

RESEARCH

Open Access



An empirical model for estimating daily atmospheric column-averaged CO₂ concentration above São Paulo state, Brazil

Luis Miguel da Costa^{1*} , Gustavo André de Araújo Santos^{1,2,3}, Alan Rodrigo Panosso¹,
Glauco de Souza Rolim¹ and Newton La Scala¹

Abstract

Background: The recent studies of the variations in the atmospheric column-averaged CO₂ concentration (X_{CO_2}) above croplands and forests show a negative correlation between X_{CO_2} and Sun Induced Chlorophyll Fluorescence (SIF) and confirmed that photosynthesis is the main regulator of the terrestrial uptake for atmospheric CO₂. The remote sensing techniques in this context are very important to observe this relation, however, there is still a time gap in orbital data, since the observation is not daily. Here we analyzed the effects of several variables related to the photosynthetic capacity of vegetation on X_{CO_2} above São Paulo state during the period from 2015 to 2019 and propose a daily model to estimate the natural changes in atmospheric CO₂.

Results: The data retrieved from the Orbiting Carbon Observatory-2 (OCO-2), NASA-POWER and Application for Extracting and Exploring Analysis Ready Samples (AppEEARS) show that Global Radiation (Qg), Sun Induced Chlorophyll Fluorescence (SIF) and, Relative Humidity (RH) are the most significant factors for predicting the annual X_{CO_2} cycle. The daily model of X_{CO_2} estimated from Qg and RH predicts daily X_{CO_2} with root mean squared error of 0.47 ppm (the coefficient of determination is equal to 0.44, $p < 0.01$).

Conclusion: The obtained results imply that a significant part of daily X_{CO_2} variations could be explained by meteorological factors and that further research should be done to quantify the effects of the atmospheric transport and anthropogenic emissions.

Keywords: Carbon cycle, Remote sensing, OCO-2, Stepwise regression analysis, Climate change, Meteorology

Background

Understanding the variability of atmospheric carbon dioxide (CO₂) concentration in time and space is a crucial task so that we can adopt mitigation strategies. In this sense, several studies analyze the average concentration of this greenhouse gas not only on a global scale [1, 2] but also to estimate anthropogenic emissions in urban

centers [3, 4]. Other studies focus on understanding the column-averaged of carbon in the atmosphere (X_{CO_2}) above tropical forests [5], or above agriculture crops in different seasons of the year [6, 7].

In a recent regional study, da Costa et al. [7] analyze the spatio-temporal variability of X_{CO_2} in a sugarcane-producing area in the southeast region of Brazil. They observed an important inverse relationship between the average carbon concentration in the atmosphere with climatic and vegetative variables. Concluding that the dependence of the natural carbon cycle is related to the predominant agriculture crop in the region and how Global Radiation (Qg), relative humidity (RH), and

*Correspondence: luism costa00@gmail.com

¹ Department of Engineering and Exact Sciences, São Paulo State University, Via de Acesso Prof. Paulo Donato Castellane s/n, Jaboticabal, São Paulo 14884-900, Brazil

Full list of author information is available at the end of the article



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

the Sun Induced Chlorophyll Fluorescence (SIF) was related to this behavior. Similarly, Morais Filho et al. [6] conducted a study that analyzed three different crops and the temporal variability of X_{CO_2} and SIF in these environments, they also found a significant negative correlation between these variables.

However, there is still a temporal gap in the X_{CO_2} data collected by remote sensing, since the measurements are not daily [8, 9]. This type of measurement is important to several factors, such as, estimate the potential capability of atmospheric CO_2 assimilation by vegetation, establishing public strategies at local levels for climate adaptation and mitigation, and even in economy incorporating daily trends in the carbon market and ecosystems services payments [10–16].

Daily CO_2 measurements can be made using the Eddy Covariance technique [17–19], although this has the disadvantage of being a point (local) study. In this sense using orbital data, such as Orbiting Carbon Observatory-2 (OCO-2), has become more common [1, 2]. Remote sensing data also can be used to estimate the daily variations of different aspects (e.g., climate, meteorological, land-use changes, ecosystems services) for a larger area [16, 20–23].

Several studies confirm that photosynthesis is the main regulator of atmospheric carbon sinks [10, 24–26]. However, photosynthesis is a process sensitive to climatic variations such as relative humidity [27], precipitation [28], evapotranspiration [29], and incident solar irradiance [30].

Therefore, the natural cycle of CO_2 is dependent on several aspects, such as vegetation and climate, being necessary data from several different bases for understanding this dynamic [7], turning pre-processing techniques and analysis of autocorrelations necessary, since the multicollinearity introduces an uncertainty due to the model overfit [31, 32]. In this sense, we aim to model the atmospheric CO_2 cycle above the state of São Paulo to estimate the time changes on a daily scale, based on vegetative and climatic variables retrieved from different orbital platforms, applying a technique to remove the collinearity and after employing a step-wise forward selection, improving in this way the regional understanding of CO_2 .

One of our assumptions, is given that we detrend the X_{CO_2} to maintain only the variability related to the natural interactions [6, 33], the transport by the wind in the atmosphere is not significant, other studies such Hakkarianen et al. [34], that proposed an anomaly model of X_{CO_2} , also disregard the atmospheric and wind transport in their study, however, this introduces a limitation of our approach that not account this aspect [35]. In the same way, the trend and increase

due to anthropogenic sources are also simplified by this detrend.

Results

Variance Inflation Factor (VIF) analysis (Table 1) shows it was possible to reduce the number of variables related to X_{CO_2} (according to the adopted criterion, $VIF < 10$) as shown comparing Fig. 1a with b, before and after the selection, respectively, and therefore reducing the overfit source of uncertainty. Despite wind speed (Ws) had a $VIF < 10$, the Pearson's correlation was not significant ($p > 0.05$). Variables most related to X_{CO_2} were the Global Radiation (Qg), Sun-Induced chlorophyll Fluorescence at 757 nm (SIF 757), and Relative Humidity (RH).

Regarding the temporal variability of X_{CO_2} , the maximum mean for the analyzed period was 393.09 ± 0.17 ppm and occurred in October 2019, while the minimum average was in November 2018, being 390.11 ± 0.15 ppm (Fig. 2a). Meanwhile, the Qg (Fig. 2b) ranged between 24.3 ± 0.09 and 13.07 ± 0.04 ($MJ\ m^{-2}\ day^{-1}$), with the maximum average occurring in December 2018 and the minimum in June of the same year.

SIF 757 (Fig. 2d) had the highest average recorded in the period in November 2015 [1.1 ± 0.05 ($Wm^{-2}\ sr^{-1}\ \mu m^{-1}$)] and the lowest in September 2017 [0.3 ± 0.06 ($Wm^{-2}\ sr^{-1}\ \mu m^{-1}$)], while the Relative Humidity (Fig. 2c) ranged from 84.86 ± 0.07 to $70.44 \pm 0.19\%$, where the highest mean was observed in March 2016 and the lowest in October 2019.

Regarding SIF 757 the minimum averages occurred in June of 2015 and 2016, September 2017, November 2018, and July 2019, ranging from 0.3 to $0.46\ Wm^{-2}\ sr^{-1}\ \mu m^{-1}$ (Fig. 2d). The minimum Qg averages vary

Table 1 Variance Inflation Factor (VIF) of the studied variables

Variable	VIF
Qg	9.35
RH	5.10
SIF 757	1.81
Prec	10.13
Temp	21.43
Ws	4.06
LST	19.54
NDVI	22.15
LAI	87.21
Fpar	65.40
ET	33.47

Qg Global radiation, *RH* Relative humidity, *SIF 757* Solar-Induced Chlorophyll Fluorescence at 757 nm, *Prec* Precipitation, *Temp* Temperature at 2 m, *Ws* Wind Speed, *LST* Land Surface Temperature (MODIS), *NDVI* Normalized Difference Vegetation Index, *LAI* Leaf Area Index, *ET* Evapotranspiration

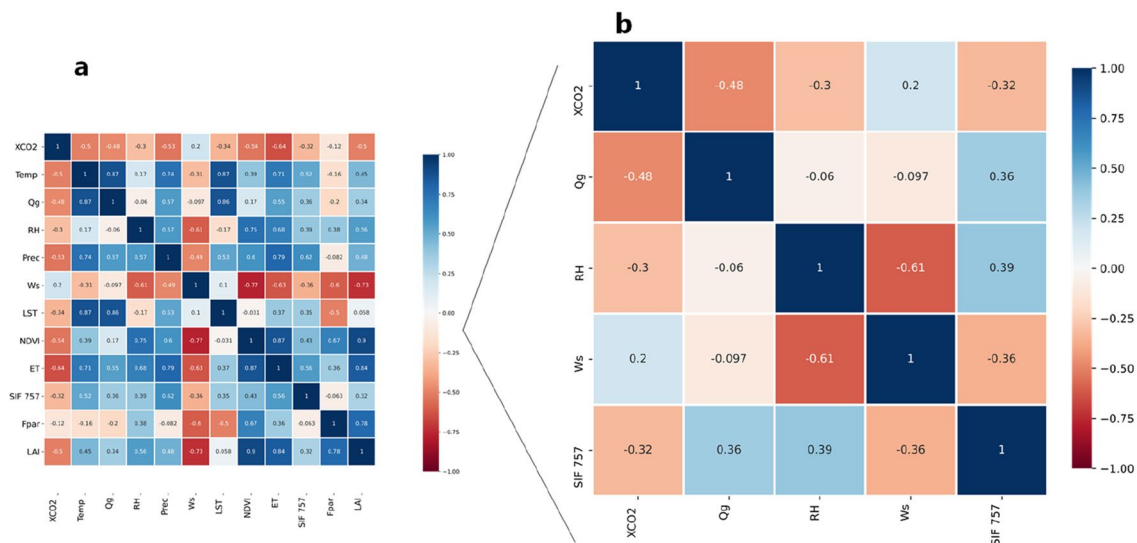


Fig. 1 Heatmap of the Pearson's correlation matrix, where: **a** before the Variance Inflation Factor (VIF) selection and **b** after the selection by Variance Inflation Factor (VIF)

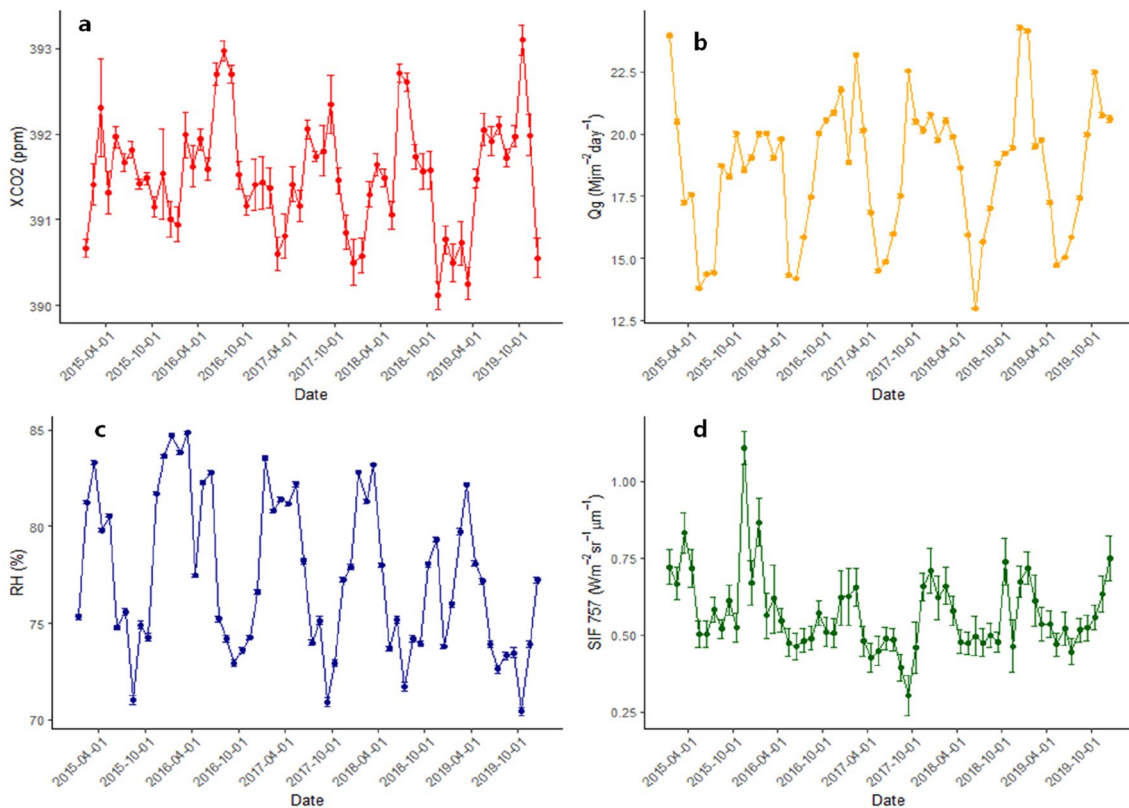


Fig. 2 Monthly variability of X_{CO_2} (**a**), Q_g (**b**), RH (**c**), and $SIF\ 757$ (**d**) over the period from January 2015 to December 2019. Where X_{CO_2} column average of carbon dioxide in the atmosphere (ppm), Q_g global radiation ($MJ\ m^{-2}\ day^{-1}$), RH relative humidity (%), $SIF\ 757$ sun-induced chlorophyll fluorescence at 757 nm ($Wm^{-2}\ sr^{-1}\ \mu m^{-1}$)

between May and June for the entire series approximately between 13.07 and 14.71 MJ m⁻² day⁻¹ (Fig. 2b). Maximum Qg averages are concentrated between December and January of each year, reaching 24 MJ m⁻² day⁻¹ in those months (Fig. 2b). The maximum average of SIF 757 occurs between November and February of each year, ranging from 0.8 to 1.1 Wm⁻² sr⁻¹ μm⁻¹ (Fig. 2d).

The stepwise forward selection method, with multiple cross-validation, had the best result with two variables, with a root mean squared error (RMSE) of ~0.60 ppm in the training sample (Fig. 3), the selected variables being Qg and RH, respectively (Eq. 1).

$$X_{CO_2 (daily)} = 391.484 (\pm 0.89) - (\pm 0.089) \times Qg - 0.263 (\pm 0.09) \times RH \quad (1)$$



Fig. 3 RMSE score for the training sample

The model built in the training (Eq. 1) was applied in the test sample of the variables cited (Qg and RH), and from the cross-validation of the estimated data with the observed data, we observe an R² of 0.44, the values of the metrics MSE, RMSE, and MAE were 0.22, 0.47, and 0.37 (ppm) respectively, and for MAPE we found a value of 1.54% (p < 0.01) (Fig. 4a), with this we were able to reduce the time scale of the OCO-2 satellite from every 15 days to a daily scale (Fig. 4b).

Discussion

The natural annual cycle of X_{CO₂} is affected by factors related to climate and vegetation aspects [6, 36, 37]. Due to the VIF analysis, we were able to summarize three main factors for São Paulo state: Global Radiation (Qg), Relative Humidity (RH) and Sun-Induced chlorophyll Fluorescence at 757 nm (SIF 757), reducing the uncertainties in the model formulation since we removed the overfit caused by multicollinearity [31, 32]. Several studies have already been conducted using this method to identify which variables select for ecological studies [38], computational studies [39], and remote sensing studies [40].

Except for wind speed (Ws), all variables studied correlated negatively with X_{CO₂} (Fig. 1), hence, related to the sink of atmospheric CO₂. The non-significant correlation between X_{CO₂} and Ws could be related to the detrending of the atmospheric CO₂ concentration (see Methods section), which removes the transport effect and simplify the X_{CO₂} variability only for the biochemical cycle [6, 7, 33]. In general, the highest concentrations of X_{CO₂} are observed in the months corresponding to the Brazilian autumn and winter (April to August) and lowest in the

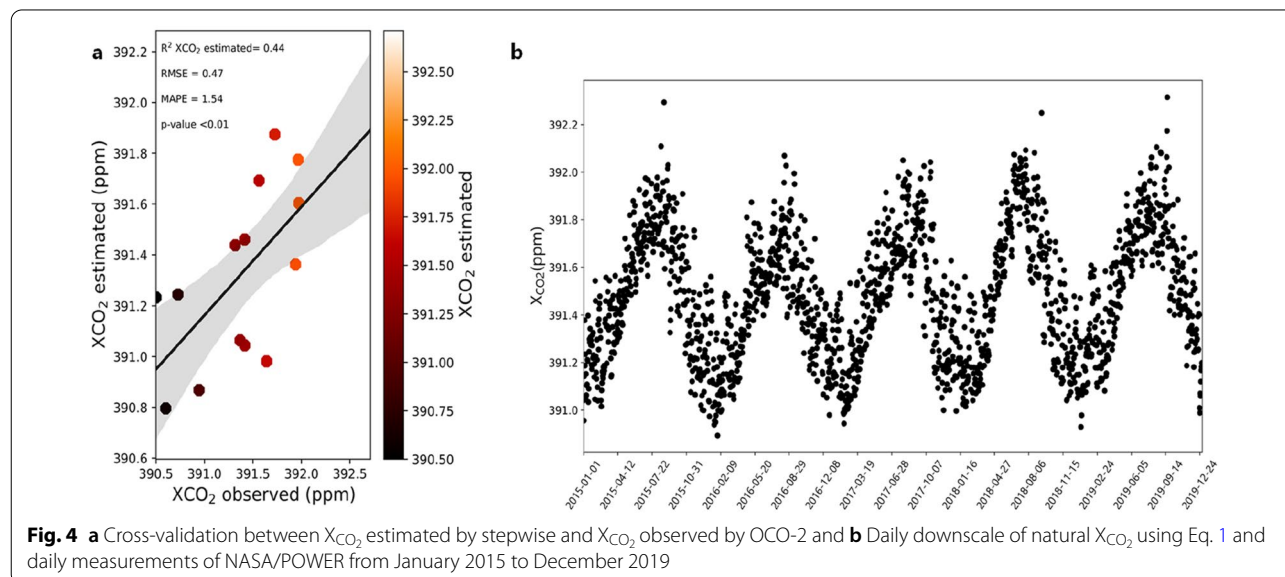


Fig. 4 a Cross-validation between X_{CO₂} estimated by stepwise and X_{CO₂} observed by OCO-2 and b Daily downscale of natural X_{CO₂} using Eq. 1 and daily measurements of NASA/POWER from January 2015 to December 2019

summer, from December to February. Studies such as by Siabi et al. [41] and Falahatkar et al. [42] reported how the different seasons affect the average CO_2 concentration in the atmosphere.

Recently, researches were conducted at regional scales in Brazil such as by Morais Filho et al. [6] and da Costa et al. [7], indicating negative correlations between X_{CO_2} and SIF over agricultural areas, approximately -0.5 and -0.8 , respectively. SIF is a variable directly related to the photosynthesis of plants, laboratory-scale experiments have demonstrated this relation [43], and remote sensing studies at the canopy and global level reported positive relations between SIF and Gross Primary Production, and also a negative correlation between SIF and the X_{CO_2} [5, 44–46].

As a result of photosynthesis, it is expected that SIF increases during summer [7, 41], as in this season, higher precipitation events and higher temperatures are observed [47]. Our results show higher SIF average values in the months when summer occurs in the São Paulo state, and an inverse relationship between SIF and X_{CO_2} . The lowest average values of X_{CO_2} usually occur during the summer period in the study region. This is due to plant CO_2 assimilation [48], printing quasi-periodical X_{CO_2} , and SIF time changes as well as observed in other studies [5, 6, 41, 49].

Most of São Paulo's state has a wet summer and dry winter [47] resulting in a positive correlation between precipitation and SIF (Pearson's correlation = 0.61 and $p < 0.05$), while negative with X_{CO_2} ($r = -0.49$, $p < 0.05$) (Fig. 1a). Precipitation is a photosynthetic control factor, so the greater availability of water that exists in the summer in São Paulo's state induces plants to perform more photosynthesis through primary productivity, which leads to a reduction of atmospheric CO_2 . The opposite is observed in the dry winter because water availability is lower resulting in less photosynthesis, or less CO_2 assimilation by plants, either in natural or agricultural areas [7, 28, 50].

Another effect observed during summer in the region is the increase of relative humidity (RH), which reduces the water transfer between soil or plant to the atmosphere [51], inducing plants to keep their stomata open, where CO_2 assimilation occurs [52]. Studies have already shown the relationship of stomata opening in periods with good water availability is related to plant growth [53, 54]. Thus, establishing the negative relationship between RH and X_{CO_2} , also previously observed by Golkar et al. [27].

In the same way, another requirement for photosynthesis occurs is sunlight, which is the source of energy to carry out the biochemical processes of this phenomenon. Therefore, as the amount of radiation (Qg) is absorbed by

the plant, photosynthesis tends to increase, and consequently higher CO_2 assimilation, decreasing in this way the concentration of this greenhouse gas in the atmosphere [7, 30]. We can observe these relationships in our results (Fig. 3b), Qg correlates positively with SIF, and these variables relate negatively with X_{CO_2} .

Since we are dealing with the natural annual cycle of CO_2 the main factor of the higher concentrations of this gas in the atmosphere is due to the lowest photosynthetic absorption by plants. The autumn and winter have low available water and sunlight for plants, leading to a decrease in photosynthesis, also another important factor is that the annual calendar for agriculture in the state of São Paulo has harvest periods between these seasons [55], and as consequence decreasing the cover area by vegetation. Shekhar et al. [56] show how the crop's grown in summer decrease the values of X_{CO_2} over the Nile Delta and when the harvest starts the values of X_{CO_2} are higher, also, they found that SIF values are higher in the grown season.

Our model was based on Qg and RH, which are two variables related to the CO_2 assimilation process, or CO_2 sink. The model has lower RMSE values than have been reported in previous studies, such as by Guo et al. [57] where the values of this metric ranged from 0.7 to 1.1 ppm. In a more recent study by Taylor et al. [58] when evaluating initial OCO-3 data results from the globe and model-related errors, they found an RMSE between 1 and 2 ppm. Another important measure is the MAPE, which shows in percentage how much we are getting wrong, studies with remote sensing have already demonstrated errors below 10% as being considered extremely low for predicting plant and climate aspects [59, 60]. With this, we can evaluate that the performance of the model proposed in this work presents a very low error.

The coefficient of determination (R^2) was 0.44, an increment of almost 20% from the simple linear fit with Qg alone, which has a higher importance in the model. Although the R^2 is moderate, studies using other orbital sensors such as MODIS to model the average CO_2 concentration in the atmosphere have reported similar results [23]. In addition, we should consider that although OCO-2 and NASA-POWER are two high quality and validated databases [8, 9, 61], the difference between grids and spatial resolution (see Table 2 in Methods and Fig. 5b) cannot be disregarded, as it is an aspect that can influence these results, leading us to consider the coefficient of determination observed in this study as being high.

These differences between the databases can be suppressed by the greater temporal coverage of NASA-POWER, allowing us to estimate the daily temporal variability of the natural CO_2 cycle in the atmosphere for

the state of São Paulo, besides reducing in the future the spatial scale of X_{CO_2} obtained from OCO-2 and gaining greater spatial resolution cover. Other vegetation index-based models aimed at reducing the spatial sampling of OCO-2 data, but focused on SIF, as is the case of Zhang et al. [62] and Yu et al. [63].

Despite the errors associated with the model and the uncertainty measures due to the difference in satellite resolution, an advantage of using models similar to the one proposed here is being able to have a daily measure of the variability of atmospheric CO_2 and how the climate parameters affect this dynamic, also serving as an indirect indicator of how is the daily assimilation capacity of this gas in a region.

Conclusions

In summary, the cycle of X_{CO_2} in the state of São Paulo has higher average values during April to October, periods of lower intensity of rainfall, and is considered as the winter in the state, in the other hand the lowest averages of X_{CO_2} were usually observed between December to March, this period corresponds to the summer, and the inverse behavior was observed for SIF 757, global radiation (Qg) and relative humidity (RH). This pattern is due to the relationship between photosynthesis and Carbon assimilation, given that photosynthesis is a process sensitive to climate variation and a process that depends on water and light, in summer this process tends to be greater, leading to a decrease in CO_2 .

Concerning the daily X_{CO_2} model presented, it performed well when we looked at the set of metrics presented. Given this, we were able to estimate the daily behavior of natural X_{CO_2} in general for the state of São Paulo, a semi-periodical wave with a maximum peak between March and July, and a minimum peak between December to February. There are still challenges in this aspect, such as the transport process in the atmosphere, which was simplified due to the detrend in the dataset, that also remove the anthropogenic sources in the CO_2 cycle, however, this study was capable in advance in the temporal gap, and properly address how to estimate the natural behavior of this gas in a synthetic way using daily meteorological open access data, establishing a low-cost approach, and we believe that this study will serve as a basis for further implementations.

We suggest that for future work, the relationship between soil respiration and factor controlling organic matter decay in soil with the X_{CO_2} would be needed to better understand CO_2 dynamics, as well the addition of

variables related to activities, such as in transports or the data of fossil fuel consumption, in big cities to improve predictions, as well the atmospheric transport.

Methods

Study region

The state of São Paulo (SP) (Fig. 5b), southern Brazil, has approximately 249×10^3 km² and 645 municipalities, with a demographic density of 179.84 habitants/km² [64] being one of the main agricultural hubs of Brazil, regarding the production of sugarcane and citrus [65]. According to Rolim et al. [47] the climate of the state, in general, has its areas characterized by a humid subtropical climate with dry winter, followed by humid tropical dry winter and sub-humid tropical dry winter, according to the climate classification proposed by Camargo [66].

Products of remote sensing: acquisition and processing

Greenhouse gas, climate, and vegetation data were collected from different satellites (Table 2) for a time series from 2015 to 2019 and were aggregated on a monthly scale. The primary product of the Orbiting Carbon Observatory-2 (OCO-2) consists of georeferenced estimates of the mean atmospheric CO_2 concentration (X_{CO_2}), in addition, the Sun Induced Chlorophyll Fluorescence (SIF), retrieved due to the overlap that occurs in the SIF wavelengths with the O_2 absorption wavelength (680–850 nm) [8, 9, 43]. Data from this satellite have already been validated by Crisp et al. [8] and, according to O'Dell et al. [9], this satellite provides about 65,000 quality observations per day worldwide.

Here we used the version 9 of the OCO-2 with a bias-correction and considered only the measurements with the best quality flag (quality flag = 0, meaning that has no cloud cover) [67, 68], also, we do not consider the data with more than 12 alert level at nadir viewing [33, 69]. Concerning the SIF, we take into account only the SIF at 757 nm, this was due to previous studies that exploited the relationship in the São Paulo's State [6, 7] and, also because this wavelength is closer to the far-red peak (~740 nm) in the whole SIF signal [43].

MODIS sensor data were extracted from the "Application for Extracting and Exploring Analysis Ready Samples" (AppEEARS). This application allows users to obtain subsets of large databases using spatial and temporal parameters. Two types of sample requests are available: point samples by entering geographic coordinates and area samples using vector polygons. Sample requests submitted to AppEEARS provide users with not only data

Table 2 Studied variables, data base, temporal and spatial resolution

Variable	Data base	Temporal resolution	Spatial resolution
GHG			
X _{CO₂} (ppm)	OCO-2*OCO-2 Data product user's guide, 2016"V9	16 days	1.29 km × 2.25 km
Climate			
Surface solar shortwave irradiance (Global radiation, Q _g) (MJ m ⁻² day ⁻¹)	FLASH Flux Version 3 (A, B,C) NASA/POWER	Daily	111.3 km × 111.3 km
Average air temperature at 2 m (Temp) (°C)	GEO-5 FP-IT (NASA/POWER)	Daily	111.3 km × 111.3 km
Land surface temperature (LST) (°C)	MOD11A1.006 V6 MODIS-TERRA	Daily	1200 km × 1200 km
Wind speed at 10 m (WS) (m s ⁻¹) (°C)	GEO-5 FP-IT (NASA/POWER)	Daily	111.3 km × 111.3 km
Relative humidity (RH) (%)	GEO-5 FP-IT (NASA/POWER)	Daily	111.3 km × 111.3 km
Precipitation (Prec) (mm day ⁻¹)	GEO-5 FP-IT (NASA/POWER)	Daily	111.3 km × 111.3 km
Vegetation			
SIF 757	OCO-2*OCO-2 Data product user's guide, 2016"V9	16 days	1.29 km × 2.25 km
LAI (m ² m ⁻²)	MCD15A2H.006 V6 MODIS- CFPAR	8 days	500 m × 500 m
Fraction of Photosynthetically Active Radiation (Fpar) (%)			
Evapotranspiration (ET) (kg m ⁻² day ⁻¹)	MOD16A2.006 V6 MODIS-TERRA	8 days	500 m × 500 m
NDVI	MOD13A1.006 V6 MODIS-TERRA	16 days	500 m × 500 m

values but also associated quality data values. Interactive visualizations with summary statistics are provided for each sample within the application, which allows users to view and interact with their samples before downloading the data [70].

Nasa Power data (<https://power.larc.nasa.gov>) consists of precipitation (mm), surface solar shortwave irradiance (MJ m⁻² day⁻¹), average air temperature (°C), and relative humidity at 2 m (%). This platform consists of a NASA project entitled: Worldwide Energy Resource Forecast (POWER) and was initiated to enhance the current renewable energy dataset and create new datasets from new satellite systems [71].

To minimize the differences between the spatial and temporal resolutions of the different orbital sensors used in this study, the process described in Fig. 5a was employed, which establishes a standard for the acquisition of data from the coordinates obtained in the OCO-2 platform (Fig. 5b). We emphasize that several studies have been conducted using different time and spatial scales [6, 7, 27].

Pre-process of the data

Using the regression method proposed by Gujarati and Potter [72], we removed the trend from X_{CO₂} data, in order to understand the natural and regional variability of X_{CO₂} and its relationships with other factors [6, 7, 33]. The other variables were standardized using the function *scale* from the R language [73].

Variance Inflation Factor (VIF)

Variance Inflation Factor (VIF) analysis was performed. This analysis is a method of detecting multicollinearity within a database since the relation between the predictors for a multi-regression model can affect the estimative and the standard errors associated with the regression model [31]. The VIF is based on the R² value (Eq. 2), and should not be greater than 10, however, this can vary according to the study [31, 32].

$$VIF = \frac{1}{1 - R^2} \quad (2)$$

where R² is the coefficient of determination.

Temporal variability, Pearson's correlation, and dependency analysis

The data was processed using month averages for the analysis period, except precipitation, which consists of monthly sums for the entire state of SP (ST.1). The means were subjected to analysis of variance (F-test) to obtain the mean standard errors. Simultaneously, the basic assumptions of analysis of variance and, normality of errors, and homogeneity of variances were tested for the selected variables by VIF analysis. To understand the variation of X_{CO₂} with the other variables, Pearson correlation analyses were performed. More about the descriptive statics of selected variables in VIF, such as the number of observations (soundings) for each month, can be found in Additional file 1: Table S2

Stepwise: forward selection

The stepwise method used in this study was the forward selection method being performed in R language

[73], as can be seen in the flow chart (Fig. 6), the variables selected in the VIF analysis were separated into a training and test samples (70% and 30% of the dataset

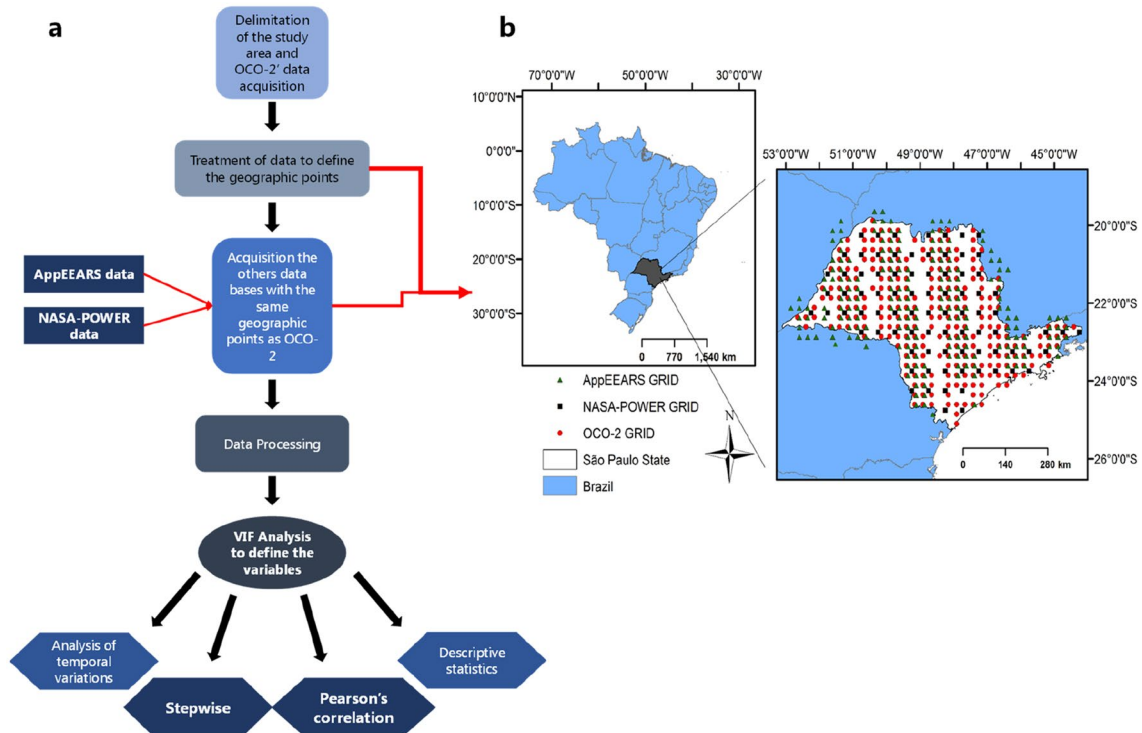


Fig. 5 Flowchart of data acquisition, processing, analysis (a), and sounding map of the satellite observations in the study region (b). Where red dots represent the OCO-2 soundings, the black dots represent the NASA-POWER and the dark green represent the MODIS

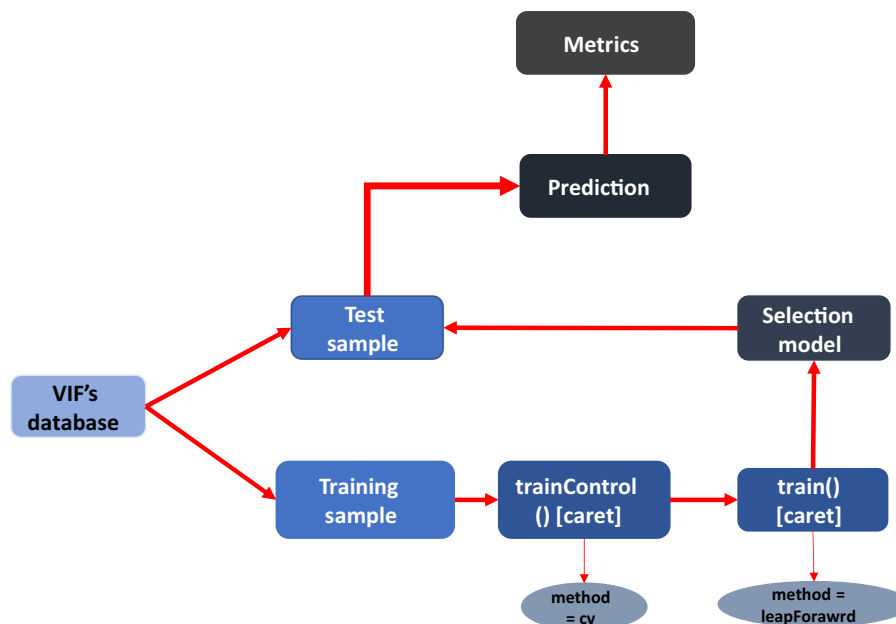


Fig. 6 Flowchart of the stepwise construction

respectively). The training sample was submitted to the *train ()* function of the *caret* package, using repeated cross-validation (cv) method. This technique consists in randomly splitting the training dataset into k-subsets, one of them is reserved and the model is trained with the others, and after is validated with the reserved subset, this process is repeated until each subset serves as a test sample, finally, the average error is how the performance is given [74]. The model is based on the lowest Root Mean Squared Error (RMSE) and, from variables selected in training, the generated model is applied to the test sample defined at begging for estimating the X_{CO_2} with these independent data. Finally, cross-validation between the estimated data and observed data in the test sample was performed and from this, we derive the metrics Mean squared error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), R^2 , and Mean absolute percentage error (MAPE).

Abbreviations

AppEEARS: Application for extracting and exploring analysis ready samples; CO_2 : Carbon dioxide; ET: Evapotranspiration; LAI: Leaf area index; LST: Land surface temperature (MODIS); MAE: Mean absolute error; MAPE: Mean absolute percentage error; MSE: Mean squared error; NASA-POWER: NASA project entitled: worldwide energy resource forecast; NDVI: Normalized difference vegetation index; OCO-2: Orbiting carbon observatory-2; OCO-3: Orbiting carbon observatory-3; Prec: Precipitation; Qg: Global radiation; RH: Relative humidity; RMSE: Root mean squared error; SIF 757: Solar-induced chlorophyll fluorescence at 757 nm; Temp: Temperature at 2 m; VIF: Variance Inflation Factor; Ws: Wind speed at 2 m; X_{CO_2} : Column average of carbon dioxide in the atmosphere.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13021-022-00209-7>.

Additional file 1: Table S1. Variables used in the VIF analyses. These variables were standardized by `scale()` in R language. **Table S2.** Descriptive statistics.

Acknowledgements

All authors are grateful to the São Paulo State University (UNESP), Jaboticabal Campus, for the academic support and the open data provide by the NASA projects: OCO-2, NASA-POWER, and AppEEARS.

Author contributions

LMC: writing-original draft, conceptualization, methodology, investigation, writing-review, and editing; GAAS: methodology, investigation, writing-review, and editing; ARP: conceptualization, methodology writing-review, editing, and supervision; GSR: conceptualization, methodology writing-review, editing, and supervision; NLS: conceptualization, methodology writing-review, editing, and supervision. All authors read and approved the final manuscript.

Funding

This work was supported by the São Paulo Research Foundation (FAPESP) [Grant number: 2019/25812-4] and CNPq—National Council for Scientific and Technological Development [Grants: 304075/2018-3, 311981/2020-8].

Availability of data and materials

The processed data can be found in Additional file tables attached to this paper. The X_{CO_2} and SIF were retrieved from: <https://co2.jpl.nasa.gov/build/>

mission=oco-2&dataset=OCO2L2Stdv10; The MODIS products from: <https://lpdaacsvc.cr.usgs.gov/appears/>; the NASA-POWER from: <https://power.larc.nasa.gov/>; The VIF and regression was made in R language and the code is available at: https://github.com/lm-costa/stepwise_fapesp.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Author details

¹Departament of Engineering and Exact Sciences, São Paulo State University, Via de Acesso Prof. Paulo Donato Castellane s/n, Jaboticabal, São Paulo 14884-900, Brazil. ²Campus Avançado Porto Franco, Instituto Federal de Educação, Ciência e Tecnologia do Maranhão – IFMA, Rua Custódio Barbosa, no 09, Centro, Porto Franco, Maranhão 65970-000, Brazil. ³Center of Agricultural, Natural and Literary Sciences, State University of the Tocantina Region of Maranhão (UEMASUL), Av. Brejo do Pinto, S/N – Brejo do Pinto, Estreito, Maranhão 65975-000, Brazil.

Received: 15 January 2022 Accepted: 20 May 2022

Published online: 11 June 2022

References

- Eldering A, O'Dell CW, Wennberg PO, Crisp D, Gunson MR, Viatte C, Avis C, Braverman A, Castano R, Chang A, Chapsky L, Cheng C, Connor B, Dang L, Doran G, Fisher B, Frankenberg C, Fu D, Granat R, Hobbs J, Lee RAM, Mandrake L, McDuffie J, Miller CE, Myers V, Natraj V, O'Brien D, Osterman GB, Oyafuso F, Payne VH, Pollock HR, Polonsky I, Roehl CM, Rosenberg R, Schwandner F, Smyth M, Tang V, Taylor TE, To C, Wunch D, Yoshimizu J. The Orbiting Carbon Observatory-2: first 18 months of science data products. *Atmos Meas Tech*. 2017;10:549–63. <https://doi.org/10.5194/amt-10-549-2017>.
- Boesch H, Baker D, Connor B, Crisp D, Miller C. Global characterization of CO_2 column retrievals from shortwave-infrared satellite observations of the orbiting carbon observatory-2 mission. *Remote Sensing*. 2011;3(2):270–304. <https://doi.org/10.3390/RS3020270>.
- Ye X, Lauvaux T, Kort EA, Oda T, Feng S, Lin JC, Yang EG, Wu D. Constraining fossil fuel CO_2 emissions from urban area using OCO-2 observations of total column CO_2 . *J Geophys Res Atmos*. 2020;125(8):e2019JD030528. <https://doi.org/10.1029/2019JD030528>.
- Kort EA, Frankenberg C, Miller CE, Oda T. Space-based observations of megacity carbon dioxide. *Geophys Res Lett*. 2012. <https://doi.org/10.1029/2012GL052738>.
- Parazoo NC, Bowman K, Frankenberg C, Lee JE, Fisher JB, Worden J, Jones DBA, Berry J, Collatz GJ, Baker IT, Jung M, Liu J, Osterman G, O'Dell C, Sparks A, Butz A, Guerlet S, Yoshida Y, Chen H, Gerbig C. Interpreting seasonal changes in the carbon balance of southern Amazonia using measurements of X_{CO_2} and chlorophyll fluorescence from GOSAT. *Geophys Res Lett*. 2013;40(11):2829–33. <https://doi.org/10.1002/grl.50452>.
- Morais Filho LFF, de Meneses KC, de Santos GA, da Bicalho AE, de Rolim GS, La Scala N. X_{CO_2} temporal variability above Brazilian agroecosystems: a remote sensing approach. *J Environ Manage*. 2021;288:112433. <https://doi.org/10.1016/j.jenvman.2021.112433>.
- da Costa LM, de Araújo Santos GA, de Mendonça GC, et al. Spatiotemporal variability of atmospheric CO_2 concentration and controlling factors over sugarcane cultivation areas in southern Brazil. *Environ Dev Sustain*. 2021. <https://doi.org/10.1007/s10668-021-01677-6>.

8. Crisp D, Fisher BM, O'dell C, Frankenberg C, Basilio R, Bösch H, Brown LR, Castano R, Connor B, Deutscher NM, Eldering A, Griffith D, Gunson M, Kuze A, Mandrake L, McDuffie J, Messerschmidt J, Miller CE, Morino I, Natraj V, Notholt J, O'Brien DM, Oyafuso F, Polonsky I, Robinson J, Salawitch R, Sherlock V, Smyth M, Suto H, Taylor TE, Thompson DR, Wennberg PO, Wunch D, Yung YL. The ACOS CO₂ retrieval algorithm—Part II: global XCO₂ data characterization. *Atmos Meas Tech*. 2012;5:687–707. <https://doi.org/10.5194/amt-5-687-2012>.
9. O'Dell CW, Connor B, Bösch H, O'Brien D, Frankenberg C, Castano R, Christi M, Eldering D, Fisher B, Gunson M, McDuffie J, Miller CE, Natraj V, Oyafuso F, Polonsky I, Smyth M, Taylor T, Toon GC, Wennberg PO, Wunch D. Description and validation against synthetic observations. The ACOS CO₂ retrieval algorithm—Part 1. *Atmos Meas Tech*. 2012;5:99–121. <https://doi.org/10.5194/amt-5-99-2012>.
10. Li X, Xiao J, Fisher JB, Baldocchi DD. ECOSTRESS estimates gross primary production with fine spatial resolution for different times of day from the International Space Station. *Remote Sensing Environ*. 2021. <https://doi.org/10.1016/j.rse.2021.112360>.
11. Wassie YT, Adaramola MS. Socio-economic and environmental impacts of rural electrification with solar photovoltaic systems: evidence from southern Ethiopia. *Energy Sustainable Develop*. 2021;60:52–66. <https://doi.org/10.1016/j.esd.2020.12.002>.
12. Rupp M, Rieke C, Handschuh N, Kuperjans I. Economic and ecological optimization of electric bus charging considering variable electricity prices and CO₂ eq intensities. *Transp Res D Transp Environ*. 2020. <https://doi.org/10.1016/j.trd.2020.102293>.
13. Smith P, Calvin K, Nkem J, et al. Which practices co-deliver food security, climate change mitigation and adaptation, and combat land degradation and desertification? *Glob Change Biol*. 2020;26:1532–75. <https://doi.org/10.1111/gcb.14878>.
14. Li Q, Wu S, Lei Y, Li S. Dynamic features and driving forces of indirect CO₂ emissions from Chinese household: a comparative and mitigation strategies analysis. *Sci Total Environ*. 2020. <https://doi.org/10.1016/j.scitotenv.2019.135367>.
15. Cernusak LA. Gas exchange and water-use efficiency in plant canopies. *Plant Biol*. 2020;22(S1):52–67. <https://doi.org/10.1111/plb.12939>.
16. Parras R, de Mendonça GC, Araújo Costa RC, Pissarra TCT, Valera CA, Fernandes LFS, Leal Pacheco FA. The configuration of forest cover in Ribeirão Preto: a diagnosis of Brazil's forest code implementation. *Sustainability*. 2020;12(14):5686. <https://doi.org/10.3390/su12145686>.
17. de Oliveira ML, dos Santos CAC, de Oliveira G, Perez-Marin AM, Santos CAG. Effects of human-induced land degradation on water and carbon fluxes in two different Brazilian dryland soil covers. *Sci Total Environ*. 2021;792: 148458. <https://doi.org/10.1016/J.SCITOTENV.2021.148458>.
18. Cabral OMR, Freitas HC, Cuadra SV, de Andrade CA, Ramos NP, Grutzmacher P, Galdos M, Packer APC, da Rocha HR, Rossi P. The sustainability of a sugarcane plantation in Brazil assessed by the eddy covariance fluxes of greenhouse gases. *Agric For Meteorol*. 2020;282–283: 107864. <https://doi.org/10.1016/j.agrformet.2019.107864>.
19. Oliveira RR, Pezzi LP, Souza RB, Santini MF, Cunha LC, Pacheco FS. First measurements of the ocean-atmosphere CO₂ fluxes at the Cabo Frio upwelling system region, Southwestern Atlantic Ocean. *Cont Shelf Res*. 2019;181:135–42. <https://doi.org/10.1016/J.CSR.2019.05.008>.
20. Tedesco D, de Oliveira MF, dos Santos AF, Costa Silva EH, de Souza Rolim G, da Silva RP. Use of remote sensing to characterize the phenological development and to predict sweet potato yield in two growing seasons. *Eur J Agron*. 2021;129: 126337. <https://doi.org/10.1016/J.EJA.2021.126337>.
21. de Aparecido LE, O., Lorençone, J. A., Lorençone, P. A., Rolim, G. de S., Meneses, K. C. de, Moraes, J. R. da S. C. de, & Torsoni, G. B. Can nonlinear agrometeorological models estimate coffee foliation? *J Sci Food Agric*. 2021. <https://doi.org/10.1002/JSCFA.11387>.
22. Helman D, Lensky IM, Osem Y, Rohatyn S, Rotenberg E, Yakir D. A biophysical approach using water deficit factor for daily estimations of evapotranspiration and CO₂ uptake in Mediterranean environments. *Biogeosciences*. 2017;14(17):3909–26. <https://doi.org/10.5194/BG-14-3909-2017>.
23. Guo M, Xu J, Wang X, He H, Li J, Wu L. Estimating CO₂ concentration during the growing season from MODIS and GOSAT in East Asia. *Int J Remote Sens*. 2015;36(17):4363–83. <https://doi.org/10.1080/01431161.2015.1081305>.
24. Li X, Hu X-M, Cai C, Jia Q, Zhang Y, Liu J, et al. Terrestrial CO₂ fluxes, concentrations, sources and budget in Northeast China: observational and modeling studies. *J Geophys Res Atmos*. 2020;125: e2019JD031686. <https://doi.org/10.1029/2019JD031686>.
25. Chhabra A, Gohel A. Dynamics of atmospheric carbon dioxide over different land cover types in India. *Environ Monit Assess*. 2019;191:799. <https://doi.org/10.1007/s10661-019-7681-z>.
26. Walker AP, et al. Predicting long-term carbon sequestration in response to CO₂ enrichment: how and why do current ecosystem models differ? *Global Biogeochem Cycles*. 2015;29:476–95. <https://doi.org/10.1002/2014GB004995>.
27. Golkar F, et al. Using OCO-2 satellite data for investigating the variability of atmospheric CO₂ concentration in relationship with precipitation, relative humidity, and vegetation over Oman. *Water*. 2020;12(1):101. <https://doi.org/10.3390/w12010101>.
28. Barbosa HA, Lakshmi Kumar Silva TVLRM. Recent trends in vegetation dynamics in the South America and their relationship to rainfall. *Nat Hazards*. 2015;77:883–99. <https://doi.org/10.1007/s11069-015-1635-8>.
29. Wagle P, Gowda PH, Billesbach DP, Northup BK, Torn MS, Neel JPS, Biraud SC. Dynamics of CO and HO fluxes in Johnson grass in the US Southern Great Plains. *Sci Total Environ*. 2020;739:140077. <https://doi.org/10.1016/j.scitotenv.2020.140077>.
30. Liu Q, Fu YH, Zeng Z, Huang M, Li X, Piao S. Temperature, precipitation, and insolation effects on autumn vegetation phenology in temperate China. *Global Change Biol*. 2016;22:644–55. <https://doi.org/10.1111/gcb.13081>.
31. Miles J. Tolerance and variance inflation factor. *Wiley StatsRef: Statistics Reference Online*. 2014. <https://doi.org/10.1002/9781118445112.STATO6593>.
32. Kalnins A. Multicollinearity: how common factors cause Type 1 errors in multivariate regression. *Strategic Manage J*. 2018;39(8):2362–85. <https://doi.org/10.1002/SMJ.2783>.
33. Rossi FS, de Araújo Santos GA, de Souza Maria L, Lourençoni T, Pellissari TD, Della-Silva JL, Júnior JWO, de Avila e Silva, A., Lima, M., Teodoro, P. E., Teodoro, L. P. R., de Oliveira-Júnior, J. F., La Scala, N., & da Silva Junior, C. A. Carbon dioxide spatial variability and dynamics for contrasting land uses in central Brazil agricultural frontier from remote sensing data. *J South Am Earth Sci*. 2022. <https://doi.org/10.1016/J.JSAMES.2022.103809>.
34. Hakkarainen J, Jalongo I, Tamminen J. Direct space-based observations of anthropogenic CO₂ emission areas from OCO-2. *Geophys Res Lett*. 2016;43(21):11400–6. <https://doi.org/10.1002/2016GL070885>.
35. Wu D, Lin J, Fasoli B, Oda T, Ye X, Lauvaux T, Yang E, Kort E. A Lagrangian approach towards extracting signals of urban CO₂ emissions from satellite observations of atmospheric column CO₂ (XCO₂): X-stochastic time-inverted Lagrangian transport model ("X-STILT v1"). *Geosci Model Develop*. 2018;11(12):4843–71. <https://doi.org/10.5194/GMD-11-4843-2018>.
36. Sá MMF, Schaefer CÉGR, Loureiro DC, Simas FNB, Alves BJR, de Sá Mendonça E, de Figueiredo EB, La Scala N, Panosso AR. Fluxes of CO₂, CH₄, and N₂O in tundra-covered and Nothofagus forest soils in the Argentinian Patagonia. *Sci Total Environ*. 2019;659:401–9. <https://doi.org/10.1016/J.SCITOTENV.2018.12.328>.
37. Vicentini ME, Pinotti CR, Hirai WY, de Moraes MLT, Montanari R, Filho MCMT, Milori DMBP, Júnior NLS, Panosso AR. CO₂ emission and its relation to soil temperature, moisture, and O₂ absorption in the reforested areas of Cerrado biome Central Brazil. *Plant Soil*. 2019;444(1):193–211. <https://doi.org/10.1007/S11104-019-04262-Z>.
38. Graham MH. Confronting multicollinearity in ecological multiple regression. *Ecology*. 2003;84:2809–15. <https://doi.org/10.1890/02-3114>.
39. Tamura R, Kobayashi K, Takano Y, Miyashiro R, Nakata K, Matsui T. Mixed integer quadratic optimization formulations for eliminating multicollinearity based on variance inflation factor. *J Global Optim*. 2019;73:431–46. <https://doi.org/10.1007/s10898-018-0713-3>.
40. Rafiei Sardooi E, Azareh A, Choubin B, Barkhori S, Singh VP, Shamshirband S. Applying the remotely sensed data to identify homogeneous regions of watersheds using a pixel-based classification approach. *Appl Geogr*. 2019;111: 102071. <https://doi.org/10.1016/J.APGEOG.2019.102071>.

41. Siabi Z, Falahatkar S, Alavi SJ. Spatial distribution of X_{CO_2} using OCO-2 data in growing seasons. *J Environ Manage*. 2019;244:110–8. <https://doi.org/10.1016/j.jenvman.2019.05.049>.
42. Falahatkar S, Mousavi SM, Farajzadeh M. Spatial and temporal distribution of carbon dioxide gas using GOSAT data over IRAN. *Environ Mon Assess*. 2017;189:627. <https://doi.org/10.1007/s10661-017-6285-8>.
43. Mohammed GH, Colombo R, Middleton EM, Rascher U, van der Tol C, Nedbal L, Goulasf Y, Pérez-Priego O, Dammh A, Meronij M, Joinerc J, Cogliatib S, Verhoeft W, Malenovský Z, Gastellu-Etchegorryl J-P, Millerm JR, Guantern L, Morenoo J, Moyaf I, Berryp JA, Frankenberg C, Zarco-Tejada PJ. Remote sensing of solar-induced chlorophyll fluorescence (SIF) in vegetation: 50 years of progress. *Remote Sensing Environ*. 2019;231: 111177. <https://doi.org/10.1016/j.rse.2019.04.030>.
44. Duveiller G, Filipponi F, Walthers S, Köhler P, Frankenberg C, Guantern L, Ces-catti A. A spatially downscaled sun-induced fluorescence global product for enhanced monitoring of vegetation productivity. *Earth System Sci Data*. 2020;12(2):1101–16. <https://doi.org/10.5194/essd-12-1101-2020>.
45. Campbell PKE, Huemmrich KF, Middleton EM, Ward LA, Julitta T, Daughtry CST, Burkart A, Russ AL, Kustas WP. Diurnal and seasonal variations in chlorophyll fluorescence associated with photosynthesis at leaf and canopy scales. *Remote Sens*. 2019;11:488. <https://doi.org/10.3390/rs11050488>.
46. Sun Y, Frankenberg C, Jung M, Joiner J, Guantern L, Köhler P, Magney T. Overview of solar-induced chlorophyll fluorescence (SIF) from the orbiting Carbon observatory-2: retrieval, cross-mission comparison, and global monitoring for GPP. *Remote Sens Environ*. 2018;209:808–23. <https://doi.org/10.1016/j.rse.2018.02.016>.
47. Rolim GS, Aparecido LEO, Camargo, Köppen and Thorntwaite climate classification systems in defining climatical regions of the state of São Paulo, Brazil. *Int J Climatol*. 2016;36:636–43. <https://doi.org/10.1002/joc.4372>.
48. Taiz L, Zeiger E. *Fisiologia Vegetal*. 4th ed. Editora Artmed; 2009.
49. Frankenberg C, O'Dell C, Guantern L, McDuffie J. Remote sensing of near-infrared chlorophyll fluorescence from space in scattering atmospheres: implications for its retrieval and interferences with atmospheric CO_2 retrievals. *Atmos Meas Tech*. 2012;5(8):2081–94. <https://doi.org/10.5194/amt-5-2081-2012>.
50. Ishizawa M, Mabuchi K, Shirai T, Inoue M, Morino I, Uchino O, Yoshida Y, Belikov D, Maksyutov S. Inter-annual variability of summertime CO_2 exchange in Northern Eurasia inferred from GOSAT X_{CO_2} . *Environ Res Lett*. 2016;11(10): 105001. <https://doi.org/10.1088/1748-9326/11/10/105001>.
51. Hansen R, Mander Ü, Soosaar K, et al. Greenhouse gas fluxes in an open air humidity manipulation experiment. *Landscape Ecol*. 2013;28:637–49. <https://doi.org/10.1007/s10980-012-9775-7>.
52. Snyder RL, Spano D. Phenology and evapotranspiration. In: Schwartz M, editor. *Phenology: an integrative environmental science*. Dordrecht: Springer; 2013.
53. Aparecido LEDO, Ferreira RB, Rolim GDS, SouzaDe BS, SouzaDe PS. Non-linear agrometeorological models for estimating lychee fruit growth. *Rev Bras Frutic*. 2017. <https://doi.org/10.1590/0100-29452017169>.
54. Kovalsky V, Henebry GM, Roy DP, Adusei B, Hansen M, Senay G, Mocko DM. Evaluation of a coupled event-driven phenology and evapotranspiration model for croplands in the United States northern Great Plains. *J Geophysical Res Atmos*. 2013;118(11):5065–81. <https://doi.org/10.1002/JGRD.50387>.
55. CONAB, Companhia Nacional de Abastecimento. Calendário de plantio e colheita de grãos no Brasil 2019. 2019. <https://www.conab.gov.br/institucional/publicacoes/outras-publicacoes/item/7694-calendario-agricola-plantio-e-colheita>.
56. Shekhar A, Chen J, Paetzold JC, Dietrich F, Zhao X, Bhattacharjee S, Ruisinger V, Wofsy SC. Anthropogenic CO_2 emissions assessment of Nile Delta using X_{CO_2} and SIF data from OCO-2 satellite. *Environ Res Lett*. 2020;15(9): 095010. <https://doi.org/10.1088/1748-9326/AB9CFE>.
57. Guo M, Wang X, Li J, Yi K, Zhong G, Tani H. Assessment of global carbon dioxide concentration using MODIS and GOSAT data. *Sensors*. 2012;12(12):16368–89. <https://doi.org/10.3390/s121216368>.
58. Taylor TE, Eldering A, Merrelli A, Kiel M, Somkuti P, Cheng C, Rosenberg R, Fisher B, Crisp D, Basilio R, Bennett M, Cervantes D, Chang A, Dang L, Frankenberg C, Haemmerle VR, Keller GR, Kurosu T, Laughner JL, Yu S. OCO-3 early mission operations and initial (vEarly) X_{CO_2} and SIF retrievals. *Remote Sens Environ*. 2020;251: 112032. <https://doi.org/10.1016/J.RSE.2020.112032>.
59. Rolim GS, de Oliveira Aparecido LE, de Souza PS, et al. Climate and natural quality of Coffea Arabica L. drink. *Theor Appl Climatol*. 2020;141:87–98. <https://doi.org/10.1007/s00704-020-03117-3>.
60. de Oliveira Aparecido LE, de Souza Rolim G, da Silva Cabral de Moraes, JR. Validation of ECMWF climatic data, 1979–2017, and implications for modelling water balance for tropical climates. *Int J Climatol*. 2020;40:6646–65. <https://doi.org/10.1002/joc.6604>.
61. White JW, Hoogenboom G, Stackhouse PW, Hoell JM. Evaluation of NASA satellite- and assimilation model-derived long-term daily temperature data over the continental US. *Agric For Meteorol*. 2008;148(10):1574–84. <https://doi.org/10.1016/J.AGRFORMET.2008.05.017>.
62. Zhang Y, Joiner J, Hamed Alemohammad S, Zhou S, Gentine P. A global spatially contiguous solar-induced fluorescence (CSIF) dataset using neural networks. *Biogeosciences*. 2018;15(19):5779–800. <https://doi.org/10.5194/bg-15-5779-2018>.
63. Yu L, Wen J, Chang CY, Frankenberg C, Sun Y. High-resolution global contiguous SIF of OCO-2. *Geophys Res Lett*. 2019;46(3):1449–58. <https://doi.org/10.1029/2018GL081109>.
64. SEADE—Fundação estadual de Análise de dados—Perfil dos municípios paulista. 2020. <https://perfil.seade.gov.br/>
65. Camargo FP et al (2020). Previsões e Estimativas das Safras Agrícolas do Estado de São Paulo, Ano Agrícola 2019/20. Análises e Indicadores do Agronegócio, São Paulo, v. 15, n. 9, set. <http://www.iea.sp.gov.br/out/TerTexto.php?codTexto=14839>
66. Camargo AP. Classificação climática para zoneamento de aptidão agroclimática. In: Heldwein AB, Schneider FM, Burial GA, Petter Medeiros SL, Estefanel V, editors. *Congresso Brasileiro de Agrometeorologia*, vol. 7. Brazil: Viçosa Sociedade Brasileira de Agrometeorologia/Universidade Federal de Viçosa; 1991. p. 126–31.
67. Nikitenko AA, Timofeev YM, Berezin IA, Poberovskii AV, Virolainen YA, Polyakov AV. The analysis of OCO-2 satellite measurements of CO_2 in the vicinity of Russian cities. *Atmos Ocean Opt*. 2020;33(6):650–5. <https://doi.org/10.1134/S1024856020060111/FIGURES/3>.
68. Massie ST, Cronk H, Merrelli A, O'Dell C, Sebastian Schmidt K, Chen H, Baker D. Analysis of 3D cloud effects in OCO-2 X_{CO_2} retrievals. *Atmos Meas Tech*. 2021;14(2):1475–99. <https://doi.org/10.5194/amt-14-1475-2021>.
69. Mandrake L, O'Dell C, Wunch D, Wennberg PO, Fisher B, Osterman GB, Eldering A. (2015). Lite files, warn level and bias correction determination. Tech. Rep., Jet Propul. Lab., California Inst. of Technol., Pasadena, Calif.
70. AppEEARS Team. 2020. Application for extracting and exploring analysis ready samples (AppEEARS). Ver. 6. NASA EOSDIS Land processes distributed active archive center (LP DAAC), USGS/Earth resources observation and science (EROS) center, Sioux Falls, South Dakota, USA. <https://lpdaa.cr.usgs.gov/appeears/>. Accessed Nov 11 2020.
71. Stackhouse PW Jr, Westberg D, Chandler WS, Zhang T, Hoell JM (2015). Prediction of worldwide energy resource (POWER) agroclimatology methodology, version 1.1.0, May 30. NASA Langley Research Center.
72. Gujarati DN, Porter DC. 2011. *Econometria básica-5*. Amgh Editora.
73. R Core team 2021. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. www.R-project.org/.
74. Gareth J, Daniela W, Trevor H, Robert T. An introduction to statistical learning: with applications in R. New York: Springer; 2013.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.