany** | DE_KLI | 13.522E, 50.893N | 10.18140/FLX/1440149 | Feb 2014 – Oct 2014 |
| | DE_GEB | 10.914E, 51.099N | 10.18140/FLX/1440146 | Apr 2014 – Oct 2014 |
| **Italy** | IT_CA2 | 12.026E, 42.377N | 10.18140/FLX/1440231 | Sep 2013 – Dec 2014 |
| **United States** | US_Bi1 | 121.499W, 38.099N | 10.17190/AMF/1480317 | Jan 2017 – Dec 2021 |
| | US_Bi2 | 121.535W, 38.109N | 10.17190/AMF/1419513 | Jan 2018 – Dec 2021 |
| | US_Ro6 | 93.058W, 44.695N | 10.17190/AMF/1419509 | Feb 2017 – Dec 2017 |
| | US_TW3 | 121.646W, 38.115N | 10.18140/FLX/1440110 | Sep 2013 – Oct 2014 |
| | US_TWT | 121.653W, 38.108N | 10.18140/FLX/1440106 | Sep 2013 – Oct 2014 |

**Table S3: Highest $R^2$ values for the step-wise regressions between ET/aKc of corn, soybeans and winter and various variables from 2013 to 2021. AT, DT, RN, and WIND represent air temperature; dewpoint temperature, surface net solar radiation and wind speed, respectively. SPEI-14 and SPEI-30 are SPEI calculated on 14 days and 30-days, respectively.**

| | Highest $R^2$ | | | | |
|---|---|---|---|---|---|
| | One variable | Two variables | Three variables | Four variables | Five variables |
| **Corn aKc** | WIND (0.2) | WIND + DT (0.24) | WIND + DT + AT (0.31) | WIND + DT + AT + SPEI-90 (0.34) | WIND + DT + AT + SPEI-30 + RN (0.35) |
| **Soybeans aKc** | WIND (0.14) | WIND + DT (0.2) | WIND + DT + AT (0.26) | WIND + DT + AT + SPEI-14 (0.29) | WIND + DT + AT + SPEI-14 + RN (0.31) |
| **Winter Wheat aKc** | WIND (0.01) | WIND + DT (0.11) | WIND + DT + AT (0.14) | WIND + DT + AT + SPEI-30 (0.19) | WIND + DT + AT + SPEI-30 + RN (0.20) |
| **Corn ET** | RN (0.71) | RN + WIND (0.75) | RN + WIND + DT (0.77) | RN + WIND + DT + SPEI-14 (0.79) | RN + WIND + DT + SPEI-14 + AT (0.8) |
| **Soybeans ET** | RN (0.73) | RN + WIND (0.78) | RN + WIND + DT (0.79) | RN + DT + AT + SPEI-14 (0.81) | RN + WIND + DT + AT + SPEI-14 (0.84) |
| **Winter Wheat ET** | RN (0.54) | RN + SPEI-30 (0.60) | RN + SPEI-30 + AT (0.60) | RN + SPEI-30 + AT + DT (0.63) | RN + SPEI-30 + AT + DT + WIND (0.65) |

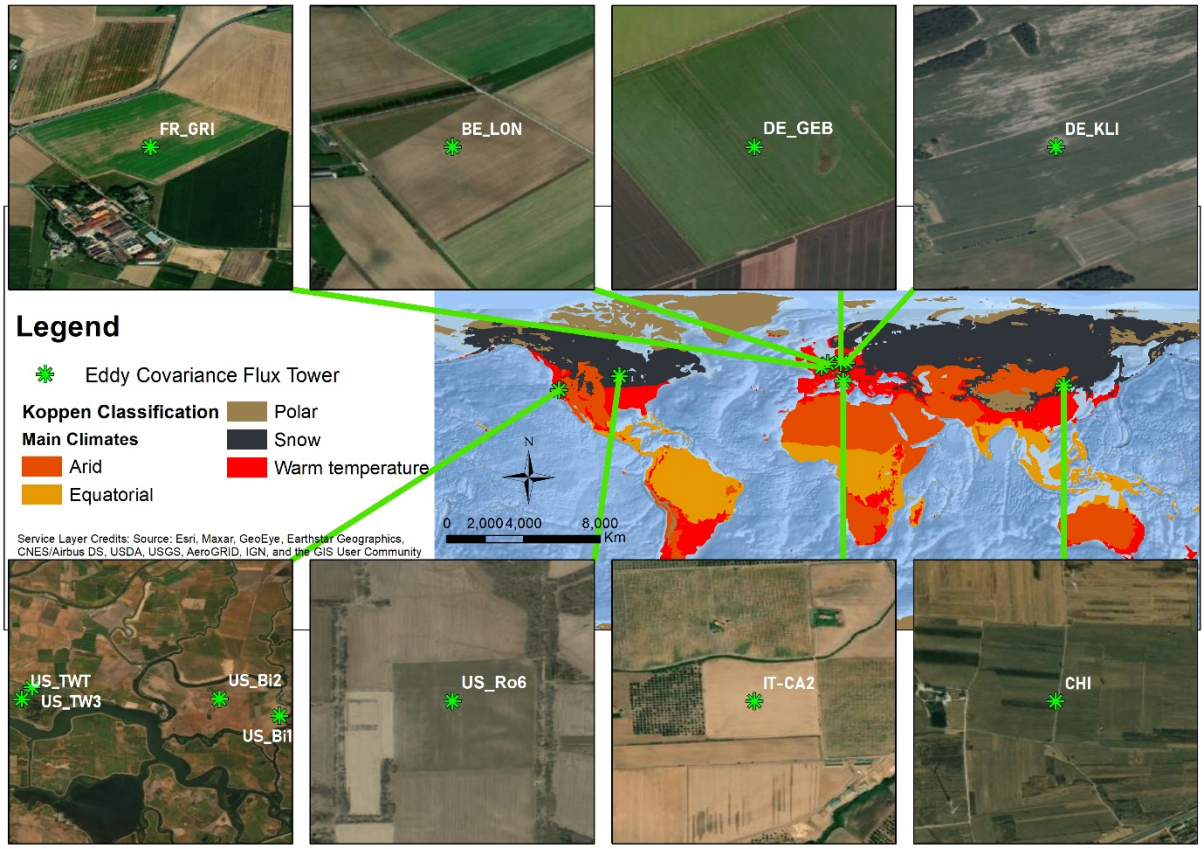


**Figure S1: Location of the flux towers in six countries as well as a Google Earth satellite view above each tower. The world map at the middle corresponds to Koppen climate classification**

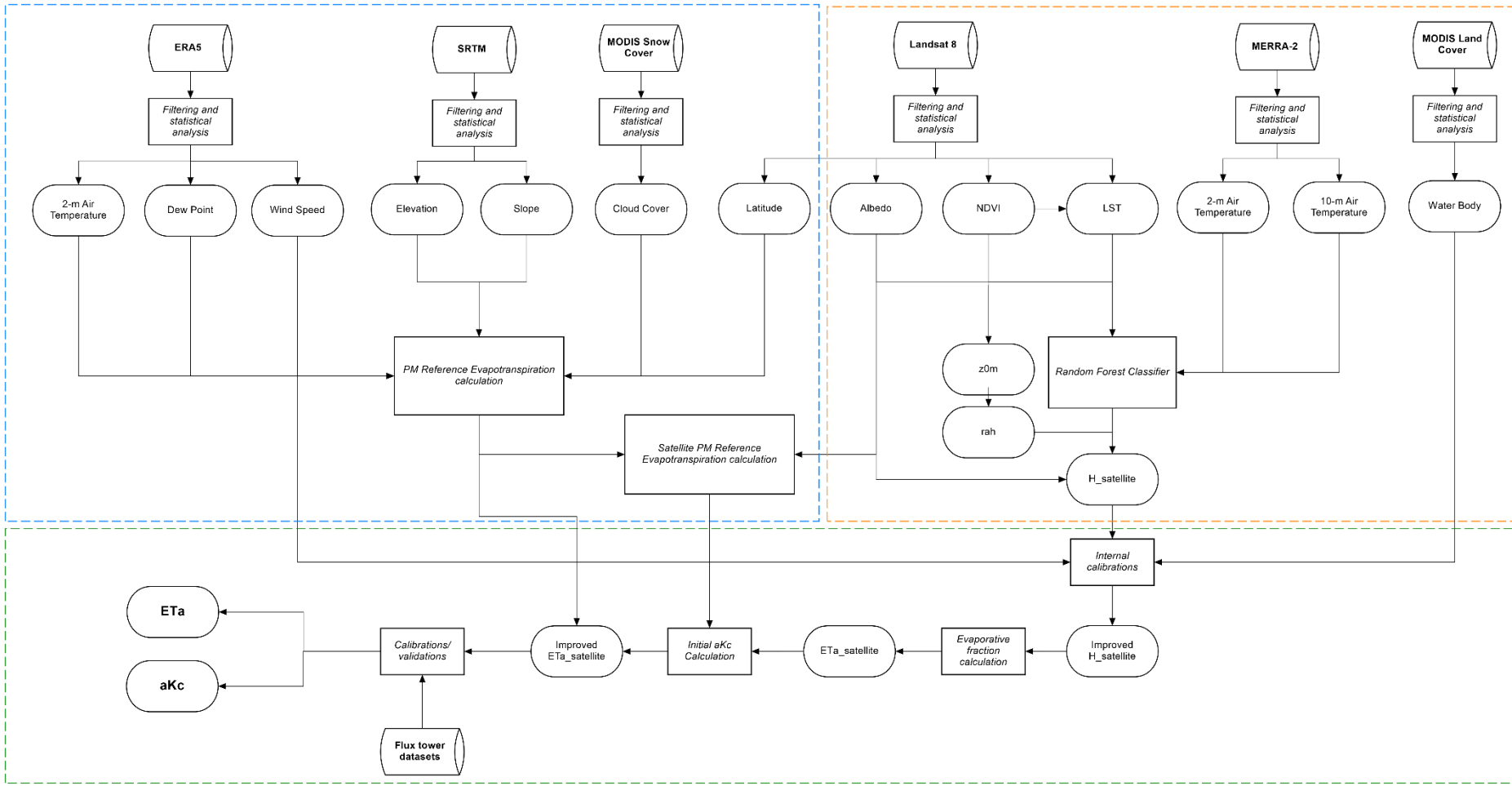


**Figure S2: SEBALIGEE v2 simplified flowchart of the used inputs and generated output**

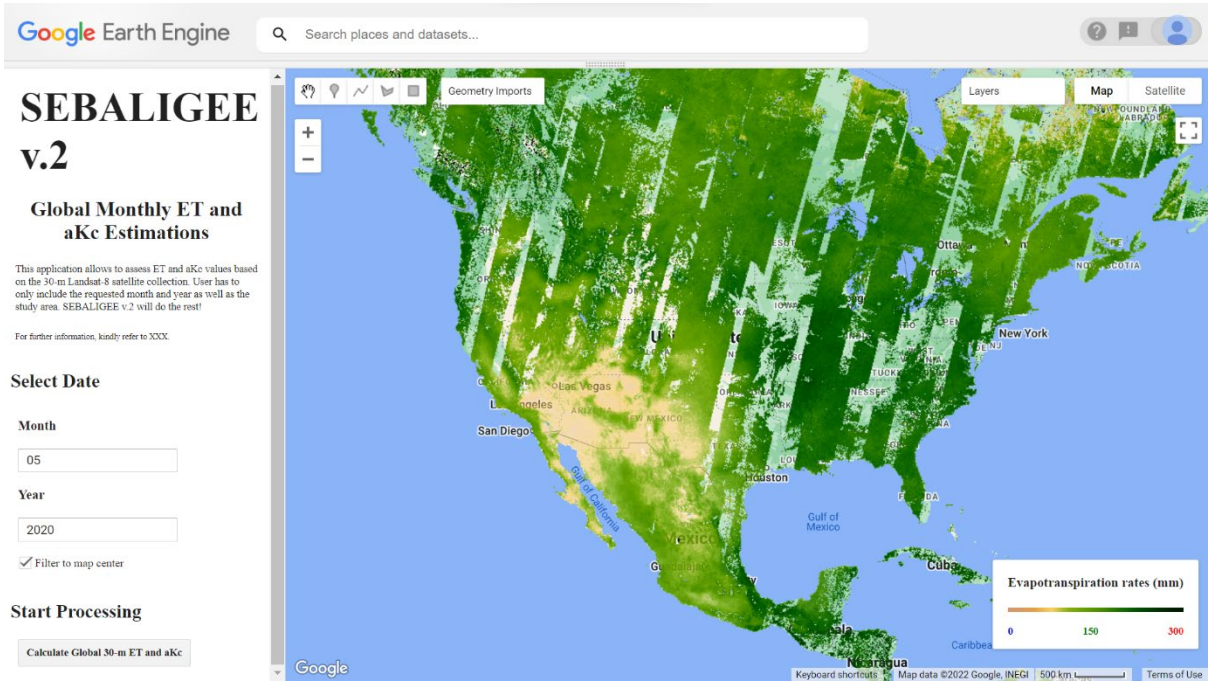


**Figure S3: Snapshot of the SEBALIGEE v.2 system interface hosted over the GEE platform**