

Estimation of XCO₂ in Rice and Sugarcane Crops with an Ensemble of Wofost and Random Forest

Henrique Fontellas Laurito

henrique.f.laurito@unesp.br

São Paulo State University

Thaís Rayane Gomes da Silva

São Paulo State University

Newton La Scala

São Paulo State University

Alan Rodrigo Panosso

São Paulo State University

Glauco de Souza Rolim

São Paulo State University

Research Article

Keywords: machine learning, remote sensing, climate change, modeling, python, sustainability

Posted Date: June 10th, 2024

DOI: <https://doi.org/10.21203/rs.3.rs-4432811/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

ESTIMATION OF XCO₂ IN RICE AND SUGARCANE CROPS WITH AN ENSEMBLE OF WOFOST AND RANDOM FOREST

Authors: Henrique Fontellas Laurito ^{a,1}, Thaís Rayane Gomes da Silva ^a, Newton La Scala Jr. ^a, Alan Rodrigo Panosso ^a, Glauco de Souza Rolim ^a

Affiliation: ^a Department of Exact Science (FCAV/Unesp)

São Paulo State University Faculty of Agricultural and Veterinary Sciences, Jaboticabal, São Paulo, Brazil. Via de Acesso Prof. Paulo Donato Castellane s/n. **Zip-code:** 14884-900

ABSTRACT – The concentration of carbon dioxide (CO₂) in the air is one of the main factors affecting climate change, which can impact climate zoning and, consequently, agricultural production, affecting food security and the quality of life of people, especially the most vulnerable. Therefore, it is important to monitor and estimate this quantity in agricultural areas. Thus, this study used productivity data (1984-2022) from rice (*Oryza sativa*) and sugarcane (*Saccharum officinarum*) areas, from the main producing regions of Brazil, under the hypothesis that the daily XCO₂ could be estimated from the ensemble of the WOFOST and Random Forest (RF) models, using remote sensing data as a source of measured XCO₂. The results confirmed the hypothesis, with high accuracy ($R^2 > 0.75$) for the models under sugarcane and rice cultivation, respectively. The daily total gross assimilation (DTGA) has a negative correlation with the calculated CO₂ measured in the atmosphere ($\rho = -0.45$, $p\text{-value} < 0.001$). The DTGA profile did not change throughout the historical series. Key factors influencing XCO₂ estimates were identified, including water stress and vegetation parameters for sugarcane, and radiation and leaf area for rice. This work contributed to the understanding that climate impacts can exert on agroecosystems, with a focus on XCO₂.

Keywords: machine learning, remote sensing, climate change, modeling, python, sustainability.

¹ Corresponding author. Telephone: +5516997048254

E-mail addresses:

henrique.f.laurito@unesp.br (LAURITO, H. F., ORCID: 0000-0001-7576-0731);

trg.silva@unesp.br (SILVA, T. R. G., ORCID: 0000-0002-67873-3717);

la.scala@unesp.br (LA SCALA, N., ORCID: 0000-0002-1575-9875);

alan.panosso@unesp.br (PANOSSO, A. R. ORCID: 0000-0001-9916-1696);

glauco.rolim@unesp.br (ROLIM, G. S. ORCID: 0000-0003-4683-3203)

1 INTRODUCTION

Human activities related to the emission of greenhouse gases (GHGs) have raised the planet's temperature by 1.1 °C since the end of the 19th century (IPCC, 2023) in the last decade. These gases come from a variety of sources, such as energy use, deforestation, agriculture, and waste (Li et al., 2022). The climate altered by these emissions affects food and water, the health of people, especially the most vulnerable, economies and societies, and harms nature (IPCC, 2023; Semenov, 2021).

Agriculture is responsible for more than one-third of these gases, which come from land use, forestry, and livestock (Khatri-Chhetri et al., 2021)). The main gases are: (i) carbon dioxide (CO₂), from deforestation, biomass burning and mechanization; (ii) methane (CH₄), from ruminant digestion and manure; (iii) nitrous oxide (N₂O), from fertilizers and soil (Chen et al., 2020; Crippa et al., 2021; Hajima et al., 2020). Therefore, we need alternative tools to measure GHGs and know their effects on agriculture.

Sugarcane and rice differ from each other, but they are of great relevance in Brazil and in the world, as they provide renewable energy and food. There are many studies in the literature that evaluate the climatic effects on these crops. For sugarcane, they point out that climate change does not affect production in southeastern Brazil (Marin et al., 2013), but that high temperatures reduce production in São Paulo, and CO₂ has little impact on it (Flack-Prain et al., 2021). For irrigated rice, they show that monthly CO₂ emissions in Chinese growing areas are 0.21 – 1.54 ppm (Li et al., 2020) and that high temperatures can shorten the vegetative phase of this crop (Holzworth et al., 2015). More detailed studies are still needed on how CO₂ influences these crops.

Crop models help you understand crop dynamics (Holzworth et al., 2015). One such model is World Food Studies (WOFOST), which simulates the growth and production of annual crops, such as sugarcane (Hu et al., 2019) and rice (Confalonieri et al., 2009), for quantitative analysis (de Wit and Boogaard, 2021). This model considers the assimilation of CO₂ as a function of various climatic and eco physiological factors, which include soil properties, atmospheric carbon, and water availability; standing out among several models (de Wit and Boogaard, 2021; Holzworth et al., 2015).

In addition, remote sensing is a tool used to study GHGs. It employs drones and satellites, for example, which collect data relevant to the object of study (Weiss et al., 2020). To analyze sources and sinks of the mean molar concentration of CO₂ in dry air column (XCO₂), the main system used is the Orbital Carbon Observatory-2 (OCO-2/NASA) (Fang et al., 2023) because it has more data and better quality than the other satellites, validated by TCCON (Fang et al., 2023; Wunch et al., 2017).

One of the OCO-2 products provides daily XCO₂ measurements with a resolution of 0.5 x 0.625 decimal degrees. This product uses interpolations and data assimilation by the CoDAS model, generating daily polls even with clouds, for example (Ott and Weir, 2022). This allows for several studies, such as the global carbon cycle (Eldering et al., 2017) and agroecosystems, such as pastures, soybeans, and sugarcane (Morais Filho et al., 2021).

One way to combine these two views is to use machine learning (ML) methods. ML is computational algorithms that learn from data to make predictions, classifications, or groupings (Paixão et al., 2022). There are several types of ML models. The most widely adopted for cause-and-effect studies is the supervised one, such as Artificial Neural Networks and Random Forest (RF) (Bentéjac et al., 2021).

RF (Breiman, 2001) has already been adopted to relate agricultural variables, such as climate and sugarcane productivity (Dos Santos and De Fiori, 2005) and to estimate the environmental impacts of GHGs on rice (Xu et al., 2021). However, ML models may not succeed in multivariate systems, as one of their problems is that it may not capture the complexity and uncertainty of geophysical processes, nor provide a physical interpretation of the results (Reichstein et al., 2019). For this reason, some researchers suggest combining physical models based on laws and principles with machine learning (ensemble) models, which can learn patterns and relationships from data. This ensemble can exploit the benefits of both types of models, increasing the accuracy, robustness, and explainability of geophysical phenomena, such as the behavior of GHG sources and sinks.

The ensemble technique combines several models that have different perspectives on the phenomenon studied. The union of ML models with mechanistic models (e.g., WOFOST) is important for research on climate change (Kristiansen et al., 2024), GHGs (Tian et al., 2023), and productivity estimates (Li et al., 2022). This union allows for a deeper view, because: (i) it improves parametrization; (ii) simplifies physical phenomena with ML; (iii) reduces physical model errors (Reichstein et al., 2019).

Studies using ML models and WOFOST have already taken place to improve leaf area index estimation (Tian et al., 2017, p. 201) for rice cultivation. However, there is still no research that uses an ensemble to estimate XCO₂ from the WOFOST carbon balance.

Thus, the objective of our study is to advance this theoretical issue, aiming to adopt an ensemble of WOFOST and RF, proposing a new approach to estimate XCO₂ in rainfed sugarcane and flood-irrigated rice systems in the main producing areas of Brazil, improving the knowledge of the effect of XCO₂ on agricultural systems, collaborating with the study of different climate change scenarios.

2 MATERIAL AND METHODS

2.1 Delimitation of study locations

Consulting harvested area and production data from the Automatic Recovery System - IBGE (SIDRA, 2024), we chose the most relevant producing regions in Brazil according to the average production between 2015 and 2022 (Figure 1). We selected for sugarcane: (i) Morro Agudo (-20,705; -48,091) and Barretos (-20,553; -48,617), in São Paulo; (ii) Quirinópolis (-18,477; -50,485) and Goiatuba (-18,028; -49,320), in Goiás. And for rice: (i) Uruguaiana (-29,807; -57,122) and Santa Vitória do Palmar (-33,477; -53,396), in Rio Grande do Sul; (ii) Turvo (-28,959; -49,712) and Forquilha (-28,746; -49,523), in Santa Catarina.



Figure 1. Studied localities.

2.2 Data collection

The following values were queried for the same points in the appropriate databases: (1) Meteorological (NASA Power)²; (2) XCO₂ (GES DISC)³; (3) Yield (SIDRA)⁴; (4) Soil (Soilgrids, Poggio et al., 2021)⁵. The historical series analyzed covered the harvest years from 1984 to 2022 of harvested area and productivity. XCO₂ measurements began in 2015 and lasted until the end of the time series.

2.3 Crop model - WOFOST

The environment used to create the features of the RF model was made in Wageningen, initially in the FORTRAN language and later in Python, resulting in the Python Crop Simulation Environment (PCSE) (de Wit and Boogaard, 2021). One of its components is WOFOST, which uses different input files with the necessary parameters. This model performs the quantitative analysis of annual crop development, simulating phenological development, productivity, biomass of the different parts of the cultivated plant, and use of water resources, based on knowledge about soil properties, climate, and crop management (de Wit and Boogaard, 2021).

Three hierarchical levels of production are established within the WOFOST: in the first level, called potential, growth is limited by climatic variables and plant characteristics; at the second, limited level, the effect of water availability is added to the first level, taking into account the occurrences of rainfall and evapotranspiration; and at the third level, reduced, factors related to reducing effects, such as pests, diseases and weeds. The cultivation of sugarcane was considered in the second level of hierarchy, as it is a crop without irrigation, while rice was considered in the first level, as it was assumed to be managed by flooding.

2.3.1 Soil conditions

Each soil type has a file that adjusts the simulated crops, describing the physical and retention properties in the soil, to calculate the water balance and estimate the daily moisture used in the photosynthesis and evapotranspiration of the crop (de Wit and Boogaard, 2021) (Table 1).

² <https://power.larc.nasa.gov/>

³ https://disc.gsfc.nasa.gov/datasets/OCO2_GEOS_L3CO2_DAY_10r/summary?keywords=OCO-2

⁴ <https://sidra.ibge.gov.br/Tabela/5457>

⁵ <https://soilgrids.org/>

Table 1. Classification of soils (WRB, 2024) in each locality, selected according to the criterion, consulted in the Soil grids database.

Locality	Soil	WOFOST Archive
Uruguaiiana	Planosols	EC3
Santa Vitória do Palmar		
Turvo		
Forquilha		
Morro Agudo	Ferralsols	EC4
Barretos		
Quirinópolis		
Forquilha		

2.3.2 Cultivation and Management

Using the files with the parameters of the crops mentioned above, "sugarcane.yaml" was chosen for sugarcane and "rice.yaml" for rice. The cycle of these crops was 365 and 140 days, respectively (Table 2). The initial parameters and values, present in the files "sugarcane.yaml" and "rice.yaml," are documented in Hu et al. (2019) and Confalonieri et al. (2009), respectively. The uncalibrated values remained at standard values (Van Diepen et al., 1988; Van Heemst, 1988). The maximum and minimum values for the calibrated parameters are described in (de Wit and Boogaard, 2021).

Table 2. Planting calendar for sugarcane (S) and rice (R) crops. DVS = Stage of development of the crop. Values from 0 to 1 are equivalent to the period from vegetation to anthesis (green). Values from 1 to 2 are equivalent to the anthesis period at harvest (yellow).

	A	M	J	J	A	S	O	N	D	J	F	M
S	0 < DVS < 0.5									0.5 ≤ DVS ≤ 1		
R							0 < DVS < 1				1 ≤ DVS ≤ 2	

2.4 WOFOST model calibration

For calibration, the python Non-Linear optimization (NLOpt) package (Johnson, 2021) was used with the Constrained Optimization BY Linear Approximations (COBYLA) algorithm (Powell, 1994). The optimal selection of the

parameters was obtained by the deviation of the productivity of the simulation in relation to the yields observed in SIDRA. This calibration was done for each year of the historical series (1984 – 2022), with their respective yields and management.

COBYLA is an algorithm for optimizing functions without derivatives and nonlinear equality and inequality constraints. It uses $n+1$ points (in n dimensions) to form linear approximations of function and constraints and optimizes them into a region of confidence at each step. The algorithm operates in an adaptive trust region, ensuring convergence to an optimal location, updating the confidence region for each iteration of the optimization, expanding when the predicted solution is promising, or contracting in case of unsatisfactory results. The simplex steps are pseudo-random to make it more robust, avoiding bad steps that worsen conditioning, but using an initial value to be deterministic. The use of this algorithm is important, as it can deal with different types of objective functions and constraints, adapting to the specific characteristics of each problem, giving it a characteristic of flexibility and versatility.

Using the fitted WOFOST model, shows the carbon balance results, used as features (Table 3), and the measured XCO_2 (OCO-2) as an independent variable to train the RF model.

Table 3. Description of the results of the WOFOST productivity simulation and their respective magnitude scales.

Result	Description	Scale
DVS	Stage of cultivation development	-
LAI	Leaf area index	m m ⁻¹
TAGP	Total above-ground production	kg ha ⁻¹
TWSO	Total dry mass of reserve organs	
TWL	Total dry mass of leaves	
TWST	Total dry mass of twigs/stems	
TRA	Transpiration rate	mm d ⁻¹
RFTRA	Transpiration rate reduction factor	-
RD	Root Zone Depth	cm
WWLOW	Total Soil Profile Moisture	cm ³ cm ³
SM	Soil moisture	
GASST	Total Gross Assimilation	CO ₂ kg ha ⁻¹
MREST	Total Maintenance Breathing	
MRES	Actual maintenance respiratory rate	
PGASS	Potential Assimilation Rate	
GASS	Assimilation rate corrected for water stress	CO ₂ kg ha ⁻¹ d ⁻¹
ASRC	Available liquid assimilates	
PMRES	Potential Respiration Rate for Maintenance	
DTGA	Daily total gross assimilation	

2.5 Conversion of XCO₂ numerical basis from ppm_v to kg ha⁻¹

To connect the daily total gross assimilation (DTGA) of the WOFOST model with the XCO₂, measured by OCO-2, the ideal gas equation (1 and 2) was used to convert ppm_v into kg of CO₂ ha⁻¹ (cCO₂). Temperature and atmospheric pressure data came from the NASA-Power bank.

$$PV = nRT \quad 1$$

$$n_{total} = \frac{PV}{RT} \quad 2$$

where: P = atmospheric pressure (kPa); V = volume occupied by the gas (m³); n_{total} = total number of moles in 1 m³ (molar); R = ideal gas constant (kPa m³ mol⁻¹ K⁻¹) equal to 8.3145; T = average temperature (Kelvin).

Once the n_{total} has been calculated, i.e. the total molar number of gases in 1 m³, it is enough to calculate the mass of cCO₂ (kg), adopting the result of XCO₂ (ppm_v) in proportion to the amount of moles of CO₂ contained in the volume (3).

$$cCO_2 = n_{total}XCO_2 10^{-5} \quad 3$$

where: XCO_2 = molar mean CO_2 in dry air column (ppm_v); cCO_2 = converted CO_2 ($kg\ ha^{-1}$).

The DTGA represents the amount of gross CO_2 that a hectare assimilates in a day, according to WOFOST. By comparing XCO_2 with DTGA on the same numerical scale, we can see how CO_2 is assimilated in different agroecosystems and detect possible changes in this process over time.

2.5.1 Spearman's correlation

To correlate the DTGA with the cCO_2 concentration, the following formula was used to determine Spearman's correlation coefficient (4):

$$\rho = 1 - \frac{6 \sum_{i=1}^N d_i^2}{n(n^2 - 1)} \quad 4$$

where: ρ is Spearman's correlation; d_i is the difference between the ranks of each pair of dice; n is the total number of data pairs.

2.6 Machine learning - Random Forest

We used the Random Forest Regressor package from Scikit-Learn (Pedregosa et al., 2011) to train and test sugarcane and rice crops. The WOFOST results, used as features, were separated into 70% for training and 30% for testing the RF model.

To find the best combination of parameters, the GridSearchCV⁶ method was used. This is an exhaustive method that checks the best possibilities on specified parameter values for an estimator. The parameters tested were: (i) number of estimators (n decision trees); (ii) criterion, which measures the quality of the estimators' divisions; (iii) maximum depth of the estimators. The number of estimators was 1400 and 1350 for the RF models of sugarcane and rice, respectively; the criterion selected was "*friedman_mse*"; and the maximum depth of the estimators was 40 for both. See the method documentation for more information⁷.

The models were trained and tested with these parameters and the metrics were obtained. The adjusted coefficient of determination (R^2 adj, 5), root mean squared error (RMSE, 6) and mean absolute error (MAE, 7) were used to validate the models.

⁶ https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html

⁷ <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html>

$$R^2_{adj.} = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad 5$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (\hat{y}_i - y_i)^2}{N - 1}} \quad 6$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad 7$$

Where: $\hat{y}_i$ is the estimated value; y_i is the observed value; $\bar{y}$ is the mean of the observed values; N is the total number of observations

The RF model for sugarcane and rice was submitted to Shapley Additive Explanations SHAP analysis (Lundberg and Lee, 2017) to verify the importance of the variables in the estimation of XCO₂. The most influential variables in the estimation of XCO₂ are arranged in descending order on the Y axis. The interpretation of the variables will depend on the color of the dot (red to blue) and the position in which it is inserted in the graph (left or right). Values on the right have a positive impact on the estimation of XCO₂, while values on the left have a negative impact. The colors represent the value of the variable observed on the Y-axis. Therefore, high values of a variable can have positive or negative impacts on the estimation of XCO₂.

3 RESULT AND DISCUSSION

3.1 WOFOST model calibration

The parameters of each crop, location and crop year were adjusted for their respective yields. After calibration, yields were compared between that estimated by WOFOST and observed by SIDRA, with R²=1 in all cases (Figure 2).

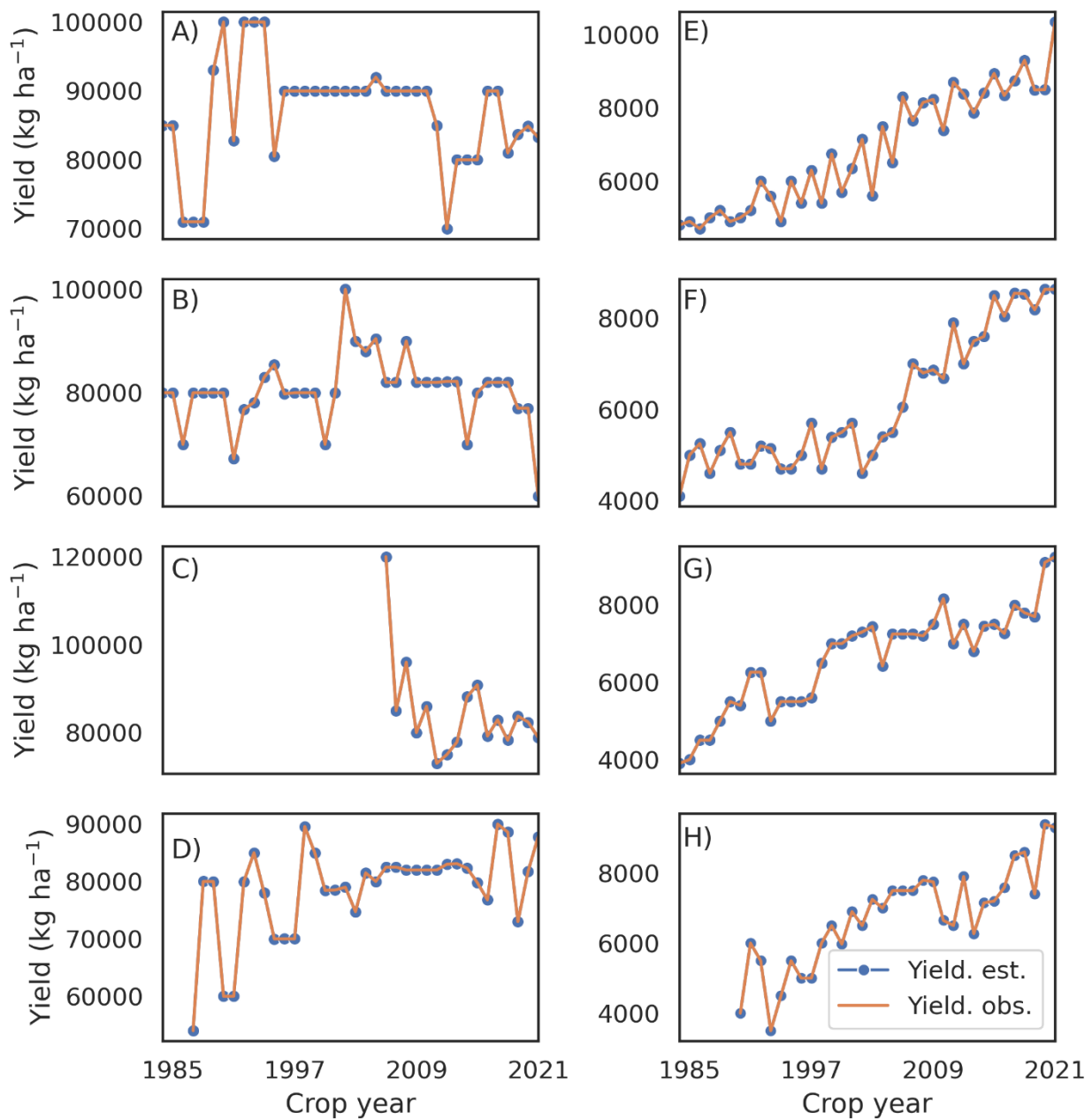


Figure 2. Productivity estimated by the WOFOST model and observed by SIDRA for the municipalities of: (A) Barretos; (B) Morro Agudo; (C) Quirinópolis; (D) Goiatuba; (E) Uruguaiana; (F) Santa Vitória do Palmar; (G) Turvo; (H) Forquilha.

Moon et al. (2023) conducted a calibration study of bell pepper (*Capsicum annuum* var. *Annuum*) in South Korea, using Bayesian optimization by HyperOpt, comparing it with nonlinear optimization by NLOpt. To do this, they used all 21 available food source crops. They concluded that, in this case, NLOpt was able to calibrate only the LAI parameter, finding it difficult to calibrate 40 culture parameters (Moon et al., 2023) demonstrating a superiority of HyperOpt for this situation. However, our study showed high efficiency in the calibration of 20 and 39 parameters of sugarcane and rice

crops, respectively, showing high efficiency during the procedure. In addition, the adjusted parameters were applied in the WOFOST yield simulation automatically and throughout the time series, solving the problems of different cultivars and technological limitation (Bassu et al., 2021).

3.2 WOFOST model results

The carbon balance of the sugarcane and rice yield simulations is shown in Figure 3.

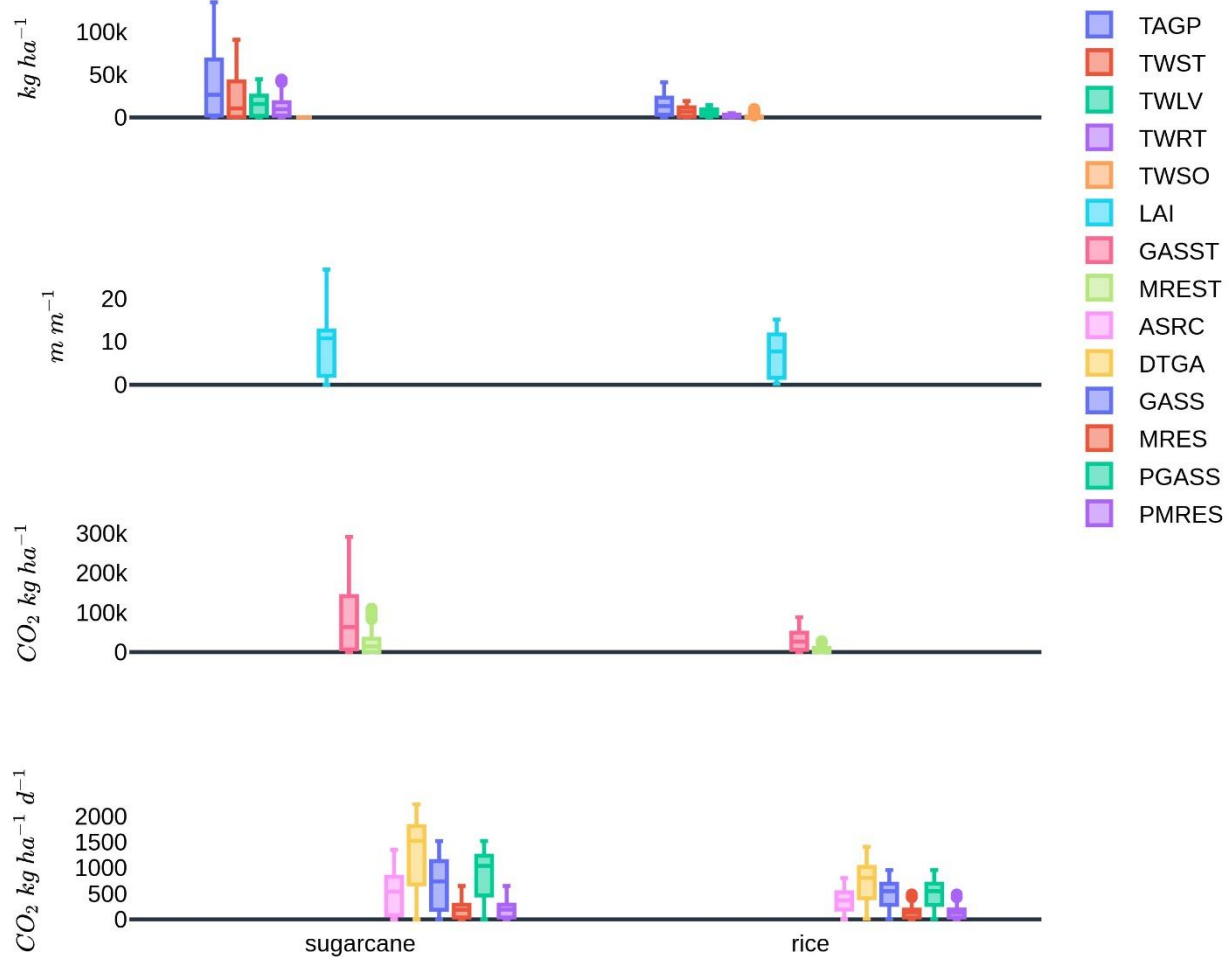


Figure 3. Result of the WOFOST carbon balance for sugarcane and rice crops

DTGA and cCO₂ showed a negative Spearman's correlation ($\rho=-0.45$; $p\text{-value}<0.001$). This shows that the relationship between the DTGA model, derived from a simulation based on physical and meteorological factors, and cCO₂

can be used to weave causal relationships of assimilation in the face of climate change or meteorological events, for example.

In sugarcane (Figure 4-A) and rice (Figure 4-B) crops, the influence of absorption on $c\text{CO}_2$ differs. Despite similar behaviors, sugarcane cultivation ($\rho=-0.4$; $p\text{-value}<0.001$) has more influence on $c\text{CO}_2$ uptake than rice cultivation ($\rho=0.03$; ns). In both scenarios, the concentration of $c\text{CO}_2$ increases in the off-season (Table 4) and decreases during the vegetative development of the crops.

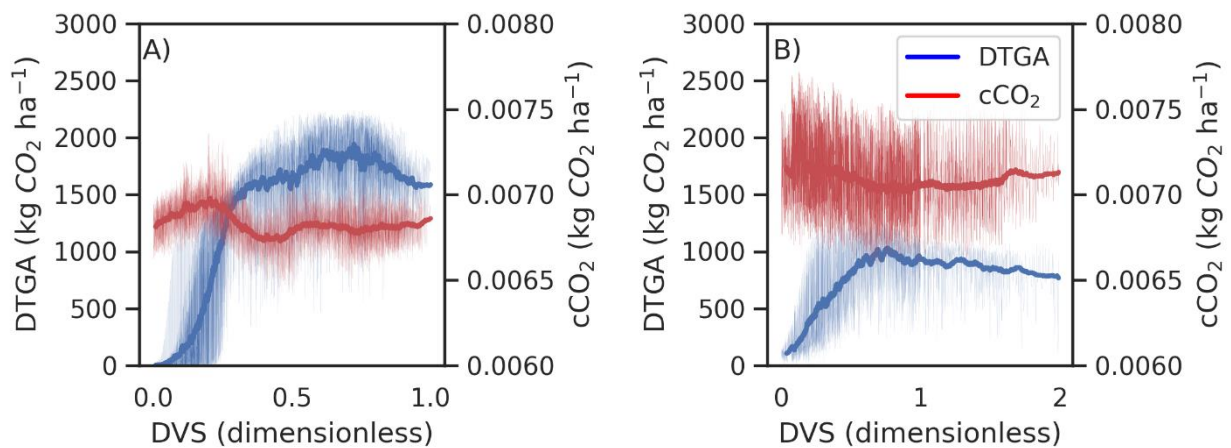


Figure 4. Comparison between the average gross $c\text{CO}_2$ assimilation of the studied sites, in the historical series, as a function of the development of sugarcane (A) and rice (B) cultivation. The center lines correspond to the moving average in the 80-day window.

Table 4. Descriptive statistics of the DTGA and $c\text{CO}_2$ variables of the locations studied, throughout the time series.

	Crop			
	Sugarcane		Rice	
	DTGA	$c\text{CO}_2$	DTGA	$c\text{CO}_2$
Sample size	10479		4184	
Mean	1273.09	6.84e^{-3}	718.12	7.10e^{-3}
Max.	2239.61	7.36e^{-3}	1405.24	7.75e^{-3}
Min.	0.18	6.43e^{-3}	12.82	6.59e^{-3}
Standard desv.	722.87	1.19e^{-4}	359.44	2.27e^{-4}
Standard error	7.06	1.16e^{-6}	5.56	3.52e^{-6}

The difference in CO_2 absorption can be attributed to the difference in C_3 and C_4 metabolism of rice and sugarcane, respectively. Sugarcane has the enzyme PEPcase (phosphoenolpyruvate carboxylase), which is more efficient

in capturing CO₂ for prefixing malate or aspartate, two 4-carbon molecules (Taiz and Zeiger, 2010). On the other hand, rice has RUBISCO activity, which is highly dependent on temperature conditions and stomatal conductance (Bunce, 2000). As a result, it is possible to observe a greater influence of CO₂ absorption in sugarcane (La Scala Júnior et al., 2012), when compared to rice.

Regarding the CO₂ content present in the different systems, sugarcane cultivation presents lower average concentrations of cCO₂ when compared to rice (6.84e^{-3} ; 7.10e^{-3} kg ha⁻¹, respectively). Machado et al. (2017) show that crops such as corn and sugarcane have a greater mitigation potential in the sequestration of CO₂ in the form of soil organic matter, while rice cultivation has the lowest averages of this soil carbon stock. The crops have an important CO₂ sink mainly in their vegetative stage ($0 < \text{DVS} < 0.5$; $0 < \text{DVS} \leq 1$, sugarcane, and rice, respectively), but they differ according to the location. In periods of senescence, this characteristic is slowed down and the net emission of CO₂ exceeds assimilation, but during the off seasons, it makes use of preparations and intensive use of agricultural inputs, acting as a source of GHGs (Machado et al., 2017).

3.3 Random Forest Model

The test phase of the RF models showed the following results: $R^2_{\text{adj}} = 0.85$ and 0.76 ; $\text{RMSE} = 2.05$ and 2.63 ppm; $\text{MAE} = 1.37$ and 1.70 ppm for the models under sugarcane (Figure 5) and rice (Figure 6).

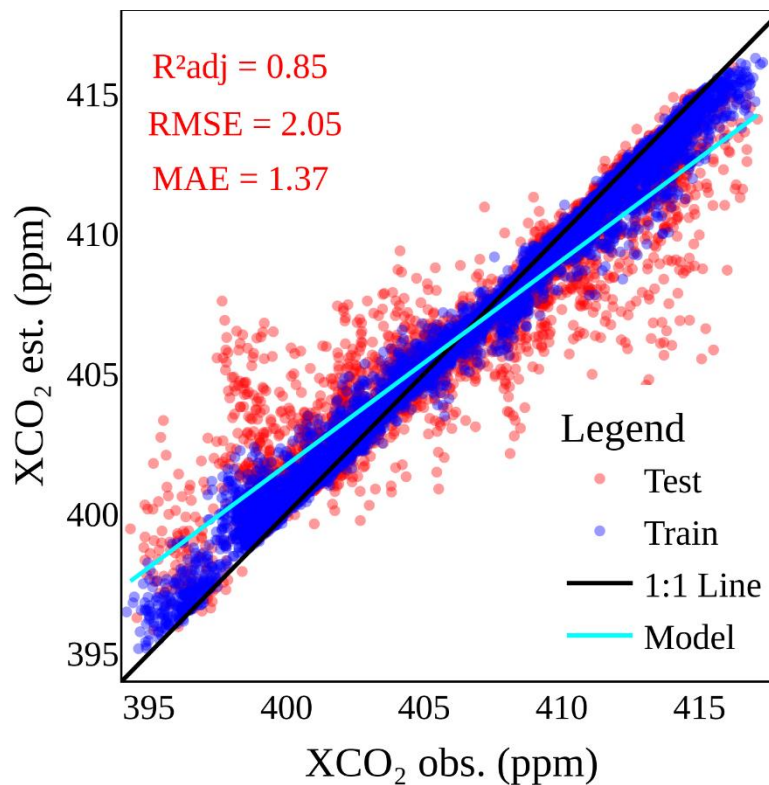


Figure 5. Performance of the testing of the Random Forest model for sugarcane crop

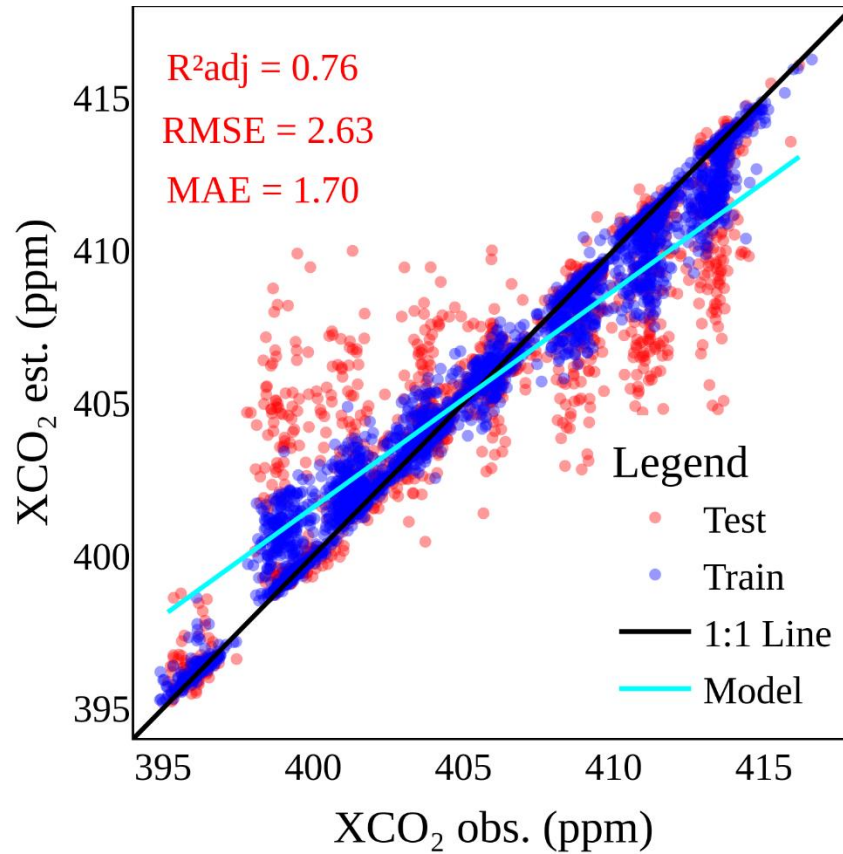


Figure 6. Testing performance of the Random Forest model for rice crop.

In general, the models showed an overestimation for XCO_2 values lower than 405 ppm and underestimation for values greater than 405 ppm. Models locally can be developed to minimize RMSE and MAE, making it possible to compose historical XCO_2 data that predate the start of the OCO-2/NASA project.

For sugarcane cultivation (Figure 7), the variables that most contributed to the estimate were: DVS, SM, MRES, TWLV and LAI. For this cultivation, DVS was the most important variable. Low DVS values, i.e., when the crop is budding, have a negative impact on the estimate, while high values (vegetative phase) have a positive impact on the XCO_2 estimation. In a study comparing the content of XCO_2 and sun-induced chlorophyll fluorescence (SIF), the results indicate a significant decrease in the concentration of XCO_2 in phases of high photosynthetic activity (Morais Filho et al., 2021). This fact is confirmed when we observe the MRES, which has a positive impact when the values are low/intermediate while it has a negative impact when they are high, i.e., periods of photo respiration can compromise the estimate.

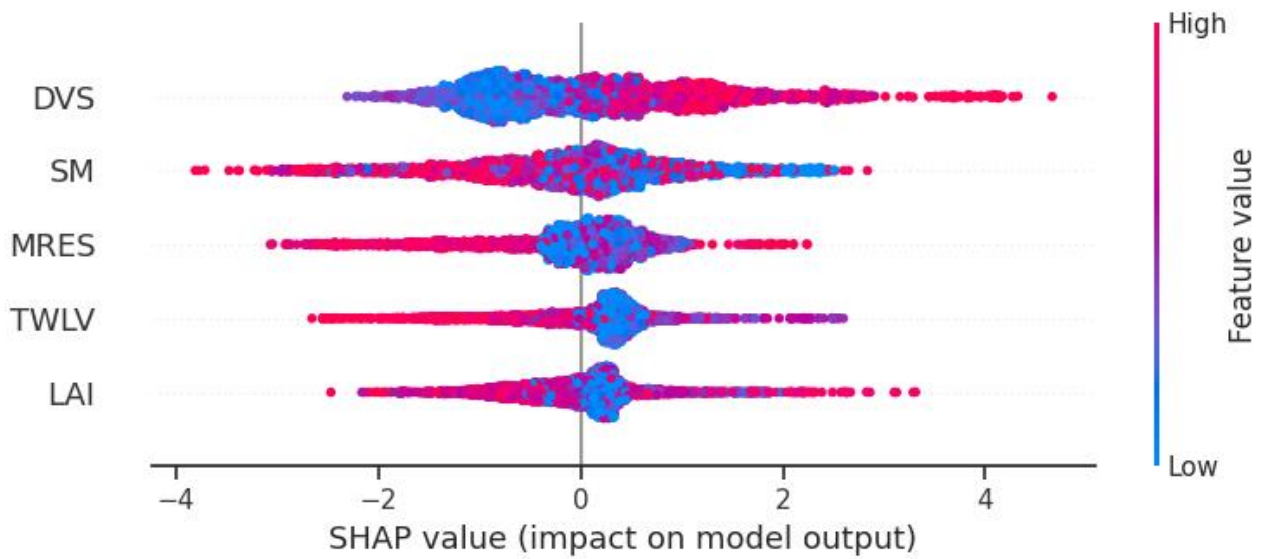


Figure 7. Impact of the best features of the RF model for sugarcane crop. Variables of the Y axis show the importance of the variables in descending order. X-axis (SHAP values) represent the impact of each variable on the XCO₂ prediction.

High soil moisture (SM) had a negative relationship with the XCO₂ estimate. More intermediate or low values tend to have a positive relationship in the estimate. This result is in line with La Scala et al. (2006) concluding that soil moisture has a more relevant influence, when compared to soil temperature, on CO₂ emissions in no-tillage systems. This is related to the microbial activity of the soil that decomposes materials deposited in the soil and the organic matter of the soil.

Low TWLV and LAI values have a positive impact on the XCO₂ estimate. High values are related to high leaf density, resulting in shading, water stress, and senescence, negatively impacting the estimate (Berger et al., 2019). Since LAI is directly affected by the quantity of plants, and is related to CO₂ absorption, high values of this index may indicate a competition for nutrients and water resources (Figueiredo et al., 2005), impacting the XCO₂ estimate.

For rice cultivation (**Error! Reference source not found.**), the most impactful variables were RD, TWLV, LAI, TWRT and TRA. High values of RD and TWRT had both a positive and negative impact on the estimate, while low/intermediate values tended to have a positive impact. The root morphology of rice and its relationship with water availability are still not widely understood (Kato et al., 2006), even in flooded crops. However, studies indicate that the vertical distribution is more related to the length of the deep root under flooded cultivation (Kato et al., 2006).

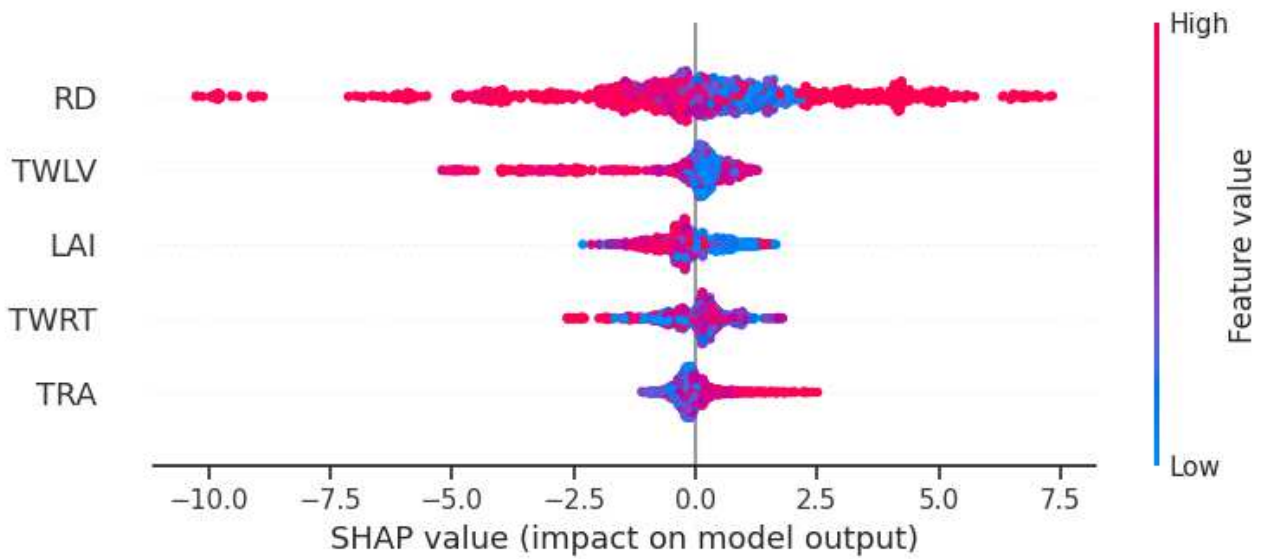


Figure 8. Impact of the best features in the RF model for rice crop.

LAI and TWLV, unlike sugarcane, are less affected by shading, as the water depth that rests on the soil reflects the light that falls on the shaded leaves (Embrapa, 2023), and high values of both variables represent an increase in TRA, positively influencing the estimate of XCO_2 when values are high. This indicates that the transpiration process of plants is associated with photosynthetic activity, since the absorption of CO_2 by the stomata leads to plant transpiration (Taiz and Zeiger, 2010).

4 CONCLUSION

In this study, we developed an alternative method to estimate XCO_2 in rainfed sugarcane and flood-irrigated rice farming systems in the main producing areas of Brazil, collaborating with the study of different climate change scenarios. To this end, we propose a new approach to estimate XCO_2 in systems from a WOFOST ensemble with RF.

The main conclusions of the research are: (1) DTGA, the result of the WOFOST model, has a significant negative Spearman's correlation ($\rho=-0.45$; $p\text{-value}<0.001$) with cCO_2 , obtained by OCO-2/NASA and converted to $kg\ ha^{-1}$; (2) Sugarcane (C_4) has an influence on the concentration of cCO_2 ($\rho=-0.4$; $p\text{-value}<0.001$), while rice (C_3) has no significant influence, demonstrating the mitigating potential that sugarcane cultivation has in relation to rice cultivation, especially during the vegetative stage ($0>DVS<0.5$); (3) The ensemble between the WOFOST mechanistic model and the Random Forest machine learning model were able to estimate the atmospheric XCO_2 for each study site with $R^2\ adj. = 0.85$ and 0.76 ; $RMSE=2.05$ and $2.63\ ppm$; $MAE = 1.37$ and 1.70 ; (4) The features that most contributed to the RF model of sugarcane and rice were, respectively: (i) DVS, SM, MRES, TWLV and LAI; (ii) RD, TWLV, LAI, TWRT and TRA.

Finally, the present work contributed to the UN's sustainable development (SDGs) with goals 2, 13, 15, that is, in the fight for zero hunger and sustainable agriculture, actions against climate change and, consequently, life on land (ONU, 2015).

5 CREDIT AUTHORSHIP CONTRIBUTION STATEMENT

Henrique Fontellas Laurito: Conceptualization, Writing-original draft, Software, Formal analysis, Visualization, Validation, Investigation, editing. **Thais Rayane Gomes da Silva:** Review. **Newton La Scala Jr.:** Review, Conceptualization. **Alan Rodrigo Panosso:** Review, Conceptualization. **Glauco de Souza Rolim:** Review, Supervision, Conceptualization, Methodology, Project administration, Funding acquisition.

6 DECLARATION OF COMPETING INTEREST

The authors declare no competing interests.

7 DATA AVAILABILITY

The python package developed for XCO₂ data extraction is available either in GitHub⁸ or PyPi⁹. The WOFOST simulation data will be available in Mendeley database:

Fontellas Laurito, Henrique (2024), "ESTIMATION OF XCO₂ IN RICE AND SUGARCANE CROPS WITH AN ENSEMBLE OF WOFOST AND RANDOM FOREST", Mendeley Data, V1, doi: 10.17632/m2gnp7j7cn.1

8 ETHICS APPROVAL DECLARATION

Nothing to declare.

9 ACKNOWLEDGMENTS

The present study was supported by the National Council for Scientific and Technological Development (CNPq) (Nos. 130068/2022-5.). We would like to express our gratitude to Newton La Scala, Alan Rodrigo Panosso and Thaís Rayane Gomes da Silva for the insightful considerations and contributions to our work. We would also like to thank Glauco de Souza Rolim for this invaluable mentoring and guidance throughout this research.

⁸ <https://github.com/henriquefl24/xcodex>

⁹ <https://pypi.org/project/xcodex/>

10 REFERENCES

- Bassu, S., Fumagalli, D., Toreti, A., Ceglar, A., Giunta, F., Motzo, R., Zajac, Z., Niemeyer, S., 2021. Modelling potential maize yield with climate and crop conditions around flowering. *Field Crops Research* 271, 108226–108226. <https://doi.org/10.1016/j.fcr.2021.108226>
- Bentéjac, C., Csörgő, A., Martínez-Muñoz, G., 2021. A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review* 54, 1937–1967. <https://doi.org/10.1007/s10462-020-09896-5>
- Berger, R., Silva, J.A.A. da, Ferreira, R.L.C., Candeias, A.L.B., Rubilar, R., 2019. Índices de vegetação para a estimativa do Índice de Área Foliar em plantios clonais de *Eucalyptus saligna* Smith. *Ciência Florestal* 29, 885–899. <https://doi.org/10.5902/1980509816942>
- Breiman, L., 2001. Random Forests. *Machine Learning* 45, 5–32. <https://doi.org/10.1023/A:1010933404324>
- Bunce, J.A., 2000. Acclimation of photosynthesis to temperature in eight cool and warm climate herbaceous C3 species: Temperature dependence of parameters of a biochemical photosynthesis model. *Photosynthesis Research* 63, 59–67. <https://doi.org/10.1023/A:1006325724086>
- Chen, J., Xiao, W., Zheng, C., Zhu, B., 2020. Nitrogen addition has contrasting effects on particulate and mineral-associated soil organic carbon in a subtropical forest. *Soil Biology and Biochemistry* 142, 107708–107708. <https://doi.org/10.1016/j.soilbio.2020.107708>
- Confalonieri, R., Acutis, M., Bellocchi, G., Donatelli, M., 2009. Multi-metric evaluation of the models WARM, CropSyst, and WOFOST for rice. *Ecological Modelling* 220, 1395–1410. <https://doi.org/10.1016/j.ecolmodel.2009.02.017>
- Crippa, M., Solazzo, E., Guizzardi, D., Monforti-Ferrario, F., Tubiello, F.N., Leip, A., 2021. Food systems are responsible for a third of global anthropogenic GHG emissions. *Nature Food* 2, 198–209. <https://doi.org/10.1038/s43016-021-00225-9>
- de Wit, A., Boogaard, H., 2021. A gentle introduction to WOFOST. Wageningen Environmental Research. 47–47.
- Dos Santos, J.E., De Fiori, A., 2005. Perception of environmental impacts in relation to land use. *International Journal of Environment and Sustainable Development*. <https://doi.org/10.1504/IJESD.2005.007238>
- Eldering, A., Wennberg, P.O., Crisp, D., Schimel, D.S., Gunson, M.R., Chatterjee, A., Liu, J., Schwandner, F.M., Sun, Y., O'Dell, C.W., Frankenberg, C., Taylor, T., Fisher, B., Osterman, G.B., Wunch, D., Hakkarainen, J., Tamminen, J., Weir, B., 2017. The Orbiting Carbon Observatory-2 early science investigations of regional carbon dioxide fluxes. *Science* 358. <https://doi.org/10.1126/science.aam5745>
- Embrapa, 2023. Métodos de irrigação - Portal Embrapa.
- Fang, J., Chen, B., Zhang, H., Dilawar, A., Guo, M., Liu, C., Liu, S., Gemechu, T.M., Zhang, X., 2023. Global Evaluation and Intercomparison of XCO₂ Retrievals from GOSAT, OCO-2, and TANSAT with TCCON. *Remote Sensing* 15. <https://doi.org/10.3390/rs15205073>
- Figueiredo, J., Neto, D., de Oliveira, R., Manfron, P., Martin, T., 2005. MODELO PARA ESTIMATIVA DO ÍNDICE DE ÁREA FOLIAR DA CULTURA DE MILHO. *Revista Brasileira de Milho e Sorgo* 4, 8–13.
- Flack-Prain, S., Shi, L., Zhu, P., Rocha, H.R., Cabral, O., Hu, S., Williams, M., 2021. The impact of climate change and climate extremes on sugarcane production. *GCB Bioenergy* 13, 408–424. <https://doi.org/10.1111/gcbb.12797>
- Hajima, T., Watanabe, M., Yamamoto, A., Tatebe, H., Noguchi, M.A., Abe, M., Ohgaito, R., Ito, Akinori, Yamazaki, D., Okajima, H., Ito, Akihiko, Takata, K., Ogochi, K., Watanabe, S., Kawamiya, M., 2020. Development of the MIROC-ES2L Earth system model and the evaluation of biogeochemical processes and feedbacks. *Geoscientific Model Development* 13, 2197–2244. <https://doi.org/10.5194/gmd-13-2197-2020>
- Holzworth, D.P., Snow, V., Janssen, S., Athanasiadis, I.N., Donatelli, M., Hoogenboom, G., White, J.W., Thorburn, P., 2015. Agricultural production systems modelling and software: Current status and future prospects. *Environmental Modelling & Software* 72, 276–286. <https://doi.org/10.1016/j.envsoft.2014.12.013>
- Hu, S., Shi, L., Huang, K., Zha, Y., Hu, X., Ye, H., Yang, Q., 2019. Improvement of sugarcane crop simulation by SWAP-WOFOST model via data assimilation. *Field Crops Research* 232, 49–61. <https://doi.org/10.1016/j.fcr.2018.12.009>
- IPCC, 2023. AR6 Synthesis Report: Climate Change 2023.

- Johnson, J., 2021. NLOpt Python Reference - NLOpt Documentation.
- Kato, Y., Abe, J., Kamoshita, A., Yamagishi, J., 2006. Genotypic Variation in Root Growth Angle in Rice (*Oryza sativa* L.) and its Association with Deep Root Development in Upland Fields with Different Water Regimes. *Plant and Soil* 287, 117–129. <https://doi.org/10.1007/s11104-006-9008-4>
- Khatri-Chhetri, A., Sapkota, T.B., Sander, B.O., Arango, J., Nelson, K.M., Wilkes, A., 2021. Financing climate change mitigation in agriculture: assessment of investment cases. *Environmental Research Letters* 16, 124044–124044. <https://doi.org/10.1088/1748-9326/ac3605>
- Kristiansen, T., Butenschön, M., Peck, M.A., 2024. Statistically downscaled CMIP6 ocean variables for European waters. *Scientific Reports* 2024 14:1 14, 1–20. <https://doi.org/10.1038/s41598-024-51160-1>
- La Scala Júnior, N., De Figueiredo, E.B., Panosso, A., 2012. A review on soil carbon accumulation due to the management change of major Brazilian agricultural activities. *Brazilian Journal of Biology* 72, 775–785. <https://doi.org/10.1590/S1519-69842012000400012>
- La Scala, N., Bolonhezi, D., Pereira, G.T., 2006. Short-term soil CO₂ emission after conventional and reduced tillage of a no-till sugar cane area in southern Brazil. *Soil and Tillage Research* 91, 244–248. <https://doi.org/10.1016/j.still.2005.11.012>
- Li, X., Hu, X., Cai, C., Jia, Q., Zhang, Y., Liu, J., Xue, M., Xu, J., Wen, R., Crowell, S.M.R., 2020. Terrestrial CO₂ Fluxes, Concentrations, Sources and Budget in Northeast China: Observational and Modeling Studies. *Journal of Geophysical Research: Atmospheres* 125. <https://doi.org/10.1029/2019JD031686>
- Li, Z., Ding, L., Xu, D., 2022. Exploring the potential role of environmental and multi-source satellite data in crop yield prediction across Northeast China. *Science of The Total Environment* 815, 152880–152880. <https://doi.org/10.1016/J.SCITOTENV.2021.152880>
- Lundberg, S.M., Lee, S.-I., 2017. A Unified Approach to Interpreting Model Predictions, in: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (Eds.), . Presented at the Advances in Neural Information Processing Systems, Curran Associates, Inc.
- Machado, K.S., Seleme, R., Maceno, M.M.C., Zattar, I.C., 2017. Carbon footprint in the ethanol feedstocks cultivation – Agricultural CO₂ emission assessment. *Agricultural Systems* 157, 140–145. <https://doi.org/10.1016/j.agsy.2017.07.015>
- Marin, F.R., Jones, J.W., Singels, A., Royce, F., Assad, E.D., Pellegrino, G.Q., Justino, F., 2013. Climate change impacts on sugarcane attainable yield in southern Brazil. *Climatic Change* 117, 227–239. <https://doi.org/10.1007/s10584-012-0561-y>
- Moon, T., Sim, S., Son, J.E., 2023. Calibration of food and feed crop models for sweet peppers with Bayesian optimization. *Horticulture, Environment, and Biotechnology* 64, 615–625. <https://doi.org/10.1007/s13580-022-00510-x>
- Morais Filho, L.F.F., Meneses, K.C. de, Santos, G.A. de A., Bicalho, E. da S., Rolim, G. de S., La Scala Jr, N., 2021. xCO₂ temporal variability above Brazilian agroecosystems: A remote sensing approach. *Journal of Environmental Management* 288, 112433–112433. <https://doi.org/10.1016/j.jenvman.2021.112433>
- ONU, 2015. Objetivos de Desenvolvimento Sustentável no Brasil.
- Ott, Lesley., Weir, Brad., 2022. OCO-2 GEOS Level 3 daily, 0.5x0.625 assimilated CO₂ V10r. Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC) 10–10. <https://doi.org/10.5067/Y9M4NM9MPCGH>
- Paixão, G.M. de M., Santos, B.C., Araujo, R.M. de, Ribeiro, M.H., Moraes, J.L. de, Ribeiro, A.L., 2022. Machine Learning na Medicina: Revisão e Aplicabilidade. *Arquivos Brasileiros de Cardiologia* 118, 95–102. <https://doi.org/10.36660/abc.20200596>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12, 2825–2830.
- Poggio, L., de Sousa, L.M., Batjes, N.H., Heuvelink, G.B.M., Kempen, B., Ribeiro, E., Rossiter, D., 2021. SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. *SOIL* 7, 217–240. <https://doi.org/10.5194/soil-7-217-2021>

- Powell, M.J.D., 1994. A Direct Search Optimization Method That Models the Objective and Constraint Functions by Linear Interpolation, in: *Advances in Optimization and Numerical Analysis*. Springer Netherlands, Dordrecht, pp. 51–67. https://doi.org/10.1007/978-94-015-8330-5_4
- Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., Prabhat, 2019. Deep learning and process understanding for data-driven Earth system science. *Nature* 566, 195–204. <https://doi.org/10.1038/s41586-019-0912-1>
- Semenov, V.A., 2021. Modern Arctic Climate Research: Progress, Change of Concepts, and Urgent Problems. *Izvestiya, Atmospheric and Oceanic Physics* 57, 18–28. <https://doi.org/10.1134/S0001433821010114>
- SIDRA, 2024. Tabela 5457 - Área plantada ou destinada à colheita, área colhida, quantidade produzida, rendimento médio e valor da produção das lavouras temporárias e permanentes. Sistema Automatico de Recuperação - IBGE.
- Taiz, L., Zeiger, E., 2010. Fisiologia vegetal. *Cell*. <https://doi.org/10.4995/WRS.2010.7744>
- Tian, L., Liu, X., Zhang, B., Liu, M., Wu, L., 2017. Extraction of Rice Heavy Metal Stress Signal Features Based on Long Time Series Leaf Area Index Data Using Ensemble Empirical Mode Decomposition. *International Journal of Environmental Research and Public Health* 2017, Vol. 14, Page 1018 14, 1018–1018. <https://doi.org/10.3390/IJERPH14091018>
- Tian, Y., Hu, S., Deser, C., 2023. Critical role of biomass burning aerosols in enhanced historical Indian Ocean warming. *Nature Communications* 2023 14:1 14, 1–8. <https://doi.org/10.1038/s41467-023-39204-y>
- Van Diepen, C.A., Rappoldt, C., Wolf, J., Van Keulen, H., 1988. Crop growth simulation model WOFOST. Documentation version 4.
- Van Heemst, H.D.J., 1988. Plant data values required for simple crop growth simulation models: review and bibliography.
- Weiss, M., Jacob, F., Duveiller, G., 2020. Remote sensing for agricultural applications: A meta-review. *Remote Sensing of Environment* 236, 111402–111402. <https://doi.org/10.1016/j.rse.2019.111402>
- WRB, 2024. World Reference Base for Soil Resources (WRB) | ISRIC.
- Wunch, D., Wennberg, P.O., Osterman, G., Fisher, B., Naylor, B., Roehl, C.M., O'Dell, C., Mandrake, L., Viatte, C., Kiel, M., Griffith, D.W.T., Deutscher, N.M., Velazco, V.A., Notholt, J., Warneke, T., Petri, C., De Maziere, M., Sha, M.K., Sussmann, R., Rettinger, M., Pollard, D., Robinson, J., Morino, I., Uchino, O., Hase, F., Blumenstock, T., Feist, D.G., Arnold, S.G., Strong, K., Mendonca, J., Kivi, R., Heikkinen, P., Iraci, L., Podolske, J., Hillyard, P.W., Kawakami, S., Dubey, M.K., Parker, H.A., Sepulveda, E., García, O.E., Te, Y., Jeseck, P., Gunson, M.R., Crisp, D., Eldering, A., 2017. Comparisons of the Orbiting Carbon Observatory-2 (OCO-2) X_{CO_2} measurements with TCCON. *Atmospheric Measurement Techniques* 10, 2209–2238. <https://doi.org/10.5194/amt-10-2209-2017>
- Xu, X., Ouyang, X., Gu, Y., Cheng, K., Smith, P., Sun, J., Li, Y., Pan, G., 2021. Climate change may interact with nitrogen fertilizer management leading to different ammonia loss in China's croplands. *Global Change Biology* 27, 6525–6535. <https://doi.org/10.1111/gcb.15874>

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [GraphicalAbstract.jpeg](#)