

## Article

# Modeling Actual Evapotranspiration with MSI-Sentinel Images and Machine Learning Algorithms

Robson Argolo dos Santos <sup>1,\*</sup>, Everardo Chartuni Mantovani <sup>1</sup>, Elpídio Inácio Fernandes-Filho <sup>2</sup>, Roberto Filgueiras <sup>1</sup>, Rodrigo Dal Sasso Lourenço <sup>1</sup>, Vinícius Bof Bufon <sup>3</sup> and Christopher M. U. Neale <sup>4,5</sup>

<sup>1</sup> Department of Agricultural Engineering, Federal University of Viçosa (UFV), Peter Henry Rolfs Ave., Viçosa 36570-900, MG, Brazil

<sup>2</sup> Department of Soil Science, Federal University of Viçosa (UFV), Peter Henry Rolfs Ave., Viçosa 36570-900, MG, Brazil

<sup>3</sup> Brazilian Agricultural Research Corporation (EMBRAPA) Cerrados, BR-020, Km 18, Planaltina 73310-970, DF, Brazil

<sup>4</sup> Department of Biological Systems Engineering, University of Nebraska Lincoln, Lincoln, NE 68583, USA

<sup>5</sup> Daugherty Water for Food Global Institute, University of Nebraska, Lincoln, NE 68588, USA

\* Correspondence: robson.argolo@ufv.br

**Abstract:** The modernization of computational resources and application of artificial intelligence algorithms have led to advancements in studies regarding the evapotranspiration of crops by remote sensing. Therefore, this research proposed the application of machine learning algorithms to estimate the  $ET_{rF}$  (Evapotranspiration Fraction) of sugar cane crop using the METRIC (Mapping Evapotranspiration at High Resolution with Internalized Calibration) model with data from the Sentinel-2 satellites constellation. In order to achieve this goal, images from the MSI sensor (MultiSpectral Instrument) from the Sentinel-2 and the OLI (Operational Land Imager) and TIRS (Thermal Infrared Sensor) sensors from the Landsat-8 were acquired nearly at the same time between the years 2018 and 2020 for sugar cane crops. Images from OLI and TIR sensors were intended to calculate  $ET_{rF}$  through METRIC (target variable), while for the MSI sensor images, the explanatory variables were extracted in two approaches, using 10 m (approach 1) and 20 m (approach 2) spatial resolution. The results showed that the algorithms were able to identify patterns in the MSI sensor data to predict the  $ET_{rF}$  of the METRIC model. For approach 1, the best predictions were XgbLinear ( $R^2 = 0.80$ ; RMSE = 0.15) and XgbTree ( $R^2 = 0.80$ ; RMSE = 0.15). For approach 2, the algorithm that demonstrated superiority was the XgbLinear ( $R^2 = 0.91$ ; RMSE = 0.10), respectively. Thus, it became evident that machine learning algorithms, when applied to the MSI sensor, were able to estimate the  $ET_{rF}$  in a simpler way than the one that involves energy balance with the thermal band used in the METRIC model.

**Keywords:** irrigation scheduling; sugarcane; remote sensing



**Citation:** dos Santos, R.A.; Mantovani, E.C.; Fernandes-Filho, E.I.; Filgueiras, R.; Lourenço, R.D.S.; Bufon, V.B.; Neale, C.M.U. Modeling Actual Evapotranspiration with MSI-Sentinel Images and Machine Learning Algorithms. *Atmosphere* **2022**, *13*, 1518. <https://doi.org/10.3390/atmos13091518>

Academic Editors: Demetrios E. Tsesmelis, Nikolaos Skondras and Nikolaos Proutsos

Received: 1 August 2022

Accepted: 12 September 2022

Published: 17 September 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Evapotranspiration is the phenomenon of transferring water from liquid to gas to the atmosphere from the evaporation of water from the soil and water bodies, as well as the transpiration of plants through the leaves [1]. According to Allen et al. [2], quantifying water consumption in large areas, particularly in extended irrigated agricultural areas, is of great relevance for planning and managing water resources, mitigating impacts on the water bodies' streamflow rate, establishing use rights and consumption regulation, and avoiding conflicts of water use.

Evapotranspiration can be determined directly in the field by lysimeter techniques, which are quite reliable but costly, or estimated with (1) full-physical models based on the principles of conservation of mass and energy; (2) semi-physical models that use conservation of mass or energy; and (3) black-box models based on artificial neural networks,

empirical relationships, fuzzy, genetic [3] and machine learning algorithms. Remote sensing can be included in the physical models because it uses the radiation reflected and emitted by the Earth's surface to estimate the properties of the Earth's surface when subjected to radiation interaction models. Vegetation cover biomass, and indexes for each image can also be spatially and temporally modeled [4,5]. Several models for estimating evapotranspiration based on remote sensing were developed and have been widely used, especially in scientific research focused on the water planning field. Some examples are the Surface Energy Balance Algorithm for Land (SEBAL) [6], Mapping Evapotranspiration at High Resolution with Internalized Calibration (METRIC) [7], and Simple Algorithm for Evapotranspiration Retrieving (SAFER) [8]. However, these models estimate a fraction of the evapotranspiration, which corresponds to a coefficient that must be multiplied by the reference evapotranspiration to obtain the actual daily evapotranspiration.

The METRIC model [7] estimates the evapotranspiration fraction ( $ET_r/F$ ) through the instantaneous evapotranspiration ( $ET_{ints}$ ) and the reference potential evapotranspiration of the alfalfa ( $ET_r$ ) or grass ( $ET_o$ ) on a daily scale. This model is very complex, as it demands, a priori, the estimation of parameters for the energy balance calculations, which can be prone to errors since this model presents an interactive method of selecting "hot" and "cold" pixels to calculate aerodynamic resistance to heat transport and exchange derived from the SEBAL model [6].

The aforementioned models were developed using data from the Landsat satellite constellation sensors with dependence on visible, near, mid and thermal infrared wavelengths. Satellites from the Landsat series provide information about the Earth's surface with a temporal resolution of 16 days and a spatial resolution of 30 m. Temporal resolution limits the model's application, as it presents a limited number of images available by the sensors on these satellites. In contrast, the Sentinel-2 satellites constellation arises as an alternative to this limitation. Formed by two satellites, Sentinel-2A and Sentinel-2B, with sensors with similar waveband characteristics, this constellation supplies images every 5 days, with spatial resolution fluctuating between 10, 20, and 60 m, hence, producing a larger number of images of the Earth's surface than the Landsat constellation. The sensor from the Sentinel-2 satellites constellation is the MultiSpectral Instrument (MSI). It acquires information from the Earth's surface in the visible, red edge, near, and mid-infrared wavelengths. However, it does not have a sensor in the thermal infrared portion of the spectrum.

By considering the differences between the satellites mentioned above, the development of models with a prediction capacity of the crop's actual evapotranspiration ( $ET_a$ ) using the MSI sensor will be of great value for the scientific community and field professionals, enabling determining  $ET_a$  with a better spatial and temporal resolution during crop cycles, and can even be integrated with sensors on other satellite platforms forming a multi-sensor suite, which, in the absence of information from one sensor, allows it to be complemented by the other, in concordance with the approach described by Filgueiras et al. [9]. Another important aspect is that, even though the MSI does not include the thermal infrared band,  $ET_a$  can be estimated without the surface temperature information, which can reduce the introduction of additional errors as models involving thermal information have, in their structure, the complex energy balance to quantify the latent energy of the system. Furthermore, thermal bands have a coarser spatial resolution, so interpolation is often necessary to align pixel sizes with those of the other bands, such as Landsat 5.7 and 8. Alternatively, modeling evapotranspiration in the absence of data from the thermal band can generate unsatisfactory results due to the intrinsic relationship between vegetation canopy temperature derived from the thermal band with stomatal conductance [10,11].

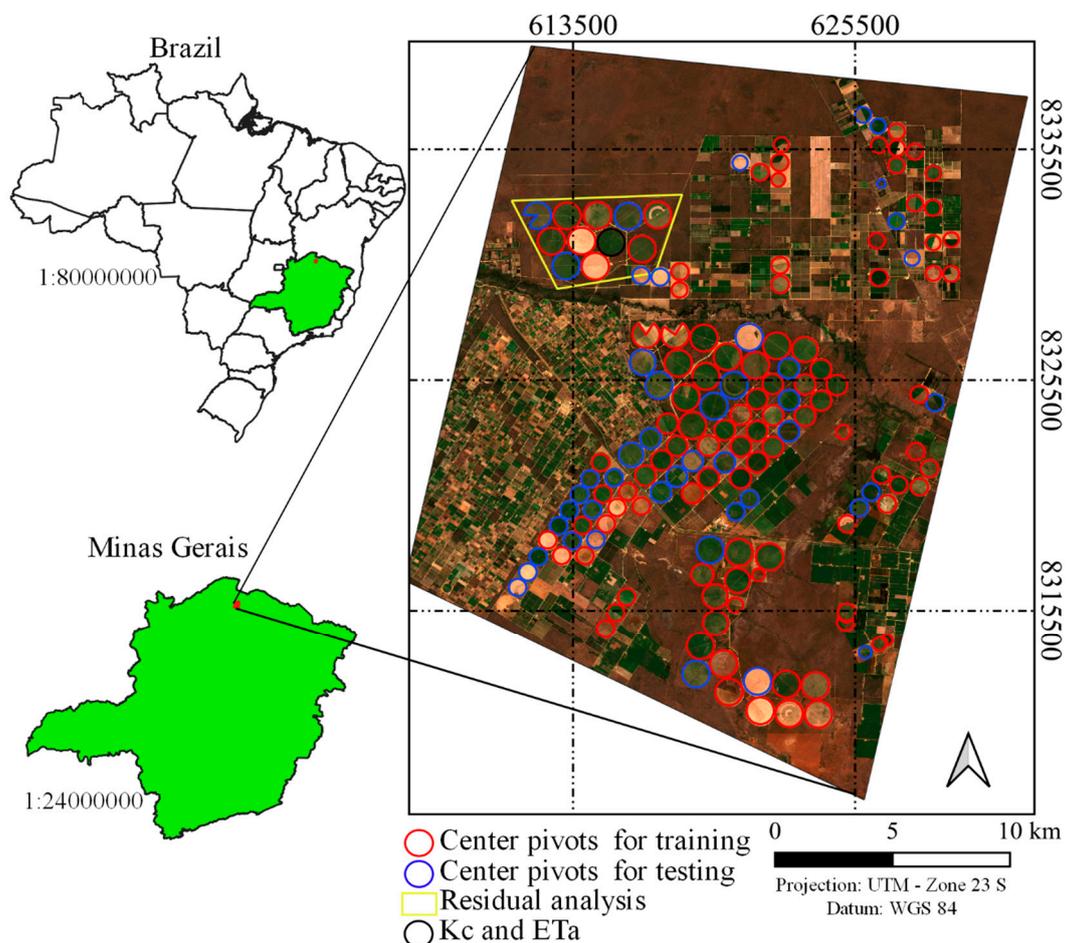
By aiming to develop models with higher prediction capacity and reliability, this work, through the application of machine learning (ML) algorithms, sought to model variables of interest. Such algorithms, based on artificial intelligence, have a robust structure that allows the identification of relationship patterns between the variables to be modeled and the so-called predictor variables (independent). According to Cervantes et al. [12], ML is an interdisciplinary area based on computer science, statistics, mathematics, and optimization,

among several other areas. Many ML algorithms, each with different characteristics, are used for prediction and classification, being, in recent years, applied in research focused on agricultural sciences and remote sensing to develop models with greater ability to represent phenomena [13–17]. In evapotranspiration prediction, these models used remote sensing data and climate data to estimate actual evapotranspiration [18,19], as well as only climate data to estimate reference evapotranspiration [20–23]. The various available studies applied the algorithms in predicting reference evapotranspiration or actual evapotranspiration using remote sensing with thermal data. Thus, this work sought to train machine learning algorithms to estimate the evapotranspiration fraction ( $ET_rF$ ) used in the METRIC model from the data of the Sentinel-2 satellites constellation that does not use thermal data on the sugar cane crop used in this case study.

## 2. Materials and Methods

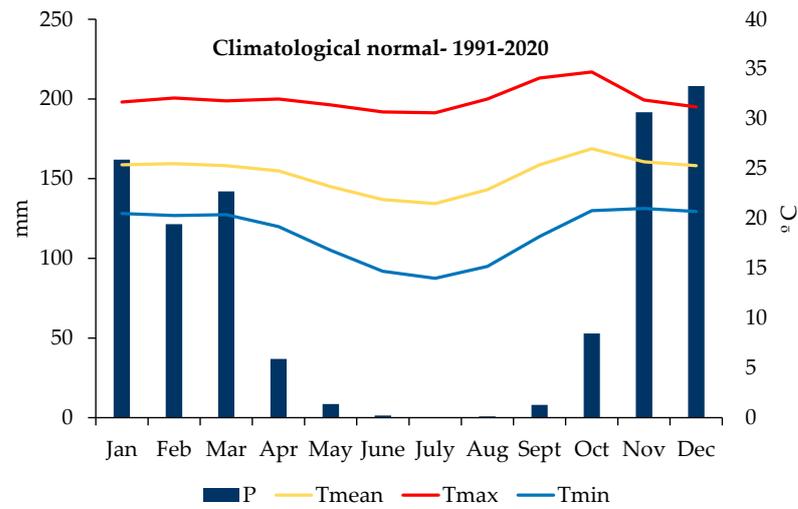
### 2.1. Study Area

The study was performed in the northern region of the state of Minas Gerais, Brazil, classified as AS climate—tropical with dry summer [24] (Figure 1). The sugar cane crop is under 150 irrigated fields using center pivot irrigation used in this study.



**Figure 1.** Location of the study area with the central pivots used for training, testing, and applying the models highlighted.

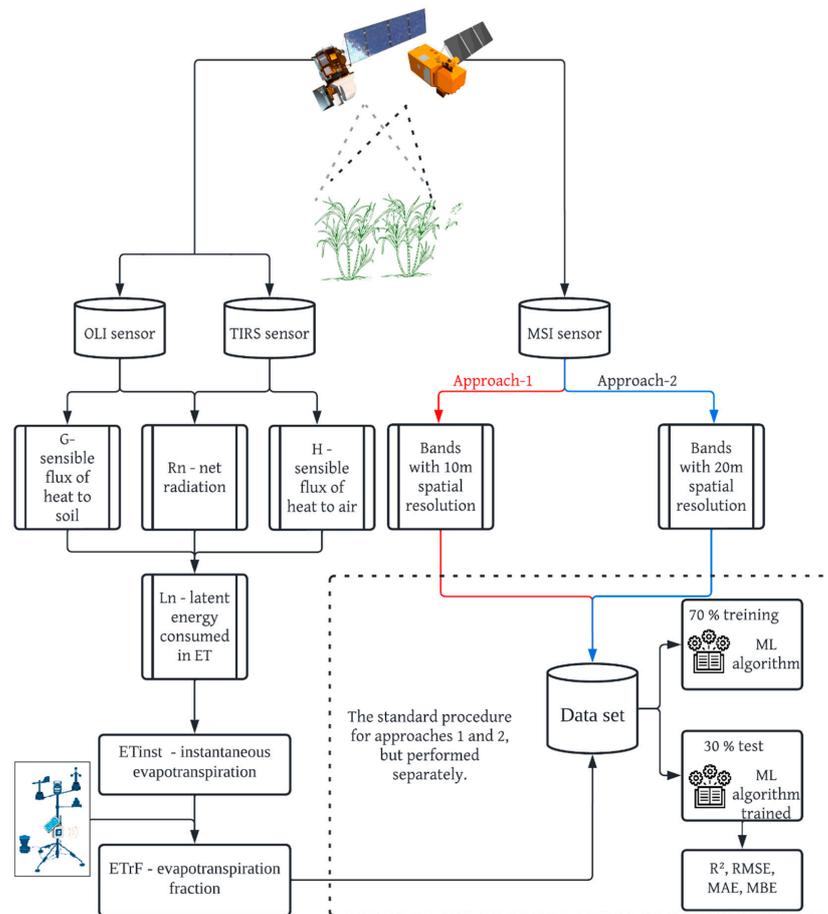
This area was chosen because it is a research partnership area with little to no precipitation (Figure 2) that favors the acquisition of a larger number of cloud-free images, thus being able to acquire a volume of spectral information of the area, mainly coinciding with Landsat-8 and Sentinel-2.



**Figure 2.** Climatological normal of the study region extracted from station 83,386 of INMET (Instituto Nacional de Meteorologia).

2.2. Landsat-8 and Sentinel-2 Data

The data from the sensors onboard the Landsat-8 and Sentinel-2 satellites were acquired to generate the observed variable through the METRIC model, and the explanatory variables used for training and testing the machine learning algorithms can be seen in a simplified way in the flowchart in Figure 3.



**Figure 3.** Acquisition of data from sensors onboard Landsat-8 and Sentinel-2 for training and testing in machine learning algorithms.

A total of 56 satellite images were acquired, of which 22 were from OLI (Operational Land Imager) sensor and TIRS (Thermal Infrared Sensor) from the Landsat-8 satellite, and 34 came from the MSI (Multispectral Instrument) sensor of the Sentinel-2A and 2B satellites. Images from OLI and TIRS sensors were necessary to estimate the evapotranspiration fraction ( $ET_rF$ ) using the METRIC model, this being the response variable, whereas data from the MSI sensor were used as predictor variables in the machine learning models. The spectral and spatial characteristics of the sensors used are available in Tables 1 and 2.

**Table 1.** Spectral and spatial characteristics of the OLI and TIRS sensor.

| Spectral Band |                              | Wavelength ( $\mu\text{m}$ ) | Spatial Resolution (m) |
|---------------|------------------------------|------------------------------|------------------------|
| OLI           |                              |                              |                        |
| B1            | Coastal aerosol (Ca)         | 0.43–0.45                    | 30                     |
| B2            | Blue (B)                     | 0.45–0.51                    |                        |
| B3            | Green (G)                    | 0.53–0.59                    |                        |
| B4            | Red (R)                      | 0.64–0.67                    |                        |
| B5            | Near-infrared (NIR)          | 0.85–0.88                    |                        |
| B6            | Shortwave infrared 1 (SWIR1) | 1.57–1.65                    |                        |
| B7            | Shortwave infrared 2 (SWIR2) | 2.11–2.29                    |                        |
| B8            | Panchromatic (PCh)           | 0.50–0.68                    |                        |
| B9            | Cirrus (C)                   | 1.36–1.38                    |                        |
| TIRS          |                              |                              |                        |
| B10           | Thermal infrared 1 (TIRS1)   | 10.60–11.19                  | 100 *                  |
| B11           | Thermal infrared 2 (TIRS2)   | 11.50–12.51                  |                        |

\* The sensor's resolution is 100 m, but images are resampled and made available in 30 m.

**Table 2.** Spectral and spatial characteristics of the MSI sensor.

| Spectral Band |                              | Wavelength ( $\mu\text{m}$ ) | Spatial Resolution (m) |    |
|---------------|------------------------------|------------------------------|------------------------|----|
| MSI           |                              |                              |                        |    |
| B2            | Blue (B)                     | 0.459–0.525                  | 10                     |    |
| B3            | Green (G)                    | 0.542–0.578                  |                        |    |
| B4            | Red (R)                      | 0.650–0.680                  |                        |    |
| B8            | Near-infrared (NIR)          | 0.781–0.887                  |                        |    |
| B5            | red edge 1 (Re1)             | 0.697–0.712                  | 20                     |    |
| B6            | red edge 2 (Re2)             | 0.733–0.748                  |                        |    |
| B7            | red edge 3 (Re2)             | 0.773–0.793                  |                        |    |
| B8A           | Near-infrared narrow (NIRn)  | 0.856–0.876                  |                        |    |
| B11           | Shortwave infrared 1 (SWIR1) | 1.569–1.660                  |                        |    |
| B12           | Shortwave infrared 2 (SWIR2) | 2.115–2.290                  |                        |    |
| B1            | Coastal aerosol              | 0.433–0.453                  |                        | 60 |
| B9            | Water vapor                  | 0.935–0.955                  |                        |    |
| B10           | Cirrus                       | 1.359–1.390                  |                        |    |

Images were obtained between the years 2018 and 2020. During this period, eight images from the Landsat-8 satellite and eight from the MSI sensor covered the study area on the same day. Images of those dates were used to train and test the models and the application of residual analysis. The choice for images with coinciding dates between the two satellites was made to avoid different reflectance of the same study area since vegetation, especially crops, change quickly. Thus, obtaining images on different days to perform the modeling could have introduced bias and not a faithful representation of the modeled product. In Table 3, images with date, time (GMT-3), and path/row used to train the models are available.

**Table 3.** Characteristics of the eight images selected to develop the models.

| Date<br>(mm/dd/yyyy) | Landsat-8       |          | Sentinel-2      |             |
|----------------------|-----------------|----------|-----------------|-------------|
|                      | Time (hh:mm:ss) | Path/Row | Time (hh:mm:ss) | Tile Number |
| Training and Test    |                 |          |                 |             |
| 07/06/2018           | 09:55:17.157    | 218/71   | 10:12:41.024    | T23LPD      |
| 05/22/2019           | 09:55:49.714    | 218/71   | 10:12:51.024    | T23LPD      |
| 10/04/2019           | 10:02:19.874    | 219/70   | 10:12:49.024    | T23LPD      |
| 10/29/2019           | 09:56:34.656    | 218/71   | 10:12:49.024    | T23LPD      |
| 01/17/2020           | 09:56:20.969    | 218/71   | 10:12:41.024    | T23LPD      |
| 05/31/2020           | 10:01:27.427    | 219/70   | 10:12:49.024    | T23LPD      |
| 08/19/2020           | 10:02:00.995    | 219/70   | 10:12:49.024    | T23LPD      |
| Residual analysis    |                 |          |                 |             |
| 12/02/2020           | 09:56:32.117    | 218/71   | 10:12:41.024    | T23LPD      |

From 11/23/2019 to 09/04/2020, 17 images were obtained from OLI and TIRS sensors (including 4 images in Table 3) and 30 from MSI sensors (including 4 images in Table 3). Those images were used to determine crop coefficient ( $K_c$ ) and actual evapotranspiration ( $ET_a$ ) during the sugar cane cycle. The estimation was performed by both the METRIC model and the suggested model.

### 2.3. Response Variable

The response variable, evapotranspiration fraction ( $ET_rF$ ), was obtained for each pixel, with a  $30\text{ m} \times 30\text{ m}$  spatial resolution, from the METRIC model.  $ET_rF$  is reckoned by dividing instantaneous evapotranspiration ( $ET_{inst}$ ) in each pixel by the hourly reference evapotranspiration ( $ET_r$ ) given by the meteorological station. Allen [7] standardized  $ET_r$  for alfalfa at 0.5 m high. According to the authors, when using this condition, the  $ET_rF$  can be considered equivalent to the crop coefficient ( $K_c$ ). It also enables the extrapolation of the crop's actual evapotranspiration when the satellite switches to the daily 24 h level. Thus, the  $ET_rF$  is determined by Equation (1).

$$ET_rF = \frac{ET_{inst}}{ET_r} \quad (1)$$

where  $ET_{inst}$  is the instantaneous evapotranspiration ( $\text{mm}\cdot\text{h}^{-1}$ ) and  $ET_r$  is the reference evapotranspiration ( $\text{mm}\cdot\text{h}^{-1}$ ) standardized to alfalfa at 0.5 m height at the moment the satellite is passing.

$ET_{inst}$  is calculated from the latent energy consumed in the evapotranspiration (ET) process and the latent heat of vaporization Equation (2).

$$ET_{inst} = 3600 \frac{LE}{\lambda \rho_w} \quad (2)$$

where 3600 is the conversion from seconds to hours of the duration of the satellite pass; LE is the latent energy consumed in ET ( $\text{W m}^{-2}$ );  $\rho_w$  is the density of water ( $\sim 1000\text{ kg m}^{-3}$ );  $\lambda$  is the latent heat of vaporization ( $\text{J kg}^{-1}$ ).

$\lambda$  represents the heat absorbed when one kg of water is evaporated, and it is determined by Equation (3).

$$\lambda = [2.50 - 0.00236(T_s - 273.15)] \times 10^6 \quad (3)$$

where  $T_s$  is the surface temperature ( $^\circ\text{K}$ ) determined by band 10 of the TIRS sensor (Table 1).

LE is calculated from the Earth's surface energy balance, Equation (4), which involves net radiation ( $R_n$ ), a sensible flux of heat transferred to the ground (G), and a sensible flux of heat convected to air (H). These three components, responsible for determining LE, are usually expressed in  $\text{W m}^{-2}$ .

$$LE = R_n - G - H \quad (4)$$

$R_n$  is the radiant energy of the surface that is partitioned into H, G, and LE and is determined by Equation (5).

$$R_n = R_{s\downarrow} - \alpha R_{s\downarrow} + R_{L\downarrow} - R_{L\uparrow} - (1 - \varepsilon_0)R_{L\downarrow} \quad (5)$$

where  $R_{s\downarrow}$  is the input of shortwave radiation ( $W\ m^{-2}$ );  $\alpha$  is the surface albedo (adim.) determined by bands 2, 3, 4, 5, 6, and 7 of the Landsat-8 (Table 1);  $R_{L\downarrow}$  and  $R_{L\uparrow}$  are the input and output of long waves ( $W\ m^{-2}$ ), respectively; and  $\varepsilon_0$  is the surface thermal emissivity.

G is the heat storage flux in the ground due to heat conduction. When vegetation is present, G tends to have lower values, which means that the rate of heat storage in the soil is lower. The METRIC model provides two methodologies to quantify G, one developed by Bastiaanssen [4] and another by Tasumi [25] (the details can be seen in Allen [7]). For this work, the methodology of Tasumi [25] was chosen, as it was developed in irrigated areas. Thus, G was quantified by Equation (6a,b).

$$\frac{G}{R_n} = 0.05 + 0.18e^{-0.521\ LAI} \text{ for } LAI \geq 0.5 \quad (6a)$$

$$\frac{G}{R_n} = \frac{1.80(T_s - 273.15)}{R_n + 0.084} \text{ for } LAI < 0.5 \quad (6b)$$

where LAI is the leaf area index ( $m^2\ m^{-2}$ ) estimated by the methodology applied by Allen [2].

Finally, H represents the rate of heat loss to air by convection and conduction influenced by a temperature difference. H was calculated by Equation (7).

$$H = \rho_{air} C_p \frac{dT}{r_{ah}} \quad (7)$$

where  $\rho_{air}$  is the air density ( $kg\ m^{-3}$ );  $C_p$  is the specific heat of air at constant pressure ( $J\ kg^{-1}\ K^{-1}$ );  $dT$  is the temperature difference between two heights ( $z_1$  and  $z_2$ ) in an area close to the surface;  $r_{ah}$  is the aerodynamic drag ( $s\ m^{-1}$ ) between these two heights.

All calculations to reach  $ET_rF$  described in this topic were performed using the Water package developed by Olmedo [26] for the R language environment [27].

#### 2.4. Data Extraction for Training

Images from MSI sensors were divided into two groups according to their spatial resolution. Two approaches were applied: approach 1 using a 10 m spatial resolution and approach 2 with a 20 m spatial resolution. This division aimed to understand the effect of the number of predictor variables and spatial resolution on the performance of predictive models.

The primary variables selected for approach 1 were reflectance of blue ( $\rho_B$ ), green ( $\rho_G$ ), red ( $\rho_R$ ), and near-infrared ( $\rho_{NIR}$ ) bands; the variables for approach 2 were reflectance of blue ( $\rho_B$ ), green ( $\rho_G$ ), red ( $\rho_R$ ), red edge 1 ( $\rho_{Re1}$ ), red edge 2 ( $\rho_{Re2}$ ), red edge 3 ( $\rho_{Re3}$ ), near-infrared ( $\rho_{NIR}$ ), shortwave infrared 1 ( $\rho_{SWIR1}$ ), and shortwave infrared 2 ( $\rho_{SWIR2}$ ). As approaches 1 and 2 have, respectively, a 10 and 20 m spatial resolution, it was necessary to match their dimensions with the dependent variable, that is, reduce their resolution to  $30\ m \times 30\ m$ . Therefore, the resample function was used with the bilinear method of the raster package [28] for the R language based on the images of the dependent variable. With images in the same spatial resolution, the central irrigation pivots were cut within each scene using a shapefile, and then the pivots were separated into two groups. The first group encompassed 70% of the total pivots and was used to train the model. The remaining 30% of center pivots were used to test the model (Figure 1).

Due to computational limitations, it was not possible to use all pixels of each pivot of the seven images. Therefore, 25,000 pieces of information, totaling 175,000 pieces of data, were randomly extracted to train the model (Table 3). To test the model, 10,714 pieces

of data were also randomly extracted from each image, totaling 74,998 pixels, maintaining the 30% data proportion for model training. Afterward,  $ET_{rF}$  values lower than 0 that eventually contained some pixels generated by noise in some images were filtered out. In the end, data frame files with geographic coordinates, spectral bands, and  $ET_{rF}$  were obtained.

## 2.5. Training and Statistical Evaluation of Models

### 2.5.1. Production and Selection of Predictor Variables

By aiming to increase the number of predictor variables, the NRPB (normalized ratio procedure between bands) technique was applied to the primary predictor variables in approaches 1 and 2. The NRPB normalized all primary variables, therefore increasing the explanatory variable number for the models (Equation (8)). Hence, the number of variables produced on each approach was higher than the primaries, especially in approach 2, which has the biggest variable number.

$$NRPB = \frac{\rho_x - \rho_y}{\rho_x + \rho_y} \quad (8)$$

where  $\rho_x$  and  $\rho_y$  correspond to the surface reflectance of the wavelengths of the MSI sensor.

The NRPB process was performed using the band ratio function of the labgeo package [29] for the R language, successfully applied by Filgueiras [14,30].

A two-step selection was applied to the NRPB variables to select the most important model variables. The first step removed correlated variables. Explanatory variables with a correlation above 95% were removed to decrease information redundancy. The second step removed the least important variables. Therefore, the recursive feature elimination algorithm (RFE) was applied through the caret package [31] in the R language. This algorithm removes the least important predictor variables from the model by building a base model with all predictors. Then, the importance of each predictor for the base model was calculated, and subsequently, the algorithm classifies these predictors and removes the least important ones, leaving a minor number of variables for the training steps. This is a key process, as it reduces unnecessary variables, aiding both training steps and future applications.

### 2.5.2. Training and Test

The machine learning model training consisted of the input of the remaining predictor variables after the variables selection process into four regression algorithms: Linear regression (LM); Cubist; eXtreme gradient boosting—linear method (Xgblinear); and eXtreme gradient boosting—tree method (Xgbtree). These models were chosen for having an elevated predictive potential and for performing fast on the training, which was carried out with the aid of the caret package [21]. After training, 8 possible predictions were obtained, 4 for approach 1 and 4 for approach 2, which were submitted to testing for evaluating the assertiveness and statistical error.

Thirty percent of data extracted from the pivots were used to test the trained algorithms. Statistical analyses for the test were: the coefficient of determination ( $R^2$ ), Equation (9); root mean squared error (RMSE), Equation (10); mean absolute error (MAE), Equation (11); mean bias error (MBE), Equation (12).

$$R^2 = \frac{[\sum_{i=1}^n (P_i - \bar{P})(O_i - \bar{O})]^2}{[\sum_{i=1}^n (P_i - \bar{P})^2][\sum_{i=1}^n (O_i - \bar{O})^2]} \quad (9)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (P_i - O_i)^2}{n}} \quad (10)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |P_i - O_i| \quad (11)$$

$$\text{MBE} = \frac{1}{n} \sum_{i=1}^n (P_i - O_i) \quad (12)$$

where  $P_i$  is the value predicted by the model;  $O_i$  is the observed value;  $\bar{O}$  is the average observed value;  $n$  is the number of data pairs.

### 2.5.3. Residual Analysis

In addition to statistical metrics, a residual analysis was performed to evaluate the error behavior of predicted values for all models trained for both approaches. Therefore, an image was selected exclusively for this purpose (Table 3). In this image and inside the study area, 11 pivots were selected (highlighted in yellow in Figure 1) for residual analysis between the observed and the predicted values. The 11 pivots were chosen for being close to each other, making it easier to elaborate thematic maps, and because there was no cloud coverage near them. It is noteworthy that the image destined for residual analysis was not in the training process, so the data from this image are new information for the models already trained. Thus, models were applied. Then the predicted  $ET_{rF}$  (approaches 1 and 2) and the observed  $ET_{rF}$  (METRIC) were extracted from the 11 pivots (Figure 1) to reckon the residues (Equation (13)).

$$\text{Residual} = \text{Observed value} - \text{Predicted value} \quad (13)$$

### 2.6. Application

Among the selected pivots (Figure 1), one pivot (highlighted in black) was used for the application of the trained models. This pivot was chosen due to having the largest number of Sentinel-2 images during the crop cycle. It had images from before sugarcane budding to harvest. The application consisted in quantifying the crop coefficient ( $K_c$ ) and the actual evapotranspiration ( $ET_a$ ) during the sugarcane crop cycle for the best METRIC model. In addition to  $K_c$ , water consumption during the sugar cane crop cycle also was estimated using 29 images of the MSI sensor from 11/23/2019 (DAE-days after emergency 05) to 09/03/2020 (DAE 290) and 17 images of the OLI and TRIS sensor between 12/07/2019 (DAE 19) and 09/04/2020 (DAE 291).

From the 29 images of the MSI sensor, the model with the best statistical results on approaches 1 and 2 was applied to determine the  $ET_{rF}$ . According to Allen [7], the  $ET_{rF}$  is equivalent to the  $K_c$  when using the  $ET_r$  of the alfalfa, which is the condition in which the  $ET_{rF}$  was modeled. The 17 images of the OLI and TIRS sensor were used to calculate  $K_c$  through the METRIC model. After quantifying the  $K_c$ , the crop's actual evapotranspiration was established. Therefore, the accumulated reference evapotranspiration for a 24 h ( $ET_{r-24}$ ) period was calculated on the same day the image was acquired.

The  $ET_{r-24}$  in each day was calculated using the Penman–Monteith equation standardized by the ASCE (American Society of Civil Engineers) [32], and meteorological data were acquired in the A539 station, Mocambinho from the INMET (National Institute of Meteorology). With all the  $ET_{r-24}$  and  $K_c$ , the  $ET_a$  in each pixel was established by Equation (14).

$$ET_a = K_c \times ET_{r-24} \quad (14)$$

Finally, the sugar cane crop total evapotranspiration was determined during the cycle through METRIC and approaches 1 and 2 applying integral function through time, where  $x$  was the dates referring to images, and  $y$  was the  $ET_a$  of each date. Thus, the area under the curve, which corresponds to the total evapotranspiration during the cycle, was calculated. This process was performed using function `auc` from the MESS package [33] for the R language.

### 3. Results

#### 3.1. Models

In approach 1, 11 predictors were generated by normalizing the spectral bands of spatial resolution of 10 m. Only seven predictors remained after removing predictors with a correlation above 95% and five when using the recursive feature elimination (RFE). Figure 4 shows the results after the application of the RFE algorithm. Notably, the potential number of predictor variables (ideas) is five. Thus, in the case of more than five variables, the ones with low importance can be removed without significantly impacting the trained models. On the other hand, less than five variables negatively affect the model prediction.

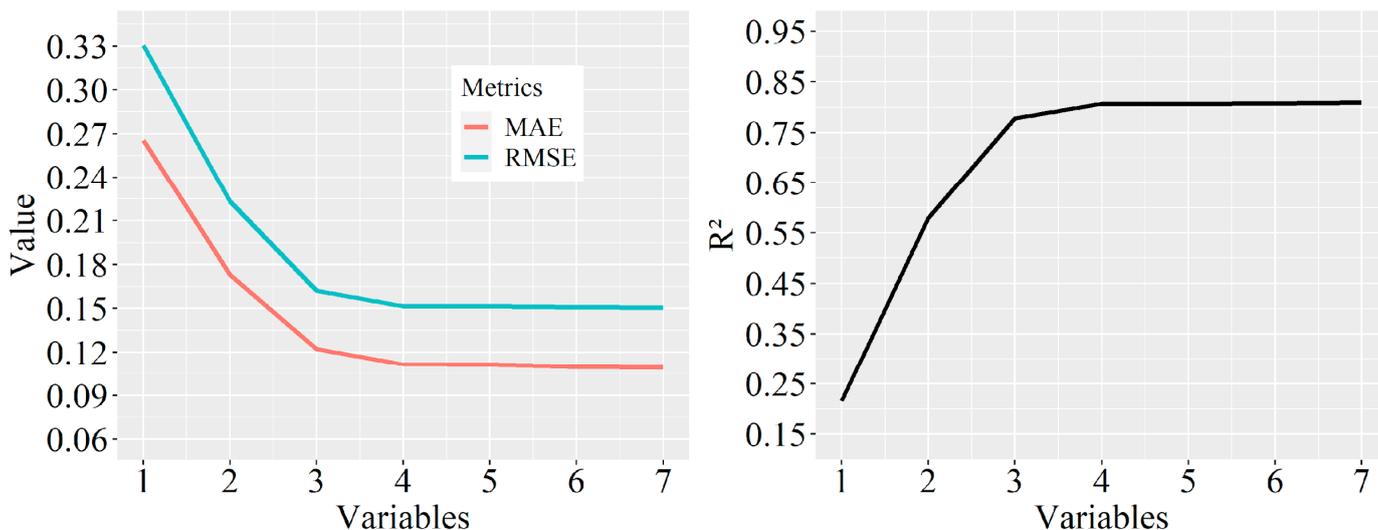


Figure 4. Statistical results for the selection of predictor variables when applying the RFE in approach 1.

In approach 2, the number of predictors generated by normalizing the bands is 45, higher than in approach 1 because of the bigger number of spectral bands for the special resolution of 20 m. When establishing the baseline of 95% of correlation, 22 variables remained and were later reduced to 12 after the application of the RFE. Figure 5 shows the stability of the statistical metrics using 12 predictable variables to train the model.

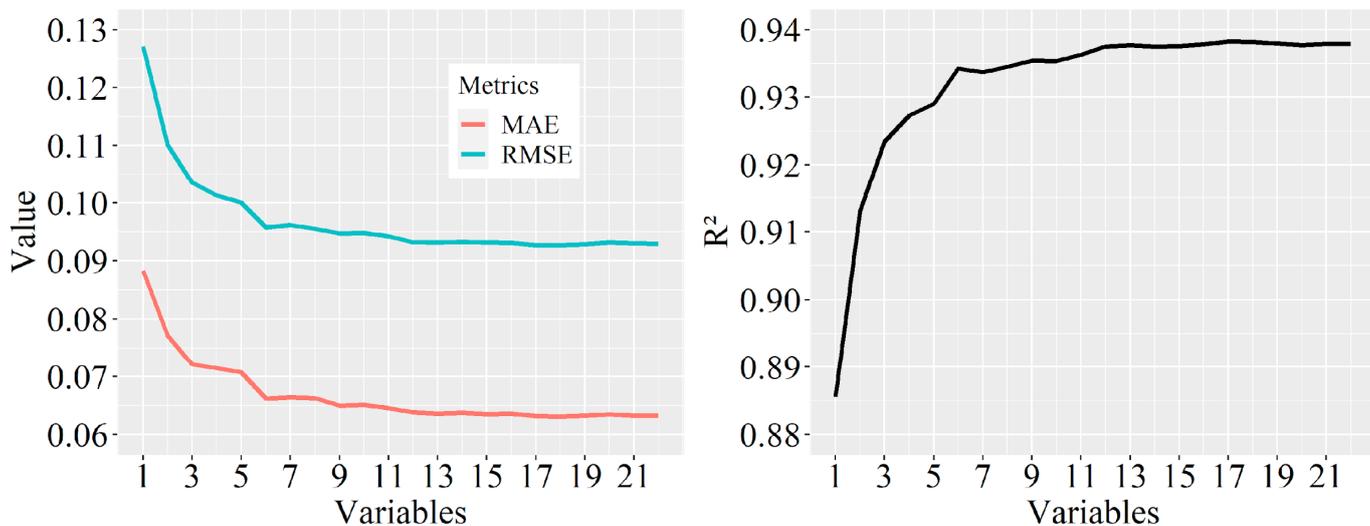


Figure 5. Statistical results for the selection of predictor variables when applying the RFE in approach 2.

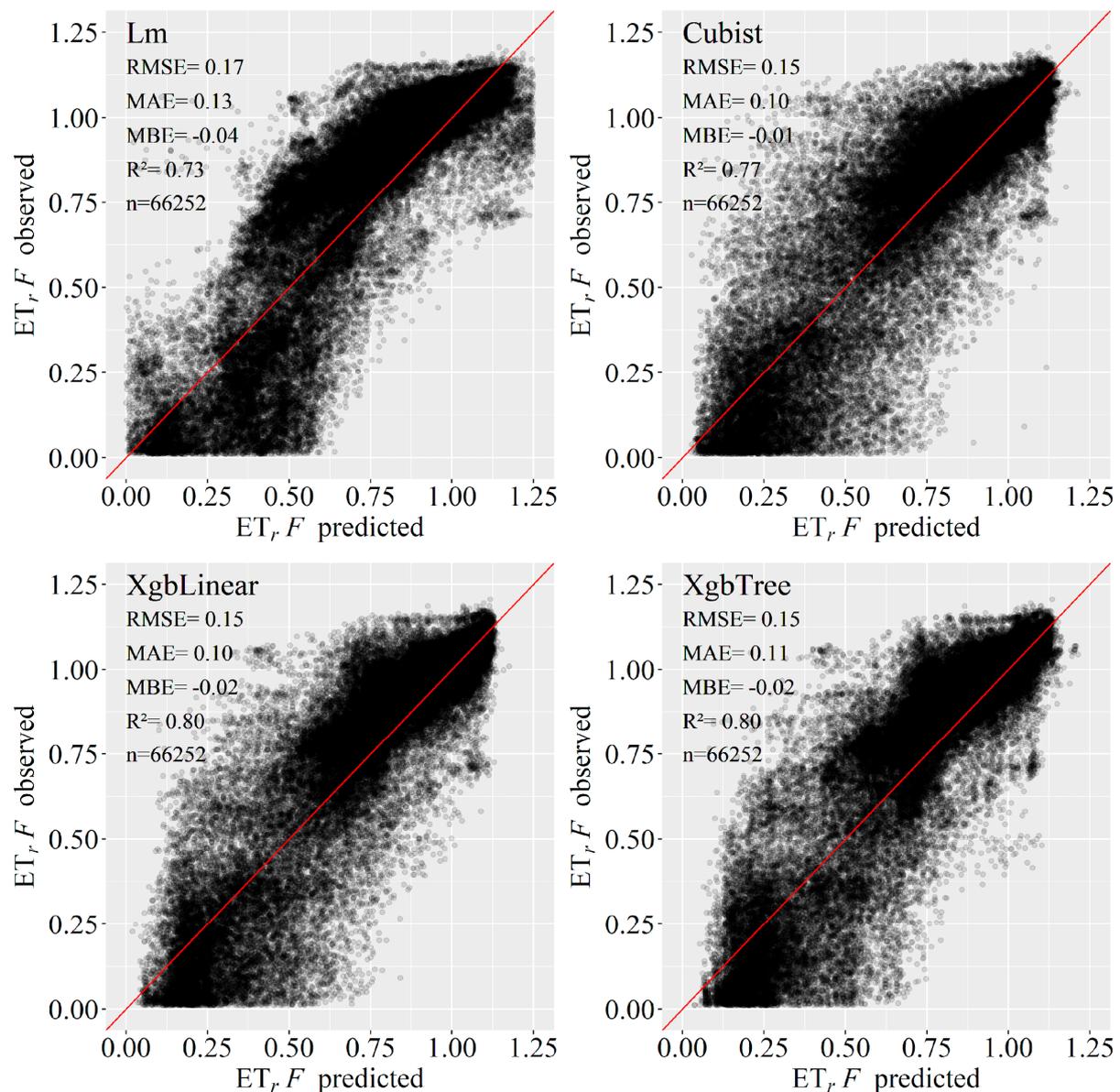
In Table 4, the 5 predictor variables selected for approach 1 and the 12 selected for approach 2 can be found. It is worth mentioning that the predictors in this table are not rated by importance, as each algorithm assigns different importance after the training. In supplementary materials are the figures with the importance of the predictors for each model in approaches 1 and 2.

**Table 4.** Selected predictor variables for training models in approaches 1 and 2.

| N° | Approach 1            | Approach 2                |
|----|-----------------------|---------------------------|
| 1  | B2                    | B6                        |
| 2  | B4                    | B8                        |
| 3  | B8                    | B12                       |
| 4  | $(B2 - B3)/(B2 + B3)$ | $(B2 - B3)/(B2 + B3)$     |
| 5  | $(B2 - B4)/(B2 + B4)$ | $(B2 - B4)/(B2 + B4)$     |
| 6  | -                     | $(B2 - B5)/(B2 + B5)$     |
| 7  | -                     | $(B2 - B12)/(B2 + B12)$   |
| 8  | -                     | $(B5 - B11)/(B5 + B11)$   |
| 9  | -                     | $(B5 - B12)/(B5 + B12)$   |
| 10 | -                     | $(B6 - B8)/(B6 + B8)$     |
| 11 | -                     | $(B8 - B12)/(B8 + B12)$   |
| 12 | -                     | $(B11 - B12)/(B11 + B12)$ |

For approach 1, the most common wavelengths among the predictors were B2 (blue) and B4 (red), both appearing twice, counted individually and in the NRPB, followed by B3 (green) and finally B8 (near infrared). B2 and B4, which stood out in this approach, are absorbed in higher intensity by plants from photosynthesizing pigments [11]. B3 is also absorbed by the same pigments, however, at a much lower intensity. While B8 tends to be reflected or transmitted because when there is absorption, the physiological structures of the plant heat up, and the plant tends to dissipate this energy through its structures. Therefore, the more nourished the dissipation structures are, the more efficient the reflectance and/or transmittance are. Thus, for approach 1, the wavelengths B2, B3, and B4 correspond to the pigments of the leaves, while B8 (Near-infrared) is related to the structural part. As a result, the selected predictors are important for both the photosynthesis process and the structure of the cellular components of the plants studied. Hence, three wavelengths have characteristics related to the photosynthesizing pigments and one related to the structure. In approach 2, the bands that stood out were B12 (short infrared), appearing five times, and B2, appearing four times. B12 is a short wavelength related to water content, which is an object of study in this work, through the physical–biological process of evapotranspiration. B5 and B6 bands, both red edges, also stood out, representing a transition among the visible region and near-infrared, and B8 and B11, shortwave infrared. Thus, it can be noted in this approach that the selected predictors are related not only to the photosynthetic and structural parts but also to water content. Hence, it can be inferred that three wavelengths with photosynthetic characteristics were assigned, one structural, two in the transition from photosynthetic to structural, and two with water content.

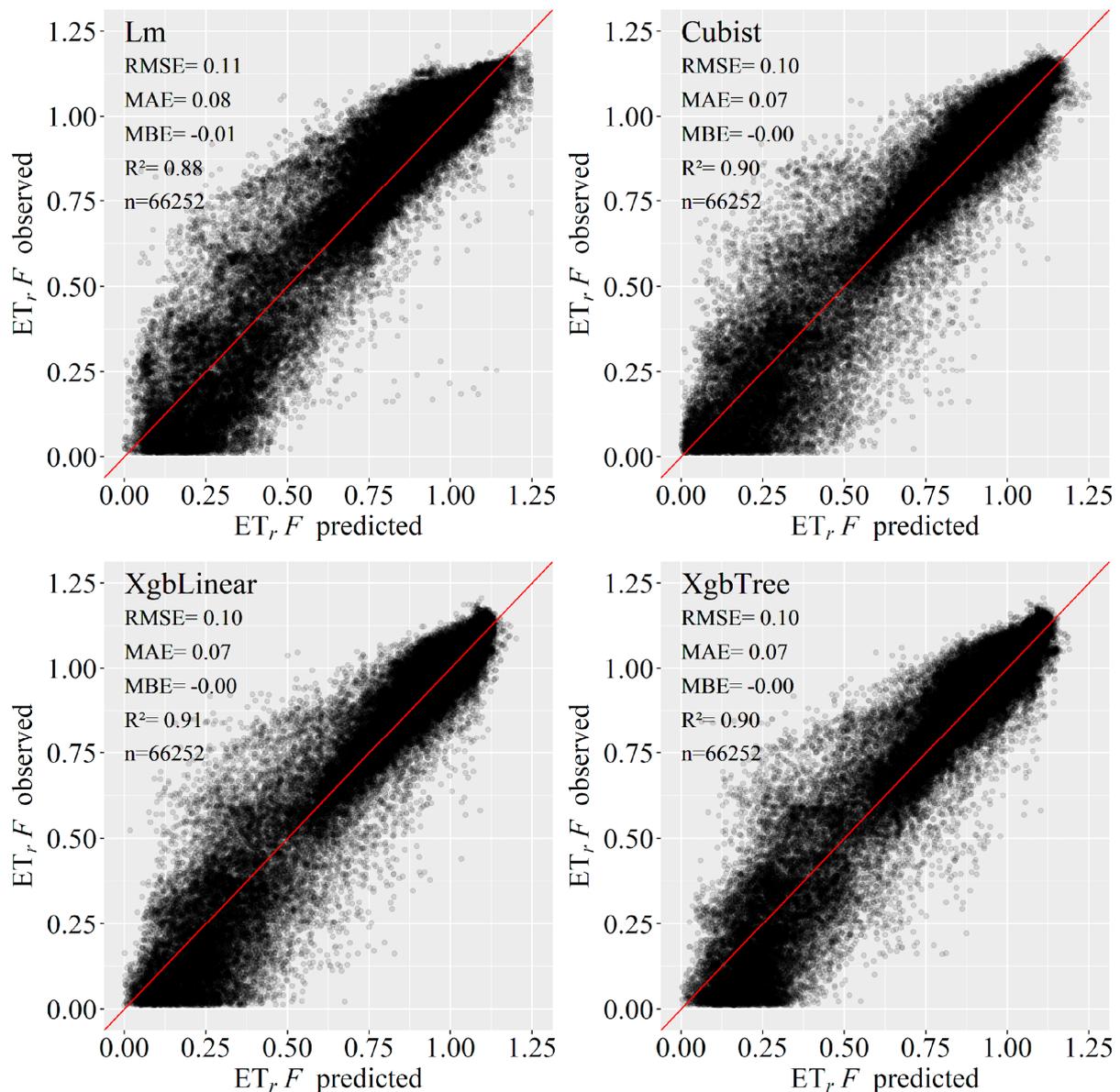
Test performances for approach 1 for each machine learning model can be seen in Figure 6.



**Figure 6.** Statistical results of the test of approach 1.

Among the four models chosen for this approach, XgbLinear and XgbTree obtained the best statistical results, with  $R^2 = 0.80$  and  $RMSE = 0.15$ ; followed by the Cubist with  $R^2 = 0.77$  and  $RMSE = 0.15$ ; lastly, and with the worst result, the multiple linear regression (Lm) with the minimum  $R^2 = 0.73$  and the major  $RMSE = 0.17$ . Regardless of the model, the high dispersion shown on the test draws attention, especially in values lower than 0.6 (Figure 6). Ke et al. [34], when estimating evapotranspiration using machine learning in Landsat-8 data and MODIS for a heterogeneous environment, noted that in areas with crops, there was a higher dispersion between predicted vs. observed values than in areas with forest, grazing, and bushes. Thus, results from these authors agree with the ones found here, as, in crop areas, there is a higher surface movement, being more dynamic both in terms of vegetation cover and in terms of soil moisture.

Figure 7 shows the statistical results of the test for approach 2, which used only a 20 m spatial resolution as a source of predictor variables.

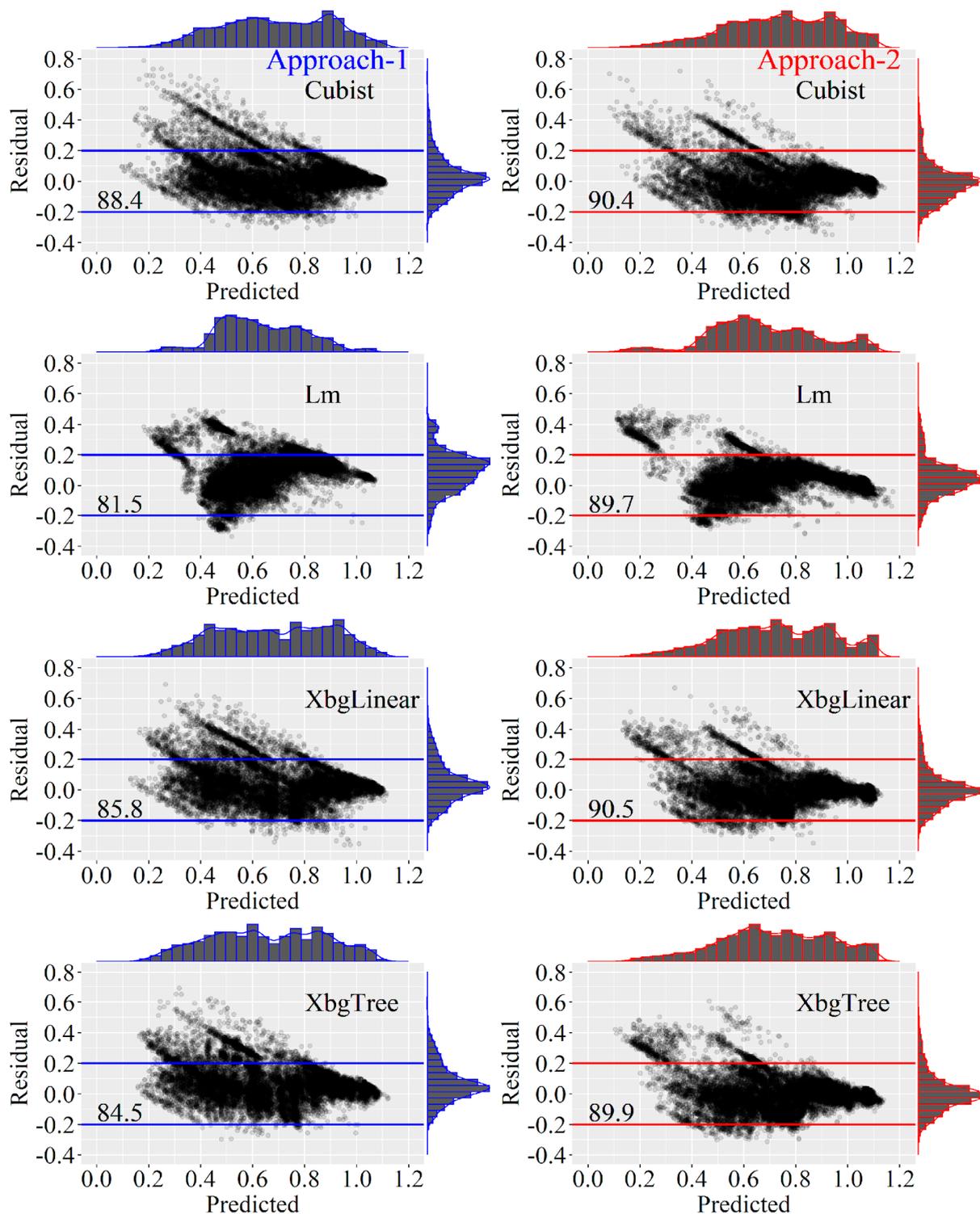


**Figure 7.** Statistical results of the test of approach 2.

It is observed in Figure 7 that Lm, as well as in approach 1, obtained the worst result, while Cubist, XgbLinear, and XgbTree were the ones that best explained the response variable, with lower RMSE (0.10) and higher  $R^2$ .  $R^2$ , as in approach 1, had different values among these last three models, being 0.91 for XgbLinear and 0.90 for Cubist and XgbTree. When comparing Figures 6 and 7, approach 2 obtained more satisfying results than approach 1, with models with  $R^2$  higher than 0.88 and RMSE lower than 0.11, whereas approach 1 had  $R^2$  lower than 0.80 and RMSE higher than 0.15. It seems that approach 2 even reduced the dispersion seen in Figure 6 significantly.

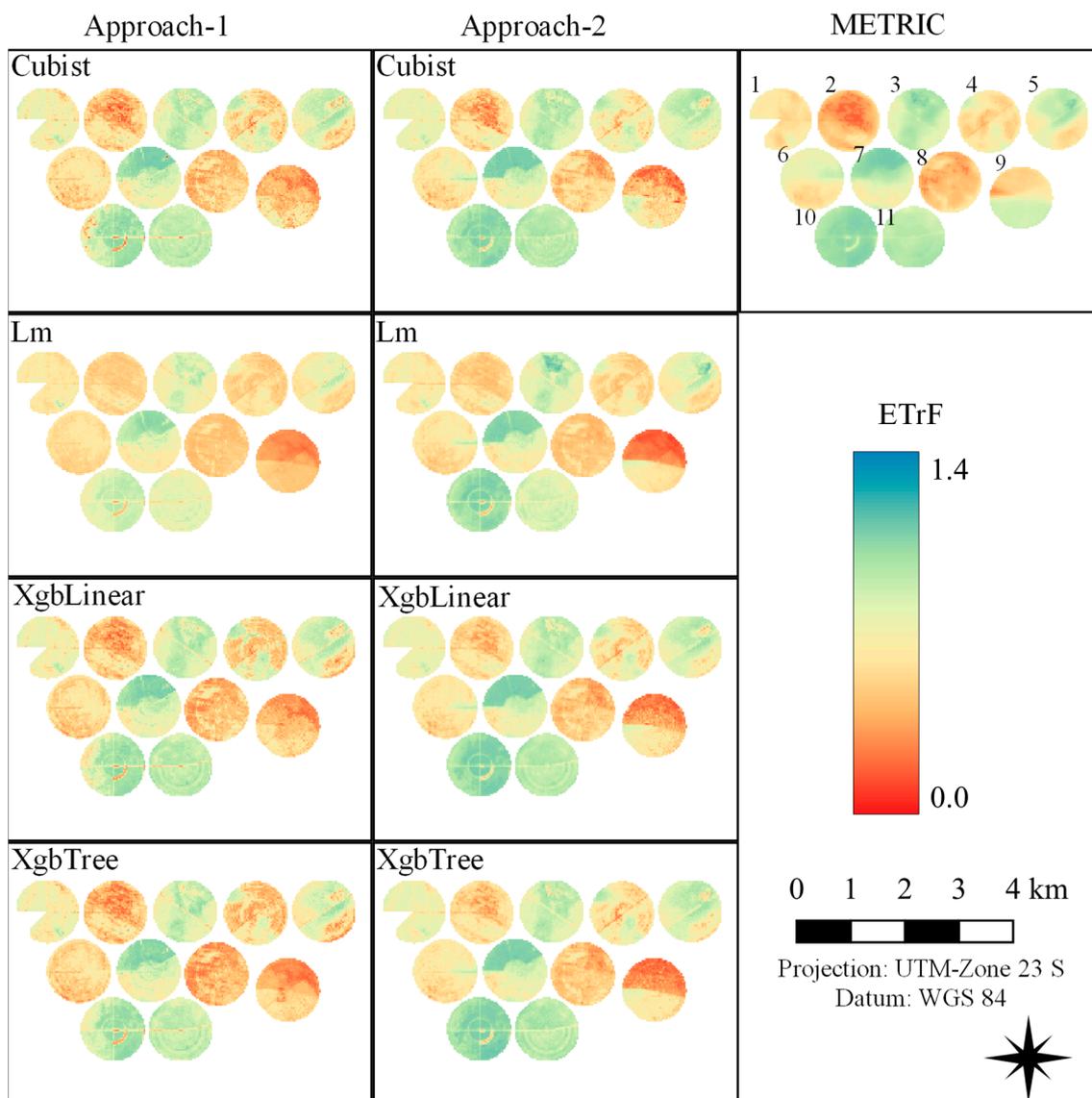
### 3.2. Residual Analysis

The elaborated models were applied to the image of 02/12/2020 (Table 3), which was destined for the residual analysis. The results are shown in Figure 8.



**Figure 8.** Residual analysis of the estimated  $ET_rF$  through machine learning models.

Notably, Figure 8 maintains the same pattern found in Figures 6 and 7, in which approach 1 presented a result inferior to approach 2 due to the minor concentration of residual points close to the zero value. For Approach 2, a good performance was observed, as less than 10.3% of the values distanced from  $-0.2$  to  $0.2$ , which indicates a low prediction error, mainly XbgLinear. In both approaches, the explanatory variable tends to be less precise when  $ET_rF$  is low, especially for values under 0.6, as previously discussed; this is better shown in Figure 9.

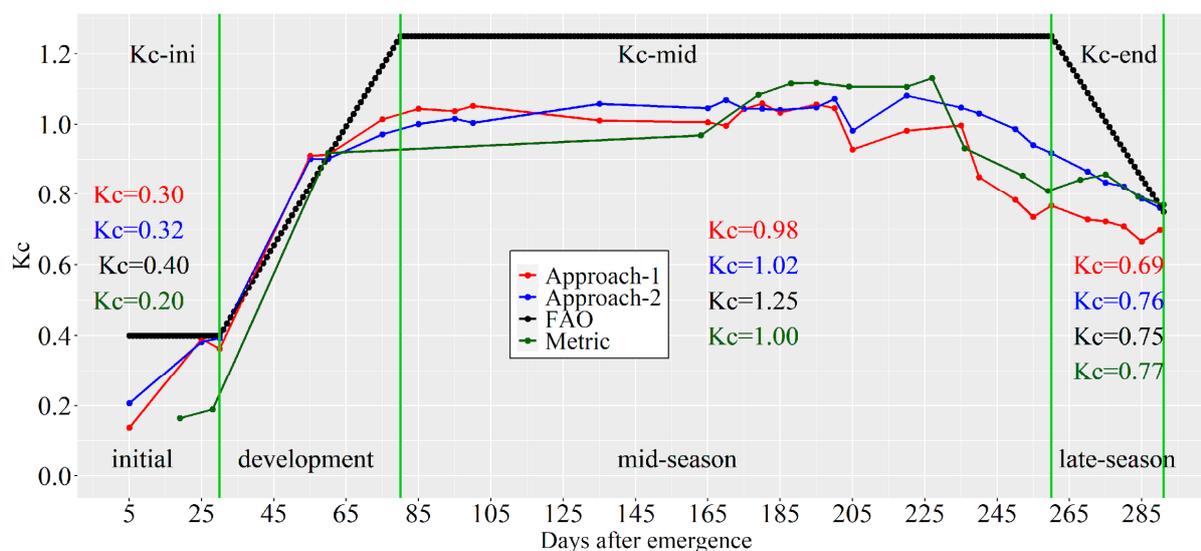


**Figure 9.** Spatial variability of  $ET_{rF}$  for approaches 1, 2, and Metric for 02/12/2020.

Figure 9 shows the lower precision during the prediction of values lower than 0.6 in greater detail, with particular reference to pivot 9. On this pivot, when METRIC showed values close to 0.6, approaches 1 and 2 obtained predicted values close to zero, which can be seen in Figure 8 with the residual analysis. Nevertheless, Figure 9 also reveals the great assertiveness of the models, which can be seen especially in pivots 7 and 10, where prediction models were able to capture a wealth of details regarding METRIC estimates. On pivot 10, the southeast region shows low  $ET_{rF}$  values through METRIC and machine learning models. However, for that pivot, this result was more prominent when using Sentinel-2 images, as they showed greater spatial resolution. Additionally, the thermal band on Landsat-8, which has a coarser resolution (100 m), can mask nuances of the surface of the monitored area.

### 3.3. Models Application

Figure 10 shows the distribution of the crop coefficient ( $K_c$ ) during the sugar cane cycle, determined through the XgbLinear, which presented better metrics for approaches 1 and 2 through the METRIC model and also through the FAO-56 report [1].

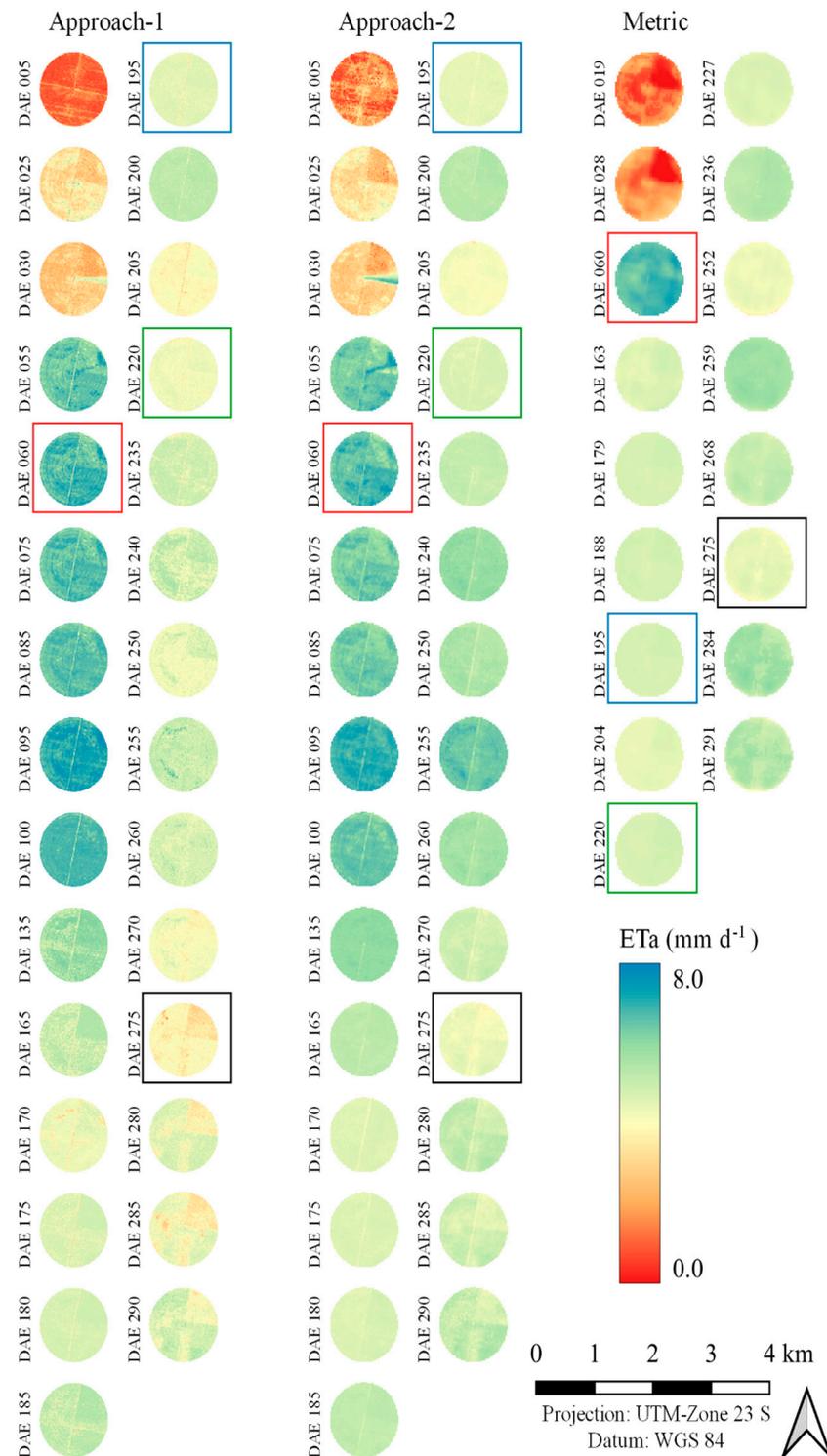


**Figure 10.** Comparison of  $K_c$  values between approaches 1 and 2, METRIC, and the ones recommended by the FAO's 56 reports.

Figure 10 evidences what was previously discussed; for the initial phases of the crop, with low-density plant coverage, machine learning prediction model values showed uncertainties when compared to METRIC, but, as the culture develops, this error tends to reduce. For phase I (initial), when there is a crop emergency, the  $K_c$  for approaches 1 and 2 were 0.3 and 0.32, respectively, while METRIC estimated a  $K_c$  of 0.20, and the  $K_c$  determined by the 56 FAO's report [1] is 0.40. In phase II (development), the  $K_c$  curves in approaches 1 and 2 become closer to the ones of the METRIC as the phenological state progresses, as well as the  $K_c$ -FAO curve becomes closer to both approaches and distances a little from the METRIC curve. For phase III (mid-season) crop growth, the average  $K_c$  was 0.98 for approach 1 and 1.00 for METRIC, while for approach 2, it was 1.02, and FAO's 56 report recommends using a 1.25 value. Moreover, in phase III, the METRIC curve behaves similarly to approach 2, which corroborates the results found in the models' tests. In phase IV (late-season), maturation, approach 1 was the most distant from both FAO's and METRIC's  $K_c$ , with a value equal to 0.69, while approach 2 had a  $K_c$  of 0.76 and METRIC of 0.77. Dingre and Gorantiwar [35], when quantifying  $K_c$  for sugar cane through the water balance method, found medium values for phases I, III, and IV of 0.36, 1.20, and 0.78, respectively. Silva et al. [36], for the Brazilian semiarid region, obtained a  $K_c$  for ratoon cane in phases I, III, and IV equal to 0.65, 1.10, and 0.85, different from FAO's recommendation. This evinces a divergence in  $K_c$  values shown in the literature from the ones recommended, which can be explained by the specificity of the local weather in which each study was performed, as well as the physiological conditions of the crop at the time of the satellite imaging. Phase III, for this work, reached the maximum value of 1.09; however, it showed variations throughout the whole phase.

Regarding Figure 10, it is noticeable that with Sentinel-2, due to its better temporal resolution, it was possible to obtain a larger number of  $K_c$  information throughout the crop cycle than Landsat-8, which is important to obtain a major temporal variability of this coefficient. This can lead to more assertive crop management than if only Landsat-8 data were used. Saleem and Awange [37] mention that Sentinel-2 represents a new age for obtaining more precise information about the Earth's surface, as it has a greater spatial and temporal resolution among satellites that provide images for free. Nevertheless, information referring to  $K_c$  can be expanded when joining the two orbital platforms, as more information will be obtained with a larger frequency of images, as highlighted in Filgueiras et al. [9].

Crop actual evapotranspiration in a spatial resolution of 10, 20, and 30 m for approach 1, approach 2, and METRIC, are shown in Figure 11.



**Figure 11.** Sugar cane temporal-spatial actual evapotranspiration in three different spatial resolutions evincing, through rectangles, coincident dates between Sentinel-2 and Landsat-8.

In Figure 11, attention is drawn to DAE 030 with a ray with an ET<sub>a</sub> larger than for the rest of the pivot. On that day, crops were in the initial emergency phase, and soil was exposed in most of the area, receiving water from irrigation. This ray corresponds to the

moisture in the exposed soil, and it is displaced clockwise from the irrigation equipment. Such information is only visible with 10 and 20 m resolution, and thus, the METRIC model, due to the use of thermal images, can not show. Spatially detailed information, as seen in the Sentinel-2 images of Figure 11, has great value for field professionals. Coinciding dates between Sentinel-2 and Landsat-8 show high similarities in the spatial  $ET_a$  between approaches 1 and 2 with METRIC. Approach 2 stands out for having larger spatial proximity. Such proximity is evidenced in Table 5, in which the averages  $ET_{as}$  in approach 2 have the smallest differences in the estimated averages by the METRIC model. In addition, the standard deviation for approach 2 was also smaller than approach 1 but larger than the METRIC model. This fact might be attached to the more detailed spatial resolutions of approaches 1 and 2 when compared to the method using the METRIC model.

**Table 5.** Evapotranspiration averages for coinciding dates between Sentinel-2 and Landsat-8.

| DAE      | Approach 1<br>(mm) | Approach 2<br>(mm) | METRIC<br>(mm) | Difference (%) |            |
|----------|--------------------|--------------------|----------------|----------------|------------|
|          |                    |                    |                | Approach 1     | Approach 2 |
| 060      | 6.71 ± 0.44        | 6.63 ± 0.36        | 6.75 ± 0.25    | 0.59           | 1.77       |
| 195      | 4.52 ± 0.22        | 4.48 ± 0.09        | 4.78 ± 0.06    | 5.44           | 6.28       |
| 220      | 4.25 ± 0.21        | 4.68 ± 0.11        | 4.79 ± 0.08    | 11.27          | 2.30       |
| 275      | 3.75 ± 0.36        | 4.32 ± 0.20        | 4.44 ± 0.17    | 16.54          | 2.77       |
| ET-Total | 1417.77            | 1474.26            | 1544.11        | 8.18           | 4.52       |

After calculating the integral of  $ET_a$ , it was found that the sugar cane total water demand along the crop cycle was 1417.77 mm for approach 1, 1474.26 mm for approach 2, and 1544.11 mm for METRIC. It is perceptible that total evapotranspiration estimated for approaches 1 and 2 were close to the total evapotranspiration by METRIC, where percentual differences were 8.18 for approach 1 and 4.52 for approach 2 when compared to METRIC. Approach 2 had a closer value to the one estimated by METRIC, going against the values found for dates indicated in Figure 11. Sugar cane total evapotranspiration during its cycle, found by Dingre and Gorantiwar [35], was 1388 mm. The one found by Silva et al. [38] for the Brazilian Northeast conditions was 1600 mm. Thus, it can be noticed that the total evapotranspiration found in this work is close to the values found in Brazil.

#### 4. Discussions

The spectral bands of blue and red stand out in approach 1 because these wavelengths are absorbed mostly by alpha and beta-chlorophylls, resulting in the release of oxygen from photosynthesis [11,39,40]. It means that the magnitude of the release of this chalcogen occurs in them. Thus, the greater the absorption of blue and red wavelengths, the greater the oxygen release and, consequently, the greater the water vapor release since water vapor is released along with oxygen during photosynthesis, explaining why these spectral bands stood out. The green wavelength of the MSI sensor ranges from 0.53 to 0.59  $\mu\text{m}$ , and the wavelengths in this range are less absorbed by pigments [11]; thus, it impacts little on photosynthesis and water release by plants. The near-infrared is reflected and/or transmitted by the structure of the mesophyll as a form of protection of its physiological and molecular structure. Thus, this length is not related to photosynthesis but to the physiological structure that influences photosynthesis indirectly. As a result, this wavelength has less prominence than the others. The wavelengths in the short infrared are the ones that stood out in the second approach because they are spectral bands strongly absorbed by the water content present in plants [41,42]. Therefore, when the leaves are in hydric comfort, these wavelengths tend to be more absorbed. On the other hand, when occurring hydric stress, these wavelengths tend to be less absorbed and more reflected. Several studies demonstrated the importance of shortwaves for monitoring and quantifying water stress on plants [43–46]. The red edge wavelengths are also important bands in approach 2 for the same principles as the short infrared. They are absorbed by water content in plants as a

result of the sensitivity of red edge spectral responses to water content in plant leaves [47]. These authors mention that vegetation indices that have red edges in their equation have a higher sensitivity to water content in maize crops. Santos et al. [42], studying water stress and spectral response in sugarcane crops, identified that the near-infrared wavelength was sensitive to the water content in leaves. Chandel et al. [43] also obtained similar results with wheat crops. Thus, it strengthens the evidence for the importance of these wavelengths in quantifying water content, even more so when they are normalized with other bands through the NRPB. This is one of the arguments stating the superior predictive capacity of approach 2 compared to approach 1. The presence of the short infrared and red edge spectral bands in approach 2 significantly improves the statistical metrics of the models.

The dispersion found in  $ET_{rF}$  values lower than 0.6 in both approaches may be related to greater heterogeneity of the cultivated area. When  $ET_{rF}$  values are lower than 0.6, the crop does not completely cover the soil, and there may be patches of soil with different colors, moisture, etc., a proliferation of weeds in parts of the area, and even a greater degree of the vigor of the crop in parts of the plot. The fact that approach 1 is more dispersed than approach 2 is linked, quantitatively and qualitatively, to their predictors. Approach 2 had a total of 12 predictors: the same as approach 1 and some others, mainly the short infrared wavelength and red edge.

The main results and discussion found in the literature when applying METRIC to predict the evapotranspiration fraction is to consider  $ET_{rF}$  to be equal to  $K_c$  when using the reference evapotranspiration of alfalfa [7,48–50]. However, this interpretation is not correct because sensors on board satellites or other platforms capture the information that is occurring in the field under natural conditions, and one cannot be sure that the plant is in maximum water comfort and with nutritional, pest, disease, and weed management adequate for maximum water uptake. Thus both METRIC and the models trained in this study estimate the product between  $K_c$  and  $K_s$  (stress coefficient), the latter being responsible for reducing  $ET_{rF}$  to values lower than  $K_c$  when the crop is under stress or making it equal to  $K_c$  when the plant is in favorable conditions for maximum water uptake.

## 5. Conclusions

Approach 2 had statistical results superior to approach 1, mainly for the XgbLinear model, which obtained  $R^2$  of 0.91 and RMSE of 0.10, whereas the metrics for the same model considering approach 1 were 0.80 and 0.15 for the  $R^2$  and RMSE, respectively. This result was mainly influenced by the greater number of spectral bands that are strongly related to water content as the short infrared and the red edge, thus being the model to be applied. However, all models developed showed limitations when the dependent variable presents values lower than 0.6, a condition in which the crop canopy has not completely covered the soil, and there is greater variability.

Despite the limitation when the soil is not fully covered, the application proved efficient in predicting  $ET_{rF}$  since the analysis had values close to zero and the maximum distance only in the linear regression algorithms for both approach 1 and approach 2. Furthermore, it was possible to note that  $ET_{rF}$  cannot be considered the same as  $K_c$  because the onboard sensors capture the actual condition of the crop in the field, hydric comfort or not. Hence,  $ET_{rF}$  can be understood as the product between  $K_c$  and  $K_s$  that best represents field conditions.

The use of the MSI sensor and machine learning techniques proved to be a new and simple alternative to estimate  $ET_{rF}$  through spectral information, complementing the METRIC model estimation using OLI and TIRS sensors and increasing the frequency that information is generated for areas of interest. The combination of these sensors is useful to obtain the highest temporal resolution of the crop, especially for irrigated agriculture, which requires  $K_c$  and  $K_s$  coefficients to be determined daily for adequate replenishment of the irrigation blade.

Finally, this work highlights the possibility of using this methodology for other remote sensors to calculate spatial and temporal evapotranspiration, enriching the scientific debate.

Further, it can be applied in other locations and different cultures since the data needed are from the orbital sensors and meteorological data from the area of interest. Thus, the methodology, despite being extensive, is easy to apply with a low financial cost. However, in hopes of advancements, this methodology can be performed with field instrumentation to collect the actual evapotranspiration, as this evapotranspiration would be replaced by the evapotranspiration of the METRIC model, but more funding would be required.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/atmos13091518/s1>, Figure S1: Importance of the variables for the approach 1 algorithm; Figure S2: Importance of the variables for the approach 1 Cubist algorithm; Figure S3: Importance of the variables for the approach 1 XgbLinear algorithm; Figure S4: Importance of the variables for the approach 1 XgbTree algorithm; Figure S5: Importance of the variables for the approach 2 LM algorithm; Figure S6: Importance of the variables for the approach 2 Cubist algorithm; Figure S7: Importance of the variables for the approach 2 XgbLinear algorithm; Figure S8: Importance of the variables for the approach 2 XgbTree algorithm.

**Author Contributions:** Conceptualization, R.A.d.S., E.I.F.-F. and E.C.M.; methodology, R.A.d.S., R.F., E.I.F.-F., E.C.M. and R.D.S.L.; software, R.A.d.S.; validation, R.A.d.S., R.F., E.I.F.-F. and E.C.M.; formal analysis, R.A.d.S., E.I.F.-F. and E.C.M.; investigation, R.A.d.S.; resources, E.C.M.; data curation, R.A.d.S.; writing—original draft preparation, R.A.d.S.; writing—review and editing, R.A.d.S., R.F., R.D.S.L., V.B.B. and C.M.U.N.; visualization, E.I.F.-F., R.F., R.D.S.L., V.B.B. and E.C.M.; supervision, E.C.M. and E.I.F.-F.; project administration, E.C.M.; funding acquisition, E.C.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), finance code: 001, the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPQ—In English: National Council for Scientific and Technological Development)—Grant number 140416/2020-0 and Federal University of Viçosa—UFV.

**Data Availability Statement:** The dataset, as well as the models produced, can be found and visualized in dos Santos (2022), “Modelling actual evapotranspiration with MSI-Sentinel images and machine learning algorithms”, Mendeley Data, v1 <http://dx.doi.org/10.17632/j74vdswh86.1>.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

|                    |   |
|--------------------|---|
| ET <sub>a</sub>    | Actual Evapotranspiration   |
| ET <sub>ints</sub> | Instantaneous Evapotranspiration  |
| ET <sub>o</sub>    | Reference Potential Evapotranspiration of the Grass                         |
| ET <sub>r</sub>    | Reference Potential Evapotranspiration of the Alfalfa                       |
| ET <sub>r</sub> F  | Evapotranspiration Fraction   |
| METRIC             | Mapping Evapotranspiration at High Resolution with Internalized Calibration |
| ML                 | Machine Learning  |
| MSI                | MultiSpectral Instrument  |
| OLI                | Operational Land Imager   |
| R <sup>2</sup>     | Coefficient of Determination  |
| RMSE               | Root Mean Squared Error   |
| MAE                | Mean Absolute Error   |
| MBE                | Mean Bias Error   |
| SAFER              | Simple Algorithm for Evapotranspiration Retrieving                          |
| SEBAL              | Surface Energy Balance Algorithm for Land                                   |
| TIRS               | Thermal Infrared Sensor   |
| P                  | Precipitation   |
| T <sub>mean</sub>  | Mean Air Temperature  |
| T <sub>max</sub>   | Maximum Air Temperature   |
| T <sub>min</sub>   | Minimum Air Temperature   |
| µm                 | Micrometer  |

|                   |   |
|-------------------|---|
| m                 | Meter   |
| $K_c$             | Crop Coefficient                                |
| LE                | Latent Energy                                   |
| $\rho_w$          | Density of Water                                |
| $\lambda$         | Latent Heat of Vaporization                     |
| $T_s$             | Surface Temperature                             |
| $R_n$             | Net Radiation                                   |
| G                 | Sensible Flux of Heat Transferred to the Ground |
| H                 | Sensible Flux of Heat Convected to Air          |
| $R_{s\downarrow}$ | Input of Shortwave Radiation                    |
| $\alpha$          | Surface Albedo                                  |
| $R_{L\downarrow}$ | Input of Long Waves                             |
| $R_{L\uparrow}$   | Output of Long Waves                            |
| $\varepsilon_0$   | Surface Thermal Emissivity                      |
| LAI               | Leaf Area Index                                 |
| $\rho_{air}$      | Air Density                                     |
| $C_p$             | Specific Heat of Air at Constant Pressure       |
| dT                | Temperature Difference Between Two Heights      |
| $r_{ah}$          | Aerodynamic Drag                                |
| $\rho_B$          | Reflectance of Blue                             |
| $\rho_G$          | Reflectance of Green                            |
| $\rho_R$          | Reflectance of Red                              |
| $\rho_{NIR}$      | Reflectance of Near-Infrared                    |
| $\rho_{Re}$       | Reflectance of Red Edge                         |
| $\rho_{SWIR}$     | Reflectance of Shortwave Infrared               |
| NRPB              | Normalized Ratio Procedure Between Bands        |
| LM                | Linear Regression                               |
| Xgblinear         | eXtreme Gradient Boosting-linear method         |
| XgbTree           | eXtreme Gradient Boosting-tree method           |
| DAE               | Days After Emergency                            |
| FAO               | Food and Agriculture Organization               |

## References

- Allen, R.G.; Pereira, L.S.; Raes, D.; Smith, M. *Table of Contents Originated by: Agriculture Crop Evapotranspiration—Guidelines for Computing Crop Water Requirements—FAO Irrigation and Drainage Paper 56*, 9th ed.; Food and Agriculture Organization of the United Nations: Rome, Italy, 1998.
- Allen, R.; Irmak, A.; Trezza, R.; Hendrickx, J.M.H.; Bastiaanssen, W.; Kjaersgaard, J. Satellite-Based ET Estimation in Agriculture Using SEBAL and METRIC. *Hydrol. Process.* **2011**, *25*, 4011–4027. [[CrossRef](#)]
- Srivastava, A.; Sahoo, B.; Raghuvanshi, N.S.; Singh, R. Evaluation of Variable-Infiltration Capacity Model and MODIS-Terra Satellite-Derived Grid-Scale Evapotranspiration Estimates in a River Basin with Tropical Monsoon-Type Climatology. *J. Irrig. Drain. Eng.* **2017**, *143*, 1. [[CrossRef](#)]
- Bastiaanssen, W.G. SEBAL-Based Sensible and Latent Heat Fluxes in the Irrigated Gediz Basin, Turkey. *J. Hydrol.* **1999**, *229*, 87–100. [[CrossRef](#)]
- Teixeira, A.D.C.; Bastiaanssen, W.; Ahmad, M.-U.; Bos, M. Reviewing SEBAL Input Parameters for Assessing Evapotranspiration and Water Productivity for the Low-Middle São Francisco River Basin, Brazil. Part B: Application to the Regional Scale. *Agric. For. Meteorol.* **2009**, *149*, 477–490. [[CrossRef](#)]
- Bastiaanssen, W.G.M.; Pelgrum, H.; Wang, J.; Ma, Y.; Moreno, J.F.; Roerink, G.J.; van der Wal, T. A Remote Sensing Surface Energy Balance Algorithm for Land (SEBAL), Part 1: Formulation. *J. Hydrol.* **1998**, *212–213*, 213–229. [[CrossRef](#)]
- Allen, R.G.; Tasumi, M.; Trezza, R. Satellite-Based Energy Balance for Mapping Evapotranspiration with Internalized Calibration (METRIC)—Applications. *J. Irrig. Drain. Eng.* **2007**, *133*, 395–406. [[CrossRef](#)]
- Teixeira, A.H.d.C. Determining Regional Actual Evapotranspiration of Irrigated Crops and Natural Vegetation in the São Francisco River Basin (Brazil) Using Remote Sensing and Penman-Monteith Equation. *Remote Sens.* **2010**, *2*, 1287–1319. [[CrossRef](#)]
- Filgueiras, R.; Mantovani, E.C.; Dias, S.H.B.; Fernandes Filho, E.I.; da Cunha, F.F. Optimizing the Monitoring of Natural Phenomena Through the Coupling of Orbital Multi-Sensors. *Geo UERJ.* **2020**, *37*, 37832. [[CrossRef](#)]
- Berni, J.A.J.; Zarco-Tejada, P.J.; Sepulcre-Cantó, G.; Fereres, E.; Villalobos, F. Mapping Canopy Conductance and CWSI in Olive Orchards Using High Resolution Thermal Remote Sensing Imagery. *Remote Sens. Environ.* **2009**, *113*, 2380–2388. [[CrossRef](#)]
- Taiz, L.; Zeiger, E.; Moller, I.M.; Murphy, A. *Plant Physiology & Development*, 6th ed.; Sinauer Associates Incorporated: Sunderland, MA, USA, 2015; ISBN 9781605353531.

12. Cervantes, J.; Garcia-Lamont, F.; Rodríguez-Mazahua, L.; Lopez, A. A Comprehensive Survey on Support Vector Machine Classification: Applications, Challenges and Trends. *Neurocomputing* **2020**, *408*, 189–215. [[CrossRef](#)]
13. Adab, H.; Morbidelli, R.; Saltalippi, C.; Moradian, M.; Ghalhari, G.A.F. Machine Learning to Estimate Surface Soil Moisture from Remote Sensing Data. *Water* **2020**, *12*, 3223. [[CrossRef](#)]
14. Filgueiras, R.; Mantovani, E.C.; Dias, S.H.B.; Fernandes Filho, E.I.; da Cunha, F.F.; Neale, C.M.U. New Approach to Determining the Surface Temperature without Thermal Band of Satellites. *Eur. J. Agron.* **2019**, *106*, 12–22. [[CrossRef](#)]
15. Granata, F. Evapotranspiration Evaluation Models Based on Machine Learning Algorithms—A Comparative Study. *Agric. Water Manag.* **2019**, *217*, 303–315. [[CrossRef](#)]
16. Tikhmarine, Y.; Malik, A.; Kumar, A.; Souag-Gamane, D.; Kisi, O. Estimation of Monthly Reference Evapotranspiration Using Novel Hybrid Machine Learning Approaches. *Hydrol. Sci. J.* **2019**, *64*, 1824–1842. [[CrossRef](#)]
17. Virnodkar, S.S.; Pachghare, V.K.; Patil, V.C.; Jha, S.K. Remote Sensing and Machine Learning for Crop Water Stress Determination in Various Crops: A Critical Review. *Precis. Agric.* **2020**, *21*, 1121–1155. [[CrossRef](#)]
18. Carter, C.; Liang, S. Evaluation of Ten Machine Learning Methods for Estimating Terrestrial Evapotranspiration from Remote Sensing. *Int. J. Appl. Earth Obs. Geoinf.* **2019**, *78*, 86–92. [[CrossRef](#)]
19. Xu, T.; Guo, Z.; Xia, Y.; Ferreira, V.G.; Liu, S.; Wang, K.; Yao, Y.; Zhang, X.; Zhao, C. Evaluation of Twelve Evapotranspiration Products from Machine Learning, Remote Sensing and Land Surface Models over Conterminous United States. *J. Hydrol.* **2019**, *578*, 124105. [[CrossRef](#)]
20. Chai, R.; Sun, S.; Chen, H.; Zhou, S. Changes in Reference Evapotranspiration over China during 1960–2012: Attributions and Relationships with Atmospheric Circulation. *Hydrol. Process.* **2018**, *32*, 3032–3048. [[CrossRef](#)]
21. Kadkhodazadeh, M.; Anaraki, M.V.; Morshed-Bozorgdel, A.; Farzin, S. A New Methodology for Reference Evapotranspiration Prediction and Uncertainty Analysis under Climate Change Conditions Based on Machine Learning, Multi Criteria Decision Making and Monte Carlo Methods. *Sustainability* **2022**, *14*, 2601. [[CrossRef](#)]
22. Raza, A.; Shoaib, M.; Faiz, M.A.; Baig, F.; Khan, M.M.; Ullah, M.K.; Zubair, M. Comparative Assessment of Reference Evapotranspiration Estimation Using Conventional Method and Machine Learning Algorithms in Four Climatic Regions. *J. Pure Appl. Geophys.* **2020**, *177*, 4479–4508. [[CrossRef](#)]
23. Tang, D.; Feng, Y.; Gong, D.; Hao, W.; Cui, N. Evaluation of Artificial Intelligence Models for Actual Crop Evapotranspiration Modeling in Mulched and Non-Mulched Maize Croplands. *Comput. Electron. Agric.* **2018**, *152*, 375–384. [[CrossRef](#)]
24. Alvares, C.A.; Stape, J.L.; Sentelhas, P.C.; De Moraes Gonçalves, J.L.; Sparovek, G. Köppen’s Climate Classification Map for Brazil. *Meteorol. Zeitschrift* **2013**, *22*, 711–728. [[CrossRef](#)]
25. Tasumi, M. *Progress in Operational Estimation of Regional Evapotranspiration Using Satellite Imagery*; University of Idaho: Moscow, ID, USA, 2003.
26. Olmedo, G.F.; Ortega-Farías, S.; Fonseca-Luengo, D.; de la Fuente-Sáiz, D.; Fuentes-peñailillo, F.; Munafó, M.V. Water: Actual Evapotranspiration with Energy Balance Models. Available online: <https://cran.r-project.org/web/packages/water/water.pdf> (accessed on 3 January 2021).
27. R Core Team. *R: A Language and Environment for Statistical Computing*, version 3.3.1; R Foundation for Statistical Computing: Vienna, Austria, 2019.
28. Hijmans, R.J.; van Etten, J.; Cheng, J.; Mattiuzzi, M.; Sumner, M.; Greenberg, J.A.; Lamigueiro, O.P.; Bevan, A.; Racine, E.B.; Shortridge, A.; et al. Raster: Geographic Data Analysis and Modeling. Available online: <https://cran.r-project.org/web/packages/raster/index.html> (accessed on 10 January 2021).
29. Fernandes Filho, E.I. Labgeo: Collection of Functions to Fit Models with Emphasis in Land Use and Soil Mapping. Available online: <https://rdr.io/github/elpidiofilho/labgeo/> (accessed on 8 January 2021).
30. Filgueiras, R.; Mantovani, E.C.; Althoff, D.; Fernandes Filho, E.I.; Cunha, F.F. da Crop NDVI Monitoring Based on Sentinel 1. *Remote Sens.* **2019**, *11*, 1441. [[CrossRef](#)]
31. Kuhn, M.; Wing, J.; Weston, S.; Williams, A.; Keefer, C.; Engelhardt, A.; Cooper, T.; Mayer, Z.; Kenkel, B.; Team, R.C. Classification and Regression Training. Available online: <https://cran.r-project.org/web/packages/caret/index.html> (accessed on 8 January 2021).
32. Allen, R.G.; Walter, I.A.; Elliott, R.; Howell, T.; Itenfisu, D.; Jensen, M. *The ASCE Standardized Reference Evapotranspiration Equation*; American Society of Civil Engineers: Reston, VA, USA, 2005.
33. Ekström, C.T. MESS: Miscellaneous Esoteric Statistical Scripts. Available online: <https://cran.r-project.org/web/packages/MESS/index.html> (accessed on 8 January 2021).
34. Ke, Y.; Im, J.; Park, S.; Gong, H. Spatiotemporal Downscaling Approaches for Monitoring 8-Day 30 m Actual Evapotranspiration. *ISPRS J. Photogramm. Remote Sens.* **2017**, *126*, 79–93. [[CrossRef](#)]
35. Dingre, S.K.; Gorantiwar, S.D. Determination of the Water Requirement and Crop Coefficient Values of Sugarcane by Field Water Balance Method in Semiarid Region. *Agric. Water Manag.* **2020**, *232*, 106042. [[CrossRef](#)]
36. da Silva, T.G.F.; de Moura, M.S.B.; Zolnier, S.; Soares, J.M.; Vieira, V.J.d.S.; Júnior, W.G.F. Water Requirement and Crop Coefficient of Irrigated Sugarcane in a Semi-Arid Region. *Rev. Bras. Eng. Agric. Ambient.* **2012**, *16*, 64–71. [[CrossRef](#)]
37. Saleem, A.; Awange, J.L. Coastline Shift Analysis in Data Deficient Regions: Exploiting the High Spatio-Temporal Resolution Sentinel-2 Products. *CATENA* **2019**, *179*, 6–19. [[CrossRef](#)]
38. de Silva, V.D.P.; Garcêz, S.L.A.; Silva, B.B.D.; De Albuquerque, M.F.; Almeida, R.S.R. Métodos de Estimativa Da Evapotranspiração Da Cultura Da Cana-de-Açúcar Em Condições de Sequeiro. *Rev. Bras. Eng. Agric. Ambient* **2015**, *19*, 411–417. [[CrossRef](#)]

39. Li, D.; Li, W.; Zhang, H.; Zhang, X.; Zhuang, J.; Liu, Y.; Hu, C.; Lei, B. Far-Red Carbon Dots as Efficient Light-Harvesting Agents for Enhanced Photosynthesis. *ACS Appl. Mater. Interfaces* **2020**, *12*, 21009–21019. [[CrossRef](#)]
40. Wang, Y.; Xie, Z.; Wang, X.; Peng, X.; Zheng, J. Fluorescent Carbon-Dots Enhance Light Harvesting and Photosynthesis by Overexpressing PsbP and PsbK Genes. *J. Nanobiotechnol.* **2021**, *19*, 260. [[CrossRef](#)]
41. Ponzoni, F.J.; Shimabukuro, Y.E.; Kuplich, T.M. *Sensoriamento Remoto Da Vegetação*, 2nd ed.; Oficina de Textos: São Paulo, Brazil, 2012; ISBN 978-85-7975-053-3.
42. dos Santos, N.V.; Demattê, J.A.M.; Silvero, N.E.Q. Improving the Monitoring of Sugarcane Residues in a Tropical Environment Based on Laboratory and Sentinel-2 Data. *Int. J. Remote Sens.* **2020**, *42*, 1768–1784. [[CrossRef](#)]
43. Chandel, N.S.; Rajwade, Y.A.; Golhani, K.; Tiwari, P.S.; Dubey, K.; Jat, D. Canopy Spectral Reflectance for Crop Water Stress Assessment in Wheat (*Triticum aestivum*, L.). *Irrig. Drain.* **2020**, *70*, 321–331. [[CrossRef](#)]
44. Chen, H.; Wang, P.; Li, J.; Zhang, J.; Zhong, L. Canopy Spectral Reflectance Feature and Leaf Water Potential of Sugarcane Inversion. *Phys. Procedia* **2012**, *25*, 595–600. [[CrossRef](#)]
45. Gates, D.M.; Keegan, H.J.; Schleter, J.C.; Weidner, V.R. Spectral Properties of Plants. *Appl. Opt.* **1965**, *4*, 11. [[CrossRef](#)]
46. Muller, S.J.; Sithole, P.; Singels, A.; Van Niekerk, A. Assessing the Fidelity of Landsat-Based FAPAR Models in Two Diverse Sugarcane Growing Regions. *Comput. Electron. Agric.* **2020**, *170*, 105248. [[CrossRef](#)]
47. Zhang, F.; Zhou, G. Estimation of Vegetation Water Content Using Hyperspectral Vegetation Indices: A Comparison of Crop Water Indicators in Response to Water Stress Treatments for Summer Maize. *BMC Ecol.* **2019**, *19*, 18. [[CrossRef](#)] [[PubMed](#)]
48. Jamshidi, S.; Zand-Parsa, S.; Pakparvar, M.; Niyogi, D. Evaluation of Evapotranspiration over a Semiarid Region Using Multiresolution Data Sources. *J. Hydrometeorol.* **2019**, *20*, 947–964. [[CrossRef](#)]
49. Nisa, Z.; Khan, M.S.; Govind, A.; Marchetti, M.; Lasserre, B.; Magliulo, E.; Manco, A. Evaluation of SEBS, METRIC-EEFlux, and QWaterModel Actual Evapotranspiration for a Mediterranean Cropping System in Southern Italy. *Agronomy* **2021**, *11*, 345. [[CrossRef](#)]
50. Ortega-Salazar, S.; Ortega-Farías, S.; Kilic, A.; Allen, R. Performance of the METRIC Model for Mapping Energy Balance Components and Actual Evapotranspiration over a Superintensive Drip-Irrigated Olive Orchard. *Agric. Water Manag.* **2021**, *251*, 106861. [[CrossRef](#)]