Effect of the net radiation substitutes on maize and soybean evapotranspiration estimation using machine learning methods

Venturini, V., Walker, E., Fonnegra Mora, D. C. and Fagioli, G.

DOI: 10.31047/1668.298x.v39.n2.37104

SUMMARY

Accurate evapotranspiration (ET) estimation is essential for water management in crops, but it is not an easy task. Empirical ET methodologies require precise net radiation (Rn) measurements to obtain accurate results. Nevertheless, Rn measurements are not easy to obtain from meteorological stations. Thus, this study explored the use of machine learning algorithms with two Rn substitutes, to estimate daily ET: the extraterrestrial solar radiation (Ra) and a modelled Rn (RnM). Support Vector Machine (SVM), Kernel Ridge (KR), Decision Tree (DT), Adaptive Boosting (AB), and Multilayer Perceptron (MLP) were applied to model FLUXNET Rn and ET observations. Adaptive Boosting produced the best field Rn measurements (RnO), yielding a Root Mean Square Error of about 16 % of the mean observed Rn. The resulting Rn (AB RnM) was used to model daily crops ET employing the above-mentioned machine learning methods with RnO, AB RnM, and Ra, in conjunction with meteorological variables and the NDVI index. The evaluated methods were suitable to estimate ET, yielding similar errors to those obtained with RnO, when contrasted with ET observations. These results demonstrate that AB and KR are applicable with rutinary meteorological and satellite data to estimate ET.

Keywords: water stress, net radiation, crops, machine learning, Adaptive Boosting

Venturini, V., Walker, E., Fonnegra Mora, D. C. and Fagioli, G. (2022). Efecto de los sustitutos de radiación neta en la estimación de la evapotranspiración del maíz y la soja mediante métodos de aprendizaje automático. *Agriscientia 39 (2)*: 1-17

RESUMEN

La estimación precisa de la evapotranspiración (ET) es esencial para gestionar el riego en cultivos, pero no es una tarea fácil. Las metodologías empíricas de ET requieren mediciones precisas de la radiación neta (Rn) para obtener resultados confiables. Sin embargo, estas mediciones no son rutinarias en las estaciones meteorológicas. Este trabajo exploró el uso de aprendizaje

Fecha de recepción: 21/03/2022; fecha de aceptación: 19/09/2022.

automático para estimar la ET diaria con dos sustitutos de Rn: la radiación solar extraterrestre (Ra) y la Rn modelada (RnM). Se utilizó Support Vector Machine (SVM), Kernel Ridge (KR), Decision Tree (DT), Adaptive Boosting (AB) y Multilayer Perceptron (MLP) para modelar observaciones de FLUXNET. Adaptive Boosting brindó los mejores resultados con observaciones de Rn (RnO), con un valor para la raíz del error cuadrático medio de aproximadamente el 16 % de Rn medio observado. La Rn resultante (AB RnM) se utilizó para modelar la ET, usando RnO, AB RnM y Ra, junto a variables meteorológicas y el índice NDVI. Los métodos evaluados estimaron adecuadamente la ET, arrojando errores similares a los obtenidos con RnO, cuando se contrastan con las observaciones de ET. Estos resultados demuestran que AB y KR son aplicables con datos rutinarios meteorológicos y de satélite para estimar la ET.

Palabras clave: estrés hídrico, radiación neta, cultivos, aprendizaje automático, acelerador adaptativo

Venturini, V. (ORCID: 0000-0003-3040-9918) and Walker, E. (ORCID: 0000-0002-4287-4828): Universidad Nacional del Litoral (UNL), Facultad de Ingeniería y Ciencias Hídricas (FICH). Ciudad Universitaria. Ruta Nacional Nº 168 – Km 472,4. (3000) Santa Fe, Argentina. Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Godoy Cruz 2290 (C1425FQB) CABA – Argentina. Fonnegra Mora, D. C. (ORCID: 0000-0003-3780-622X): Universidad Nacional del Litoral (UNL), Facultad de Ingeniería y Ciencias Hídricas (FICH). Ciudad Universitaria. Ruta Nacional Nº 168 – Km 472,4. (3000) Santa Fe, Argentina. Fagioli, G (ORCID: 0000-0002-9548-2756): KILIMO S.A., Zenón López 1121, Pilar – Córdoba (X5972), Argentina.

Correspondence to: vventurini@fich.unl.edu.ar

INTRODUCTION

Actual evapotranspiration (ET) is a crucial process that links the terrestrial water and the energy balance (Xu et al*.,* 2019). Thus, during the last decades researchers have been studying the relationship between ET and its controlling factors to better estimate this process. In general, there are three identified ET driving forces, i.e., net radiation (Rn), atmospheric variables, and surface properties (Qiu et al., 2019). As Rn is the main forcing variable, the soil moisture content (SM) is the most important surface state variable on which ET depends. SM controls the exchange of latent and sensible heat between the surface and the atmosphere (Kim et al., 2018; Purdy et al., 2018). In fact, ET empirical and semiempirical methods, such as Penman (1948), Priestley and Taylor (1972) are based upon Rn; however, neither ET nor Rn are readily available observations (Jain et al., 2008; Chen, J. et al., 2020).

ET most precise field measurements are made with lysimeters, flux towers, and scintillometers (Tikhamarine, Malik, Pandey et al. 2020); these observations are scarce outside the northern hemisphere. Moreover, it can be roughly estimated that there are 100 meteorological stations for each ET observation device in developed countries, and this ratio might decrease in developing countries (Tikhamarine, Malik, Souag-Gamane, and Kisi, 2020).

On the other hand, the solar radiation fluxes are observed using pyranometers and solarimeters, between other instruments. These instruments are also rare to find in most of the meteorological stations due to the high cost of installation and maintenance (Yadav and Chandel, 2014). The scarcity of solar radiation data, available from few meteorological stations, led to the development of several radiation models for clear sky using remotely sensed data (Bisht et al*.,* 2005; Zhang et al*.,* 2020) and machine learning algorithms (ML) (Alizamir et al., 2020). Besides, other hybrid forecasting models are reported in the literature (Si et al*.,* 2020).

The impact of the radiation variables on the ET estimations has been the focus of different studies. For instance, error analysis on Rn and ET models

were done by Llasat and Snyder (1998). The authors concluded that Rn and ET estimates are sensitive to the soil temperature (Ts) errors and insensitive to the air temperature (Ta), however Ta is readily available everywhere on the Earth while Ts is available at coarser temporal resolution. Trnka et al. (2007) analysed the effect of solar radiation estimates on crop yield models and transpiration models. They reported a Root Mean Square Error (RMSE) of about 15 % in crop yield estimates with Ångström-Prescott global radiation formula and RMSEs up to 33% with the formulation presented by Hargreaves, Hargreaves and Riley (1985). Jain et al. (2008) highlighted the importance of solar radiation to compute the reference evapotranspiration (ETo). Their results showed that solar radiation has the major impact (more than 30 %) on the ETo estimation compared to other meteorological variables. However, Majidi et al. (2015) concluded that the effect of calculated missing solar radiation data on the Penman–Monteith (P-M) estimates is negligible, in both semi-humid and semi-arid climate conditions. Mokhtari et al. (2018) also investigated the impact of different solar radiation estimations (using empirical, physically-based data assimilation, and satellite observation models) on P-M equation in a semiarid region. They found that the ETo error was related to solar radiation error with a fourth-degree equation.

Carter and Liang (2019) evaluated the effects of remote sensing or observed radiation data, along with vegetation indexes, in 10 ET ML algorithms for diverse ecosystem types around the world. They showed that Global Land Surface Satellite (GLASS) solar radiation produced similar ET errors compared to that obtained with Rn measurements. Granata (2019) explored four ML algorithms, i.e., Regression Tree (RT), Bootstrap Aggregating (BA), Random Forest (RF), and Support Vector Machine (SVM), to model ET from a grassland site in Florida. The author built three ET models, combining Rn, soil moisture content, relative humidity (RH), Ta, wind speed, and sensible heat fluxes data. Their results emphasized the importance of taking into account Rn data to obtain satisfactory ET results; however, the author did not evaluate the impact of Rn in their results. More recently, Yamaç and Todorovic (2020) analysed the influence of meteorological variables in the potato ET estimation using the k-Nearest Neighbour (kNN), Artificial Neural Network (ANN) and Adaptive Boosting (AB) techniques. They improved ET estimates, by more than 50 % in terms of RMSE, when observed solar radiation data was included in the input dataset along with Ta, HR, and wind speed data. Tang et al. (2018) investigated two ML algorithms (ANN and SVM) to estimate ET in a rainfed maize field under mulching and non-mulching condition, using meteorological and crop data. The authors tested two different input combinations to model ET, i.e., one combination considered meteorological observations (maximum, minimum, mean Ta, maximum, minimum, mean RH, solar radiation, wind speed) and crop data (leaf area index, plant height), and the other combination of input variables had only meteorological data. They did not analyse the influence of radiation data in ET errors.

Most of the aforementioned studies were applied using observations of different shortwave and longwave radiations, which are not easily available in developing countries at field scale. Also, some of these studies analysed the influence of the radiation terms on estimating ETo with meteorological variables, and only a few of them used vegetation indexes or crop characteristics along with ML algorithms.

Thus, this work aims to analyse the errors of different Rn substitutes on ML models to estimate ET from two crops: maize (*Zea mays L.*) and soybean (*Glycine max (L.) Merr.*). Therefore, in the present study, ET was estimated using five ML methods, including Support Vector Machine (SVM), Kernel Ridge (KR), Decision Tree (DT), Adaptive Boosting (AB), and Multilayer Perceptron (MLP) with three different radiation inputs, i.e., observed Rn (RnO), modelled Rn (RnM), and computed extraterrestrial solar radiation (Ra), in conjunction with meteorological variables and the normalized difference vegetation index (NDVI).

MATERIALS AND METHODS

Meteorological and satellite data

Meteorological data provided by the FLUXNET ground observations network and a Moderate-Resolution Imaging Spectroradiometer (MODIS) satellite product were used in this study.

FLUXNET ground tower sites data labelled as croplands (CRO) were selected here, but only those with maize and soybean were processed. Besides, operative stations were considered in this analysis according to Purdy et al. (2018), Walker and Venturini (2019) criteria, i.e., stations with Rn and ET high-quality measurements, with more than 60 % of reliable Rn data and few ET outliers. The selected meteorological stations, crop information source, location, time span, the Köppen-Geiger climate class (Beck et al., 2018), the dominant soil group, and the mean observed ET and Rn for each site are listed in Table 1. Data source references

can be found on FLUXNET website. Six of the processed stations are in temperate climates and only two are in continental climates. The mean Rn varies according to the latitude, as expected, and the mean ET varies from approximately 1 to 7.5 mm/d. The dominant soil group information for each site was obtained from the Harmonized World Soil Database (HWSD), published in 2012 by Food and Agriculture Organization of the United Nations (FAO), International Institute for Applied Systems Analysis (IIASA), ISRIC-World Soil Information, Institute of Soil Science – Chinese Academy of Sciences (ISSCAS), and Joint Research Centre of the European Commission (JRC). Table 1 shows that most of the stations are in different soil types, except for US-Ne1, US-Ne2 and US-Ne3 which, due to their proximity, are in the same dominant soil group. However, based on the USDA texture classification (Shirazi et al., 1988) most of these soils have loam textures.

FLUXNET network was established for quantifying carbon, water vapor, and energy fluxes (Miralles et al., 2011), and fluxes data are available at https://fluxnet.org/data/fluxnet2015-dataset/. For this work, the meteorological variables mean Ta (Ta), minimum Ta (Tamin), maximum Ta (Tamax), Ta range (Tar), mean RH (RH), minimum RH (RHmin), maximum RH (RHmax), Rn, and latent heat flux (LE) were considered. LE was converted to water loss measure, i.e., ET in mm/d. Raw FLUXNET data were pre-processed for removing missing or wrong data and outliers using Tukey's methodology (Schwertman et al., 2004). Then, mean daily values were calculated by integrating only the quality checked measurements of the daylight hours.

NDVI index was obtained from MODIS products using the Google Earth Engine platform. NDVI was estimated with MOD09Q1 V6 product, an eight-day composite dataset, which provides an estimate of the surface spectral reflectance of EOS-Terra MODIS bands 1 and 2 corrected for atmospheric conditions such as gasses, aerosols, and Rayleigh scattering. The NDVI has a spatial resolution of about 250m, comparable to the FLUXNET towers footprint. The time series of NDVI index were obtained for each FLUXNET site and linearly interpolated after passing a moving average filter, to estimate daily values.

ML algorithms were calibrated and validated with FLUXNET ET as the output variable and FLUXNET meteorological data, Rn substitutes, and NDVI were used as input variables.

Machine learning

As mentioned in the Introduction, in this work five ML algorithms were used, considering the results published in Carter and Liang (2019). The regressor methods applied here are SVM, KR, DT, AB, and MLP (Carter and Liang, 2019). These methodologies are briefly described below.

Table 1. Summary of the general information for FLUXNET tower sites used in this study

Country (Site ID)	Crops	Time Span	Latitude (degree)	Longitude (degree)	Köppen-Geiger Climate class	FAO HSWD Dominant soil group	Mean observed ET (mm/d)	Mean observed Rn (W/m^2)
Germany (DE-Kli)	spring barley, maize, winter barley, rapeseed, winter wheat	2004-2014	50.8930	13.5223	Dfb	Cambisols	3.08	282.23
Denmark (DK-Fou)	Maize	2005	56.4842	9.5872	Dfb	Podzols	1.16	232.79
France (FR-Gri)	winter wheat, winter barley, mustard, maize	2004-2014	48.8442	1.9519	Cfb	Podzoluvisols	5.15	294.79
USA (US-CRT)	soybean, winter wheat	2011-2013	41.6284	-83.3470	Dfa	Luvisols	7.07	362.55
USA (US-Ne1)	irrigated maize	2001-2013	41.1650	-96.4766	Dfa	Phaeozems	7.39	382.22
USA (US-Ne2)	irrigated maize-soybean rotation	2001-2013	41.1648	-96.4701	Dfa	Phaeozems	7.15	375.17
USA (US-Ne3)	rainfed maize-soybean rotation	2001-2013	41.1796	-96.4396	Dfa	Phaeozems	6.43	359.42
USA (US-Tw2)	Maize	2012-2013	38.1047	-121.6433	Csa	Fluvisols	6.19	467.04

1. Support Vector Machine (SVM)

SVM was developed by Vapnik (1999). The SVMs are derived from the concept of structural risk minimization theory to minimize the empirical risk and the confidence interval of the learning machine. The strength of these methodologies is their solid mathematical bases in statistical theory and have demonstrated accurate results in a wide range of real-world problems. Initially developed for solving classification problems, SVM techniques can also be successfully applied in regression problems.

A regression is estimated by using SVM for a given data set $\{(x_i, y_i)\}_n$, where x_i are the input vectors, y_i is the output value and n is the total number of data sets (Tang et al., 2018). So, the regression equation can be formulated as:

$$f(x) = \omega\varphi(x) + b \quad (1)$$

where ω is weight vector; φ(x) is the nonlinear transfer function and b is the bias.

The parameters ω and b can be expressed by minimizing the regularized risk function as:

$$R = C\sum_{i=1}^{N} L_\varepsilon [f(x_i), y_i] + \frac{1}{2} \| \omega \|^2 \quad (2)$$

$$L_\varepsilon[f(x_i), y_i] = \begin{cases} 0 \text{ for } |f(x_i) - y_i| < \varepsilon \\ |f(x_i) - y_i| - \varepsilon \text{ otherwise} \end{cases} \quad (3)$$

where C is a positive constant, $\|\omega\|^2$ is the regularization term which denotes the Euclidean norm, and Lε is called ε-insensitive loss function. Then, a nonlinear regression function can be derived as:

$$f(x, a, a^*) = \sum_{i=1}^{N} (a_i - a^*) k(x_i, x) + b \quad (4)$$

where a_i-$a_i^* = 0$, a_i-$a_i^* \geq 0$ i = 1, …., N, and the kernel function $k(x_i, x)$ describes the inner product in the D-dimensional feature space.

$$k(x_i, x) = \sum_{i=1}^{D} \phi_j(x_i)\, \phi_i(x) \quad (5)$$

2. Kernel Ridge (KR)

The KR method (Saunders et al., 1998) is a special case of SVM, which combines ride regression with kernel techniques for capturing nonlinear relationships (You et al., 2018). Specifically, in the KR method, the predictor variables are mapped nonlinearly into a high-dimensional space, where the estimation of the predictive regression model is based on a shrinkage estimator to avoid overfitting (Exterkate et al., 2016).

For a given dataset $\{(x_i, y_i)\}$ with N samples, the KR regression can be estimated as:

$$\hat{f} = \underset{f \in \mathcal{H}}{\operatorname{argmin}} \left\{ \frac{1}{N} \sum_{i=1}^{N} (f(x_i) - y_i)^2 + \lambda \| f \|_{\mathcal{H}}^2 \right\} \quad (6)$$

where $\lambda > 0$ is a regularization parameter, and $\mathcal{H}$ is a reproducing kernel Hilbert space.

This technique is suitable for estimating nonlinear models with many predictors and it is widely used in different applications (Hofmann et al., 2008; Zhang et al., 2013).

3. Decision Tree (DT)

DTs are very popular ML techniques since they have a simple format. Also, they are efficient methods for solving classification and regression problems (Xu et al., 2005). Basically, DT algorithms construct a tree with leaves that are labelled with a specific class property and with inner nodes that represent the class attribute. Given an inner node, the breeding of that node follows different values of a descriptive attribute. The result of this process is a decision tree, that classifies the new information following a track beginning from the root to a leaf according to the selected descriptive attributes. These models generate a set of rules which can be used for prediction through the repetitive process of splitting.

In a regression problem, $X = X_1, X_2, \ldots, X_{pn}$ are the predictor variables and pn is the total number of predictor variables. Let n be the number of observations and $Y = Y_1, Y_2, \ldots, Y_n$ a target variable that takes continuous values, vf is a feature variable and th is a threshold value. Let t and $g=(vf, th_t)$ be a node and a candidate split, respectively. Then:

$$Q1\,(g) = (x, y) | x_{vf} \leq th_t \quad (7)$$

$$Q2\,(g) = (x, y) | x_{vf} > th_t \quad (8)$$

$$\hat{Y}_t = \frac{1}{n} \sum_{i \in n} Y_i \quad (9)$$

where equation 7 and 8 shows that Q1 (that is the left side in the decision tree) and Q2 (the right side in the decision tree) are found by splitting the data into g candidate split. Then, formulation

9 presents the calculation of the mean predicted value at terminal node.

The ability to track and evaluate every step in the DT process is an important advantage that makes these methods applicable to different problems. Indeed, DTs have been applied to remote sensing (Zhang et al., 2017), biology (Darnell et al., 2007), hydrology (Nourani et al., 2019), among other applications.

4. Adaptive Boosting (AB)

The AB is one of the most used boosting methods given its simplicity and accurate estimation (Wu et al., 2008). It is an ensemble learning algorithm in which weak learners are combined into a weighted sum. The success of this method lies in looking for a strong learner by lineal combinations of weak learners:

$$H(x) = \text{sign}\left(\sum_{k=1}^{K} a_k h_k(x)\right) \quad (10)$$

where $h_k(x)$ denotes the kth weak learner; K is the number of weak learners; a_k denotes the coefficient of the kth weak learner, and H(x) denotes a strong learner.

The training process is done in three steps: first, a training dataset is randomly selected to begin; secondly, the model is repetitively trained to select the training set based on the errors of the last results, and finally, the model assigns the higher weight to weaker estimations. The algorithm iterates until the training data is estimated with the minimum error and reaches the maximum number of iterations. AB algorithms can be used both in classification and regression problems (Yamaç and Todorovic, 2020).

5. Multilayer Perceptron Regressor (MLP) – Artificial Neuronal Network (ANN)

The ANN models are the most well-known ML methods, used for modelling soil moisture (García et al., 2019), water balance (Kumar et al., 2011), and solar radiation (Yadav and Chandel, 2014). The method connects neurons (input variables), by assigning weight to each of them, to find the pattern that explains the output variable. The training ANN process defines the relationship among the input neurons, so that it can be applied to a new dataset to estimate the output variable (Yamaç and Todorovic, 2020).

The MLP model consists of multiple layers, classified as input, hidden, and output. Input neurons are the explanatory variables, the output layer is the estimated unknown variable, while the hidden layers are artificial neurons needed to connect the input and the output layers. Hidden layers are critical for modelling nonlinear processes. The MLP model can be mathematically formulated as:

$$Y = F\left(\sum_{j=1}^{m} w_{kj} F\left(\sum_{i=1}^{n} w_{ji} X_i + B_j\right) + B_k\right) \quad (11)$$

where w_{kj} are weights between hidden and output layers; w_{ji} are weights between input and hidden layers; X_i are input variables; m is the number of neurons in a hidden layer; n is the number of neurons in an input layer, B_j and B_k are the bias values of the neurons in the hidden and output layers, respectively; F is the transfer function; and Y is the output.

Model implementation and hyperparameters selection

In this work, the data were normalized using the mean and the standard deviation, as suggested by Yamac and Todorovic (2020). Then, the amount of data was randomly partitioned into training and testing data sets. Specifically, 80 % of the data were used for the parameter tunning of each ML method, and the remaining 20% were used for testing. Table 2 presents the main statistics (minimum, maximum, mean, standard deviation and median) for each variable for the training and test, maize and soybean dataset. It can be noted that the proposed variables have similar statistics in both datasets, suggesting that training and test sets are not significantly different. ET maximum values in both datasets are about 20 mm/d, although the mean values are around 6 mm/d. It was observed that stations US-Ne1 and US-Ne2 (watered field) presents maize ET values as high as 21 mm/d during summer while in US-Ne3 (rainfed) ET reaches 15 mm/d. The other stations show values lower than 10 mm/d.

A k-fold cross-validation method was applied in prediction error estimation, and to set up the hyperparameters. This method is an iterative process, consisting of randomly splitting the dataset into k groups of approximately equal size, k-1 groups are used to train the model and one of the groups is used for testing. This process is repeated k times using a different group for testing in each iteration. The process generates k error estimates whose average is used as the final estimate. Here, the k-fold method was applied dividing the dataset into five subsets, i.e., k=5, (Anguita et al., 2005).

Table 2. Minimum (Min), maximum (Max), mean, standard deviation (SD), and median statistics for each used variable for the training and test dataset

Variable	Training data					Test data				
	Min	Max	Mean	SD	Median	Min	Max	Mean	SD	Median
Ta (°C)	0.10	35.74	22.74	6.25	23.73	2.46	35.64	23.23	6.29	24.46
Tamin (°C)	0.04	32.52	18.30	6.17	19.06	0.75	30.59	18.66	6.11	19.72
Tamax (°C)	0.10	38.92	25.24	6.42	26.42	3.18	38.16	25.72	6.48	26.98
Tar (°C)	0.00	26.54	6.94	3.07	6.62	0.00	19.20	7.06	2.78	6.93
RH (%)	13.20	100.00	59.11	16.33	58.25	20.72	100.00	59.57	16.61	58.68
RHmin (%)	9.59	100.00	48.99	16.91	47.80	12.80	100.00	49.02	17.32	47.35
RHmax (%)	16.84	100.00	76.86	14.48	77.70	30.40	100.00	77.84	14.76	79.39
NDVI	0.06	0.93	0.59	0.23	0.61	0.17	0.95	0.53	0.26	0.46
RnO (W/m^2)	28.35	632.41	366.39	128.35	385.06	28.30	646.28	356.37	119.95	381.08
ET (mm/d)	0.01	21.94	6.69	4.45	5.72	0.40	22.29	6.54	4.31	5.55

Rn substitutes

As aforesaid, routinely Rn measurements are not easy to obtain from meteorological stations. Hence, maize and soybean ET errors were tested with two Rn substitutes i.e., RnM and Ra.

The machine learning techniques presented here were applied to model Rn using the meteorological data from the stations listed in Table 1. Ra, the solar radiation received at the top of the atmosphere on a horizontal surface, is calculated as a function of the latitude, date, and time of day. Here, Ra was computed according to the methodology proposed by FAO 56 (Allen et al., 1998), through the following formulation:

$$\mathrm{Ra} = \left(\frac{24*(60)}{\pi} G_{sc} d_r [\omega_s \sin(\varphi)\sin(\delta) + \cos(\varphi)\cos(\delta)\sin(\omega_s)]\right) * 11.6 \quad (12)$$

where Ra is the extraterrestrial solar radiation (W/m^2), G_{sc} is the solar constant (0.0820 $MJ/m^2/d$), d_r is the inverse relative distance Earth-Sun, ω_s is the sunset hour angle (radians), φ is the latitude (radians) and δ is the solar declination (radians).

Models performance

In order to analyse the performance of the SVM, KR, DT, AB, and MLP algorithms, the RMSE, the Mean Absolute Error (MAE), Bias, and the determination coefficient (R^2) were quantified. Besides, Taylor's diagram was plotted (Taylor, 2001), which comprised the standard deviation (SD), correlation coefficient (r), and the centered Root Mean Square difference (RMS) statistics. The equations of the statistics RMSE, MAE, Bias, R^2, SD, r, and RMS used here are the following:

$$\mathrm{RMSE} = \sqrt{\sum \frac{(O-M)^2}{n}} \quad (13)$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|O - M| \quad (14)$$

$$\mathrm{Bias} = \sum (O - M)/n \quad (15)$$

$$R^2 = \frac{\sum(M-\bar{O})^2}{\sum(O-\bar{O})^2} \quad (16)$$

$$\mathrm{SD_O} = \sqrt{\frac{\sum(O-\bar{O})^2}{n}} \quad (17)$$

$$\mathrm{SD_M} = \sqrt{\frac{\sum(M-\bar{M})^2}{n}} \quad (18)$$

$$r = \frac{\frac{1}{n}\sum(M-\bar{M})*(O-\bar{O})}{\mathrm{SD_M}*\mathrm{SD_O}} \quad (19)$$

$$\mathrm{RMS} = \sqrt{\frac{\sum[(M-\bar{M})-(O-\bar{O})]^2}{n}} \quad (20)$$

where n is the number of observations. O is the observed data, M is the modelled data, $\bar{O}$ and $\bar{M}$ are the mean values of O and M, respectively. Also, $\mathrm{SD_M}$ and $\mathrm{SD_O}$ are the standard deviations of M and O, respectively.

RESULTS AND DISCUSSION

Radiation analysis

Year 2005 was randomly selected to present results from Ra calculation and RnO in Figure 1. In this Figure, those FLUXNET stations with maize and soybean data during 2005 were plotted. It is clear that Ra is about twice RnO, however both variables present similar trends. Ra is the incident solar radiation outside of the atmosphere, so it is reasonable to consider it as a radiation input, substituting Rn, in ET ML calculation.

The meteorological variables Ta, Tamin, Tamax, RH, RHmin, RHmax, and Ra were the inputs for the SVM, KR, DT, AB, and MLP algorithms to obtain RnM. The data were normalized using the mean and the standard deviation, and randomly partitioned into training and testing data sets. Then, the k-fold cross-validation method was applied, and the results were contrasted with RnO using the aforementioned statistics.

The mean, maximum and minimum RnO for all the processed stations were 361.39 W/m^2, 639.35 W/m^2, and 28.33W/m^2, respectively, while the same statistics for Ra were 888.57 W/m^2, 1059.83 W/m2, and 393.91 W/m^2.

Figure 2 shows the results in terms of the median, first (25 %), and third (75 %) quartiles, the data range and outliers of the RnM. The RMSE, MAE, and R^2 metrics were added to the box of each method. Clearly, AB has better performed RnO,

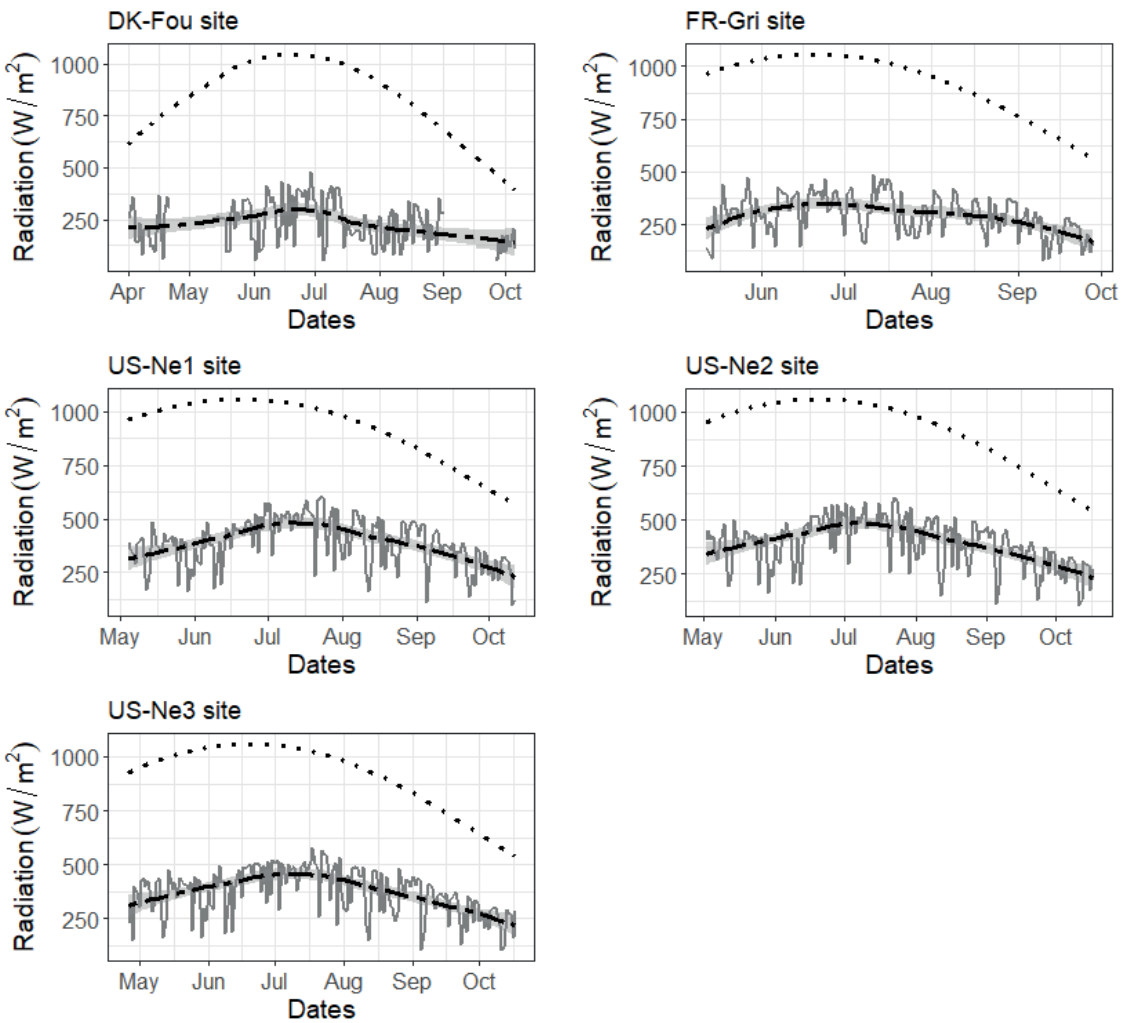

Figure 1. Comparison of calculated daily Ra (black dotted line) with mean daily RnO (grey solid line) for FLUXNET stations for a randomly selected year (2005) with maize and soybean crops. The black dashed line shows RnO trends

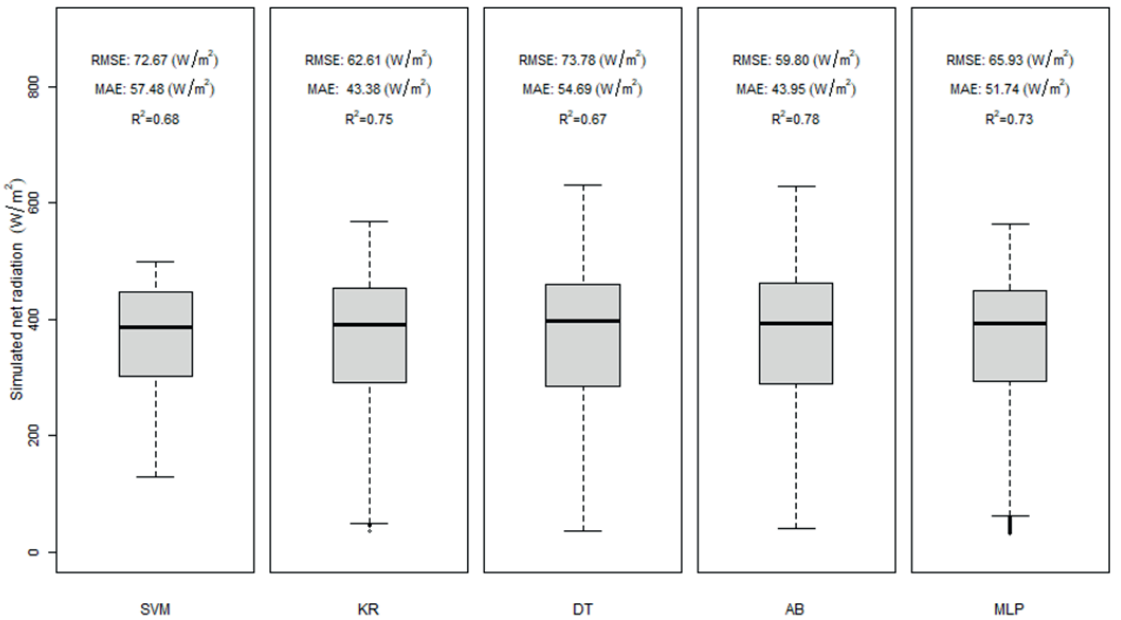

Figure 2. Boxplots of the Rn estimations for SVM, KR, DT, AB, and MLP algorithms. The RMSE, MAE, and R^2 between RnO and RnM are presented for each evaluated method

with the lowest RMSE of 59.80 W/m² (16.4 % of the mean RnO). The KR and MLP algorithms computed Rn estimations with errors similar to those obtained with the AB algorithm. The DT and SVM methods presented the poorest correlations compared with RnO, with a RMSE of 73.78 W/m² and 72.67 W/m², respectively.

Figure 3 shows the scatter plot between AB Rn (RnM) estimates and RnO, for all the studied sites. It can be observed that AB estimations presented a good correlation with field Rn measurements, although there are important differences between them.

The results of this analysis show that AB yielded the best results compared to field Rn measurements. Hence, AB Rn estimations will be used as a Rn substitute to calculate daily ET.

The results presented in Figure 2 demonstrate that all the evaluated ML methods were able to adequately model RnO. However, the AB, KR, and MLP algorithms exhibited the best RnO estimation accuracy, being AB the technique that showed the lowest RMSE (16.4 % of the mean RnO). These results are comparable to those presented by other studies. Wang et al. (2019) applied a Boosting method to estimate surface shortwave Rn, reporting a RMSE of about 11% of the observed Rn. Similarly, Jiang et al. (2014) published a RMSE of about 16% of the mean RnO, using the MLP algorithm to

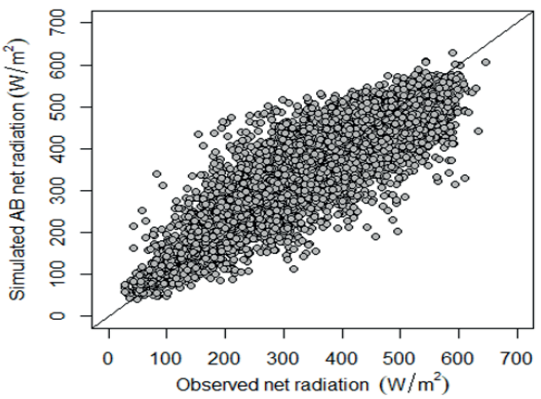

Figure 3. Relationship between AB Rn estimations and ground Rn observations (n=7051) for all the study FLUXNET stations. The solid black line represents the 1:1 line

simulate field Rn. However, these works modelled Rn from multi-source data, using remote sensing products, surface measurements, and reanalysis products.

On the other hand, Ojo et al. (2021) used MLP with observed meteorological variables for predicting RnO, obtaining RMSEs of about 8 % of the observed data. Their investigations were conducted only in tropical regions (Ojo et al., 2021). Here, RnO was estimated using routinely measured meteorological variables from stations

spatially distributed across the world, producing similar errors.

ET machine learning models for Maize and Soybean

The proposed ML algorithms were applied to estimate daily maize and soybean ET with three different radiation inputs, i.e., RnO, AB RnM, and calculated Ra, in conjunction with seven meteorological variables (Ta, Tamin, Tamax, Tar, RH, RHmin, RHmax), and the vegetation index NDVI. Thus, three scenarios were analysed in this study, as shown in Table 3, to investigate the effect of Rn substitutes estimates in maize and soybean ET errors.

The data were normalized, randomly splitted and pre-processed, as already explained in sub-section Model implementation and hyperparameters selection. The soundness of ML methods was evaluated using the proposed statistics.

Results of ML algorithms for daily maize ET, for three input combinations, were contrasted with observed ET and presented in Table 4. As expected, combination 1 yielded the lowest error for each case, given it made use of the most accurate Rn data, i.e., RnO. However, the errors and bias from the Rn substitutes are close to those of combination 1. The KR and AB methods presented the best results compared with field ET measurements for each evaluated combination (see Table 4). Using a simple estimation of the amount of incoming energy, such as Ra, would increase daily maize ET errors by 6% of the mean observed ET (6.72 mm/d), compared to the RMSE obtained in combination 1.

The efficiency of ML methods with the three input combinations is shown in Figure 4. Taylor's diagrams confirm that RnO produces the lowest RMS compared with field ET measurements. Even so, all the evaluated algorithms yielded correlations higher than 0.87 and similar SD for the different radiation inputs.

From the above results, AB seems to be the most precise ML algorithm, thus it was used to plot the comparison between simulated and observed ET with all the input combinations (see Figure 5).

Table 4. Statistical values of Support Vector Machine (SVM), Kernel Ridge (KR), Decision Tree (DT), Adaptive Boosting (AB), and Multilayer Perceptron (MLP) ML algorithms for daily maize and soybean ET estimation under the different input combinations

	Statistics			
	RMSE (mm/d)	MAE (mm/d)	R^2	Bias (mm/d)
Ta, Tamin, Tamax, Tar, RH, RHmin, RHmax, NDVI, RnO				
		Maize		
SVM1	1.34	0.99	0.91	0.019
KR1	**1.28**	**0.96**	**0.92**	**0.007**
DT1	1.61	1.19	0.87	0.010
AB1	**1.27**	**0.93**	**0.92**	**0.012**
MLP1	1.31	0.99	0.91	0.026
		Soybean		
SVM1	1.35	0.98	0.90	0.055
KR1	**1.26**	**0.93**	**0.91**	**0.017**
DT1	1.70	1.24	0.84	0.006
AB1	1.30	0.95	0.90	0.032
MLP1	**1.26**	**0.94**	**0.90**	**-0.007**
Ta, Tamin, Tamax, Tar, RH, RHmin, RHmax, NDVI, RnM				
		Maize		
SVM2	1.81	1.33	0.84	0.120
KR2	**1.75**	**1.30**	**0.85**	**0.005**
DT2	2.19	1.59	0.77	-0.003
AB2	**1.78**	**1.29**	**0.84**	**0.072**
MLP2	1.79	1.35	0.84	0.008
		Soybean		
SVM2	1.81	1.32	0.82	0.185
KR2	**1.75**	**1.29**	**0.83**	**0.015**
DT2	2.44	1.76	0.70	-0.024
AB2	1.80	1.31	0.82	0.098
MLP2	**1.77**	**1.32**	**0.83**	**-0.016**
Ta, Tamin, Tamax, Tar, RH, RHmin, RHmax, NDVI, Ra				
		Maize		
SVM3	1.83	1.37	0.83	0.073
KR3	**1.76**	**1.32**	**0.85**	**0.005**
DT3	2.16	1.55	0.77	0.046
AB3	**1.68**	**1.21**	**0.86**	**0.051**
MLP3	1.80	1.36	0.84	0.020
		Soybean		
SVM3	1.89	1.41	0.80	0.155
KR3	**1.82**	**1.36**	**0.82**	**0.016**
DT3	2.30	1.68	0.72	0.015
AB3	**1.78**	**1.29**	**0.83**	**0.093**
MLP3	1.84	1.38	0.81	0.014

Best statistics are highlighted in bold.

Table 3. The combinations of input variables used in Support Vector Machine (SVM), Kernel Ridge (KR), Decision Tree (DT), Adaptive Boosting (AB), and Multilayer Perceptron (MLP) ML algorithms

ML algorithms					Scenarios
SVM1	KR1	DT1	AB1	MLP1	Ta, Tamin, Tamax, Tar, RH, RHmin, RHmax, NDVI, RnO
SVM2	KR2	DT2	AB2	MLP2	Ta, Tamin, Tamax, Tar, RH, RHmin, RHmax, NDVI, RnM
SVM3	KR3	DT3	AB3	MLP3	Ta, Tamin, Tamax, Tar, RH, RHmin, RHmax, NDVI, Ra

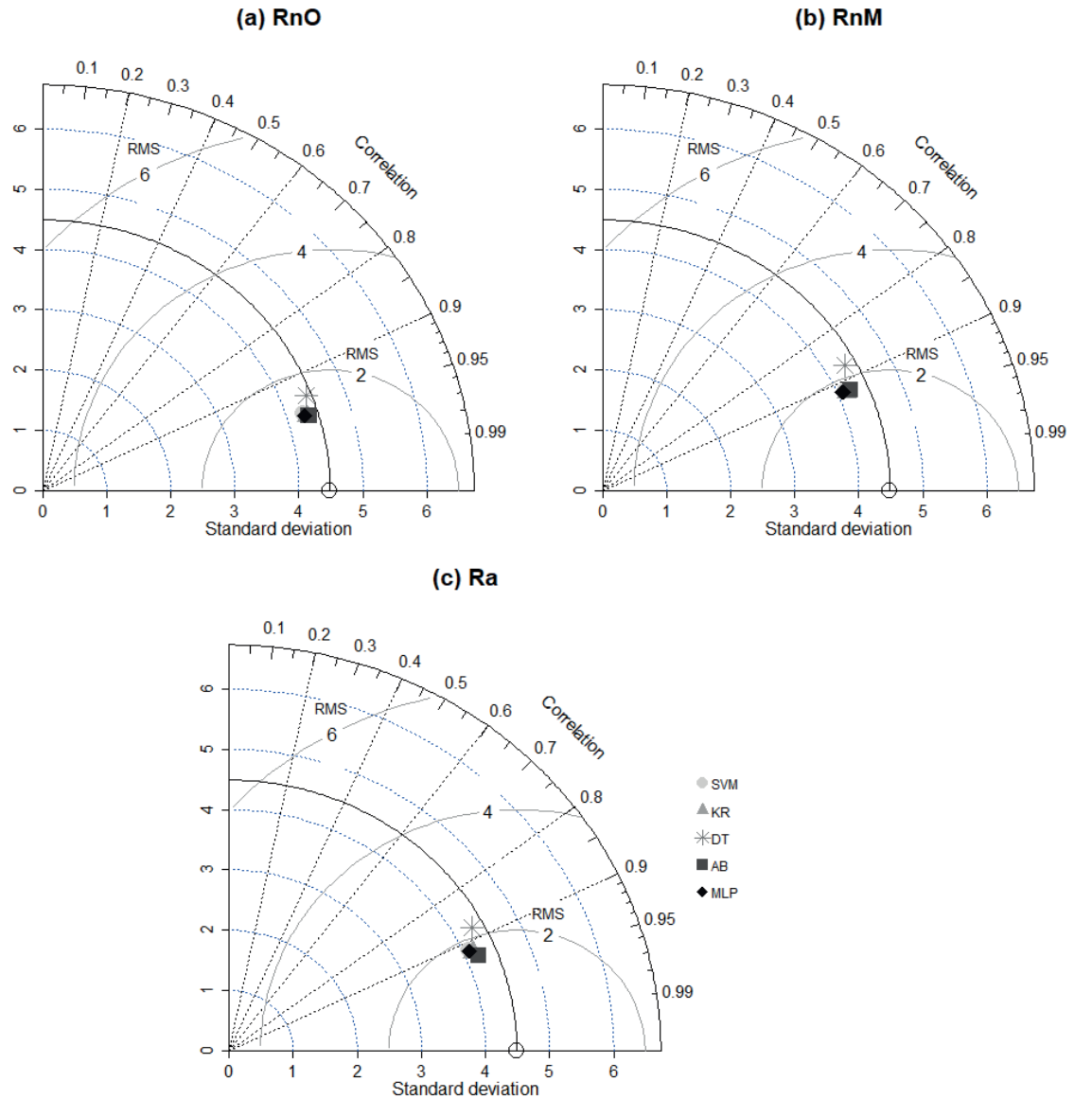

Figure 4. Taylor's diagrams for comparative assessment of SVM, KR, DT, AB, and MLP daily maize ET estimation, with three different radiation inputs, RnO (a), RnM (b) and Ra (c). The black circle on the x-axis represents observed ET statistics

Results with RnO (AB1) are closer to the 1:1 line than RnM and Ra results (AB2 and AB3, respectively). Nevertheless, the Rn substitutes performance is good, delivering similar ET estimations to AB1 in maize.

The SVM, KR, DT, AB, and MLP daily ET estimations were compared with daily soybean ET measurements. Table 4 presents a summary of the RMSE, MAE, R^2, and Bias for each evaluated ML method under the three input combinations. As was expected, combination 1 with RnO, presented the lowest errors and the highest R^2 with ground ET observations. Nevertheless, combinations 2 and 3 yielded errors and bias comparable to those obtained with RnO.

The KR, AB, and MLP algorithms exhibited the best ET estimation accuracy in soybean for each analysed input combination. Using Rn substitutes in ET estimation with ML methods, would increase ET errors up to 7 % of the RMSE obtained with RnO. Indeed, the mean ML RMSE with combinations 1 and 3 are 21.13 % and 29.63 % of the mean observed ET (6.50 mm/d), respectively. The DT method gave the worst daily soybean ET estimates

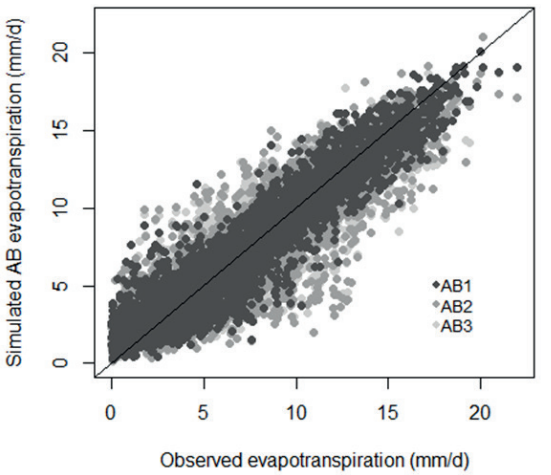

Figure 5. Relationship between simulated AB ET with all the input combinations (AB1, AB2, and AB3) and observed daily maize ET

for each case.

Taylor's diagrams were plotted for comparative assessment of SVM, KR, DT, AB, and MLP daily soybean ET estimation with the three input combinations (see Figure 6).

It can be noted that RnM and Ra produce similar RMS compared with ground ET observations; nevertheless, RnO yielded the best results. In general, the proposed ML methods were able to estimate ET with good accuracy using the three input combinations, except for DT which showed the highest errors.

According to the results presented in Table 4, KR, MLP and AB seem to be the most accurate ML methods to estimate daily soybean ET. In particular, KR1, KR2 and AB3 yielded the lowest RMSE for combinations 1, 2, and 3, respectively. So, the KR1, KR2 and AB3 ET estimations were compared with field ET measurements in Figure 7. Results from the RnO (KR1), are the closest to the 1:1 line (see Figure 7.a). However, the Rn substitutes ET estimations (KR2 and AB3) presented a good correlation and bias with observed ET, with a moderated dispersion.

ET results discussion

The proposed ML algorithms were applied to estimate daily ET under three different input combinations as show in Table 3. Considering that ET represents an important component in the energy balance, similar input variables were used to model Rn and ET. In fact, the ET process expends most of the energy absorbed by the earth surface during a year.

Results presented here demonstrate that ML algorithms are suitable to simulate complex nonlinear processes such as ET. Moreover, the advantage of ML methods, compared with traditional ET equations, is their capability to assimilate substitute variables that represent the dynamics of a process, as a surrogate of accurate field measurements.

All the implemented ML algorithms provided good performances in daily ET estimation using Ra and RnM as inputs, yielding correlations higher than 0.87 and 0.83 for maize and soybean, respectively. Nevertheless, AB and KR methods exhibited the lowest RMSs compared with observed ET in maize and soybean (see Figures 4 and 6). Comparable findings were published by Carter and Liang (2019), who demonstrated that Boosting and Kernel methods similar to those used here presented the lowest error to estimate daily cropland ET, compared with other ML algorithms. The success of AB lies in looking for a strong regressor through lineal combinations of weak samples, iterating until the training data is estimated with the minimum error (Wu et al., 2008; Yamaç and Todorovic, 2020). KR is the suitable method for estimating a nonlinear process using many variables as inputs (Hofmann et al., 2008; Zhang et al., 2013).

In this study, the AB method presented accurate results for daily maize and soybean ET. Indeed, AB yielded RMSEs of about 25 % (1.7 mm/d) in maize and 27 % (1.8 mm/d) in soybean of the mean observed ET using Rn substitutes as input variables. These results are in good agreement with Granata (2019), Yamac and Todorovic (2020), and Fan et al. (2021), who proved that Boosting techniques had a high precision for modelling daily ET and transpiration. These previous studies reported RMSEs varying from 8 to 13 % (Granata, 2019), 4 to 29 % (Yamac and Todorovic, 2020), and 20 to 33% (Fan et al., 2021) of the mean observed, when modelling daily potato ET, grassland ET, and maize transpiration, respectively, using observed radiation data. However, Fan et al. (2021) findings cannot be directly extrapolated to maize ET, since evaporation is important in early maize stages while transpiration has great importance from V6 stage.

MLP algorithm showed good predictive capabilities to model ET using Rn substitutes, with RMSEs of about 26 and 27 % of the mean observed ET for maize and soybean, respectively. Comparable results were presented by Yamac and Todorovic (2020) when they simulated ET from potatoes, with RMSEs ranging from 2 to 29% of the mean observed ET. They used solar radiation along

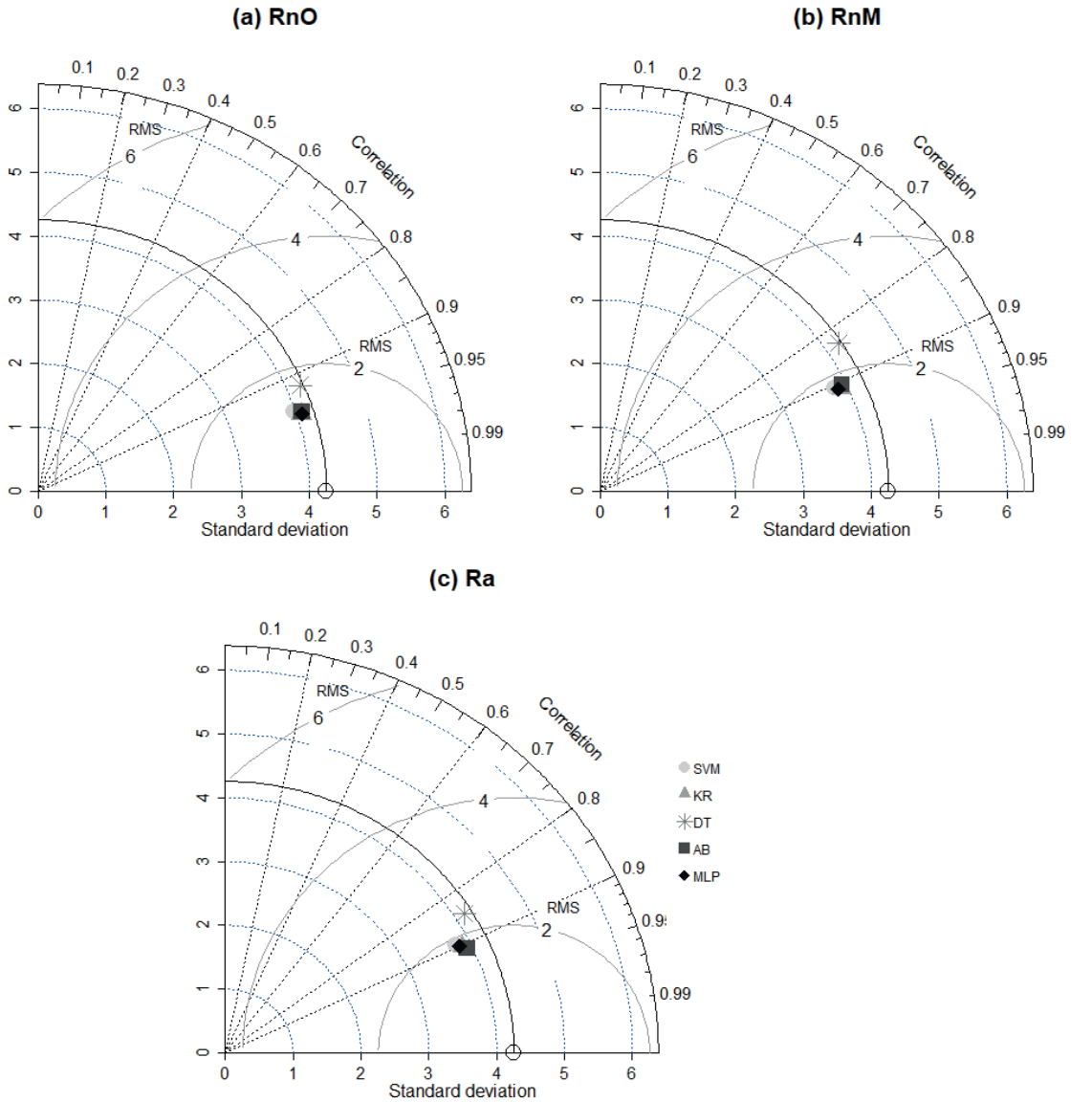

Figure 6. Taylor's diagrams for SVM, KR, DT, AB, and MLP daily soybean ET estimation under the various radiation inputs, RnO (a), RnM (b) and Ra (c). The black circle on the x-axis represents observed ET statistics

with different variables to estimate FAO 56 crop evapotranspiration, Allen et al. (1998).

The results of this study exhibited that SVM algorithm was able to adequately model ET using Rn substitutes, yielding RMSEs of about 27 and 28 % of the mean observed maize and soybean ET, respectively. These results are comparable to those published by Tang et al. (2018), Granata (2019), Chen, Z. et al. (2020), and Fan et al. (2021). Tang et al. (2018) reported RMSEs ranging from 9 to 17 % of the mean observed maize ET, using the wind speed and crop height as input variables. Granata (2019) evaluated the SVM to model ET in a grassland site, with RMSEs ranging from 7 to 14 % of the mean observed ET. Chen, Z. et al. (2020) published RMSEs higher than 20 % using SVM to estimate daily ETo under different combinations of atmospheric variables. Fan et al. (2021) used SVM for daily maize transpiration estimation, with RMSEs ranging from 6 to 31 % of the mean daily observed transpiration.

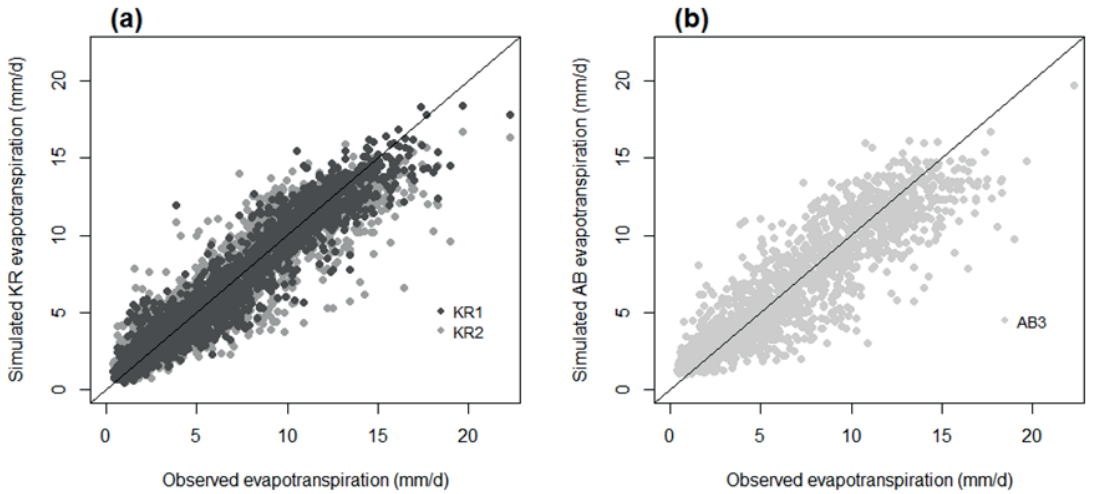

Figure 7. Relationship between KR1, KR2 (a) and AB3 (b) ET estimations and observed daily soybean ET

CONCLUSION

Empirical and semiempirical ET equations require precise net radiation measurements to obtain accurate results, at any time and space scale. Since Rn is not readily available information, a comparison of five ML methods for obtaining ET from two Rn substitutes, was performed here. Our results showed good efficiency of the ML algorithms assessed, yielding acceptable errors with easily obtainable radiation proxies, meteorological and NDVI variables. However, these errors are larger than physics-based model errors, as can be corroborated in the literature. In general, this type of ML methods is operative and flexible but the accuracy is debatable. Indeed, Rn was modelled with Support Vector Machine, Kernel Ridge, Decision Tree, Adaptive Boosting and Multilayer Perceptron methods, using meteorological variables readily available everywhere. In general, all the evaluated ML methods were able to effectively model Rn, with errors of about 16 % (60 W/m^2) of the mean observed Rn. However, Kernel Ridge, Adaptive Boosting, and Multilayer Perceptron presented the most accurate estimations. Hence, Rn substitutes computed from routinely meteorological data seem to be an effective alternative to consider in many regions where the heat flux and radiation observations are rare.

The proposed ML methods were suitable to estimate ET with extraterrestrial solar radiation and modelled net radiation. Adaptive Boosting and Kernel Ridge presented consistent results for maize and soybean using Rn substitutes, rendering RMSE lower than 26% (1.75mm/d) of the mean observed ET. Thus, it can be concluded that Adaptive Boosting and Kernel Ridge techniques can be applied with Rn substitutes for mapping ET with meteorological data and satellite NDVI images.

ACKNOWLEDGEMENTS

The authors thank the National Scientific and Technical Research Council - Argentina witch supported this investigation. Also, the authors acknowledge KILIMO S A for supporting Gianfranco Fagioli's work. Finally, the authors wish to thank the FLUXNET ground observation network for freely providing the *in-situ* data belonging to its stations.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

REFERENCES

Alizamir, M., Kim, S., Kisi, O., and Zounemat-Kermani, M. (2020). A comparative study of several machine learning based non-linear regression methods in estimating solar radiation: Case studies of the USA and Turkey regions. *Energy*, *197*, 117239. https://doi.org/10.1016/j.energy.2020.117239

Allen, R. G., Pereira, L. S., Raes, D., and Smith, M. (1998). Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56. *FAO Rome*, *300*(9), D05109.

Anguita, D., Ridella, S., and Rivieccio, F. (2005, July). K-fold generalization capability assessment for support vector classifiers. In *Proceedings. 2005 IEEE*

International Joint Conference on Neural Networks, *2* (pp. 855-858). IEEE. https://doi.org /10.1109/IJCNN.2005.1555964

Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., and Wood, E. F. (2018). Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Scientific Data, 5(*1), 180214, 1-12. https://doi.org/10.1038/sdata.2018.214

Bisht, G., Venturini, V., Islam, S., and Jiang, L. (2005). Estimation of the net radiation using MODIS (Moderate Resolution Imaging Spectroradiometer) data for clear sky days. *Remote Sensing of Environment*, *97*(1), 52-67. https://doi.org/10.1016/j.rse.2005.03.014

Carter, C. and Liang, S. (2019). Evaluation of ten machine learning methods for estimating terrestrial evapotranspiration from remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, *78*, 86-92. https://doi.org/10.1016/j.jag.2019.01.020

Chen, J., He, T., Jiang, B., and Liang, S. (2020). Estimation of all-sky all-wave daily net radiation at high latitudes from MODIS data. *Remote Sensing of Environment*, *245*, 111842. https://doi.org/10.1016/j.rse.2020.111842

Chen, Z., Zhu, Z., Jiang, H., and Sun, S. (2020). Estimating daily reference evapotranspiration based on limited meteorological data using deep learning and classical machine learning methods. *Journal of Hydrology*, *591*, 125286. https://doi.org/10.1016/j.jhydrol.2020.125286

Darnell, S. J., Page, D., and Mitchell, J. C. (2007). An automated decision-tree approach to predicting protein interaction hot spots. *Proteins: Structure, Function, and Bioinformatics*, *68*(4), 813-823. https://doi.org/10.1002/prot.21474

Exterkate, P., Groenen, P. J. F., Heij, C., and van Dijk, D. (2016). Nonlinear forecasting with many predictors using kernel ridge regression. *International Journal of Forecasting, 32(3)*, 736–753. https://doi.org/10.1016/j.ijforecast.2015.11.017

Fan, J., Zheng, J., Wu, L., and Zhang, F. (2021). Estimation of daily maize transpiration using support vector machines, extreme gradient boosting, artificial and deep neural networks models. *Agricultural Water Management*, *245*, 106547. https://doi.org/10.1016/j.agwat.2020.106547

Food and Agriculture Organization of the United Nations (FAO), International Institute for Applied Systems Analysis (IIASA), ISRIC-World Soil Information, Institute of Soil Science – Chinese Academy of Sciences (ISSCAS), and Joint Research Centre of the European Commission (JRC). (2012). Harmonized World Soil Database (version 1.2) [Software]. FAO, Rome, Italy and IIASA, Laxenburg, Austria. http://webarchive.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/index.html?sb=1

García, G. A., Venturini, V., Brogioni, M., Walker, E., and Rodríguez, L. (2019). Soil moisture estimation over flat lands in the Argentinian Pampas region using Sentinel-1A data and non-parametric methods. *International Journal of Remote Sensing, 40*(10), 3689-3720. https://doi.org/10.1080/01431161.2018.1552813

Granata, F. (2019). Evapotranspiration evaluation models based on machine learning algorithms — A comparative study. *Agricultural Water Management*, *217*, 303-315. https://doi.org/10.1016/j.agwat.2019.03.015

Hargreaves, G. L., Hargreaves, G. H., and Riley, J. P. (1985). Irrigation water requirements for Senegal River basin. *Journal of Irrigation and Drainage Engineering*, *111*(3), 265-275. https://doi.org/10.1061/(ASCE)0733-9437(1985)111:3(265)

Hofmann, T., Schölkopf, B., and Smola, A. J. (2008). Kernel methods in machine learning. *The Annals of Statistics*, *36*(3), 1171-1220. https://doi.org/10.1214/009053607000000677

Jain, S. K., Nayak, P. C., and Sudheer, K. P. (2008). Models for estimating evapotranspiration using artificial neural networks, and their physical interpretation. *Hydrological Processes: An International Journal*, *22*(13), 2225-2234. https://doi.org/10.1002/hyp.6819

Jiang, B., Zhang, Y., Liang, S., Zhang, X., and Xiao, Z. (2014). Surface daytime net radiation estimation using artificial neural networks. *Remote Sensing*, *6*(11), 11031-11050. https://doi.org/10.3390/rs61111031

Kim, H., Parinussa, R., Konings, A. G., Wagner, W., Cosh, M. H., Lakshmi, V., Zohaib, M. and Choi, M. (2018). Global-scale assessment and combination of SMAP with ASCAT (active) and AMSR2 (passive) soil moisture products. *Remote Sensing of Environment, 204,* 260-275. https://doi.org/10.1016/j.rse.2017.10.026

Kumar, M., Raghuwanshi, N. S., and Singh, R. (2011). Artificial neural networks approach in evapotranspiration modeling: a review. *Irrigation Science*, *29*(1), 11-25. https://doi.org/10.1007/s00271-010-0230-8

Llasat, M. C. and Snyder, R. L. (1998). Data error effects on net radiation and evapotranspiration estimation. *Agricultural and Forest Meteorology*, *91*(3-4), 209-221. https://doi.org/10.1016/S0168-1923(98)00070-7

Majidi, M., Alizadeh, A., Vazifedoust, M., Farid, A., and Ahmadi, T. (2015). Analysis of the effect of missing weather data on estimating daily reference evapotranspiration under different climatic conditions. *Water Resources Management*, *29*(7), 2107-2124. https://doi.org/10.1007/s11269-014-0782-0

Miralles, D. G., Holmes, T. R. H., De Jeu, R. A. M.,

Gash, J. H., Meesters, A. G. C. A., and Dolman, A. J. (2011). Global land-surface evaporation estimated from satellite-based observations. *Hydrology and Earth System Sciences*, *15*(2), 453-469. https://doi.org/10.5194/hess-15-453-2011

Mokhtari, A., Noory, H., and Vazifedoust, M. (2018). Performance of different surface incoming solar radiation models and their impacts on reference evapotranspiration. *Water Resources Management*, *32*(9), 3053-3070. https://doi.org/10.1007/s11269-018-1974-9

Nourani, V., Tajbakhsh, A. D., and Molajou, A. (2019). Data mining based on wavelet and decision tree for rainfall-runoff simulation. *Hydrology Research*, *50*(1), 75-84. https://doi.org/10.2166/nh.2018.049

Ojo, O. S., Adeyemi, B., and Oluleye, D. O. (2021). Artificial neural network models for prediction of net radiation over a tropical region. *Neural Computing and Applications*, *33*(12), 6865-6877. https://doi.org/10.1007/s00521-020-05463-9

Penman, H. L. (1948). Natural evaporation from open water, bare soil and grass. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, *193*(1032), 120-145. https://doi.org/10.1098/rspa.1948.0037

Priestley, C. H. B. and Taylor, R. J. (1972). On the Assessment of Surface Heat Flux and Evaporation Using Large-Scale Parameters. *Monthly Weather Review*, *100*(2), 81-92. https://doi.org/10.1175/1520-0493(1972)100<0081:OTAOSH>2.3.CO;2

Purdy, A. J., Fisher, J. B., Goulden, M. L., Colliander, A., Halverson, G., Tu, K., and Famiglietti, J. S. (2018). SMAP soil moisture improves global evapotranspiration. *Remote Sensing of Environment*, *219*, 1-14. https://doi.org/10.1016/j.rse.2018.09.023

Qiu, R., Liu, C., Cui, N., Wu, Y., Wang, Z., and Li, G. (2019). Evapotranspiration estimation using a modified Priestley-Taylor model in a rice-wheat rotation system. *Agricultural Water Management*, *224*, 105755. https://doi.org/10.1016/j.agwat.2019.105755

Saunders, C., Gammerman, A., and Vovk, V. (1998). Ridge Regression Learning Algorithm in Dual Variables. In *Proceeding 15th International Conference Machine Learning* (pp. 1–7).

Schwertman, N. C., Owens, M. A., and Adnan, R. (2004). A simple more general boxplot method for identifying outliers. *Computational Statistics and Data Analysis*, *47*(1), 165-174. https://doi.org/10.1016/j.csda.2003.10.012

Si, Z., Yu, Y., Yang, M., and Li, P. (2020). Hybrid Solar Forecasting Method Using Satellite Visible Images and Modified Convolutional Neural Networks. *IEEE Transactions on Industry Applications*, *57*(1), 5-16. https://doi.org/10.1109/TIA.2020.3028558

Shirazi, M. A., Boersma, L., and Hart, J. W. (1988). A unifying quantitative analysis of soil texture: improvement of precision and extension of scale. *Soil Science Society of America Journal*, 52(1), 181-190. https://doi.org/10.2136/sssaj1988.03615995005200010032x

Tang, D., Feng, Y., Gong, D., Hao, W., and Cui, N. (2018). Evaluation of artificial intelligence models for actual crop evapotranspiration modeling in mulched and non-mulched maize croplands. *Computers and Electronics in Agriculture*, *152*, 375-384. https://doi.org/10.1016/j.compag.2018.07.029

Taylor, K. E. (2001). Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres*, *106*(D7), 7183-7192. https://doi.org/10.1029/2000JD900719

Tikhamarine, Y., Malik, A., Pandey, K., Sammen, S. S., Souag-Gamane, D., Heddam, S., and Kisi, O. (2020). Monthly evapotranspiration estimation using optimal climatic parameters: efficacy of hybrid support vector regression integrated with whale optimization algorithm. *Environmental Monitoring and Assessment*, *192*(11), 1-19. https://doi.org/10.1007/s10661-020-08659-7

Tikhamarine, Y., Malik, A., Souag-Gamane, D., and Kisi, O. (2020). Artificial intelligence models versus empirical equations for modeling monthly reference evapotranspiration. *Environmental Science and Pollution Research*, *27*(24), 30001-30019. https://doi.org/10.1007/s11356-020-08792-3

Trnka, M., Eitzinger, J., Kapler, P., Dubrovský, M., Semerádová, D., Žalud, Z., and Formayer, H. (2007). Effect of estimated daily global solar radiation data on the results of crop growth models. *Sensors*, *7*(10), 2330-2362. https://doi.org/10.3390/s7102330

Vapnik, V. (1999). *The nature of statistical learning theory* (2nd Ed.). Springer.

Walker, E. and Venturini, V. (2019). Land surface evapotranspiration estimation combining soil texture information and global reanalysis datasets in Google Earth Engine. *Remote Sensing Letters*, *10*(10), 929-938. https://doi.org/10.1080/2150704X.2019.1633487

Wang, Y., Jiang, B., Liang, S., Wang, D., He, T., Wang, Q., Zhao, X., and Xu, J. (2019). Surface Shortwave net radiation estimation from Landsat TM/ETM+ data using four machine learning algorithms. *Remote Sensing*, *11*(23), 2847. https://doi.org/10.3390/rs11232847

Wu, X., Kumar, V., Ross Quinlan, J., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G. J., Ng, A., Liu, B., Yu, P. S., Zou, Z.-H., Steinbach, M., Hand, D. J., and Steinberg, D. (2008). Top 10 algorithms in data mining. *Knowledge and Information Systems*, *14*(1), 1-37. https://doi.org/10.1007/s10115-007-0114-2

Xu, M., Watanachaturaporn, P., Varshney, P. K., and Arora, M. K. (2005). Decision tree regression for soft classification of remote sensing data. *Remote Sensing of Environment*, *97*(3), 322-336. https://doi.org/10.1016/j.rse.2005.05.008

Xu, T., Guo, Z., Xia, Y., Ferreira, V. G., Liu, S., Wang, K., Yao, Y., Zhang, X., and Zhao, C. (2019). Evaluation of twelve evapotranspiration products from machine learning, remote sensing and land surface models over conterminous United States. *Journal of Hydrology*, *578*, 124105. https://doi.org/10.1016/j.jhydrol.2019.124105

Yadav, A. K. and Chandel, S. S. (2014). Solar radiation prediction using Artificial Neural Network techniques: A review. *Renewable and Sustainable Energy Reviews*, *33*, 772-781. https://doi.org/10.1016/j.rser.2013.08.055

Yamaç, S. S. and Todorovic, M. (2020). Estimation of daily potato crop evapotranspiration using three different machine learning algorithms and four scenarios of available meteorological data. *Agricultural Water Management*, *228*, 105875. https://doi.org/10.1016/j.agwat.2019.105875

You, Y., Demmel, J., Hsieh, C. J., and Vuduc, R. (2018, June). Accurate, fast and scalable kernel ridge regression on parallel and distributed systems. In *Proceedings of the 2018 International Conference on Supercomputing* (pp. 307-317). https://doi.org/10.1145/3205289.3205290

Zhang, Y., Duchi, J., and Wainwright, M. (2013, June). Divide and conquer kernel ridge regression. In *Conference on learning theory* (pp. 592-617). PMLR.

Zhang, X., Treitz, P. M., Chen, D., Quan, C., Shi, L., and Li, X. (2017). Mapping mangrove forests using multi-tidal remotely-sensed data and a decision-tree-based procedure. *International Journal of Applied Earth Observation and Geoinformation*, *62*, 201-214. https://doi.org/10.1016/j.jag.2017.06.010

Zhang, Y., Qin, X., Li, X., Zhao, J., and Liu, Y. (2020). Estimation of Shortwave Solar Radiation on Clear-Sky Days for a Valley Glacier with Sentinel-2 Time Series. *Remote Sensing*, *12*(6), 927. https://doi.org/10.3390/rs12060927