Article

Ensemble Machine Learning Outperforms Empirical Equations for the Ground Heat Flux Estimation with Remote Sensing Data

Josep Bonsoms [1] **and Gilles Boulet** [2,*]

[1] Department of Geography, Universitat de Barcelona, Barcelona, Catalonia, Spain.
[2] Centre d'Etudes Spatiales de la Biosphère (CESBIO), Université de Toulouse, CNES, CNRS, IRD, UPS, IN-RAE, Toulouse, France.
*   gilles.boulet@ird.fr

Abstract: Estimating evapotranspiration at field scale is a major component of sustainable water management. Due to the difficulty to assess some major unknowns of the water cycle at that scale, including irrigation amounts, evapotranspiration is often computed as the residual of the instantaneous surface energy budget. One of the Surface Energy Balance components with the largest uncertainties in their quantification over bare soils and sparse vegetation areas is the ground heat flux (G). Over the last decades, the estimation of G with RS data has been mainly achieved with empirical equations, on the basis of the G and net radiation (Rn) ratio, G/Rn. G/Rn empirical equations generally require vegetation data (Type I empirical equations), in combination with surface temperature (Ts) and albedo (Type II empirical equations). In this article we aim to evaluate the estimation of G with RS. For the first time, we compare eight G/Rn empirical equations against two types of machine learning (ML) methods: an ensemble ML type, the Random Forest (RF), and the Neural Networks (NN). The comparison of each method is evaluated over dense dataset, including a wide range of climate and land covers, with data of Eddy-Covariance towers extended along the mid-latitude area that encompass the European and African continent. Our results have shown evidence that the driver of G in bare soils and sparse vegetation areas (Fraction of Vegetation, Fv <= 0.25) is Ts, instead of vegetation greenness indexes. On the other hand, the estimation of G with Rn, Ts or Fv decreases at dense vegetation areas (Fv >= 0.50). There are not significant differences between the most accurate type I and II empirical equations. For bare soils and sparse vegetation areas the empirical equation that better estimates G is E8, which combines the Leaf Area Index (LAI) and Ts. In dense vegetation areas (Fv >= 0.25), an exponential empirical equation based on Fv (E4), shows the best performance. However, ML better estimates G than the empirical equations, independently of the Fv ranges. A RF model with Rn, LAI and Ts as predictor variables shows the best accuracy and performance metrics, outperforming the NN model.

Keywords: Ground heat flux; Machine learning; Remote sensing; Surface Energy Balance

1. Introduction

Subtropical and Mid-latitude regions such as the Mediterranean area has been identified as one of the climates hot-spots of the earth [e.g., 1]. Since the 1980s, in the N of Africa and S of Europe, the expansion of the Hadley cell during the warm half of the year has caused a poleward shift in the tropical high-pressures systems; triggering a decrease of precipitation and an increase of the drought frequency during the warm half of the year [2]. Positive anomalies of the anticyclonic weather types in the subtropical belt during the warm half of the year have provoked a northward migration of mid-latitude climate types [3]. In addition, increases in temperature and atmospheric water demand triggered an increase of the drought severity [4]. By the mid-end of the 21st century, climate projections for mid-latitude areas project an upward trend of temperature and heat waves [5]. In combination with the projected decreases of precipitation [e.g., 6], it will trigger an increase of

aridity and drought episodes [7]. Dwindling water resources has direct consequences in the ecological as well as socioeconomic sphere [8] and the changing climate scenarios pointed out require of further efforts for the improvement of the water estimation. In water-limited areas, evapotranspiration (ET), the flux of water returned back to the atmosphere from the soil (evaporation) or crop canopy (transpiration), is the main relevant negative flux (loss) of the hydrological cycle [9]. Generally, the estimation of ET with Remote Sensing data (RS) is performed with single-source energy and dual-source energy balance models. For the latter case, the vegetation and the soil are analyzed individually based on their respective surface energy balance (SEB). With RS data, the estimation of ET is obtained as the residual term of the SEB, or in other words, the difference between the SEB heat fluxes (Rn = LE – H – G); where Rn is the net radiation, LE is the latent heat flux, H is the sensible heat flux and G is the energy absorbed or released at the soil surface (all in W/m^2). The latter, G, is one of the SEB components most difficult to quantify [10]. Hence, further efforts for the bias reduction of the G are required in order to improve the energy fluxes closure, and in the end, provide a better estimation of ET.

The estimation of G at local scales is generally performed with the Harmonic method, based on the Fourier series [e.g., 11-16]. However, the Harmonic method cannot estimate the G for large areas, since it requires soil properties data (i.e., soil conductivity and temperature, at different soil depths), which are only measured at specific sites; usually with other SEB fluxes, in Eddy-Covariance (EC) towers. RS data can overcome the spatial limitations mentioned, providing an estimation of G with acceptable temporal and spatial resolution. For this reason, this article is focused only on the estimation of G with RS data. Over the last decades, the scientific literature has proposed several formulations for the estimation of G with RS data. The proposed methods could be grouped on two groups: (i) an adaptation of the Harmonic method, with an estimation of surface temperature (Ts) based on radiometric brightness, usually Meteosat data, but with a coarse spatial-resolution (ca. 3 km) limitation [17-18]. On the other hand, for sun-synchronous satellites with a single daytime overpass, the estimation of G is also performed with (ii) the ratio of G and Rn, the so-called G/Rn ratio [e.g., 19-20]. In fact, several empirical equations have been applied for the estimation of the G/Rn. For instance, the ratio can be calculated following a sinusoidal function depending on the hour of the day and the maximum G/Rn observed [21] or with meteorological parameters, such as the wind [22]. However, the most common approach is the estimation of the G/Rn with vegetation greenness data. The G/Rn is usually calculated with vegetation indexes, such as the well-known Normalized Difference Vegetation Index (NDVI) [23], the Leaf Area Index (LAI) [24], the Fraction of Vegetation (Fv) [25], the Modified Soil Adjusted Index (MSAVI) [26], or determined as a constant depending on vegetation height [27]. Finally, the G/Rn can also be estimated with the vegetation indexes named, in combination with Ts [28] and albedo (α) data [29]. The accuracy of the different empirical equations based on the G/Rn ratio for the estimation of G have been analyzed in previous works [e.g., 30-31], showing large uncertainties, especially in high canopies covers. Therefore, other methodologies should be proposed for the estimation of G.

In the last decades machine learning (ML) methods have been applied in the environmental science, generally obtaining accurate results. Nevertheless, up the date, only [32] modelled the G with RS data and the Artificial Neural Networks (ANN) ML algorithm. Furthermore, only [33] provided a comparison of the ANN model against two G/Rn empirical equations. To our best of knowledge, no study compared the performance of ensemble ML models, such as boosted regression trees (e.g., Random Forest (RF)), against the results obtained with neural networks (NN), and the ones obtained with the calibrated version of the G/Rn empirical equations. In addition, an estimation of G with the above-mentioned methods, evaluated over a dense dataset of EC data in the tropical and mid-latitude area is still lacking. In this article we aim to address these knowledge gaps. We provide the first systematic evaluation of the estimation of G with RF, NN and eight G/Rn empirical equations, evaluated by vegetation ranges. The analysis is performed in a

climate hot-spot area (e.g., Mediterranean basin), with a dense dataset of G measurements acquired at several EC towers, and including a wide range of land covers and climate types.

The manuscript is structured as follows: in section 2 we present a description of the data, the geographical settings of each EC, the methods implemented together with the evaluation metrics used. Subsequently, in section 3 the results are presented and discussed. Finally, the conclusions are summarized in section 4.

2. Materials and Methods

In this section we present a detailed description of the meteorological and vegetation data used in this work together with the methodology followed.

2.1. Materials

We used daily records of vegetation greenness data (LAI) and 30 minutes records of meteorological data (Ts, Rn, G and α) at 10:30 a.m and 13:30 p.m, corresponding to typical sun-synchronous satellite overpass times. Meteorological records without data were excluded of the analysis, whereas days without LAI data were gap filled with the broadly used local polynomial regression (LOESS) [31]. The dataset has been data quality checked and previously analyzed at [34], where further technical details can be found. The dataset includes EC records from different climate areas, including oceanic, semiarid and arid climate types. Barbeau (Bar) is located in the SE of Paris (N of France), in an oak-forest site (Fv > 0.95; Figure 1). Bar, as well as Auradé (Aur) and Lamasquère (Lam) EC sites, is under the Cfb climate influence. Aur and Lam are situated in the SW of Toulouse (SW of France), in an area with mid to high LAI values (ca. 0.5; Figure 1) with crops of wheat and sunflower. On the other hand, Avignon (Avi) EC is in the NE of Montpellier (SE of France), ruled by a Mediterranean climate type (Csa), including a wide range of crops types (depending on the season), such as peas, wheat and sorghum crops. Moving into the S, Kairouan (Kai) is an EC surrounded by crops of wheat in a semi-arid region (Csa climate type) in the S of Tunis (Tunisia). Haouz (Hou), in the S of Marrakesh (Morocco), is situated in an irrigated land zone of wheat crops, governed by a mid-latitude steppe-climate type (BSk). Finally, Agofou (Ago) and Wankama (Wan), are two sites located in the E of Niamey (Níger), and ruled by a mid-latitude steppe and desert (BSh) climate type. Ago is a millet crop site, whilst in Wan two EC are installed. One EC is located on the savannah, whereas another one is in a zone of millet crops. The main geographical characteristics of the EC towers included in this work are summarized in Table 1.

Table 1. Main geographical characteristics and data length of the EC towers.

EC	Code	Country	Year	Köppen climate type	Ecosystem
	Ago	Níger	2009	BSh (Mid-Latitude steppe and desert climate)	Millet
Auradé	Aur	France	From 2006 to 2013	Cfb (Marine West Coast)	Wheat, sunflower
Avignon	Avi	France	From 2005 to 2013	Csa (Mediterranean)	Peas, wheat, sorghum and sunflower
Barbeau	Bar	France	From 2014 to 2015	Cfb (Marine West Coast)	Oak forest
Haouz	Hou	Morocco	2004	BSk (Mid-Latitude steppe climate)	Wheat
Kairouan	Kai	Tunisia	From 2012 to 2015	Csa (Mediterranean)	Olive and wheat
Lamasquère	Lam	France	From 2007 to 2013	Cfb (Marine West Coast)	Wheat
Wankama	War	Níger	2009	BSh (Mid-Latitude steppe and desert climate)	Savannah and millet

2.2 Methods

2.2.1 Estimation of vegetation indexes.

The LAI values measured at each EC were converted into the other vegetation indexes, given that no in-situ measurements of NDVI and Fv were available for the whole dataset. We calculated the NDVI with the method presented at [35], successfully applied for all types of crops [e.g., 36]:

$$\text{LAI} = -\frac{1}{k}\ln\left(\frac{\text{NDVI}_{\infty} - \text{NDVI}}{\text{NDVI}_{\infty} - \text{NDVI}_{\text{soil}}}\right)$$

(1)

$$\text{NDVI} = \left(\left(\exp\left(\frac{\text{LAI}}{\text{NDVI}_{\infty} - \text{NDVI}_{\text{soil}}}\right) * 0.80\right) - 0.94\right) - 1$$

(2)

Where k is the calibrated extinction, set to 1.13 [36]. NDVI_{∞} is the NDVI value of the fully developed canopy, whereas $\text{NDVI}_{\text{soil}}$ is the NDVI of bare soils. According to previous literature, the NDVI_{∞} was set to 0.97 [36]. We tested different $\text{NDVI}_{\text{soil}}$ values, ranging from 0 to 0.2, but no significant G/Rn differences were found. Therefore the $\text{NDVI}_{\text{soil}}$ was set to 0.1 [30].

The Fv was calculated with the semi-empirical method detailed at [37], where Fv is performed with the following exponential function of LAI:

$$\text{Fv} = 1 - \exp(-b \text{ x F(o) } \text{ x I x LAI}$$

(3)

Where b, F(o) and I are constants, corresponding to -0.945, 0.5 and 1, respectively.

2.2.2 Empirical equations and Machine Learning algorithms

The empirical equations included in this work are divided into type I and type II. Type I gathers the empirical equations that estimate G with only vegetation indexes. This includes E1, E2 and E3 (based on NDVI), E4 and E5 (Fv), and finally E6 (LAI). Type II includes E7, which estimates G with Ts and LAI, and E8 which estimates G with NDVI, Ts and α. A detailed description of the empirical equations can be found in Table 2. Regarding the ML algorithms, NN I and RF I (ML I) are trained with LAI data, providing a fair comparison with type I equations. NN II and RF II (ML II) are trained with LAI and Ts. Last, NN III and RF III (ML III) are trained with LAI, Ts and Rn.

Table 2. Empirical equations and ML algorithms used for the estimation of the G.

Empirical equations and ML code	Empirical equation	Calibrated values	Reference
E1	a – b x NDVI	a = 0.3, b= 0.26	[18]
E2	a x exp(– b x NDVI)	a = 0.39, b= 1.952	[10]
E3	a x NDVI + b	a = 0.47, b= 0.43	[38]
E4	a x (1- Fv)	a = 0.23	[24]
E5	-a + (1–Fv) x (b–c)	a = 0.015, b = 0.315 c = 0.064	[10]
E6	a x exp(- b x LAI)	a = 0.23, b =0.45	[39]
E7	If LAI < 0.5 = $-\dfrac{a(Ts - 273.15)}{Rn + b}$ If LAI > 0.5 = $c + d * \exp(-e\ x\ LAI)$	a = 1.7, b = 0.079 c = 0.05, d = 0.12 e = 0.621	[27]
E8	Ts x Rn x 0.0038 + 0.074 x α x (1-0.98 x NDVI^4)	a =0.06 , b =0.012, c = 0.978	[19]
NN	**Size** From 10 to 30, increments of 5	20	[40]
	Decay From 0.001 to 0.2, increments of 0.05	0.15	
RF	**Ntree** 50,100,200,400 and 800	800	[41]
	Mtry From 1 to 15, increments of 1	1	

2.2.2.1. Neural Networks (NN)

The NN algorithm creates a distributed system of neurons with non-linear functions and several layers interconnected for the prediction of one variable, in this case the G. In this work we tuned the two principal hyper-values of the NN, named size and decay (Table 2). The parametrization of the NN size ranged from 10 to 30, with increments of 5; while the parametrization of the decay ranged between 0.001 and 0.2, with increments of 0.05. An accurate description of the mathematical formulation of the NN can be found at [40].

2.2.2.1. Random Forest (RF)

The RF is a non-parametric method based on decision trees and bootstrap aggregation [41]. A group of trees is developed using multiple decision trees trained from random subsets of the dataset. In this work we applied the regression version of the RF, where the estimated values are the result of the arithmetic average of all the tree predictions (M). The trees $\{M_1(X), M_1(X), ..., M_K(X)\}$, where X = {x1, x2, …, xβ}, being β-dimension input vector that creates a forest. The ensembles generate P values that correspond to the tree Y_p (p = 1, 2, …, P).

The RF regression is calculated as follows:

$$f(x) = \frac{1}{M}\sum_{m=1}^{M} M_k(x_j) \qquad (4)$$

The tunning of the regression RF has been analyzed at [42]. In their study they conclude that the number of decision trees randomly selected of the model, the subset of training samples, together with the variables split at each tree node, prevents the algorithm to overfitting. Two hyper-values have been tuned of the RF. The number of predictors evaluated in each node (mtry), and the number of trees (ntree) of a randomly sample. In this work, the mtry was set as 1 and the ntree was equal to 800.

We used a grid-search and a 10-fold cross validation for validate the hyper-values of both of the ML algorithms. We applied the RF with the randomForest package [43] and the NN with the Caret package [44], both of them of R Studio [45].

2.3. Evaluation

The initial dataset has been randomly split into two groups, the training dataset (conforming the 70% of the dataset) and the testing dataset (which constitutes the remaining 30%). Hence, the results obtained with the calibrated empirical equations and the trained ML algorithms are comparable, given that the evaluation and calibration of all the methods has been performed under the same conditions. In the training dataset, the empirical equations have been calibrated and the ML algorithms tuned. Table 2 shows the different ML hypervalues tested and the values set, as well as the calibrated constant values of the empirical equations. Subsequently, the trained ML algorithms and the calibrated empirical equations have been applied in the independent and testing dataset. We also evaluated the accuracy and performance of each method by Fv ranges, from 0 to 1, by steps 0.25. The accuracy and performance of the models have been evaluated using three types of statistical metrics: the residual (Eq. 5), the mean absolute error (MAE; Eq. 6), the Root Mean Square Error (RMSE; Eq. 7), the coefficient of determination (R²; Eq. 8) and the Willmott's D (D; Eq. 9).

$$\text{Residual} = \ y_i - y \quad (5)$$

$$\text{MAE} = \ \frac{1}{N}\Sigma_{i=1}^{N}|y_i - \hat{y}| \quad (6)$$

$$\text{RMSE} = \ \sqrt{\frac{1}{N}\Sigma_{i=1}^{N}(y_i - \hat{y})^2} \quad (7)$$

$$\text{R}^2 = \ 1 - \frac{\Sigma(y_i - \hat{y})^2}{\Sigma(y_i - \bar{y})^2} \quad (8)$$

$$\text{Willmott's D} = 1 \ - \sqrt{\frac{\Sigma_{i=1}^{N}(y_i - y)^2}{\Sigma_{i=1}^{N}(|y_i| - |y_i|)^2}} \quad (9)$$

where N is the number of the samples, y_i is the observed value, $\hat{y}$ is the estimated and $\bar{y}$ represents the average value of the estimated values. The residual is the difference between y_i and $\hat{y}$. The MAE and RMSE summarise the mean differences between the predicted values and the observed values. Low RMSE and MAE values are inversely related with the high accuracy of the method proposed. The RMSE is sensitive to high values or outliers [46], therefore it can be used as an indicator of the magnitude of extreme errors. On the other hand, high values of D and R² are related with a high performance of the methods.

3. Results and discussion

In the first section we determine which are the meteorological and vegetation drivers of G. Subsequently, we analyze the observed and estimated values of the heat flux. Then, we provide a detailed analysis of the accuracy and performance metrics of all the methods, together with an evaluation of the methods by Fv ranges. The results are discussed and compared with previous works. Finally, we elaborate about the reason of the good accuracy and performance metrics obtained with ML.

3.1. Characterization and drivers of G

The theoretical G values estimated by the empirical equations are presented in Figure 1a. Both of the empirical equation's types, based only on vegetation indexes (type I); as well as on vegetation indexes, ancillary remote sensing and meteorological data (type II)

estimate an overall decrease of the G depending on Fv. For mid to high vegetation ranges (Fv > 0.25), the E1, E2, E4, E5, E6 and E7 show a parallel shape for the G estimation, and hence similar values can be expected. The highest difference between equations is found between E3 and E8. Whilst the equations have a curve shape for Fv > 0.4, for bare soils and sparse vegetation areas (Fv <= 0.25) E8 assumes a constant G value, and E3 estimates G as an exponential function of Fv (Figure 1b).
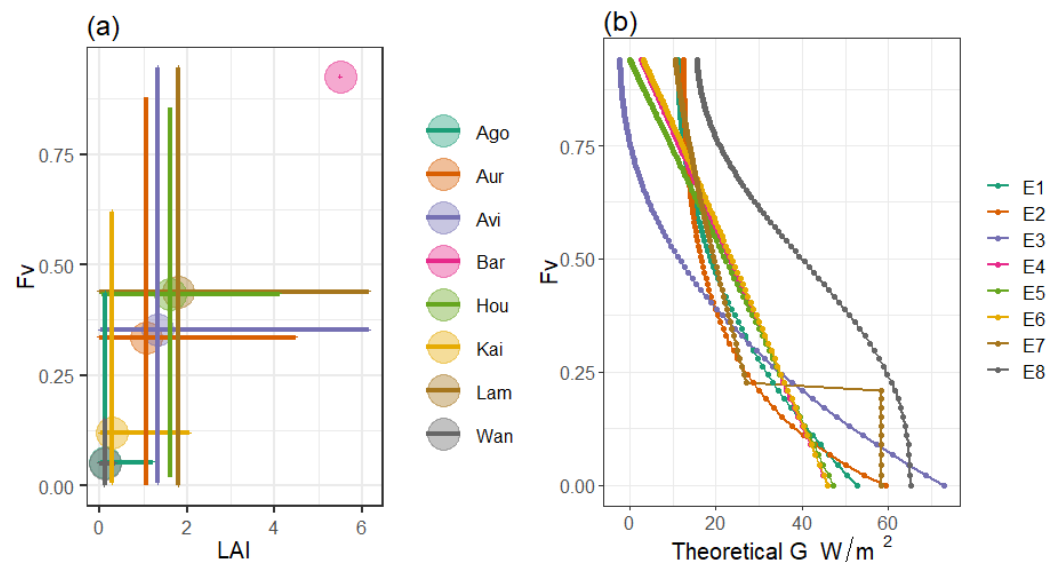

Figure 1. (a) Maximum, minimum and mean (a) Fv values (y axis) and LAI (x axis) of the EC included in this work. (b) Theoretical shape of the empirical equations based on vegetation indexes. E7 and E8 do not only rely on vegetation data. We assumed the other input parameters as constant. The Ts was = 25°C, α = 0.4 and Rn = 200 W/m².
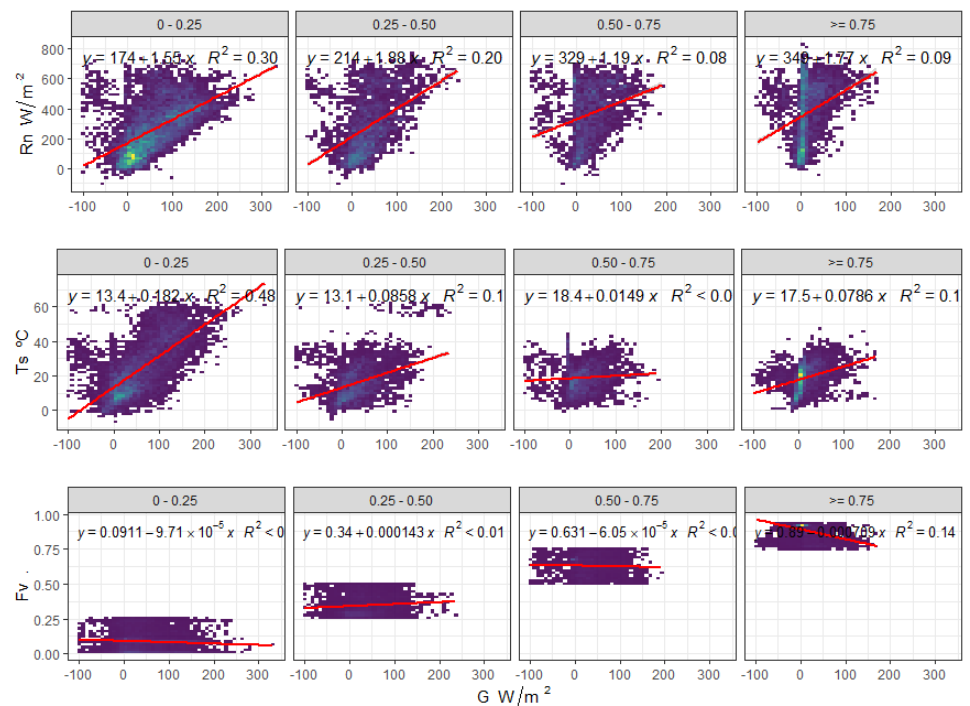

Figure 2. Regression analysis of the observed G (x axis) and the main drivers of G (y axis) grouped by Fv ranges. The first row corresponds to Rn, the second to Ts and the last one to Fv.

Figure 2 shows evidence that the different drivers of G depend on the Fv range. For bare soils and sparse vegetation areas, the best parameter for estimate G is Ts (R^2 = 0.48), followed by Rn (R^2 = 0.30). Over dense vegetated areas and high canopies (Fv >= 0.50), the minimum G values are observed (ca. 5 W/m²; Figure 2 and 3). In many days, the G can be neglected or be assumed to practically equal to 0 [19]. This is because high vegetation-covers thermal insulates the surface, reducing significantly the solar radiation that reaches the ground [48-49], and buffering temperature gradients in areas with significant rates of ET [31]. Given that the G variability under dense vegetated areas and high canopies is not mainly ruled by Rn and Ts, the incorporation of these variables as a predictor variable is redundant.
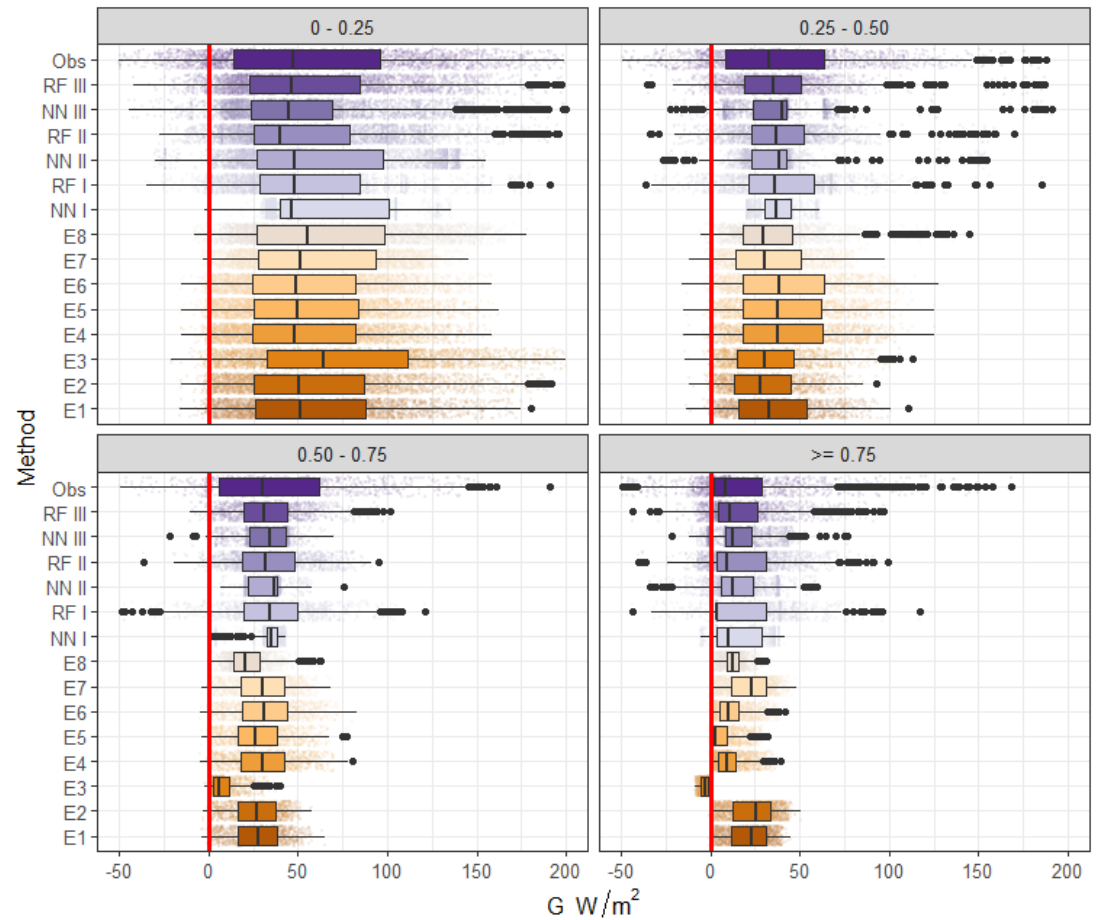

Figure 3. Box plot of the estimated G values (W/m²), grouped by the method (y axis), and Fv ranges.

The observed and estimated G values by the different empirical equations and ML algorithms are shown in Figure 3 and 4. The observed mean value of G ranged from 59 to 18 W/m². In bare soils and sparse vegetation areas, where the majority of the data is found (Figure 1a), the highest G values are measured in accordance with previous works [e.g., 21]. In these land types, characteristics of arid and semiarid climates zones, the highest values of G are generally observed during the warmest months of the year [31]. The G rules a significant part of the SEB hourly variability, with instantaneous G values ranging from the 5-10% up to 50% of the Rn [30,38,47,48, 49].

3.2. Analysis of the accuracy and performance of the different methods used for the estimation of G

The calibration of the constant values slightly improved the type I equations accuracy, but negligible difference is found for type II equations (Figure 4). Regarding the ML

models, no significant differences of the accuracy metrics have been observed between the training and the testing dataset, providing evidence of the lack of overfitting. In addition, negligible differences have been found between the different hypervalues tested during the ML tunning process (Table S1).
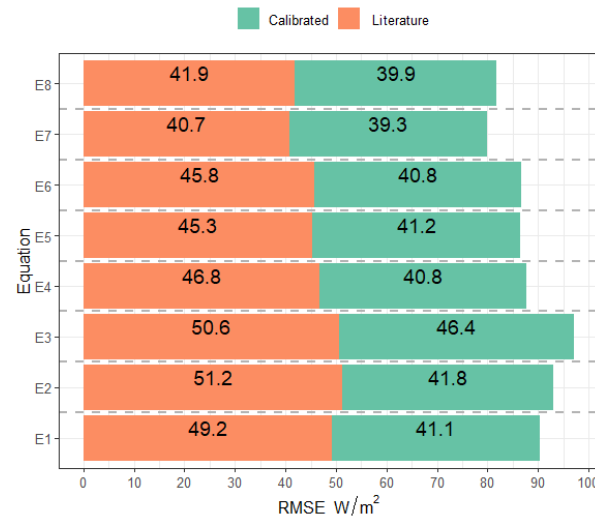

Figure 4. RMSE values observed with the constant values calibrated and proposed by the literature.

The majority of the empirical equations have shown a systematic underestimation of G, specially E3 for Fv >= 0.75. The highest bias is observed with the empirical equations that conform type I. The residual term (the difference between the estimated and the observed values) of type I increases with Fv, except for E4 and E6. On the other hand, type II overestimates G for bare soils and sparse vegetation areas (5 W/m²), and underestimates the heat flux for the other Fv ranges. The residual term for type I is larger at dense vegetated areas and high canopies covers (> 5 W/m²) than at bare soils and sparse vegetation areas, and therefore the data suggest that type I equations are not able to reproduce the decrease of G depending on Fv. This is because type I equations assumes a constant value for all the Fv ranges. In contrast, other empirical equations such as E7, use a threshold as a function of LAI, determining two empirical equations for the estimation of G (Table 2) and in the end providing better results.
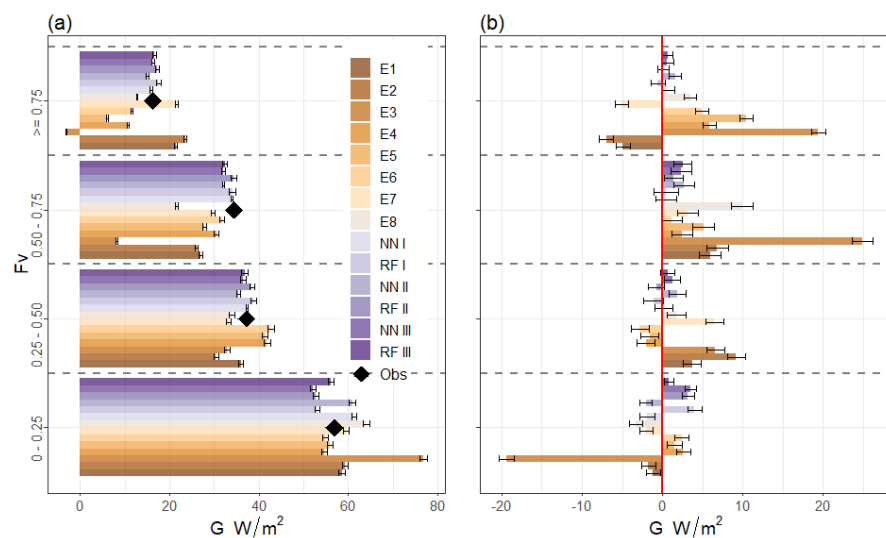

Figure 5. (a) Bar plot of the mean observed G (black points) and estimated (colour bars) values, grouped by Fv ranges. (b) Bar plot of the residual term (difference between the estimated and the observed G values).

The estimation of G with ML has shown lower residual differences than the obtained with the empirical equations. In fact, the residual term with ML is practically the same (avg. ca. 5 W/m^2) for the majority of Fv ranges. In addition, no systematic biases are estimated with ML II and ML III. Furthermore, ML II and III are able to reproduce the highest (percentile 90th) G values (Figure 3). The validation metrics by Fv ranges are shown at Table 3, Figure 6 and 7. The results provides evidence that the estimation of G based on type I empirical equations leads high uncertainties in comparison with the ML methods. The error metrics for the majority of the type I empirical equations are practically the same under bare soils and scarce vegetated areas (MAE = 36 and RMSE = 51 W/m^2). Nevertheless, the highest differences between type I equations are observed at dense vegetation areas. The accuracy of E1, E2 and E3 is worse than the group conformed by E4, E5 and E6 (avg. MAE 24 vs 18 W/m^2), confirming the E1, E2 and E3 systematic underestimation of the G. Type II improve the estimation of G to some degree. The accuracy of E8 (MAE = 28, RMSE = 39 W/m^2) is slightly better than the E7 (MAE = 28, RMSE = 37 W/m^2), but two type I equations (E4 and E6) outperformed all the empirical equations for the estimation of G at dense and high canopies covers. Therefore, these results suggest that the estimation of G is more accurate with Fv data (E4 and E5) and LAI (E6) than with NDVI (E1, E2 and E3). The results presented in this article are in accordance with previous studies that estimated the G based on vegetation indexes [e.g., 19, 31]. The lack of difference between E7 and E8 suggest that the inclusion of α in the empirical formulations that already include Ts does not substantially improves the estimation of G. This is probably explained because of the collinearity existent between Ts and α.

Table 3. Evaluation metrics recorded for every empirical equation, grouped by Fv ranges.

| Metric	FV	E1	E2	E3	E4	E5	E6	E7	E8	NN I	RF I	NN II	RF II	NN III	RF III
MAE	0-0.25	36.6	37.3	39.8	36.6	36.5	36.6	31	31.6	40.7	39	30.4	29.9	29	23.6
	0.25-0.50	30.4	31.2	31.1	30.4	30.3	30.4	30.7	31.8	34.3	37.8	31.9	30.7	29.6	24.5
	0.50-0.75	30.7	30.7	35.9	31	31.2	31	30.7	31.6	31.9	36.2	32.4	29.1	30.9	25.9
	>= 0.75	22.2	23	26	18.1	19.6	18	21.6	19.5	17.8	19.7	16.4	15.8	16.6	13.9
RMSE	0-0.25	49.7	50.9	56.9	49.2	49.3	49.2	43	43.4	52.0	51.4	41.6	41.3	39.8	33.9
	0.25-0.50	41	42	41.9	41.2	41.1	41.3	41.4	42.6	44.5	48.7	41.8	40.7	39.2	33.6
	0.50-0.75	41.8	41.9	49.3	41.9	42.2	41.9	41.6	43.4	42.5	48.4	43.1	39.7	41.7	36.4
	>= 0.75	31.3	32	37.9	30	31.5	29.7	30.9	30.6	29.5	33.2	28.5	28.3	28	25.4
D	0-0.25	0.6	0.5	0.6	0.5	0.5	0.5	0.6	0.6	0.4	0.5	0.6	0.6	0.6	0.7
	0.25-0.50	0.5	0.4	0.4	0.5	0.5	0.5	0.4	0.4	0.2	0.3	0.3	0.4	0.4	0.6
	0.50-0.75	0.3	0.3	0.4	0.3	0.3	0.3	0.3	0.3	0.2	0.3	0.2	0.4	0.3	0.5
	>= 0.75	0.3	0.3	0.4	0.4	0.4	0.4	0.4	0.3	0.5	0.5	0.5	0.6	0.5	0.6
R²	0-0.25	0.2	0.2	0.3	0.2	0.2	0.2	0.3	0.3	0.1	0.1	0.4	0.4	0.4	0.5
	0.25-0.50	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0	0	0	0.1	0.1	0.3
	0.50-0.75	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0.2
	>= 0.75	0	0	0	0.1	0	0.1	0	0	0.1	0	0.1	0.1	0.1	0.2
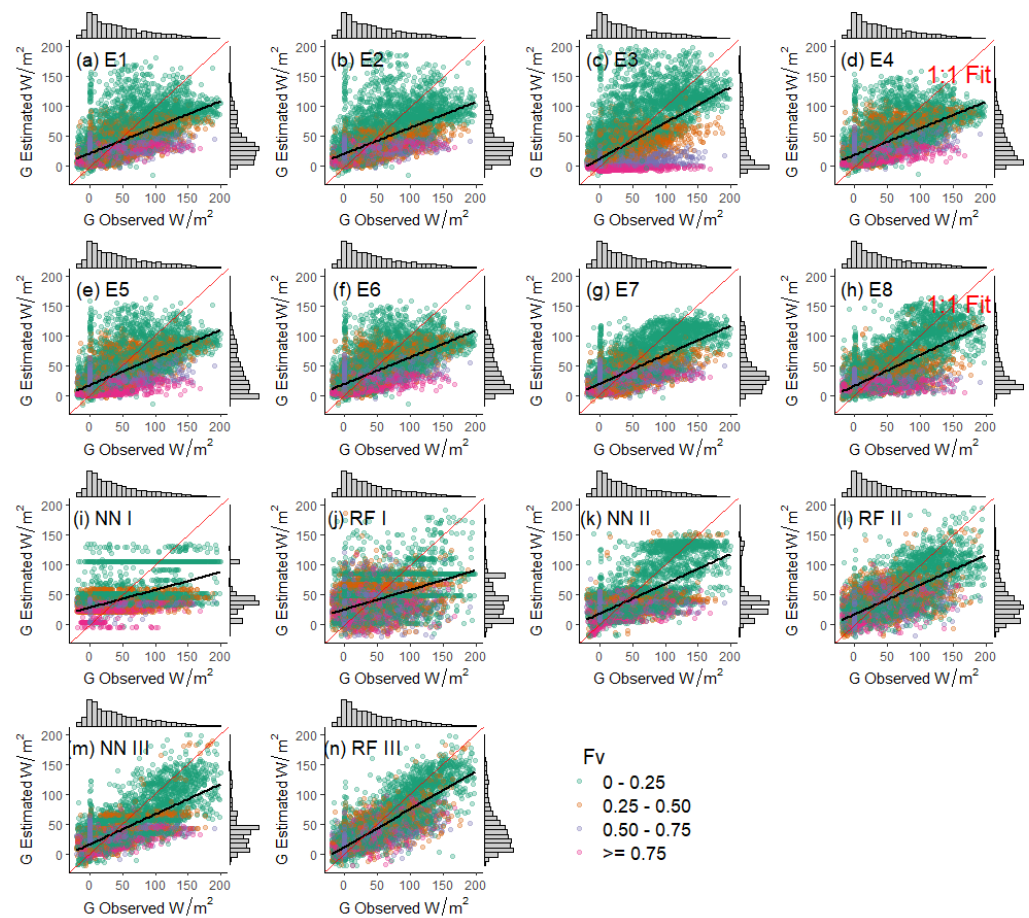
Figure 6. Probability density function (grey bars) and regression analysis of the G values observed (x axis) and estimated (y axis), grouped by method (boxes) and Fv (colors).

ML I, which only includes LAI as a predictor variable, shows similar accuracy metrics as type I and hence a low performance for the estimation of G. ML II, which includes LAI and Ts as predictor variables, outperformed all the empirical equations. Moreover, ML III, including LAI, Ts and Rn, significantly improved all the estimations of G and showed the best accuracy and performance (Figure 5 to 7). In particular, RF II shows slightly better results than NN II. Yet, the difference between algorithms increases when Rn is included in the regression models (Table 3). Except for E4, in bare soils and sparse vegetation areas, type I error is almost the double (MAE = 39.8 W/m²) than the measured with the RF III (MAE = 23.6 W/m²). Furthermore, the performance of RF III by Fv ranges are more constant than the other methods. The minimum errors are observed at bare soils and sparse vegetation areas (D = 0.7 and R^2 = 0.5), which could be explained because the majority of the data is found in this Fv range. In dense vegetated areas, the RF III performed also better than all the other methods. For instance, at Fv > 0.5, the R^2 values of RF III are 0.3, whereas the other methods rarely reach a R^2 of 0.1 (Figure 7). However, NN III and RF III do not significantly improve the large uncertainty in the highest vegetation range (Fv >= 0.75). For NN III, the D and R^2 is only 0.3 and 0.1, respectively. RF III shows a slightly better performance, with a D and a R^2 of 0.3 and 0.2, respectively.
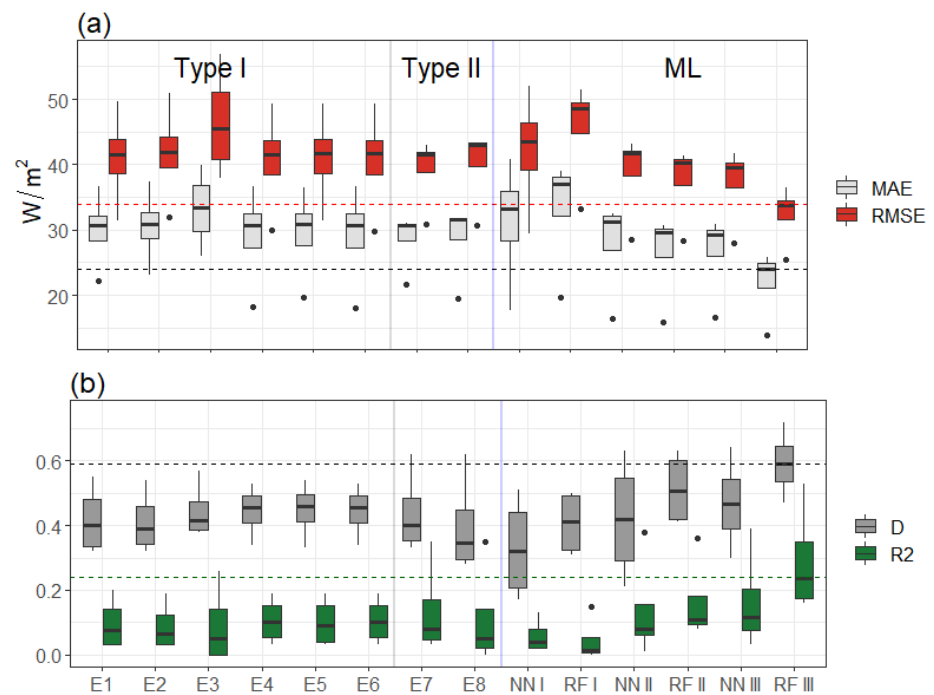
Figure 7. (a) Box plot of the MAE and RMSE values and (b) D and R^2 boxplot grouped by method.

Uncertainties in the SEB fluxes quantification can be related with measurement bias, which corresponds with ca. 10% for G, and ca. 5% for Rn [21]. However, all the methods have been evaluated under the same conditions, and hence the superiority of RF II and III it is related with the algorithm architecture. The regression-based ML methods have shown promising results over the last years for the prediction of continuous variables, and have been successfully applied in a wide range of scientific disciplines, including climate and hydrological modelling [e.g., 50]. For example, ML provide better results than Multiple Linear Regressions for the spatial interpolation of air temperature [e.g., 51]. ML also shown optimal results for the estimation of environmental variables [52], H and LE retrievals [e.g., 53], other SEB components such as Rn, H and LE [e.g., 54, 55] or ET [56, 57]. For the estimation of G, [32] modelled the heat flux with ANN and RS data, and [33] compared the ANN and two empirical equations based on NDVI, Ts and α. In accordance with our findings, in both of the works, ML outperformed the other methods for the estimation of G. The underestimation of the empirical equations and the large errors found at high Fv ranges are partly because the majority of the type I and II equations rely on a single linear relationship with the predictor variables. The main strengths of the NN and RF is that they are very flexible, adaptable to the shifting G values depending on the vegetation and meteorological conditions, and do not follow a stationary and linear assumption. The bootstrap aggregating (bagging) algorithms, such as RF, creates several models by selecting a randomly subset of variable for each decision tree. Moreover, the inclusion of Rn in the RF III improved the G estimation, showing evidence that the recursive model partitioning of RF is resistant to overlaps in the covariates, and can handle correlated variables such as Rn and Ts. In addition, RF is a non-parametric model, overcome noise and outliers [41]. On the other hand, the limitation of the ML algorithms is that generally reproduce the data of the training dataset, and could not extrapolate the values outside of the training. The estimation of G with ML also requires more computational-time than the required with the empirical equations. The training time required for train the RF algorithms was equal to 5, 8 and 11h for RF I, II and III, respectively. On the contrary, the training time of the NN was 26, 28 and 31' for NN I, II and III, respectively. Nonetheless, once the ML algorithms are trained, they are computationally efficient. In this case, the

prediction time required was lower than the minute. Finally, further research is needed in order to improve the energy balance closure and the estimation of G in high canopies covers. If the accuracy between ML methods shows small differences, the introduction of other predictor variables should be tested and further works should be also focused in combine ML with physical modelling [e.g., 50].

4. Conclusion

Anthropogenic climate change is leading changes in the water and energy fluxes of some areas of the planet, such as the tropical and mid-latitude region. Therefore, an accurate hydrological estimation is crucial in order to better quantify water balances, and support socioecological decision-makings. In this sense, this work compared different methods for the estimation of a SEB component with large uncertainties, the G. For the first time, we evaluated an ensemble ML method (the RF), against NN and several empirical equations that has been extendedly used for decades for the estimation of G with RS data. The evaluation has been performed over a dense dataset including many EC sites, with a wide range of vegetation (i.e., forests, crops of wheat or millet) and climate types (from marine west coast to desert climate types), along the mid-latitude area found between the continental Europe up to the middle of the African continent.

The data have shown that the driver of G depends of the Fv range. In bare soils and sparse vegetation areas, Rn followed by Ts rules the estimation of G, decreasing with Fv. On the contrary, meteorological and vegetation data have not shown a relevant and statistically significant link with the estimation of G in dense vegetation areas. NN and RF models, with LAI and Ts as a predictor variable, outperforms the empirical equations for the estimation of the G, independently of the Fv range considered. The inclusion of Rn in the model significantly improved the estimation of G. Despite the computational time required (8 hours), RF III have shown the highest R^2, D as well as the lowest MAE and RMSE. The accuracy values of the empirical equations are almost the double than the measured with RF III. Further works should analyze the estimation of G over large areas with the predictor variables used here, in combination with other predictor variables, for instance downscaled soil moisture data. Also, further research should test the combination of ML with physical-based models for the estimation of G.

Author Contributions: Conceptualization, J.B. and G.B.; methodology, J.B and G.B.; software, J.B.; validation, J.B. and G.B.; formal analysis, J.B; investigation, J.B.; resources, G.B.; data curation, J.B and G.B; writing—original draft preparation, J.B; writing—review and editing, J.B. and G.B; visualization, J.B. and G.B.; supervision, G.B.; project administration, J.B. and G.B.; funding acquisition, G.B. All authors have read and agreed to the published version of the manuscript.

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

References

1.   Giorgi, F. Climate change hot-spots. Geophys. Res. Lett. 2006, 33, L08707, https://doi.org/10.1029/2006GL025734
2.   Hu, Y.; Fu, Q. Observed poleward expansion of the Hadley circulation since 1979. Atmos. Chem. Phys. 2007, 7, 5229–5236. https://doi.org/10.5194/acp-7-5229-2007
3.   De Luis, M.; Brunetti, M.; Gonzalez-Hidalgo, J.C.; Longares, L.A.; Martin-Vide, J. Changes in seasonal precipitation in the Iberian Peninsula during 1946–2005. Glob. Planet Change. 2010, 74, 27–33. https://doi.org/10.1016/j.gloplacha.2010.06.006
4.   Vicente-Serrano, S. M.; Lopez-Moreno, J. I.; Beguería, S.; Lorenzo-Lacruz, J.; Sanchez-Lorenzo, A.; García-Ruiz, J. M.; et al. Evidence of increasing drought severity caused by temperature rise in southern Europe. Environ. Res. Lett. 2014, 9(4), 044001. https://doi.org/10.1088/1748-9326/9/4/044001

5.   Orlowsky, B.; Seneviratne, S.I. Global changes in extreme events: Regional and seasonal dimension. Clim. Chang. 2012, 110, 669–696. https://doi.org/10.1007/s10584-011-0122-9

6.   Giorgi, F.; Lionello, P. Climate change projections for the Mediterranean region, Global Planet. Chang. 2008, 63, 90– 104. https://doi.org/110.1016/j.gloplacha.2007.09.005

7.   Sheffield, J.; Wood. E. F. Projected changes in drought occurrence under future global warming from multi-model, multi-scenario, IPCC AR4 simulations. Climate Dynam. 2008, 31, 79–105. https://doi.org/110.1007/s00382-007-0340-z

8.   Viviroli, D.; Durr, H.; Messerli, B.; Meybeck, M.; Weingartner, R. Mountains of the world, water towers for humanity: Typology, mapping, and global significance. Water Resources Res. 2007, 43: W07447. https://doi.org/10.1029/2006WR005653

9.   Boulet, G.; Jarlan, L.; Olioso, A.; Nieto, H. Evapotranspiration in the Mediterranean region. In: Water. Brocca, L.; Tramblay, Y.; Molle, F. (eds.). Amsterdam: Elsevier, p. 23-49, 2020.

10.   Kpemlie, E. Assimilation variation nelle de donnees de télédétection dans des modèles de fonctionnements des couverts végétaux et du paysage agricole, Thèse de doctorat, Université d'Avignon et des Pays de Vaucluse, Avignon (FRA). 2009.

11.   Sauer T.J.; Horton R. Soil heat flux. In: Hatfield JL, Baker JM, editors. Micrometeorology in agricultural systems. Amer. Soc. Agron. 2005, 131–54. https://doi.org/10.2134/agronmonogr47.c7

12.   Liebethal, C.; Foken, T. Evaluation of six parameterization approaches for the ground heat flux. Theor. and Appl. Climatol. 2007, 88, 43–56. https://doi.org/10.1007/s00704-005-0234-0

13.   Shao C.; Chen J.; Li L.; Xu W.; Chen S.; Gwen T. Spatial variability in soil heat flux at three Inner Mongolia steppe ecosystems. Agric. and F. Meteor. 2008, 148, 1433–43. https://doi.org/10.1016/j.agrformet.2008.04008

14.   Venegas, P.; Grandón, A.; Jara, J.; Paredes, J. Hourly estimation of soil heat flux density at the soil surface with three models and two field methods. Theor. and Appl. Climatol. 2013, 112(1–2), 45–59. https://doi.org/10.1007/s00704-012-0705-z

15.   An, K.; Wang, W.; Wang, Z.; Zhao, Y.; Yang, Z.; Chen, L.; Zhang, Z.; Duan, L. Estimation of ground heat flux from soil temperature over a bare soil. Theor. and Appl. Climatol. 2017, 129(3–4), 913–922. https://doi.org/10.1007/s00704-016-1816-8

16.   Gao, Z.; Russell, E. S.; Missik, J. E. C.; Huang, M.; Chen, X.; Strickland, C. E.; et al. A novel approach to evaluate soil heat flux calculation: An analytical review of nine methods. J. of Geophys. Res.: Atmosph. 2017, 122, 6934–6949. https://doi.org/10.1002/ 2017JD027160

17.   Murray, T.; Verhoef, A. Moving towards a more mechanistic approach in the determination of soil heat flux from remote measurements. I. A universal approach to calculate thermal inertia. Agric. F. Meteor. 2007, 147, 80-87. https://doi.org/10.1016/j.agrformet.2007.06.009

18.   Murray, T.; Verhoef, A. Moving towards a more mechanistic approach in the determination of soil heat flux from remote measurements. II. Diurnal shape of soil heat flux. Agric. F. Meteor. 2007, 147, 88-97. https://doi.org/10.1016/j.agrformet.2007.06.009

19.   Kustas, W. P.; Daughtry, C. S. T. Estimation of the soil heat flux/net radiation ratio from spectral data. Agric. For. Meteorol. 1990, 49, 205-233. https://doi.org/10.1016/0168-1923(90)90033-3

20.   Bastiaanssen, W.G.M.; Menenti, M.; Feddes, R.A.; Holslag, A.A.M. A remote sensing surface energy balance algorithm for land (SEBAL): 2 Validation. J. of Hydrol. 1998, 212-213, 213-229. https://doi.org/10.1016/S0022-1694(98)00253-4

21.   Santanello, J. A.; Friedl, M. A. Diurnal variation in soil heat flux and net radiation. J. of Appl. Meteor. 2003, 42, 851–862. https://doi.org/10.1175/1520-0450(2003)042<0851:DCISHF>2.0.CO;2

22.   Cellier, P.; Richard, G.; Robin, P. Partition of sensible heat fluxes into bare soil and the atmosphere. Agric. F. Meteorol. 1996, 82(1), 245–265. https://doi.org/10.1016/0168 1923(95)02328-3

23.   Kustas, W.P.; Daughtry, C.S.T.; Van Oevelen, P.J. Analytical treatment of the relationship between soil heat flux/net radiation ratio and vegetation indices. Remote Sens. Environ. 1993, 46, 319-330. https://doi.org/10.1016/0034-4257(93)90052-Y

24.   Choudhury, B. J.; Idso, S. B.; Reginato, R. J. Analysis of an empirical model for soil heat flux under a growing wheat crop for estimating evapotranspiration by an infrared-temperature based energy balance empirical equation. Agric. For. Meteorol. 1987, 39(4), 283–297. https://doi.org/10.1016/0168-1923(87)90021-9

25.   Anderson, M. C.; J. M. Norman, J. R. Mecikalski, J. P. Otkin.; W. P. Kustas. A climatological study of evapotranspiration and moisture stress across the continental U.S. based on the thermal remote sensing: I. Model formulation. J. of Geophys. Res. Lett. 2007, 112, D10117. https://doi.org/10.1029/2006JD007506

26.   Sobrino, J. A.; Gómez, M.; Jiménez-Muñoz, J. C.; Olioso, A.; Chehbouni, G. A. simple algorithm to estimate evapotranspiration from DAIS data: Application to the DAISEX Campaigns. J. of Hydrol. 2005, 315, 117−125. https://doi.org/10.1016/j.jhydrol.2005.03.027

27.   Miralles, D. G.; Holmes, T. R. H.; De Jeu, R. A. M.; Gash, J. H.; Meesters A. G. C. A.; Dolman, A. J. Global land-surface evapotranspiration estimated from satellite-based observations. Hydrol. Earth Sys. Sci. 2011, 15(2), 453–469. https://doi.org/10.5194/hess-15-453-2011

28.   Allen, R. G.; Tasumi.; M.; Trezza, R. Satellite-based energy balance for mapping evapotranspiration with internalized calibration (METRIC). Model. J. of Irrig. and Drain. Eng. 2007, 133(4), 380–394. https://doi.org/10.1061/(asce)0733-9437(2007)133:4(395)

29.   Bastiaanssen, W. G. M. SEBAL-based sensible and latent heat fluxes in the irrigated Gediz Basin, Turkey. J. of Hydrol. 2000, 229, 87–100. https://doi.org/10.1016/S0022-1694(99)00202-4

30.   Sun, Z.; Gebremichael M.; Wang Q. Evaluation of empirical remote sensing-based empirical equations for estimating soil heat flux. J. Meteorol. Soc. Jap. 2013, 91-5, 627–638. https://doi.org/10.2151/jmsj.2013-505

31.   Purdy, A.; Fisher, J.; Goulden, M.; Famiglietti, J. Ground heat flux: An analytical review of 6 models evaluated at 88 sites and globally. J. Geophys. Res.: Biog. 2016, 121, 3045–3059. https://doi.org/10.1002/2016JG003591.

32.   Canelón, D.J.; Chávez, J.L. Soil heat flux modeling using artificial neural networks and multispectral airborne remote sensing imagery. Remote Sens. 2011, 3, 1627–1643. https://doi.org/10.3390/rs3081627

33.   De Andrade, B.C.C.; Pedrollo, O.C.; Ruhoff, A.; Moreira, A.A.; Laipelt, L.; Kayser, R.B.; Biudes, M.S.; dos Santos, C.A.C.; Roberti, D.R.; Machado, N.G.; et al. Artificial neural network model of soil heat flux over multiple land covers in South America. Remote Sens. 2021, 13, 2337. https://doi.org/10.3390/rs13122337

34.   Delogu, E.; Boulet, G.; Olioso, A.; Garrigues, S.; Brut, A.; Tallec, T.; Demarty, J.; Soudani, K.; Lagouarde, J.-P. Evaluation of the SPARSE Dual-Source Model for Predicting Water Stress and Evapotranspiration from Thermal Infrared Data over Multiple Crops and Climates. Remote Sens. 2018, 10, 1806. https://doi.org/10.3390/rs10111806.

35.   Clevers, J.: The Application of a Weighted Infrared-Red Vegetation Index for Estimating Leaf-Area Index by Correcting for soil moisture, Remote Sens. Environ. 1989, 29, 25–37. https://doi.org/10.1016/0034- 4257(89)90076-x

36.   Chirouze, J.; Boulet, G.; Jarlan, L.; Fieuzal, R.; Rodriguez, J.C.; Ezzahar, J.; Bigeard, G.; Merlin, O. Intercomparison of four remote-sensing-based energy balance methods to retrieve surface evapotranspiration and water stress of irrigated fields in semi-arid climate. Hydrol. Earth Sys. Sci. 2014, 18, 1165–1188. https://doi.org/10.5194/hess-18-1165- 2014

37.   Roujean, J. L.; Lacaze, R. Global mapping of vegetation parameters from POLDER multi angular measurements for studies of surface atmosphere interactions: A pragmatic method and validation, J. of Geophys. Res. 2002, 107 (12), 4150. https://doi.org/10.1029/2001JD000751

38.   Boegh, E.; Soegaard, H.; Christensen, J. H.; Hasager, C. B.; Jensen, N. O.; Nielsen, N. W. Combining weather prediction and remote sensing data for the calculation of evapotranspiration rates: application to Denmark. Intern. J. of Rem. Sensing. 2004, 25, 2553 - 2574. https://doi.org/10.1080/01431160310001647984

39.   Choudhury, B. J.; Idso, S. B.; Reginato, R. J. Analysis of an empirical model for soil heat flux under a growing wheat crop for estimating evapotranspiration by an infrared-temperature based energy balance empirical equation, Agric. For. Meteorol. 1987, 39(4), 283–297. https://doi.org/10.1016/0168-1923(87)90021-9

40.   Haykin, S. Neural networks: a comprehensive foundation. Prentice Hall, Upper Saddle River. 1998.

41.   Breiman, L. Random forests. IEEE Machine Learning. 2001, 45: 5–32.

42.   Maxwell, A.E.; Warner, T.A.; Fang, F. Implementation of machine-learning classification in remote sensing: An   applied review. Int. J. Remote Sens. 2018, 39, 2784–2817. https://doi.org/10.1080/01431161.2018.1433343

43.   Liaw, A.; Wiener, M. Random Forest: Breiman and Cutler's Random Forests for Classification and Regression. R Package Version. 2012, 4.6-7.

44.   Kuhn, M. Caret: Classification and Regression Training. R package version 6.0-30. 2014.

45.   R Core Team, R. A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna. 2018. https://www.R-project.org/.

46.   Willmott, C.J. Some comments on the Evaluation of Model Performance. Bull. of Amer. Meteorol. Soc. 1982, 63, 1309-1313. https://doi.org/10.1175/1520-0477(1982)063<1309:SCOTEO>2.0.CO;2

47.   Su, Z. The Surface Energy Balance Systems (SEBS) for estimation of turbulent heat fluxes. Hydrol. Earth Sys. Sci. 2002, 6(1), 85-99. https://doi.org/10.5194/hess-6-85-2002

48.   Tanguy, M.; Baille, A.; Gonzalez-Real, M.M.; Lloyd, C.; Cappelaere, B.; Kergoat, L.; Cohard, J.-M. A new parameterisation scheme of ground heat flux for land surface flux retrieval from remote sensing information. J. of Hydrol. 2012, 454–455. 113–122. https://doi.org/10.1016/j.jhydrol.2012.06.002

49.   Cammelleri, C.; la Loggia, G.; Loggia, A.; Maltese, A. Critical analysis of empirical ground heat flux empirical equations on a cereal field using micrometeorological data. SPIE. 2009, 7472, 747225, 1–12. https://doi.org/10.1117/12.830289

50.   Zhao, W. L.; Gentine, P.; Reichstein, M.; Zhang, Y.; Zhou, S.; Wen, Y.; et al. Physics-constrained machine learning of evapotranspiration. Geophys. Res. Lett. 2019, 46, 14,496–14,507. https://doi.org/10.1029/2019GL085291

51.   Meyer, H.; Katurji, M.; Appelhans, T.; Müller, M.U.; Nauss, T.; Roudier, P.; Zawar-Reza, P. Mapping Daily Air Temperature for Antarctica Based on MODIS LST. *Remote Sens.* 2016, *8*, 732. https://doi.org/10.3390/rs8090732

52.   Reichstein, M.; Camps-Valls, G.; Stevens, B.; Jung, M.; Denzler, J.; Carvalhais, N. Deep learning and process understanding for data-driven Earth system science. Nature. 2019, 566(7743), 195–204. https://doi.org/10.1038/s41586-019-0912-1

53.   Kuhnlein, M.; Tim, A.; Boris, T.; Thomas, N. Improving the accuracy of rainfall rates from optical satellite sensors with machine learning: A random forests-based approach applied to MSG SEVIRI. Remote Sens. of Environ. 2014, 141: 129–43. https://doi.org/10.1016/j.rse.2013.10.026

54.   Tramontana, G. et al. Predicting carbon dioxide and energy fluxes across global FLUXNET sites with regression algorithms. Biogeos. 2016, 13, 4291–4313. https://doi.org/10.5194/bg-13-4291-2016

55. Alemohammad, S.H.; Fang, B.; Konings, A.G.; Aires, F.; Green, J.K.; Kolassa, J.; Miralles, D.; Prigent, C.; Gentine, P. Water, Energy and Carbon with Artificial Neural Networks (WECANN): A statistically based estimate of global surface turbulent fluxes and gross primary productivity using solar-induced fluorescence. Biogeos. 2017, 14, 4101–4124. https://doi.org/10.5194/bg-14-4101-2017

56. 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. https://doi.org/10.1016/j.jag.2019.01.020

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