

Monitoring and modeling seasonally varying anthropogenic and biogenic CO_2 over a large tropical metropolitan area

Rafaela Cruz Alves Alberti¹, Thomas Lauvaux³, Angel Liduvino Vara-Vela^{6,7,8}, Ricard Segura Barrero², Christoffer Karoff^{6,7,8}, Maria de Fátima Andrade¹, Márcia Talita Amorim Marques¹, Noelia Rojas Benavente⁵, Osvaldo Machado Rodrigues Cabral⁴, Humberto Ribeiro da Rocha¹, and Rita Yuri Ynoue¹

¹Department of Atmospheric Sciences, University of São Paulo, Brazil

²Institute of Environmental Sciences and Technology, Universitat Autònoma de Barcelona, Spain.

³Université de Reims Champagne-Ardenne, CNRS, GSMA, Reims, France

⁴Brazilian Agricultural Research Corporation, Embrapa Environment, Brazil

⁵Physics Institute, University of São Paulo, Brazil

⁶Department of Geoscience, Aarhus University, Denmark

⁷Department of Physics and Astronomy, Aarhus University, Aarhus, Denmark

⁸iCLIMATE Aarhus University Interdisciplinary Centre for Climate Change, Aarhus, Denmark

Correspondence: Rafaela Cruz Alves Alberti (rafaela_alves@usp.br)

Abstract. Atmospheric CO_2 concentrations in urban areas reflect a combination of fossil fuel emissions and biogenic fluxes, offering a potential approach to assess city climate policies. However, atmospheric models used to simulate urban CO_2 plumes face significant uncertainties, particularly in complex urban environments with dense populations and vegetation. This study addresses these challenges by analyzing CO_2 dynamics in the Metropolitan Area of São Paulo (MASP) using the WRF-Chem model. Simulations were evaluated against ground-based observations from the METROCLIMA network, the first greenhouse gas monitoring network in South America, and column concentrations (XC_{O_2}) from the OCO-2 satellite spanning February to August 2019. To improve biogenic fluxes, we optimized parameters in the Vegetation Photosynthesis and Respiration Model (VPRM) using eddy covariance flux measurements for key vegetation types, including the Atlantic Forest, Cerrado, and sugarcane. Results show that at the urban site (IAG), the model consistently underestimated CO_2 concentrations, with a negative mean bias of -9 ppm throughout the simulation period, likely due to the complexity of vehicular emissions and urban dynamics. In contrast, at the vegetated site (PDJ), simulations showed a consistent positive mean bias of 5 ppm and closely matched observations. Seasonal analyses revealed higher CO_2 concentrations in winter, driven by greater atmospheric stability and reduced vegetation uptake estimated by VPRM, while summer exhibited lower levels due to increased mixing and higher agricultural productivity. A comparison of biogenic and anthropogenic scenarios highlights the need for integrated emission modeling and improved representation of biogenic fluxes, anthropogenic emissions, and boundary conditions for high-resolution modeling in tropical regions.

18 Urban areas, although occupying only a small fraction of the Earth's surface, exert an outsized influence on global carbon
19 emissions. Accounting for a staggering 70% of CO_2 emissions from fossil fuel burning while covering just 2% of the planet's
20 landmass (Seto et al., 2014; Change et al., 2014), cities have become focal points for climate action. The relentless pace of
21 urbanization has further exacerbated this phenomenon, driving up energy consumption and emissions levels (Seto et al., 2012).
22 Consequently, combating climate change necessitates a targeted approach, with policies increasingly tailored to address urban
23 emissions. In response to the growing need for climate action, initiatives like the International Council for Local Environmen-
24 tal Initiatives (ICLEI), the C40 Cities Climate Leadership Group (C40), and the Covenant of Mayors (CoM) have emerged to
25 coordinate global efforts and share best practices among cities. These initiatives highlight the crucial role cities play in the fight
26 against climate change and the importance of localized mitigation strategies. São Paulo, Brazil's largest municipality (IBGE,
27 2021), is a member of C40 and focuses on reducing greenhouse gas emissions, with transportation accounting for 58% of its
28 total emissions (SEEG, 2019). The city is working towards carbon neutrality through projects in green infrastructure, urban
29 planning, public transportation improvements, energy efficiency, and waste management (Caetano et al., 2021). These efforts
30 aim to reduce emissions and enhance São Paulo's resilience, fostering a more sustainable urban environment. Central to these
31 efforts is the need for accurate data and robust modeling frameworks to inform policy decisions effectively. Urban atmospheric
32 networks, such as Metroclima in the Sao Paulo Metropolitan Area (MASP), in Brazil, provide vital insights into greenhouse gas
33 concentrations and emission patterns. By leveraging these datasets alongside sophisticated atmospheric transport models and
34 statistical techniques, policymakers gain tools for designing targeted interventions and monitoring their efficacy. However, the
35 complexity of urban CO_2 dynamics presents significant challenges for modeling and analysis. Process-driven biosphere mod-
36 els and inverse modeling techniques offer complementary approaches for capturing the intricate spatio-temporal variabilities
37 inherent in urban environments (Kaiser et al.; Che et al., 2022; Zhang et al., 2023; Wilmot et al., 2024). Despite advancements
38 in modeling capabilities, gaps remain in our understanding of CO_2 dynamics, particularly at regional and national scales. South
39 America, in particular, suffers from limited data availability, and research focusing on this region is scarce. Additionally, veg-
40 etation models in tropical regions often exhibit poor performance due to inaccuracies in simulating seasonality, oversimplified
41 representations of biodiversity, and errors in carbon and water cycle interactions. These models struggle to capture the complex
42 dynamics of tropical ecosystems, leading to underestimations of productivity and poor predictions of vegetation responses to
43 climate variability (De Pue et al., 2023; He et al., 2024). This study aims to address these gaps by conducting a comprehensive
44 analysis of anthropogenic and biospheric CO_2 dynamics near the MASP. By utilizing the Weather Research and Forecasting
45 model coupled with chemistry (WRF-Chem) offline coupled to the Vegetation Photosynthesis and Respiration Model (VPRM)
46 (Mahadevan et al., 2008) integrated with vehicular emissions from VEIN model (Ibarra-Espinosa et al., 2018) and industrial
47 sector emissions (EDGAR), we seek to elucidate the underlying drivers of CO_2 variability. In addition, we utilized data from
48 the OCO-2 satellite to cover the study domain, comparing smoothed XCO₂ concentrations derived from WRF-Chem (consid-
49 ering biogenic and anthropogenic emissions). Through a combination of model simulations, field observations, and satellite
50 data analysis, this study seeks to provide an understanding of CO_2 dynamics in urban environments. This is the first study in

51 this field conducted in any city in the Global South, making it an innovative effort with significant implications. By setting a
52 precedent, this research paves the way for future studies, contributing to a more comprehensive global picture of CO_2 dynamics
53 in urban environments.

54 2 WRF-Chem

55 2.1 Model set-up

56 A set of high-resolution simulations of atmospheric Greenhouse Gas concentrations were performed with the Weather Research
57 and Forecasting model coupled with Chemistry (WRF-Chem V4.0). The WRF-Chem was used to simulate the transport of the
58 mole fraction of CO_2 , and no chemical processes or reactions have been used. The period that has been simulated was from
59 1 February to 31 August, 2019. This period was selected due to available data from monitoring stations from the Metroclima
60 network for CO_2 . The simulations were made for each month. For each run, the simulation was initiated 5 days before and these
61 5 days were discarded as spin-up time. The single modeling domain was centered at $23.5^\circ S$ and $46.3^\circ W$ with a horizontal grid
62 spacing of 3 km as shown in Figure 1, projected on a Lambert plane and consists of 166 grid points in the west-east direction,
63 106 grid points in the north-south direction, and 34 vertical levels that extend from the surface up to 50 hPa (20 km), as used in
64 previous studies for this same area (Andrade et al., 2015; Vara-Vela et al., 2016; Gavidia-Calderón et al., 2023; Benavente et al.,
65 2023). The meteorological initial and boundary conditions to drive the simulations were obtained from the European Centre
66 for Medium-Range Weather Forecasts (ECMWF) global ERA5 reanalysis data set with $0.25^\circ \times 0.25^\circ$ horizontal resolution and
67 6-hourly intervals (Hersbach, 2016), while chemical initial and boundary conditions were obtained from Carbon Tracker and
68 have a horizontal resolution of 3° in longitude and 2° in latitude, with 25 vertical layers (<http://carbontracker.noaa.gov>). This
69 global dataset was interpolated to provide the lateral boundary conditions for CO_2 simulations and ensure consistency with the
70 input data for WRF-Chem. The main physics and chemistry options used in this study are listed in Table 1.

71 2.1.1 Anthropogenic Emissions

72 In the Metropolitan Area of São Paulo (MASP), the vehicular fleet is the primary source of CO_2 emissions (CETESB, 2019).
73 For this study, we employed the Vehicle Emission Inventory model (VEIN), a tool designed to estimate emissions from mobile
74 sources. VEIN accounts for both exhaust and evaporative emissions performs speciation, and includes functions to generate
75 and spatially allocate emissions databases (Ibarra-Espinosa et al., 2018). The model enables the use of customized emission
76 factors, which in this study were derived from experimental campaigns conducted in traffic tunnels within São Paulo (Nogueira
77 et al., 2021). VEIN processes vehicle fleet age distributions extrapolates hourly traffic data, and estimates emissions with high
78 temporal and spatial resolution. For consistency with the WRF-Chem model domain, VEIN emissions were aggregated to a
79 3 km spatial resolution. Additionally, we included a figure (Fig. B4) in Appendix B illustrating the daily mean and hourly
80 temporal variation of vehicular emissions for all months in the study period. In contrast, emissions from the energy and
81 industrial sectors were obtained from the EDGAR v6.0 GHG inventory for 2018 (Crippa et al., 2021). EDGAR provides

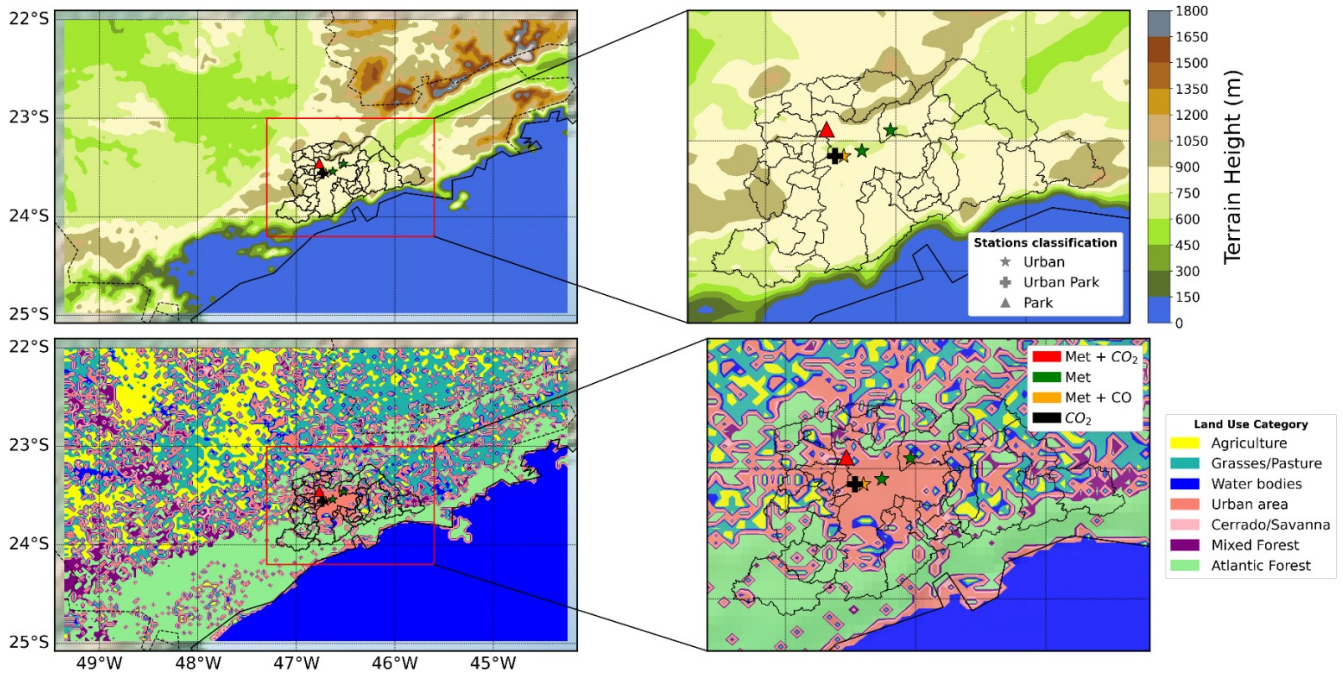


Figure 1. The first panel represents the terrain height and urban boundaries of the MASP area within the model domain (D01) used in the WRF-Chem model. The station classifications are represented by different symbols: Urban (★), Urban Park (⊕), and Park (▲). The second panel represents the land use category map for the same domain (D01), used in the VPRM model to calculate CO_2 fluxes. The station colors represent the variables measured at each location: red indicates stations measuring meteorological variables (Met) and CO_2 concentrations, green denotes stations measuring only meteorological variables, dark yellow represents stations measuring both meteorological variables and CO concentration, and black represents stations measuring only CO_2 concentrations. The IAG station is (⊕), the PDJ station is (▲), Pinheiros station is (★), Guarulhos and Parque D.Pedro II are (★).

global emission data with a spatial resolution of $0.1^\circ \times 0.1^\circ$. These emissions were processed to match the 3 km spatial resolution of the WRF-Chem model using interpolation techniques. However, EDGAR emissions lack temporal variability and were assumed constant throughout the day, as the inventory does not provide hourly profiles (Fig. in B5 Appendix B).

2.1.2 Biogenic Fluxes

Biogenic CO_2 fluxes were simulated offline using the Vegetation Photosynthesis and Respiration Model (VPRM) (Mahadevan et al., 2008), such as a flux input incorporated within the WRF-Chem simulations as input data. This model estimates net ecosystem exchange (NEE) by calculating the difference between gross ecosystem exchange (GEE) and ecosystem respiration (R), where negative fluxes indicate CO_2 absorption by ecosystems (Equation 1).

$$NEE = GEE - R \quad (1)$$

Table 1. WRF-Chem Simulation Design.

Atmosphere Schemes		
Scheme	Type	Description/Reference
Microphysics	Two-moment	Morrison scheme (Morrison et al., 2009)
Longwave radiation	RRTMG	(Iacono et al., 2008)
Shortwave radiation	RRTMG	(Iacono et al., 2008)
Boundary layer	YSU	(Hong et al., 2006)
Land surface	Noah LSM	Unified scheme (Tewari et al., 2007)
Initial and Lateral Boundary Conditions		
Meteorological	ERA5	0.25°, 34 pressure levels
Chemical	Carbon Tracker	25 vertical layers
Emissions Inventories/Model		
Anthropogenic	EDGAR v6.0	(Crippa et al., 2021) and VEIN (Ibarra-Espinosa et al., 2018)
Biogenic	VPRM	(Mahadevan et al., 2008)

91 The meteorological variables 2m air temperature (T_{2m}) and downward shortwave radiation (PAR) from WRF model simula-
92 tions were used to calculate the GEE (Equation 2) and Respiration (Equation 3) fluxes. Additionally, factors such as the light
93 use efficiency (λ), PAR saturation (PAR0), and the Enhanced Vegetation Index (EVI), which refer to the fraction of shortwave
94 radiation absorbed by leaves were used to calculate GEE. The temperature sensitivity of the photosynthesis parameter (Tscale)
95 and the effects of leaf age on canopy photosynthesis parameter (Pscale) were both calculated as functions of the land surface
96 water index (LSWI) to identify the green-up (leaf expansion) and senescence phases (Mahadevan et al., 2008). These vegeta-
97 tion indices were derived from Moderate Resolution Imaging Spectroradiometer (MODIS) reflectance data from MOD09A1
98 Version 6 (Vermote, 2021).

99
$$GEE = \lambda \times T_{scale} \times P_{scale} \times W_{scale} \times EVI \times \frac{1}{1 + \frac{PAR}{PAR_0}} \times PAR$$
 (2)

100 Respiratory fluxes (R) were estimated using a linear model based on air temperature and two parameters that represent the
101 linear sensitivity of respiration to air temperature (α) and the baseline respiration (β), as defined in Mahadevan et al. (2008).

102
$$R = \alpha \times T_{2m} + \beta$$
 (3)

103 The land cover data used by the VPRM were derived from the MapBiomass data (Souza Jr et al., 2020). The VPRM param-
104 eters (λ , PAR0, α , β) were optimized against flux tower NEE for the main land cover type over the study domain described in
105 section 2.2.2.

106 **2.1.3 Meteorological data**

107 Meteorological data from the São Paulo State Environmental Protection Agency (CETESB) air quality network were used to
108 evaluate the model’s performance in simulating meteorological fields. CETESB manages automatic and manual air quality
109 stations over São Paulo state. These stations provide hourly information on meteorological and pollutant parameters, such
110 as air temperature, wind speed, and wind direction (Table 2), as well as the concentration of air pollutants. Monitoring fol-
111 lows instrumentation standards and directives from the Environmental Protection Agency (US EPA) and the World Health
112 Organization (WHO) respectively for air pollutants, and from the World Meteorological Organization (WMO) for meteoro-
113 logical variables (CETESB, 2019). The air quality and meteorological data are continuously published on the Qualar website
114 (<https://qualar.cetesb.sp.gov.br/qualar/>). This study used data from four stations located in the MASP (Figure 1): Parque D.
115 Pedro II, PDJ, Guarulhos, and Pinheiros. Table 2 provides the location of the sites, the classification type of the stations, the
116 observed variables, and the data source.

Table 2. Location of the sites used for the model evaluation of the meteorological drivers, together with a list of the meteorological variables included in the analysis.

Sites	Location	Classification	Variables	Source Data
Parque D. Pedro II	23.54S, 46.63W	Urban	T_{2m} , WD, WS	CETESB
PDJ	23.45S, 46.76W	Park	T_{2m} , WD, WS and CO_2	CETESB/ METROCLIMA
Guarulhos	23.46S, 46.52W	Urban	T_{2m} , WD, WS	CETESB
Pinheiros	23.46S, 46.70W	Urban	T_{2m} , WD, WS and CO	CETESB
IAG	23.55S, 46.73W	Urban Park	CO_2	METROCLIMA

Note: Air temperature at 2 m (T_{2m}), wind speed (WS), and wind direction (WD).

117 **2.2 CO_2 observational data**

118 **2.2.1 Ground-based observations**

119 For the surface model evaluation, we used CO_2 data from the METROCLIMA network in São Paulo (see Table 3 and Figure 1),
120 the first conventional in situ greenhouse gas measurement network established in South America (www.metroclima.iag.usp.br).
121 The network comprises four continuously operating monitoring stations, all located within the MASP and equipped with cavity
122 ring-down spectroscopy instruments (Picarro) that measure the concentrations of CO_2 following the directives from WMO.
123 The monitoring stations are located at various locations within MASP: in a vegetated area at the extreme west (Pico do Jaraguá,
124 PDJ); in a suburban area in the center-west, inside the campus of the University of São Paulo (IAG); at the top of a 100 m
125 building (ICESP); and in an urban area in the east zone characterized by heavy traffic in the neighborhood (UNICID). However,
126 we only used data from the IAG and PDJ sites, which are 13 km apart, as these were the only two stations monitoring CO_2
127 during the selected study period, prior to the Covid-19 pandemic (Souto-Oliveira et al., 2023).

Table 3. Description of the METROCLIMA monitoring stations utilized in this study.

Station	Instrument	Inlet elevation (m)	Altitude (m)
PDJ	G2301 II	3	1079
IAG	G2301 II	15	731

2.2.2 CO_2 fluxes data and VPRM optimization

In this study, the VPRM model computed the biosphere fluxes for 5 different plant functional types (PFT), representing different vegetation land covers, and for that required a set of four model parameters for each vegetation class, dependent on the region of interest. Ideally, these parameters are optimized using a network of eddy flux towers for each PFT over the domain. The VPRM parameters were optimized for only three plant functional types (PFT) corresponding to the three ecosystems observed by eddy-covariance flux towers. However, these three PFT represent almost 60% of land covers over the domain (i.e. sugarcane - 23.86%, Atlantic Forest - 34.86%, and cerrado - 0.91%). We used a set of parameters optimized by Botía et al. (2022) for the remaining PFT's, such as grasses and mixed forest, based on measurements from sites in the Amazon region in Brazil, deployed in the context of the Large Scale Biosphere-Atmosphere Experiment (LBA-ECO) (Botía et al., 2022). The methodology for optimizing the VPRM parameters for the Atlantic Forest used data from Serra do Mar State Park in São Paulo State, Brazil (23°17'S, 45°03'W at 900 m altitude) for the period from January to 2015 to December 2015 (Freitas, 2012). For cerrado, we used observed data from Pé Gigante, in São Paulo, Brazil (21°36'S, 47°34'W at 660m) from January 2015 to January 2017 (Rocha et al., 2002). For sugarcane we used data from the municipality of Pirassununga, in São Paulo State, Brazil (21°57'S, 47°20'W at 655 m altitude) for the period from November 2016 to August 2017 (Cabral et al., 2020). The VPRM parameters were optimized separately for each PFT using half-hourly observed fluxes from the flux towers over the entire observation periods. We optimized the parameters for the GEE and R simultaneously, and for the default VPRM parameters we used non-linear least squares minimization, between the modeled NEE and the flux tower estimation of the observed NEE. In the optimization, the VPRM model is driven by the meteorological measurements of the sites and their specific land covers. The vegetation indices (EVI and LSWI) were derived from the product MOD09A1 of MODIS at 500 m resolution and 8-daily frequency using Google Earth Engine.

2.2.3 XCO_2 satellite observations

Satellite-based XCO_2 observations were utilized in addition to surface CO_2 measurements over the study domain. OCO-2, NASA's inaugural Earth remote sensing satellite dedicated to atmospheric CO_2 observations, was launched in 2014 (Crisp, 2015). Operating on a solar synchronous orbit, OCO-2 conducts global measurements of CO_2 absorption and emission at 13:30 Local Solar Time. The OCO-2 observation data utilized were ACOS L2 Lite Output Filtered with oco2-lite_fle_prefilter_b9, which were converted from Level 1 radiance to Level 2 data using the ACOS retrieval algorithm developed by O'Dell et al. (2012). Data quality assessment for OCO-2 observations can be performed using the xco2_quality_flag and warn_level parameters, as detailed in the OCO-2 Data Product User's Guide (Osterman et al., 2018). In this study, we considered only

OCO-2 data with a '0' xco2_quality_flag value that indicates "good" quality. Initially, simulated CO_2 concentrations were interpolated to match the latitude, longitude, horizontal resolution, and vertical levels of OCO-2 data. Additionally, to ensure consistency in the comparison, the simulated data were selected to correspond as closely as possible to the OCO-2 overpass time (13:30 Local Solar Time) over the study region. Due to the difference in data types and units between the simulated CO_2 concentrations and observed XCO_2 from satellites, a conversion was necessary prior to comparison. Consequently, CO_2 concentrations simulated at each pressure level in WRF-Chem were transformed into XCO_2 concentrations following the methods by Connor et al. (2008) and O'Dell et al. (2012), as follows:

$$XCO_2^{\text{model}} = XCO_{2a} + \sum_i w_i^T A_i (CO_2^{\text{interp}} - CO_{2a})_i \quad (4)$$

where XCO_{2a} is a priori XCO_2 , w_i^T is the pressure weighting function, A_i is the column averaging kernel, CO_2^{interp} is the interpolated simulated CO_2 concentrations of WRF-Chem, and CO_{2a} is a priori CO_2 .

2.3 Evaluation metrics

Several statistical metrics are available for assessing the effectiveness of atmospheric models. These include mean bias error (bias, Equation A1), indicating the average difference between the simulation and the observation; root-mean-square error (RMSE, Equation A2), which quantifies the square root of the average squared deviation between simulation and observation; and the correlation coefficient (R^2 , Equation A3), representing the degree and direction of the linear connection between the simulation and the observation. To evaluate the model results, bias, root mean square error (RMSE), and correlation (R^2), the equations are described in Appendix A.

3 Results

Hourly simulations were conducted from 1 February to 31 August 2019, with each month simulation including a five-day spin-up period. In the following sections, the performance of meteorological drivers will be first presented, followed by the terrestrial surface CO_2 fluxes and atmospheric CO_2 concentrations from the IAG and PDJ stations. These measurements were used to evaluate the model performances and to assess the local impacts of the main CO_2 sources and sinks on atmospheric CO_2 concentrations.

3.1 Model performance for meteorological drivers

The assessment of the meteorological model performances is essential for accurately simulating greenhouse gas concentrations. In this study, the model represented the temporal variability and trends of 2-meter temperature (T_{2m}), 10-meter wind speed (WS), and direction (WD) throughout the simulation period, as illustrated in Fig. 2 and the supplementary material. The WRF-Chem model effectively captured significant changes in the observed variables, although it failed to accurately represent the maximum and minimum peaks, particularly for wind speed. The simulated 2-meter temperature tended to overestimate values

at specific sites, such as Parque D. Pedro II (bias = 0.5°C), Guarulhos (bias = 0.1°C) (see figure B1a and B2a in Appendix B), and PDJ (bias = 0.7°C) (see Figure 2a). However, at the Pinheiros station, the simulated surface temperature was underestimated (bias = -0.7°C) (Figure B3a in Appendix B).

In terms of biases, the model overestimated the wind speed at all sites (bias < 1.5 ms⁻¹), with PDJ exhibiting the highest mean bias (1.4 ms⁻¹). This overestimation could be attributed to the model's misrepresentation of land use, leading to elevated wind speeds in areas classified as urban rather than vegetated. Notably, numerical models tend to lack sensitivity in simulating very low-velocity speeds due to imperfections in land surface processes and the model's ability to accurately resolve topographical features (Shimada et al., 2011; Zhang et al., 2009; Vara-Vela et al., 2018, 2021). The model's wind directions showed sufficient sensitivity, aligning accurately with observed values. Both the model and observations indicated that prevailing winds were predominantly from the southeast. In summary, the WRF model showed proficiency in reproducing atmospheric conditions in the study area, particularly concerning air temperature and wind direction, with similar performances as previous studies (Feng et al., 2016; Deng et al., 2017).

3.2 The VPRM Model: Evaluation with Flux Tower Data

The optimization results are shown in Table 4. Substituting alpha and beta back into the respiration equation led to a better model representation of NEE compared to NEE values simulated with default parameters (Mahadevan et al., 2008) for the main PFT across the domain.

Table 4. Default (Mahadevan et al., 2008) and Optimized VPRM parameters (highlighted) for atlantic forest, cerrado and sugarcane, and for mixed forest and grasses from Botía et al. (2022).

Type of Vegetation (PFTs)	PARo	Default			Optimized & Botía et al. (2022)			
		λ	α	β	PARo	λ	α	β
Atlantic forest	570	0.127	0.271	0.250	178615	0.008	-0.211	4.715
Mixed forest	629	0.123	0.244	0.240	206	0.255	0.342	0.000
Grasses	321	0.122	0.028	0.480	15475	0.056	0.312	7.337
Cerrado	3241	0.057	0.012	0.580	2300	0.616	0.070	1.665
Sugarcane	2051	0.200	0.209	0.802	14550	0.049	-0.339	10.052
Urban area	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

The optimized VPRM parameters for the Atlantic Forest exhibited the greatest discrepancies compared to other vegetation classes. The geomorphological characteristics of the Atlantic forest differ from those of the evergreen forest studied by (Mahadevan et al., 2008), where the default parameters (VPRM_default, represented by the red curve in Figure 3) were used. The optimized VPRM parameters (VPRM_optimized, shown as the green curve in Figure 3) more accurately captured the seasonal cycle in the daily average NEE for the three PFTs optimized in this study. The model was particularly successful in capturing the seasonal profile for the agricultural ecosystem, which can be attributed to the more pronounced seasonal transitions of sugarcane (as indicated by the EVI), even though the low-resolution satellite indices do not fully capture the onset of the growing

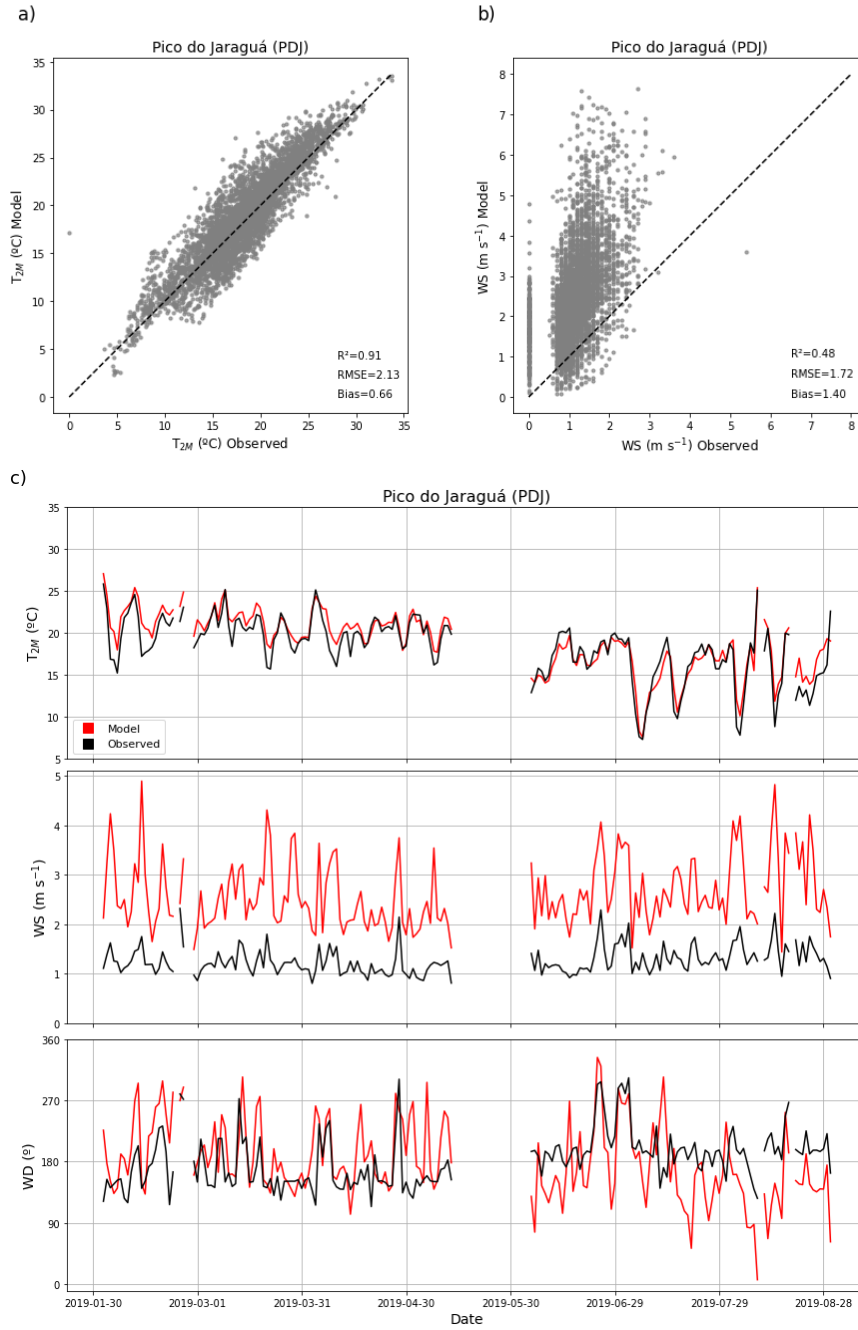


Figure 2. The panels in a) show the scatter plots of hourly measurements of 2 m air temperature (T_{2m}) and b) show 10 m wind speed (WS) compared to observed data from the PDJ station. The figure illustrates the relationship between modeled and observed data. The panels in c) show the daily averages from February to August 2019 of 2 m air temperature (T_{2m}), 10 m wind speed (WS), and wind direction (WD). Black line represents the observed data and red line represents the model simulation.

208 season. However, this allowed the model to better represent the GEE equation for this ecosystem. For the cerrado, the model
 209 smoothed the NEE peaks, and the GEE and respiration equations were also smoothed with the optimization. Optimizing the
 210 VPRM parameters improved the representation of the growing season, especially for the Atlantic Forest and sugarcane, while
 211 using either optimized or default parameters for the cerrado resulted in similar NEE simulation.

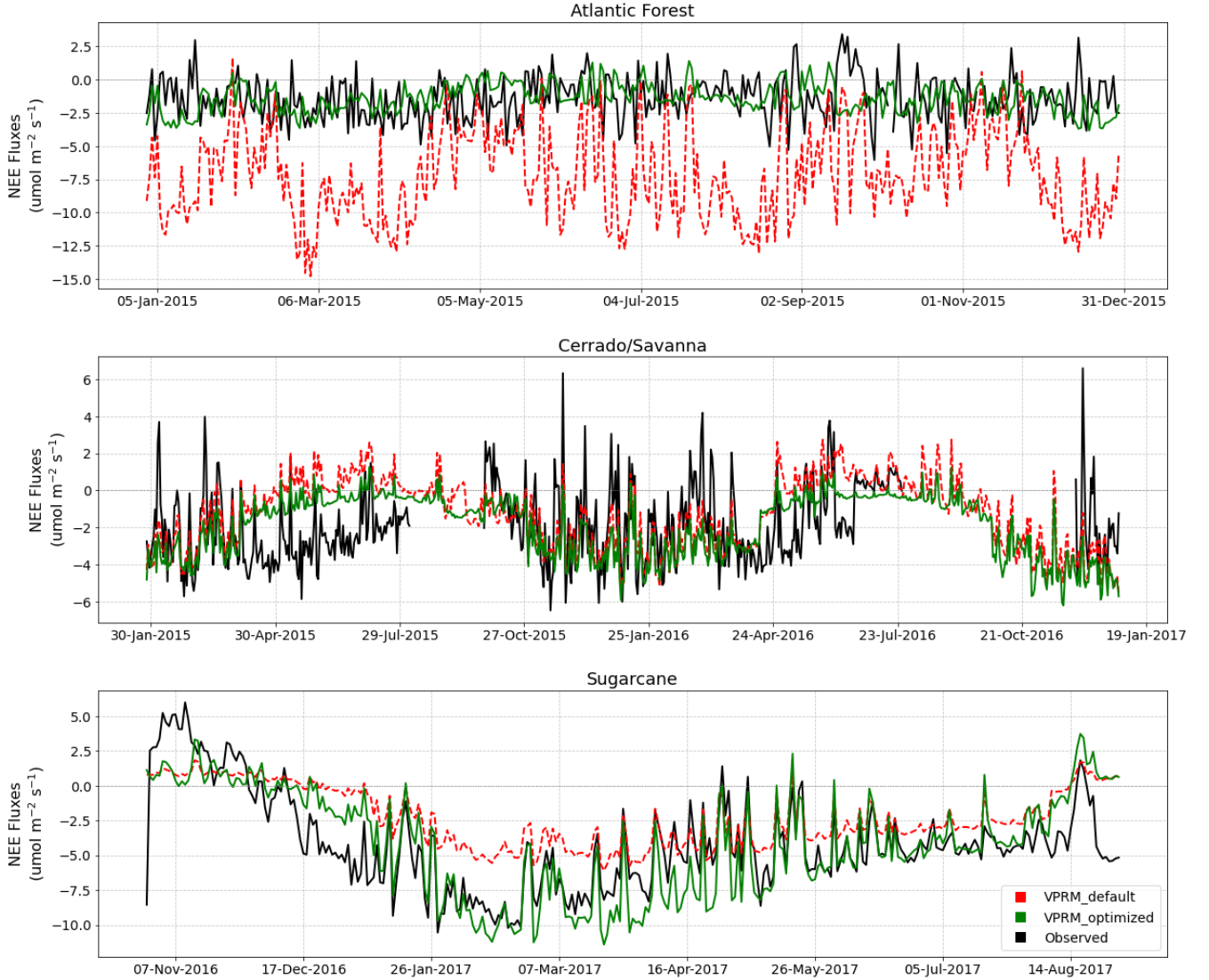


Figure 3. Daily variability of NEE fluxes ($\mu\text{mol m}^{-2} \text{s}^{-1}$) from the flux tower (black line), compared with NEE fluxes simulated by the VPRM model using default (red line) and optimized (green line) parameters for the Atlantic Forest, Cerrado/Savanna, and Sugarcane.

212 The first panel in Figure 4 shows the monthly net CO_2 flux simulated by the VPRM model for 2019. February represents a
 213 summer month, while August represents a winter month. The second panel shows the monthly daily net CO_2 flux simulated

214 at the three flux tower sites used to optimize the VPRM model parameters. In February, negative NEE values are found in the
 215 northern part of the MASP, while the southern part exhibits positive NEE fluxes in the coastal region. This gradient reflects
 216 the distribution of vegetation types, their phenology, and productivity, as well as the impact of urbanization, with null fluxes
 217 observed in highly urbanized regions such as downtown MASP. In the summer season in the Southern Hemisphere, vegetation
 218 productivity reaches its peak across all land cover classes, leading to more negative NEE values (Figure 4a) specially for the
 219 Cerrado, and sugarcane (within the state of São Paulo), but less intense in the Atlantic Forest in the southern coastal region.
 220 This results in negative NEE fluxes (dark green across the domain), indicating that these areas acted as a CO_2 sink.

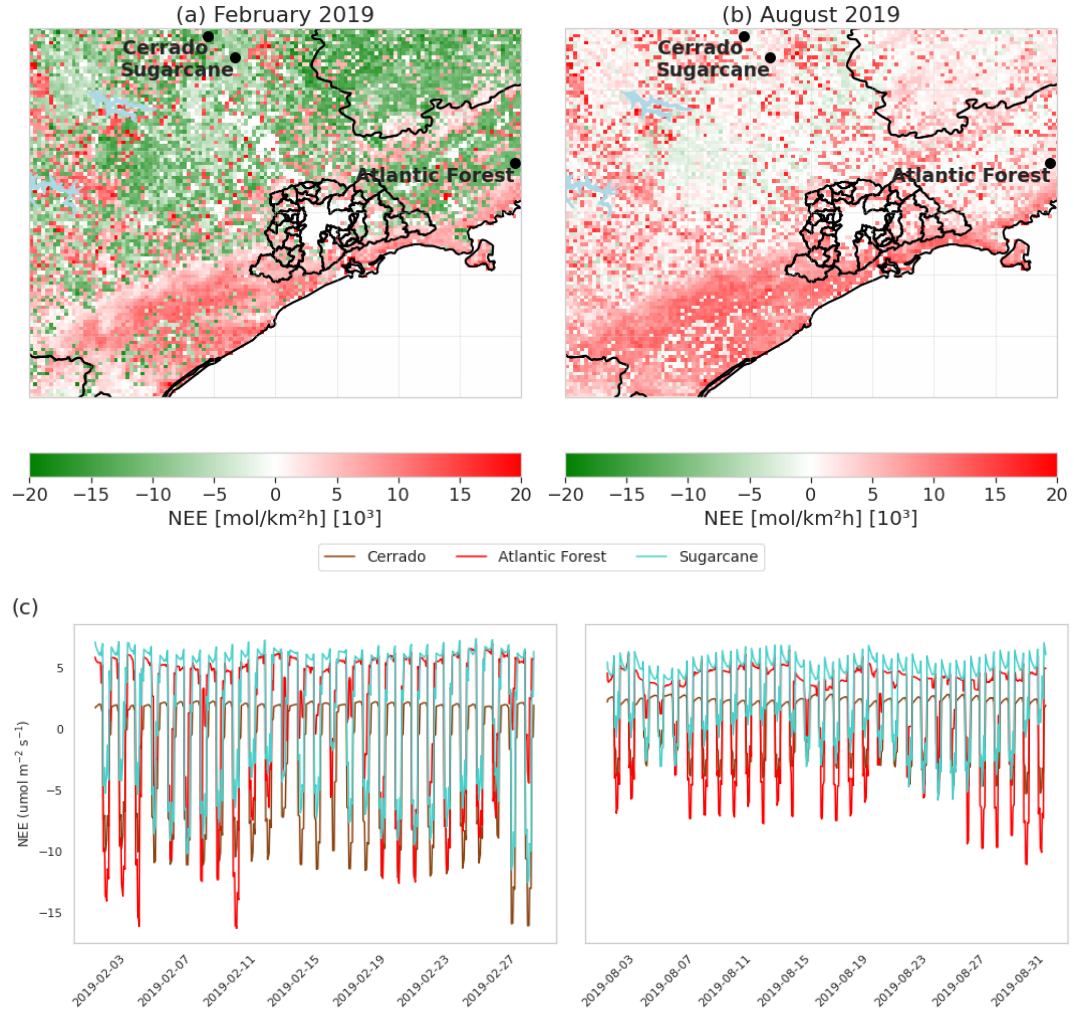


Figure 4. The first panel shows the monthly mean diurnal cycle of net ecosystem exchange (NEE) ($mol km^{-2} h^{-1}$) for February (a) and August (b) 2019. The second panel (c) presents the daily variability of NEE ($\mu mol m^{-2} s^{-1}$) for the same months (February and August) at three different PFTs: Atlantic Forest, Cerrado/Savanna, and Sugarcane.

221 In August, the cold and dry conditions, due to reduced solar radiation and a lower leaf area index, resulted in positive fluxes
 222 across most of the domain and low negative fluxes in only a few areas (Figure 4b). The highest positive NEE values are found in
 223 the southern coastal region. Generally, larger areas with negative CO_2 fluxes are observed in February compared to August for
 224 the same dominant land cover classes. This indicates greater CO_2 absorption by agriculture in February compared to forested
 225 regions. Conversely, in August, CO_2 fluxes are predominantly lower and negative across most of the domain, with higher
 226 positive values in the coastal area, especially in the south. Overall, the domain acts as a net CO_2 sink during summer, while
 227 vegetation becomes a CO_2 source in winter, except for the Atlantic Forest in the southern part of the study area. The second
 228 panel also shows simulated fluxes for the same flux tower sites, with negative net fluxes in February, particularly in the Atlantic
 229 forest, sugarcane, and cerrado. This underscores the reduction in negative fluxes during winter, as seen in the August data for
 230 all three vegetation types. Unfortunately, observed data from these flux towers for this period were not available for statistical
 231 model evaluation. However, Figure 4 illustrates the significant influence of climatic drivers on reduced flux trends, consistent
 232 with findings by Raju et al. (2023) for a tropical region. Note that the respiration equation in Mahadevan et al. (2008) is a
 233 simple linear function of temperature and does not account for seasonal or spatial variability in biomass and litter inputs to soil
 234 carbon pools Gourdji et al. (2022), which is particularly relevant for forest ecosystems like the Atlantic Forest.

235 3.3 Seasonal variations in observed and modeled CO_2 mixing ratios

236 Figure 5 and Table 5 depict the monthly mean, standard deviation, bias and RMSE of CO_2 concentrations at two sites in the
 237 MASP. In 2019, the IAG station recorded CO_2 values ranging from 406 to 464 ppm. The seasonal variation peaked during
 238 winter (June to August, 437.3 ± 32.2 ppm), followed by autumn (March to May, 433.0 ± 26.0 ppm), with the lowest values
 239 observed in summer (February, 432.7 ± 24.6 ppm). This variation in CO_2 levels is primarily influenced by factors such
 240 as the latitude of the observation site, meteorological conditions including wind speed and atmospheric stability, as well as
 241 seasonal patterns of photosynthesis and vehicular traffic see Fig.B4 in Appendix B. The maximum and minimum monthly
 242 CO_2 concentrations at IAG were recorded in June (442.5 ± 32.8 ppm), during the winter season, and March (430.2 ± 24.5
 243 ppm), during the autumn season, respectively. During this month, the MASP experiences changes in synoptic circulation and
 244 atmospheric moisture that typically reduce atmospheric stability and increase the dispersion of various gases and particles
 245 (Chiquetto et al., 2024). Meanwhile, at the PDJ station, CO_2 levels ranged from 414 ppm to 417 ppm. The seasonal variation
 246 peaked during autumn (416.8 ± 9.5 ppm), closely followed by summer (416.0 ± 10.3 ppm), with the lowest values observed in
 247 winter (414.6 ± 7.4 ppm). The maximum monthly CO_2 mean at PDJ was identified in May (417.3 ± 9.1 ppm), corresponding
 248 to the autumn season, while the minimum was recorded in July (414.0 ± 6.3 ppm), during the winter season. Monthly values at
 249 PDJ exhibited less variability and a smaller standard deviation compared to the IAG site. This result was expected, considering
 250 that the IAG site is significantly impacted by vehicular traffic in its vicinity. In contrast, PDJ is located at a higher elevation
 251 in a more vegetated area, with less influence from local anthropogenic sources. Additionally, it was expected that PDJ would
 252 show lower CO_2 concentrations during the summer due to the stronger vegetation signal at PDJ compared to the IAG site.
 253 However, PDJ actually shows peak CO_2 levels in summer and the lowest values in winter, indicating that additional ecological
 254 and ecosystem variables need to be considered for a better understanding of this location.

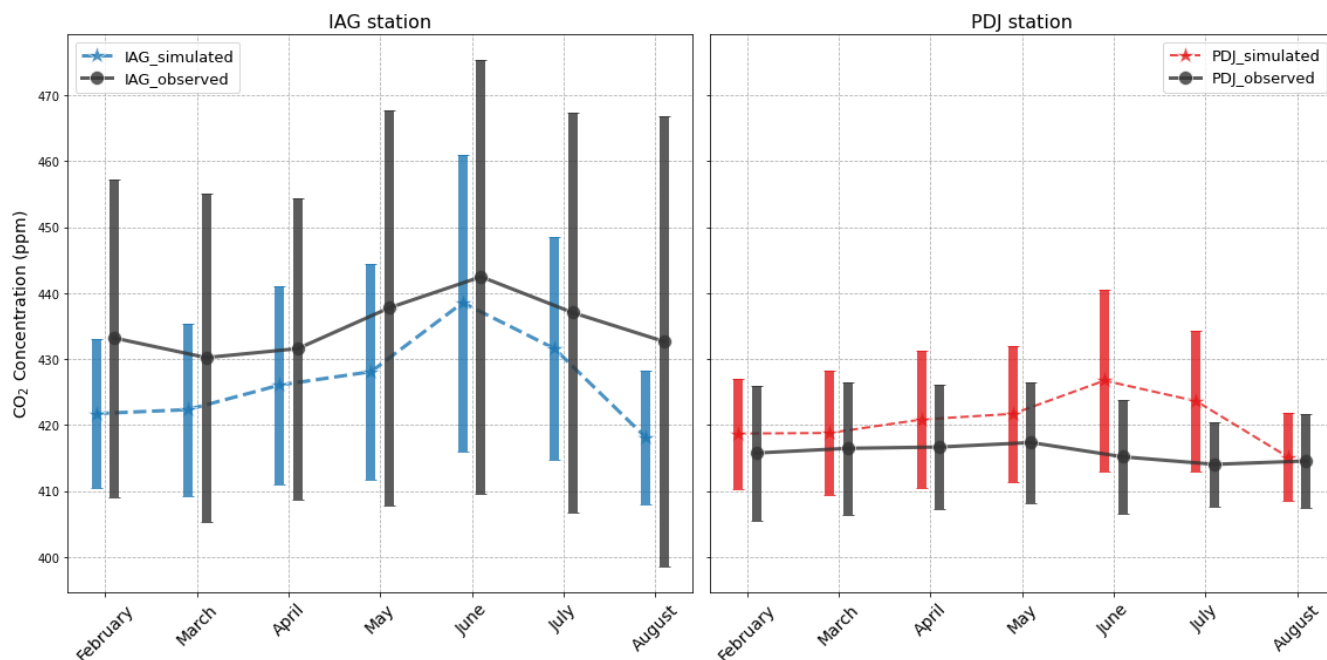


Figure 5. CO_2 concentration seasonality observed and simulated at IAG and PDJ stations in 2019. Error bars represent the monthly standard deviation.

255 The simulated CO_2 concentrations for the IAG station ranged from 410 ppm to 437 ppm, with a seasonal variation peaking
 256 in winter (429.4 ± 19.2 ppm), followed by autumn (425.2 ± 15.1 ppm), and the lowest values occurring in summer ($422.3 \pm$
 257 12.3 ppm), mirroring the observed data. Notably, the highest and lowest monthly CO_2 concentrations at IAG were identified
 258 in June (438.7 ± 22.5 ppm) and February (418.1 ± 10.0 ppm), respectively. Although the maximum monthly value from the
 259 model coincided with the observed data, the month with the minimum concentration was February, which may be attributed to
 260 gaps in measurement, which were not considered when calculating the mean, thereby influencing the observed monthly mean.
 261 The CO_2 concentrations at PDJ ranged from 415 ppm to 426 ppm, with seasonal variation peaking in winter (421.8 ± 11.8
 262 ppm), followed by autumn (420.4 ± 10.1 ppm), and the lowest values occurring in summer (419.0 ± 8.8 ppm). The model
 263 data profile for PDJ more closely resembles the simulated IAG profile than the PDJ station's observed profile, likely due to the
 264 model's resolution, its limitations in representing land use, and underestimated vehicular emissions in these areas. However,
 265 negative biases were observed for all seasonal periods at IAG, indicating an underestimation of CO_2 concentrations and higher
 266 root mean square errors compared to the statistics for the PDJ station. The PDJ station exhibited low positive biases, indicating
 267 better agreement between the model and observations across all periods and lower errors between the model and observations.
 268 The higher elevation and vegetation cover at PDJ simplify seasonal trend modeling, reducing the impact of urban factors and
 269 making model predictions more accurate (see Figure B6 in Appendix B).

Table 5. Seasonality means and standard deviation of CO_2 concentrations for IAG and Pico do Jaraguá (PDJ) stations.

Station	Season	Observed (ppm)	Simulated (ppm)	Bias (ppm)	RMSE (ppm)
IAG	Summer (February)	432.7 ± 24.6	422.3 ± 12.3	-12.1	25.2
	Autumn (MAM)	433.0 ± 26.0	425.2 ± 15.1	-7.5	24.8
	Winter (JJA)	437.3 ± 32.2	429.4 ± 19.2	-7.2	31.1
PDJ	Summer (February)	416.0 ± 10.3	419.0 ± 8.8	3.6	11.1
	Autumn (MAM)	416.8 ± 9.5	420.4 ± 10.1	3.6	12.0
	Winter (JJA)	414.6 ± 7.4	421.8 ± 11.8	7.3	13.8

3.3.1 Distribution of surface CO_2 concentrations

In addition to the simulations conducted for the period from February to August 2019, using the same configurations and input data, we performed simulations involving variable emission scenarios for the summer (February) and winter (August) seasons. The aim was to comprehensively understand the dynamics of CO_2 concentration in the metropolitan region and surrounding areas during these distinct seasonal periods. Figure 6 shows the monthly average spatial distributions of simulated CO_2 concentrations under four conditions: a) Background without emissions, considering only boundary and initial conditions (BCK); b) considering both anthropogenic emissions and biogenic fluxes (see Table 1) (ALL); c) considering biogenic fluxes only (BIO); and d) considering anthropogenic emissions (industrial and vehicular) only (ANT).

Figure 6a shows that the simulated background CO_2 concentration in February ranged around 408 ppm across most of the domain. For biogenic simulations (Fig. 6c), we observed an average increase of 14 ppm across the domain compared to the previous simulation. The increase, however, was only 6 ppm in downtown MASP. Although the VPRM model did not explicitly calculate CO_2 fluxes in urban areas due to limited vegetation coverage, the transport of biogenic signals from the surrounding vegetated regions into the urban area is evident. The southwest region of the domain, characterized by the Atlantic Forest, exhibits the highest CO_2 concentrations in this scenario, ranging from 420 to 424 ppm. This dense vegetation region and higher ecosystem respiration contribute to elevated CO_2 levels, underscoring the influence of biogenic sources on regional concentration patterns. This region has altitudes lower than 200 m and the CO_2 released to the atmosphere by the vegetation is trapped due to the Serra do Mar, with altitudes higher than 500 m. The Atlantic Forest present on the northern coast, on the other hand, is concentrated on the plateau of Serra do Mar, and thus, the CO_2 released is better dispersed to other areas. The simulation with anthropogenic emissions (Figure 6d) stands out elevated CO_2 concentrations over the center of the city of São Paulo, characterized by high vehicle emissions, as well as over other two urban areas in the north and northeast of MASP. The monthly mean CO_2 concentration in these two urban areas was roughly 420 ppm, attributed to emissions from refineries represented by the EDGAR datasets as well as vehicles. Figure 6b shows the simulated CO_2 concentration considering both vegetation fluxes and anthropogenic emissions. As expected, this simulation combines both contributions, resulting in high CO_2 concentrations over urban areas and along the coastal region. For August, it can be observed that the background concentrations (Figure 6e) were slightly higher around MASP. Additionally, the monthly mean

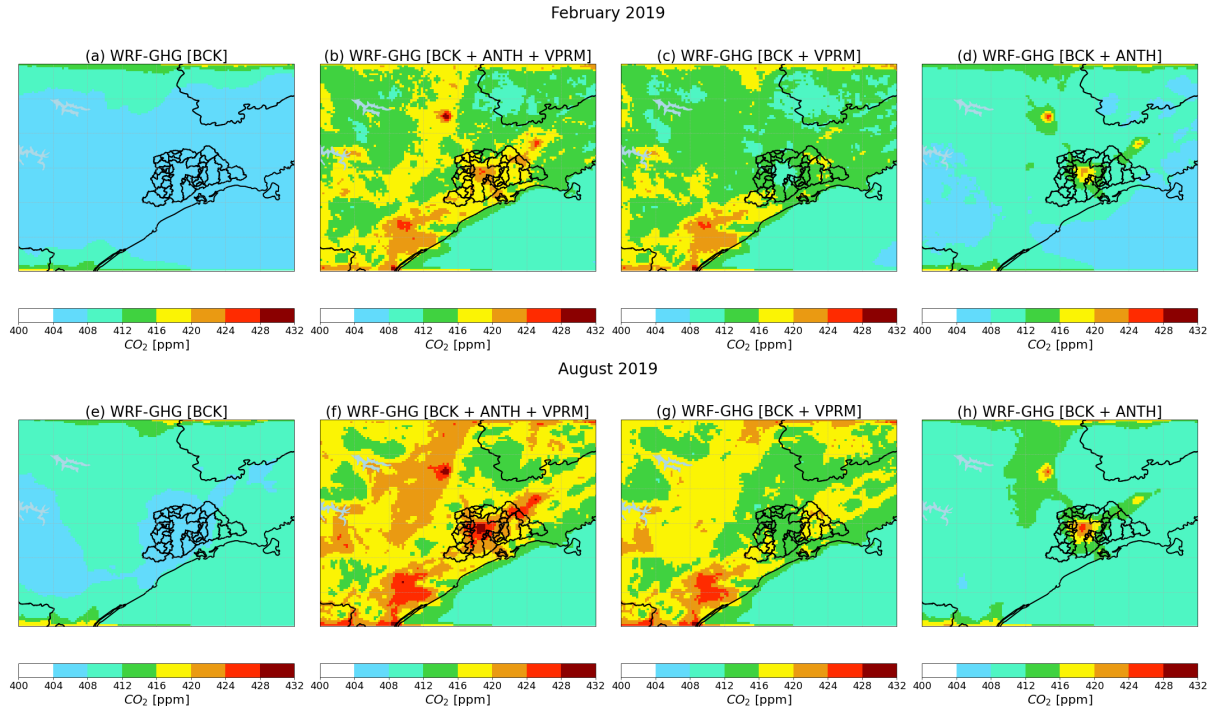


Figure 6. Atmospheric CO_2 concentrations under different emission scenarios (refer to the text). The panels in the first row represent the monthly mean concentration for February (a, b, c, d), while the panels in the second row represent the monthly mean concentration for the August period (e, f, g, h). Panels a) and e) represent the background scenario. Panels b) and f) represent simulation of total (background, anthropogenic and biogenic) emissions scenario, panels c) and g) represent simulation of only background and biogenic scenario, and d) and h) represent simulation of only background and anthropogenic scenario.

295 CO_2 concentration for the scenario in August with only biogenic sources was 8 ppm higher than that in February, which can
 296 be explained by the lower photosynthetic rates in this period, as observed in Figure 4. The Atlantic forest in the coastal region
 297 exhibits more positive CO_2 fluxes and lower photosynthetic activities, characterized by lower amounts of rainfall in the region
 298 that contribute to this reduced photosynthetic production by vegetation. The simulation with only anthropogenic emissions
 299 (Figure 6h) shows higher CO_2 concentrations compared to those in February. This increase in CO_2 levels in August is attributed
 300 to a lower planetary boundary layer height. However, it is important to point out that the EDGAR anthropogenic emission
 301 inventory generally overestimates the emissions around local anthropogenic sources (e.g., urban areas) (Seo et al., 2024). The
 302 higher simulated CO_2 concentration for August compared to February, in the scenario with both biogenic and anthropogenic
 303 sources, is largely dependent on factors such as atmospheric stability and meteorological conditions. Atmospheric stability,
 304 along with meteorological variables such as humidity, solar radiation, and temperature, plays a crucial role in determining
 305 biogenic CO_2 concentrations. In addition, under stable atmospheric conditions, such as those often observed during winter
 306 periods, CO_2 concentrations tend to accumulate near the surface, resulting in higher concentrations, especially in urban areas.

Therefore, the comparative analysis between simulations of CO_2 concentrations during summer and winter periods highlights the importance of considering not only anthropogenic emissions but also biogenic fluxes from vegetation, along with local atmospheric conditions.

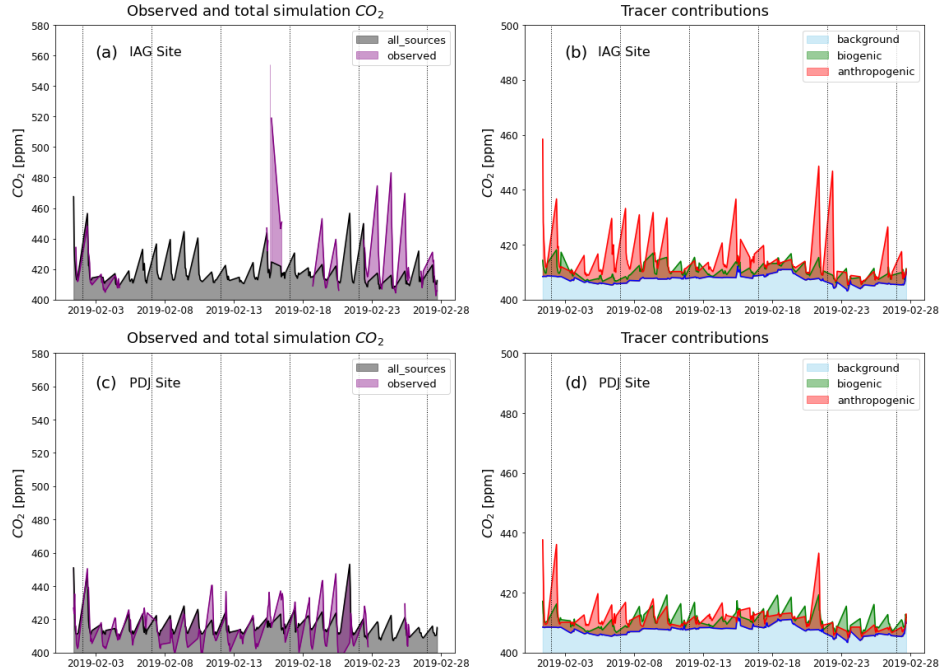
3.3.2 Evaluation of sources contribution

In Figure 7, we applied a data selection scheme to all-time series to minimize the effects of local contributions and increase the spatial representativeness of each record, it consists of retaining mid-afternoon (09–17 h local) data, when the air is well-mixed, providing a large spatial representativeness with minimum influence from local sources (Gerbig et al., 2008; Ramonet et al., 2020). Figure 7, shows the comparison of the daily mid-afternoon average CO_2 concentrations simulated by the model for February and August 2019, considering both biogenic and anthropogenic sources (see Figures 6b and 6f), at both IAG and PDJ sites. The left panels (Figures 7a, 7c, 7e, and 7g) depict the simulated CO_2 concentration considering both anthropogenic and biogenic sources (all_sources, in gray), alongside observed concentrations (observed, in purple) for both sites. Conversely, the right panels (Figures 7b, 7d, 7f, and 7h) display the different simulations considering anthropogenic and biogenic sources separately to the daily concentration. In Figure 7a, which represents the sole summer month with observed data in February 2019, the simulated values generally underestimated the observed concentrations. While the observed average CO_2 concentration stood at 424.0 ppm, this figure was somewhat compromised by missing data in the observed profile, whereas the simulated average was 416 ppm, indicating an approximate 8 ppm discrepancy below the monthly average observed in February. For the anthropogenic sources the simulation is aligned with the expectations that the emission is dominated by vehicular emissions around this vicinity (Fig. 7b). However, on February 23rd, 24th and 25th, there was a distinct peak in the observed CO_2 concentrations. This spike is absent in both the all-source and anthropogenic simulations, suggesting that other localized or transient activities, not accounted for in the emissions inventory, may have contributed. This discrepancy likely arises because the inventories assume identical emissions for all days with only hourly variations. As a result, specific events or activities that occur on these particular days are not captured in the simulations. Furthermore, on February 2nd, 21st, and 22nd, observed CO_2 peaks were captured by the model with the same magnitude only when both anthropogenic and biogenic emissions were included. Simulations considering only anthropogenic sources underestimated these peaks, highlighting the importance of biogenic contributions to accurately representing observed concentrations.

At the PDJ site, the mean observed and simulated CO_2 concentration for the study period was 414 ppm. The model captures the overall trend and major peaks of CO_2 variability during this period, with biogenic contributions more pronounced at PDJ compared to the IAG site (Figure 7d). This higher biogenic influence in PDJ is attributed to its location in a vegetated area and localized in higher altitude than IAG, relatively isolated from vehicular emissions and other anthropogenic sources typical of urban environments, as previously discussed.

In August, characterized by a drier, more stable boundary layer and lower wind speed, observed data for IAG showed an average of 426 ppm (Figure 7e), while with the model showed a monthly average of 413 ppm, resulting in a discrepancy of only 13 ppm, i.e. a closer approximation compared to February. In terms of the contributions of the sources (Figure 7f), simulations showed similar daily patterns, with a few days where CO_2 contributions from biogenic fluxes exceeded those

February 2019



August 2019

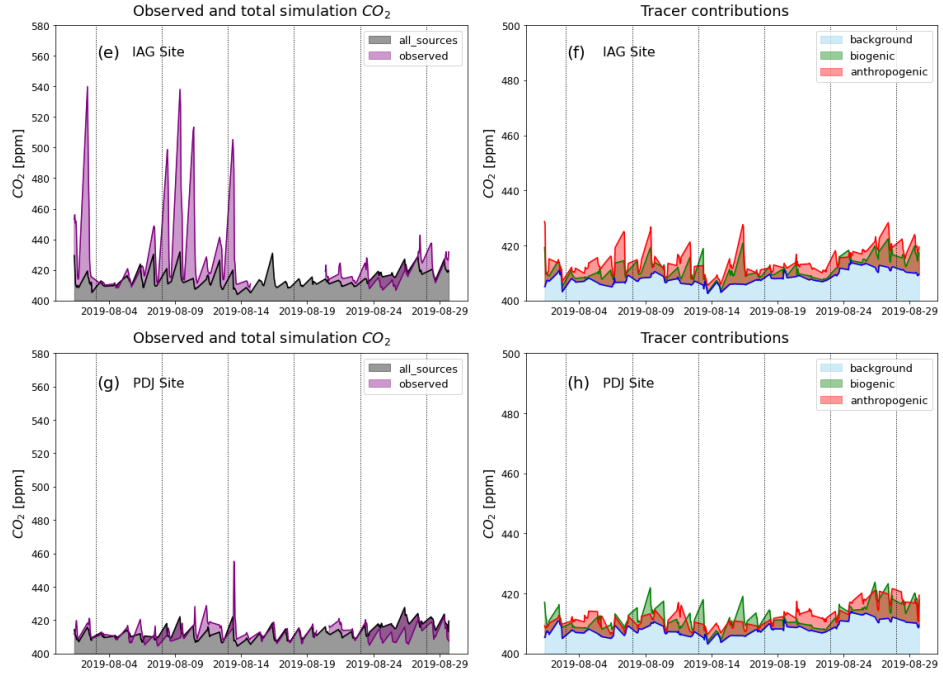


Figure 7. Daily mean CO_2 concentrations simulated and observed for the IAG site in February 2019 (a), for the PDJ site in February (c), for the IAG site in August (e), and for the PDJ site in August (g). And the daily simulated at the BCK (background), VPRM (biogenic), and ANTH (anthropogenic) scenarios for the IAG site during February (b), for the PDJ site in February (d), for the IAG site during August (f), and for the PDJ site in August (h).

341 from anthropogenic emissions. Oppositely, for PDJ (Figure 7g), the monthly average concentration stood at 412 ppm, slightly
342 surpassing the simulated average of 412 ppm. While the model slightly underestimated the monthly average, it generally
343 captured the observed variability. The higher monthly mean in the observations was influenced by a peak in mid-August, but
344 differences between observed and simulated values varied throughout the month. Before late August, observed values tended
345 to be higher than the simulations, whereas, in the final days of the month, the model overestimated CO_2 concentrations. This
346 highlights the role of both biogenic and meteorological processes in shaping CO_2 variability at this site (Fig. 7h), emphasizing
347 the importance of considering these dynamics in future simulations. Additionally, Figure 4 illustrates more positive CO_2 fluxes
348 (representing CO_2 emissions to the atmosphere) by the VPRM model during this period.

349 The bias and RMSE for each simulation at the IAG and PDJ sites for February and August 2019, are illustrated (see Figure
350 B7 in Appendix B). Overall, the bias tended to be negative across the board, indicating that the simulated surface CO_2 con-
351 centrations generally underestimated the observed values. Notably, with the exception of the ALL_PDJ simulation for August
352 (Figure 7c), which displayed a small negative bias, CO_2 simulations consistently fell below-observed levels at this site. Among
353 the six sets of simulations, PDJ exhibited the smallest bias, averaging at -3.0 ppm, while IAG displayed a larger average bias of
354 -13.3 ppm. Further analysis revealed that simulations incorporating both biogenic and anthropogenic sources (ALL_*) consis-
355 tently yielded the smallest biases. RMSE values at PDJ remained below 12.0 ppm, while those at IAG exceeded this threshold.
356 Notably, simulations focusing solely on anthropogenic sources at PDJ exhibited the poorest RMSE for both February and
357 August, highlighting the significance of vegetation fluxes at this site. On the other hand, at IAG, simulations relying solely on
358 biogenic sources in February and on anthropogenic sources in August resulted in the highest RMSE values, highlighting the
359 importance of anthropogenic emissions, especially traffic ones. In February, simulations ALL_IAG and ALL_PDJ displayed
360 the lowest RMSEs at 22.63 ppm and 10.12 ppm, respectively (Figure B7b in Appendix B). In August, these figures stood
361 at 26.90 ppm and 6.83 ppm (Figure B7d in Appendix B), respectively. Overall, simulations incorporating both biogenic and
362 anthropogenic sources yielded better results in terms of RMSE and bias, indicating a closer alignment between simulated and
363 observed surface CO_2 concentrations. Additionally, CO_2 simulations at PDJ demonstrated the closest resemblance to observed
364 values among the six simulations.

365 Considering that CO serves as a vehicular tracer, we analyzed CO concentrations at the Pinheiros site using data from
366 the CETESB network (see Figure 1 and Table 1) to compare with CO_2 concentration profiles at the IAG site for February
367 to August 2019, located less than 3 kilometers away from the Pinheiros site. The hourly correlation between observed CO_2
368 concentrations at the IAG site and observed CO concentrations at Pinheiros was determined, along with the correlation between
369 simulated CO_2 concentrations for IAG and observed CO concentrations. In Figure 8, both bar graphs of the hourly correlation
370 between CO_2 vs. CO concentrations show a correlation above 0.5 for observed CO_2 and 0.25 for simulated CO_2 during the
371 early hours of the day until 10h, and again in the late afternoon after 19h, which corresponds to periods of high vehicular traffic
372 in this region. Midday, this correlation decreases and even turns negative for the simulated CO_2 vs. CO graph, suggesting
373 the influence of the photosynthesis process on CO_2 concentrations, which is also evident in the observed data. The similarity
374 between the trend lines of the hourly correlation profiles for observed CO_2 vs. CO and simulated CO_2 vs. CO is evident.

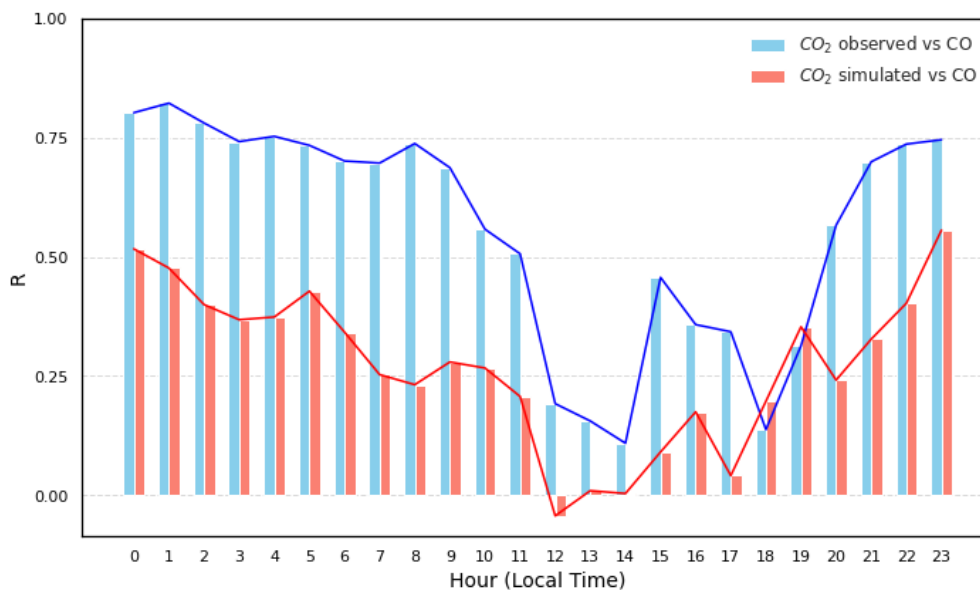


Figure 8. Hourly correlation between CO_2 concentrations observed at the IAG site and CO concentrations observed at the Pinheiros site (blue bars), and between simulated CO_2 concentrations at the IAG site and observed CO concentrations at the Pinheiros site (orange bars) for the period from February to August 2019

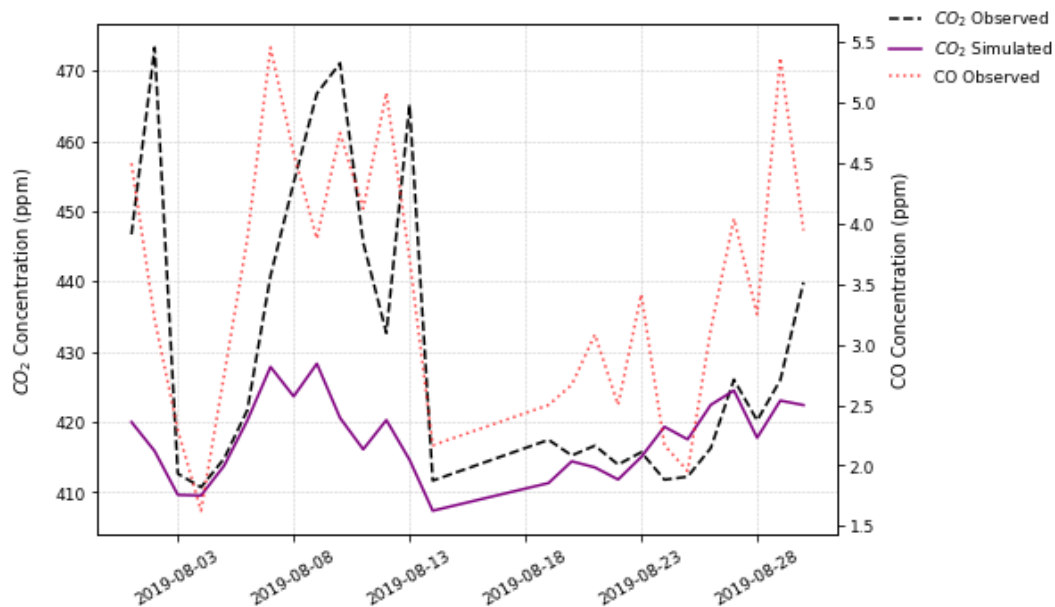


Figure 9. Daily mean concentrations of CO_2 observed concentrations (black dashed line), CO_2 simulated concentrations (purple line), and CO observed concentrations (red dotted line) at the IAG site during August 2019.

In addition to the correlation between gases, Figure 9 indicates that both the modeled and observed CO_2 profiles suggest that a significant portion of the CO_2 concentrations at the IAG site originates from vehicular sources, as carbon monoxide is a trace gas associated with traffic emissions (Nogueira et al., 2021). Peaks in the CO_2 time series at IAG are observed at the beginning, where the model fails to capture the magnitude of these concentrations. These peaks also appear in the observed CO profile at the begin of the month, confirming that a large part of the CO_2 concentrations at IAG comes from vehicular sources, particularly on days with high concentrations, which are also reflected in the CO profile. However, the model struggles to simulate this high CO_2 concentrations since it assumes that emissions follow the same diurnal variation every day of the month. Additionally, a distinct increase in CO concentrations without a corresponding rise in CO_2 was observed between August 18 and 21 and August 27 and 28, which coincided with the long-range transport of smoke plumes from Amazon forest fires to São Paulo (Bencherif et al., 2020). While biomass burning emits both CO and CO_2 , their atmospheric transport and dispersion differ significantly. CO is more prevalent in incomplete combustion and tends to be transported at altitudes that favor long-range dispersion, whereas CO_2 concentrations are more influenced by local emissions and atmospheric mixing (Gatti et al., 2010). These transport dynamics, combined with the long distance of the event's origin, likely explain why the CO peak was detected at Pinheiros but not accompanied by a significant CO_2 enhancement at the IAG site.

3.3.3 Model evaluation against OCO-2 and XCO_2 observations

Figure 10a shows the monthly boxplots of observed and all_sources simulated XCO_2 concentrations for the period from 1 April 2019 to 31 August 2019. However, due to insufficient OCO-2 data over MASP during this period, the analysis covers all simulated domains rather than solely the metropolitan area. Regarding temporal variability, a clear seasonal cycle of XCO_2 is evident from its smooth month-to-month variation (green boxes in Figure 10a). The simulated XCO_2 concentrations, i.e., the simulated profiles with smoothing, generally captured this cycle, although with a less dispersion (length of the box) compared to the observed XCO_2 concentrations. Notably, model-observation discrepancies are most pronounced during the winter months, with differences in median concentrations ranging from 0.8 to 1.5 ppm, while they are minimized during the autumn season, with differences in median concentrations between 0.5 and 0.6 ppm. The simulated XCO_2 concentrations demonstrate similar trends within the same range but tend to slightly underestimate values on most days.

When generating time-averaged modeled values, we only take into account the measurement period as previously mentioned. Regarding XCO_2 , the smoothed column concentrations (depicted by red dotted lines in Figure B8 in Appendix B) consistently fall below the observed values on a global scale. Figure 10b depicts the bias and RMSE, respectively, calculated across the pixel-by-pixel domain. Higher RMSE values are evident in the eastern region of MASP and along the border of São Paulo and Rio de Janeiro states. In these areas, characterized by heavy vehicular traffic, the model tends to overestimate XCO_2 concentrations. Conversely, for the central region of the domain, we observe slightly negative bias values accompanied by higher RMSE values, indicating an underestimation of XCO_2 concentrations. The uncertainties surrounding XCO_2 simulation stem from various factors, including potential biases in the model's wind representation, particularly in urban areas, consideration of emissions solely at the surface rather than at different pressure levels, as well as errors in the initial and boundary conditions of

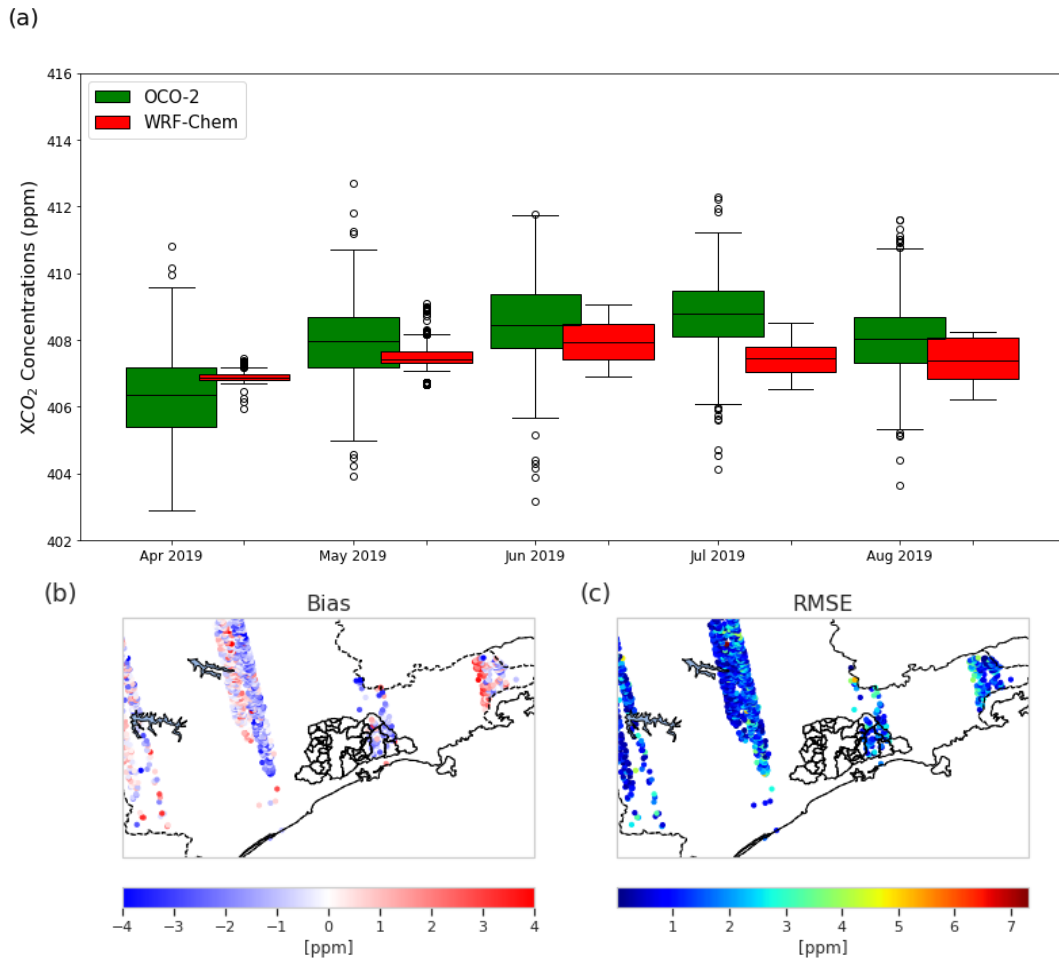


Figure 10. a) Monthly boxplots of observed and simulated XCO_2 concentrations for the period from 1 April 2019 to 31 August 2019, b) Bias and c) RMSE calculated by pixel over the study domain.

concentration provided by the Carbon Tracker, which has also been seen in other studies (Chen et al., 2019; Lian et al., 2021; Peiro et al., 2022).

4 Conclusions

A comprehensive assessment of atmospheric CO_2 concentrations in the metropolitan region of São Paulo (MASP) and its surroundings was conducted, utilizing the WRF model coupled to a greenhouse gas module. Given the burgeoning demand for research in this domain, particularly in South America, where urban areas are marked by significant emission sources, this study aimed to furnish a broad understanding of the key characteristics of CO_2 concentrations. To ensure an accurate estimation of CO_2 levels in MASP, the initial focus of the evaluation was on the model's capability to simulate meteorological

variables. Biogenic fluxes were derived from the VPRM model, which was fine-tuned with flux tower data. Our results show that using this local data significantly improved simulated biogenic CO_2 fluxes, highlighted the model's capacity to represent key seasonal dynamics, with negative net ecosystem exchange (NEE) values predominating in February (summer) and positive values in August (winter). However, we recommend the deployment of additional flux towers and targeted measurement campaigns to improve the characterization other ecosystems. A more comprehensive representation of PFTs is essential, as vegetation processes play a fundamental role in shaping CO_2 patterns in tropical regions. The availability of additional flux tower data would enable a more refined optimization approach, enhancing the characterization of parameters for each vegetation type. Anthropogenic emissions were curated from vehicular model and global inventory to provide a comprehensive representation of urban emissions, incorporating spatial and temporal resolution for key sources such as vehicular traffic for our domain. Boundary and initial conditions were scrutinized using global products. The WRF-Chem model demonstrated skill in simulating meteorological variables, particularly temperature; however, discrepancies in local wind speed and direction persisted. These differences are attributed to the region's complex topography and the model's resolution (3 km), which limits its ability to capture fine-scale dynamical processes.

Simulated CO_2 concentrations exhibited distinct diurnal cycles influenced by local emissions, boundary layer dynamics, and vegetation fluxes. The model's performance varied between monitoring stations, highlighting the interplay between urban and vegetative environments. At the IAG site, CO_2 concentrations were consistently underestimated, with negative biases of -9.17 ppm in February and -12.83 ppm in August. This underestimation was closely linked to the model's difficulty in capturing the impact of high vehicular emission densities, as indicated by the correlation with CO concentrations. Conversely, at the vegetated and elevated PDJ site, the model closely matched observational data, with minimal biases of 0.73 ppm in February and -0.61 ppm in August. In suburban locations such as the PDJ site, distant from urban sources, anthropogenic emissions diminish, and the vertical gradient of CO_2 concentration generated by city emissions attenuates through atmospheric convection and diffusion processes. However, during the growing season, the contribution of biogenic flux to CO_2 concentration warrants attention, especially concerning the simulation of nocturnal CO_2 concentrations and ecosystem respiration. Improvements in the respiration equation of the VPRM model (Gourdji et al., 2022) could enhance the accuracy of these simulations. Importantly, the modeled CO_2 concentrations exhibited high sensitivity not only to atmospheric vertical mixing near the surface but also to the prescribed temporal profiles of anthropogenic and biogenic emissions, highlighting the underestimation of vehicular emissions. These sources of error, particularly pronounced in winter, present challenges in accurately quantifying city emissions.

In general, the WRF-Chem model demonstrated proficiency in simulating seasonal variations, including XC_{O_2} , with profiles akin to OCO-2 data. This study underscores the imperative for further investigations and applications of the WRF-Chem model in uncharted regions such as the MASP, showcasing its prowess in simulating meteorological fields and CO_2 observations.

Code availability. The WRF-Chem model code version 4.0 is freely distributed by NCAR at <https://www2.mmm.ucar.edu/wrf/users/download/> (Skamarock et al., 2019). The VPRM code adapted from <https://github.com/Georgy-Neroblov/VPRM-code> (Neroblov et al., 2021). VEIN

can be installed from CRAN, and it is also available on Zenodo <https://doi.org/10.5281/zenodo.3714187> (Ibarra-Espinosa et al., 2018). Run control files, preprocessing and postprocessing scripts, and relevant primary input/output data sets needed to replicate the modelling results are available upon request from the corresponding author.

Data availability. All datasets and model results corresponding to this study are available upon request from the corresponding author.

Author contributions. RA performed the simulations and prepared the manuscript with the support of all co-authors. RA and RY design the experiment. TL and RB provided support to set up and run VPRM parameters optimization. OC, MM, and HR provided the observed data used in this work. RA, RY, TL, RB, AV, MA, NR, and CK contributed to the analysis and interpretation of the results.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. This work was supported by the National Council for Scientific and Technological Development (CNPq) fellowship process number 141962/2019-4, the FAPESP (process number 2016/18438-0 and 2021/11762-5), the French Ministry of Research (Junior Chair professor CASAL) and the Innovation Fund Denmark through the INNO-CCUS project MONICA, the National Institute of Science and Technology – INCT Klimapolis, which is funded by the Brazilian Ministry of Science, Technology, and Innovation (MCTI), and the National Council for Scientific and Technological Development (CNPq) under project number 406728/2022-4.

Appendix A: Metrics evaluation

$$Bias = \frac{\sum_{i=1}^N (pred_i - obs_i)}{N} \quad (A1)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (pred_i - obs_i)^2}{N}} \quad (A2)$$

$$R^2 = \frac{\sum_{i=1}^N (pred_i - \overline{pred_i})(obs_i - \overline{obs_i})}{\sqrt{\sum_{i=1}^N (pred_i - \overline{pred_i})^2 \sum_{i=1}^N (obs_i - \overline{obs_i})^2}} \quad (A3)$$

where $pred_i$ is the model simulation value, obs_i is the observed value, and N is the number of observations.

467 **Appendix B: Supplementary figures**

468 This appendix contains figures that give some additional insight to the conclusions given in the sections above and are refer-
469 enced in the text.

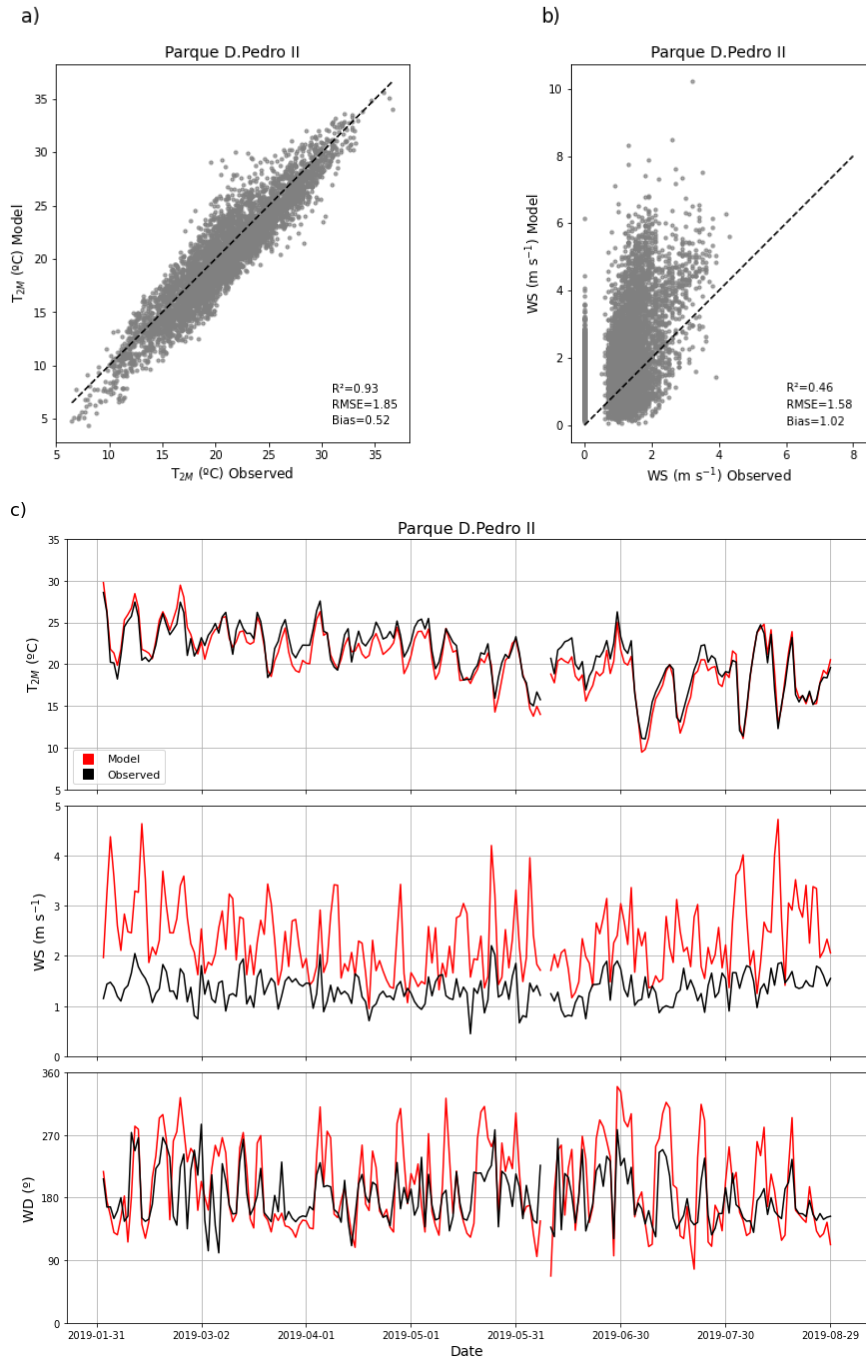


Figure B1. The panels in a) show the scatter plots of hourly measurements of 2 m air temperature (T_{2m}) and b) show 10 m wind speed (WS) compared to observed data from the Parque D. Pedro II station. The figure illustrates the relationship between modeled and observed data. The panels in c) show the daily averages from February to August 2019 of 2 m air temperature (T_{2m}), 10 m wind speed (WS), and wind direction (WD). Black line represents the observed data and red line represents the model simulation.

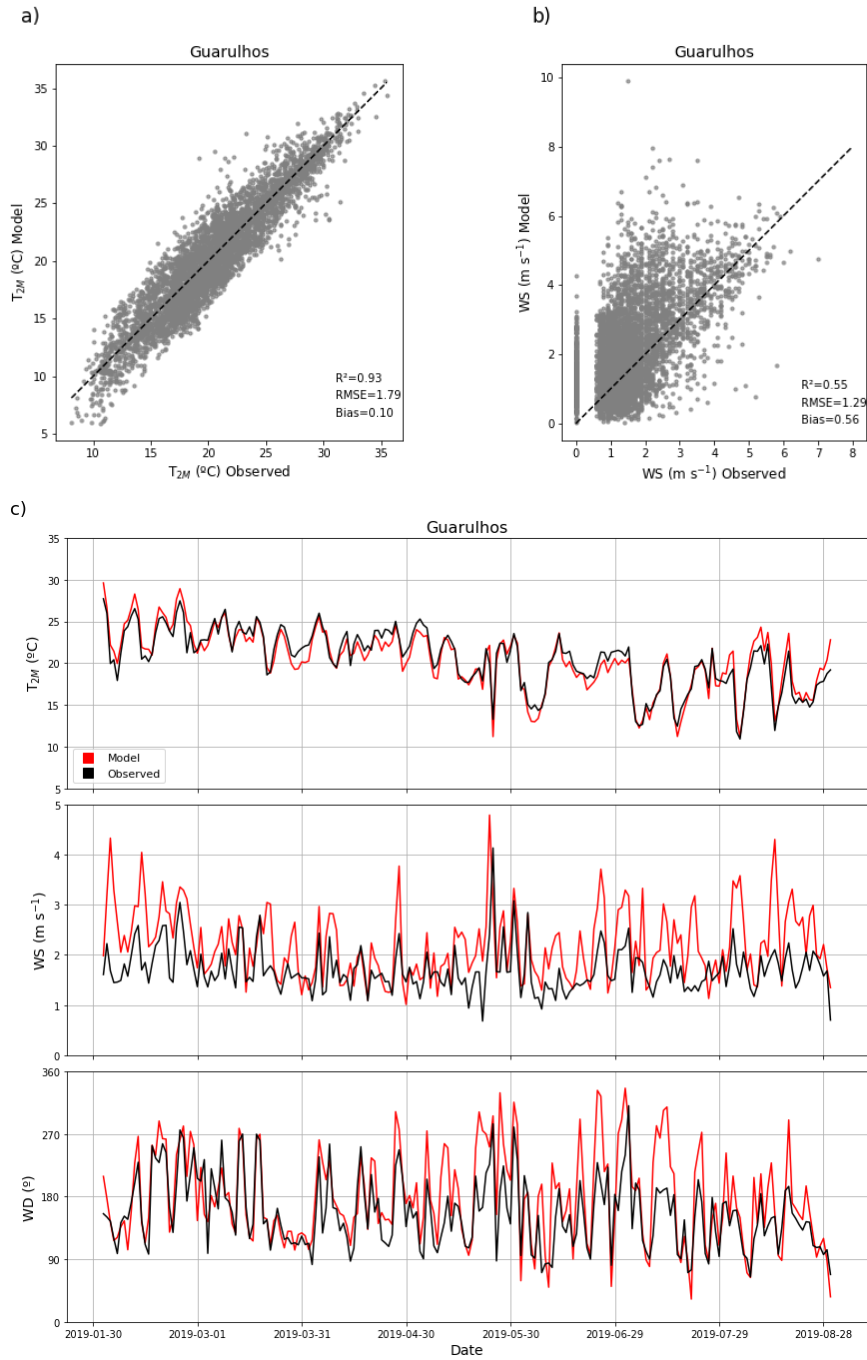


Figure B2. The panels in a) show the scatter plots of hourly measurements of 2 m air temperature (T_{2m}) and b) show 10 m wind speed (WS) compared to observed data from the Guarulhos station. The figure illustrates the relationship between modeled and observed data. The panels in c) show the daily averages from February to August 2019 of 2 m air temperature (T_{2m}), 10 m wind speed (WS), and wind direction (WD). Black line represents the observed data and red line represents the model simulation.

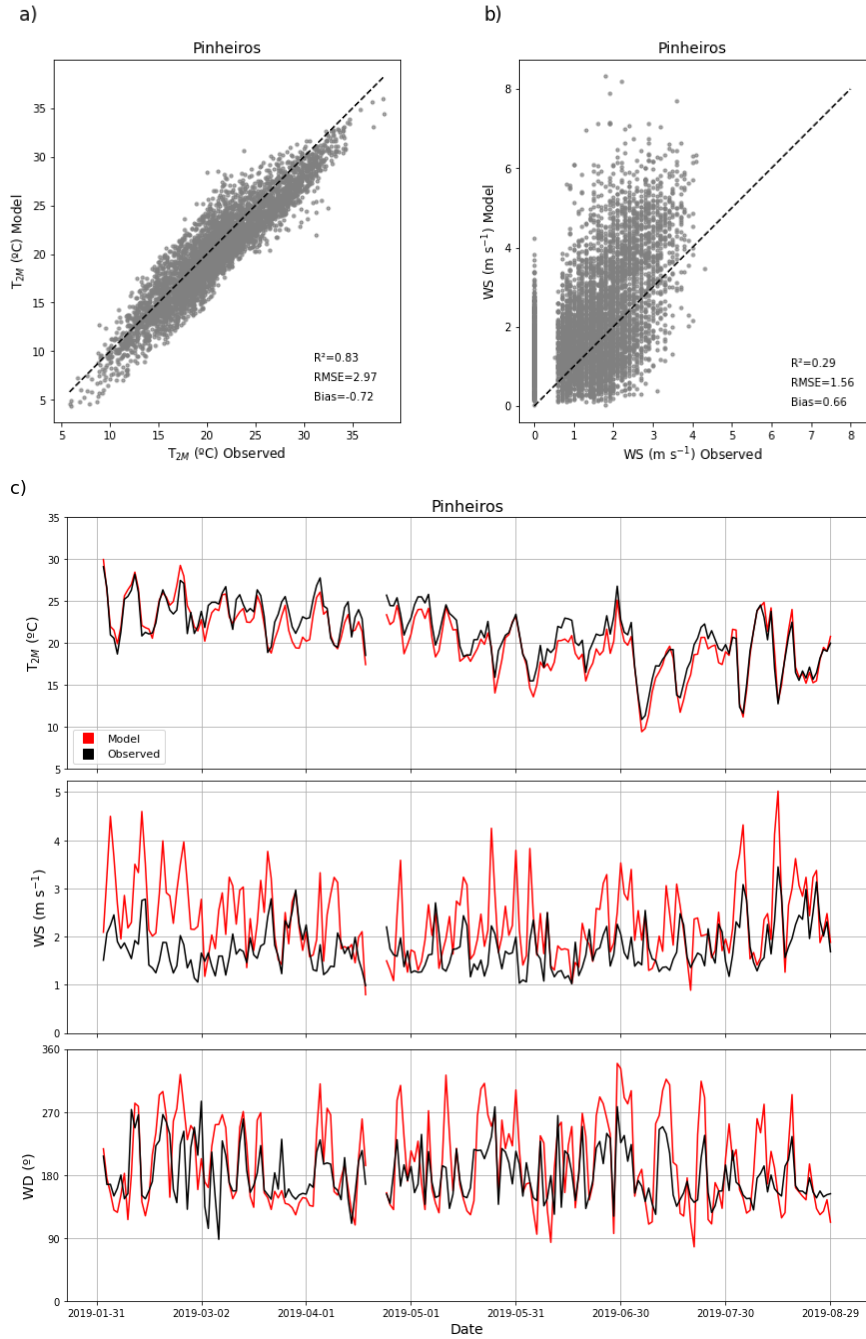


Figure B3. The panels in a) show the scatter plots of hourly measurements of 2 m air temperature (T_{2m}) and b) show 10 m wind speed (WS) compared to observed data from the Pinheiros station. The figure illustrates the relationship between modeled and observed data. The panels in c) show the daily averages from February to August 2019 of 2 m air temperature (T_{2m}), 10 m wind speed (WS), and wind direction (WD). Black line represents the observed data and red line represents the model simulation.

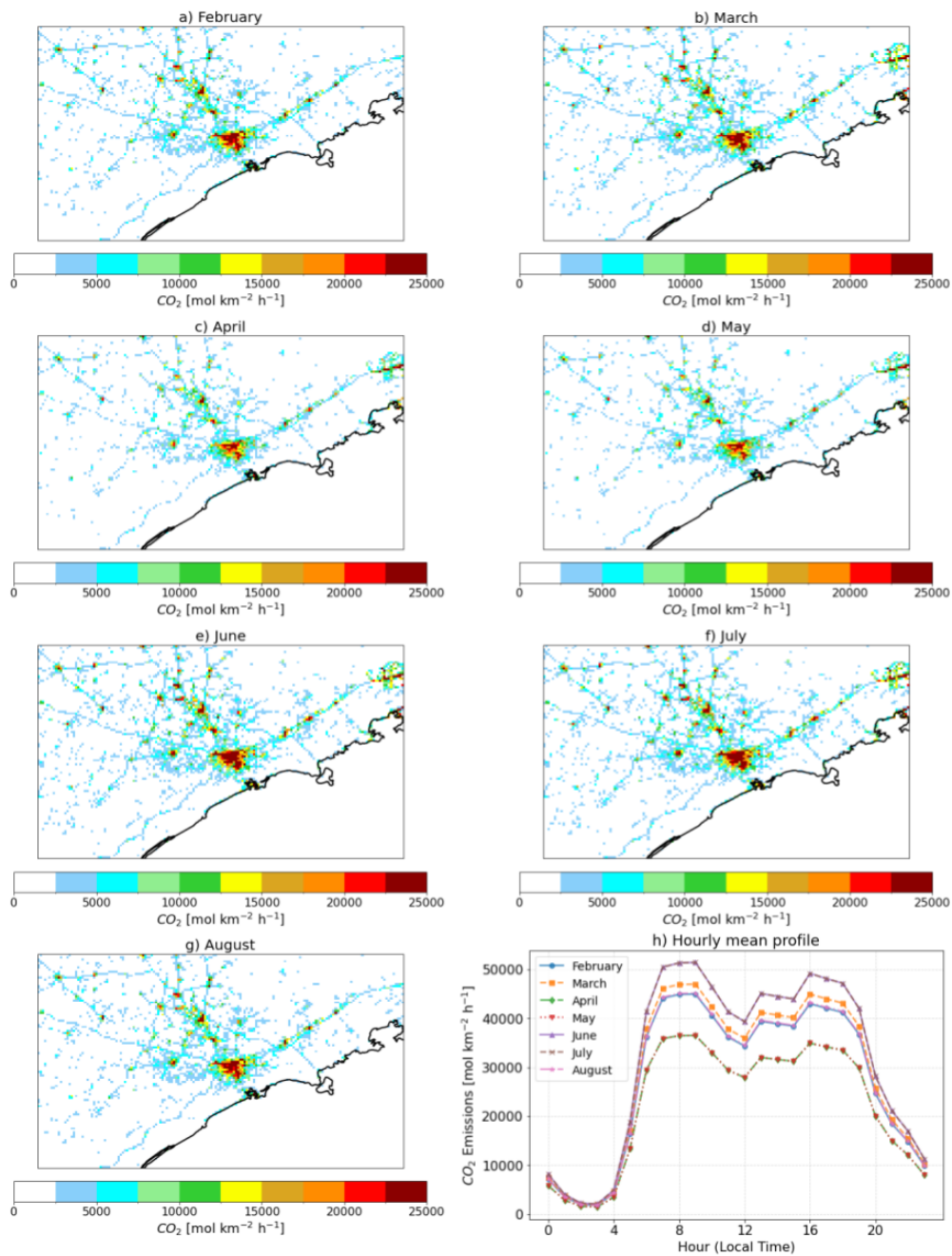


Figure B4. Daily mean of CO_2 emissions by the VEIN model for each month (a) to (g) and (h) Hourly mean profile at IAG site.

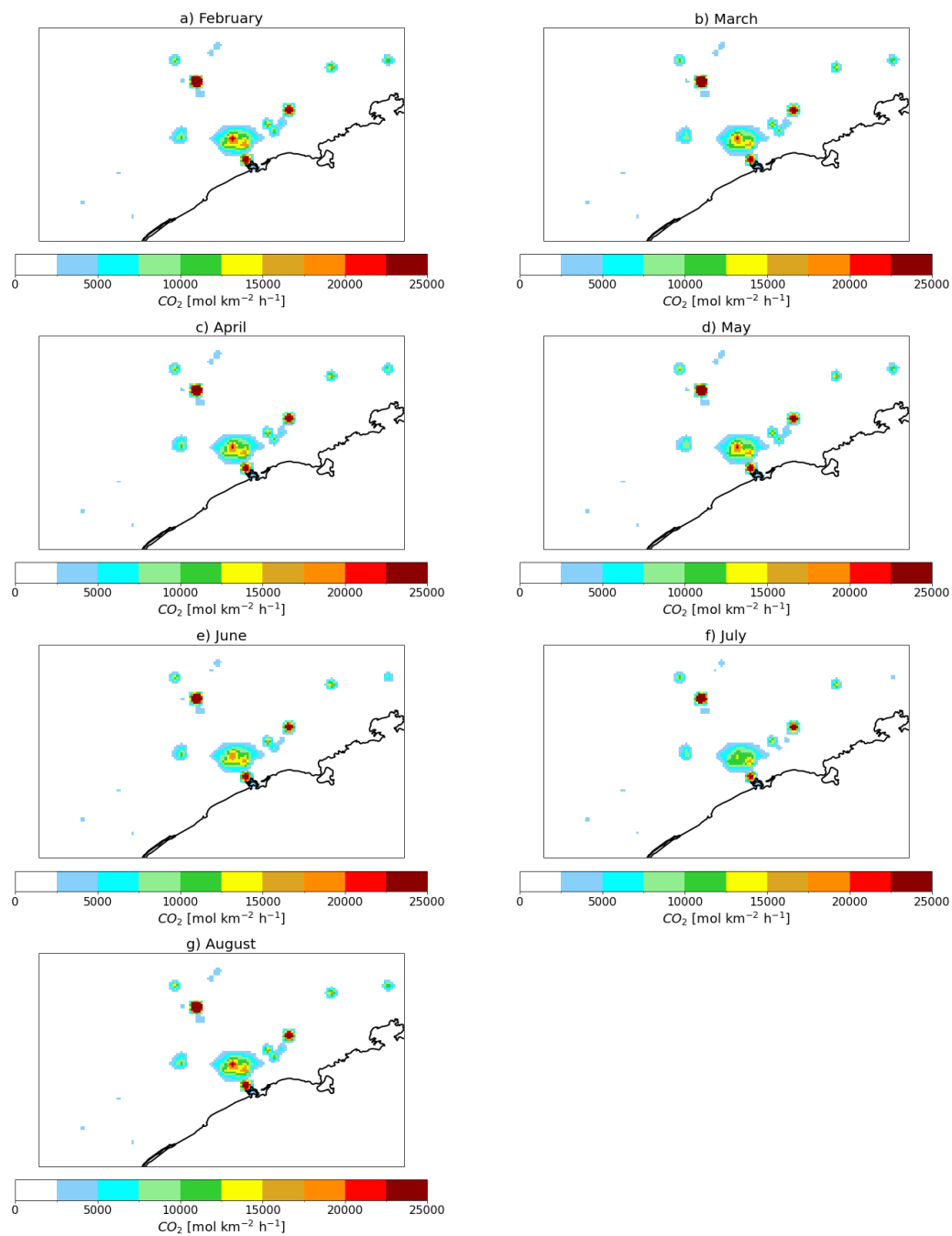


Figure B5. Daily mean of CO_2 emissions by EDGAR for each month.

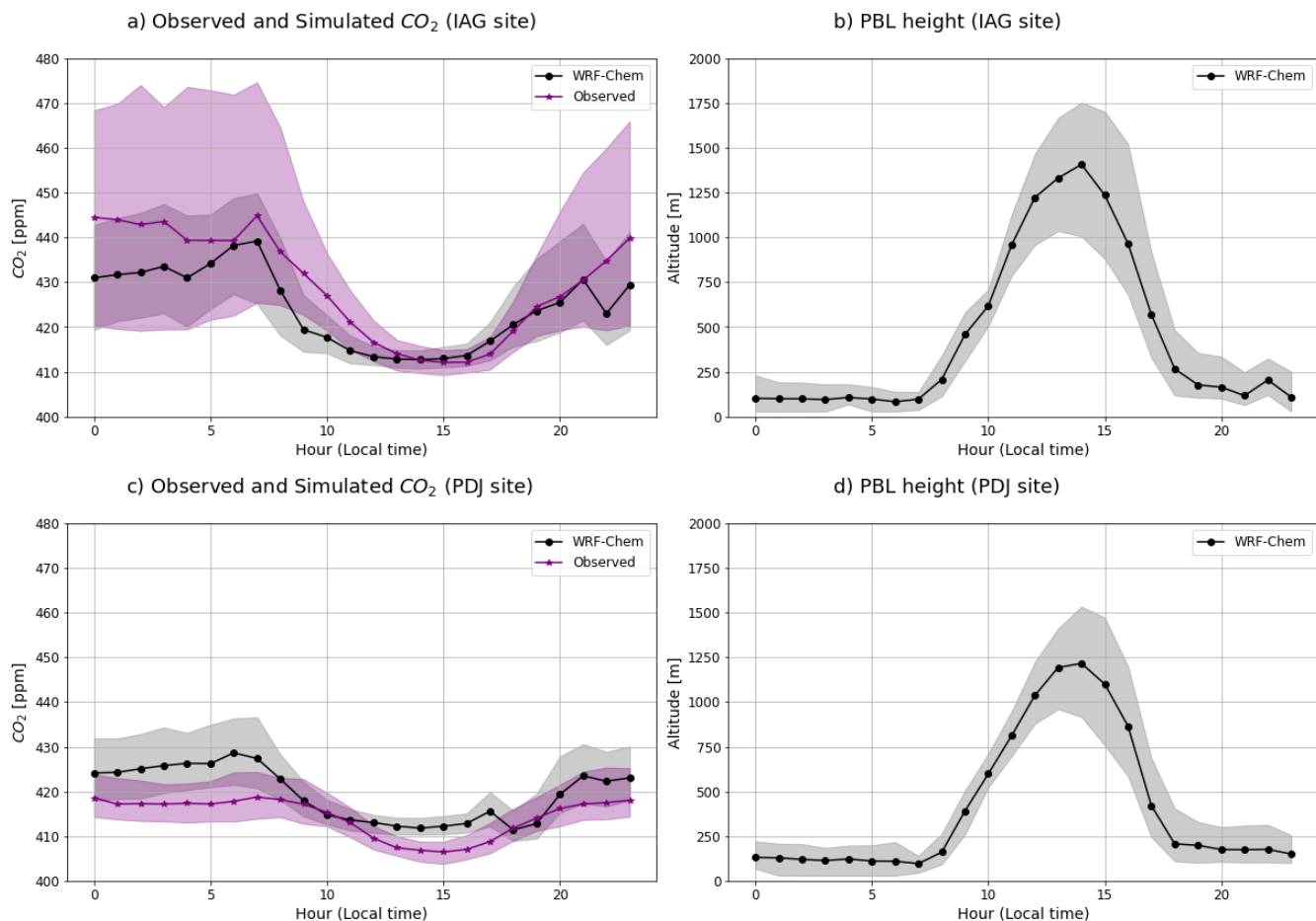


Figure B6. Diurnal cycle of in situ CO_2 concentration and planetary boundary layer (PBL) height for the entire simulated period. The black line represents the median hourly concentrations from WRF-Chem, while the purple line corresponds to the observed values. The shaded areas indicate the interquartile ranges. Panel a) shows the observed and simulated surface CO_2 concentration at the IAG site; b) the simulated PBL height at the IAG site; c) the observed and simulated surface CO_2 concentration at the PDJ site; and d) the simulated PBL height at the PDJ site.

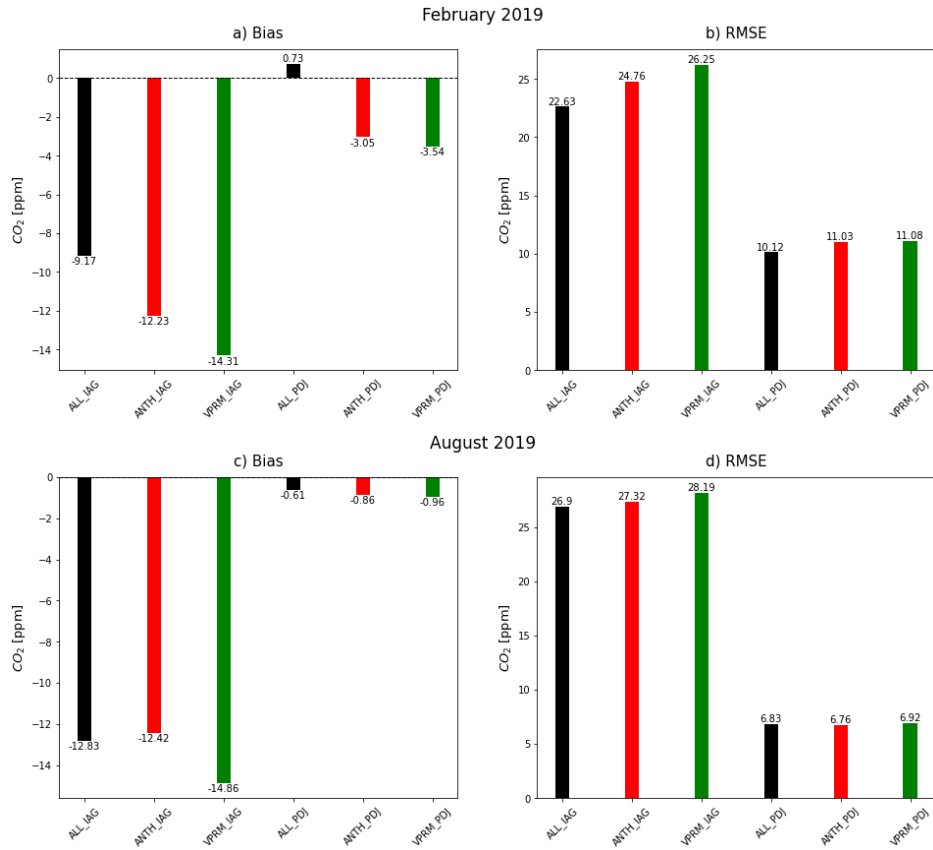


Figure B7. Bias (ppm) and RMSE (ppm) for each simulation at the surface CO_2 observation sites. Panels (a) and (b) represent the simulations for February, while panels (c) and (d) represent the simulations for August (ALL_*: black, ANTH_*: red, VPRM_*: green) *Represents the observation sites, e.g. IAG and PDJ.

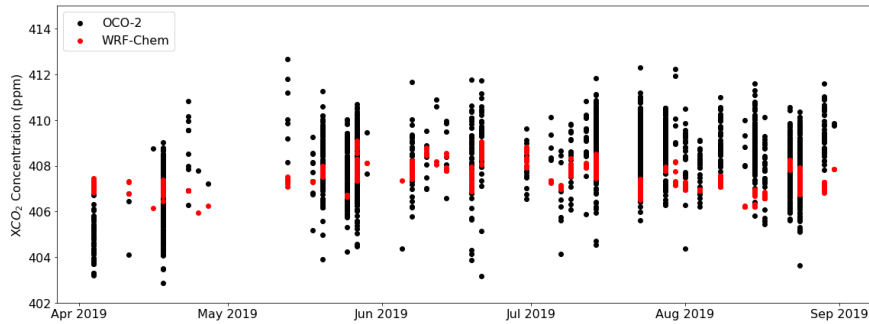


Figure B8. Time series of smoothed column concentrations observed (black) and modeled (red) for the period from 1 April 2019 to 31 August 2019.

470 References

- 471 Andrade, M. d. F., Ynoue, R. Y., Freitas, E. D., Todesco, E., Vara Vela, A., Ibarra, S., Martins, L. D., Martins, J. A., and Carvalho, V. S. B.:
472 Air quality forecasting system for Southeastern Brazil, *Frontiers in environmental Science*, 3, 9, 2015.
- 473 Benavente, N. R., Vara-Vela, A. L., Nascimento, J. P., Acuna, J. R., Damascena, A. S., de Fatima Andrade, M., and Yamasoe, M. A.: Air
474 quality simulation with WRF-Chem over southeastern Brazil, part I: Model description and evaluation using ground-based and satellite
475 data, *Urban Climate*, 52, 101 703, 2023.
- 476 Bencherif, H., Bègue, N., Kirsch Pinheiro, D., Du Preez, D. J., Cadet, J.-M., da Silva Lopes, F. J., Shikwambana, L., Landulfo, E., Vescovini,
477 T., Labuschagne, C., et al.: Investigating the long-range transport of aerosol plumes following the Amazon fires (August 2019): a multi-
478 instrumental approach from ground-based and satellite observations, *Remote Sensing*, 12, 3846, 2020.
- 479 Botía, S., Komiya, S., Marshall, J., Koch, T., Gałkowski, M., Lavric, J., Gomes-Alves, E., Walter, D., Fisch, G., Pinho, D. M., et al.: The CO₂
480 record at the Amazon Tall Tower Observatory: A new opportunity to study processes on seasonal and inter-annual scales, *Global Change*
481 *Biology*, 28, 588–611, 2022.
- 482 Cabral, O. M., Freitas, H. C., Cuadra, S. V., de Andrade, C. A., Ramos, N. P., Grutzmacher, P., Galdos, M., Packer, A. P. C., da Rocha,
483 H. R., and Rossi, P.: The sustainability of a sugarcane plantation in Brazil assessed by the eddy covariance fluxes of greenhouse gases,
484 *Agricultural and Forest Meteorology*, 282, 107 864, 2020.
- 485 Caetano, P. M. D., Pereira, H. M. S. B., Figueiredo, L. C. R., Sepe, P. M., and Giatti, L. L.: The City of São Paulo's Environmental Quota:
486 A Policy to Embrace Urban Environmental Services and Green Infrastructure Inequalities in the Global South, *Frontiers in Sustainable*
487 *Cities*, 3, 685 875, 2021.
- 488 CETESB: Relatório de Qualidade do Ar no Estado de São Paulo, Tech. rep., CETESB, São Paulo, Brasil, 2019.
- 489 Change, I. C. et al.: Impacts, adaptation and vulnerability, Part A: global and sectoral aspects. Contribution of working group II to the fifth
490 assessment report of the intergovernmental Panel on Climate Change, 1132, 2014.
- 491 Che, K., Cai, Z., Liu, Y., Wu, L., Yang, D., Chen, Y., Meng, X., Zhou, M., Wang, J., Yao, L., et al.: Lagrangian inversion of anthropogenic
492 CO₂ emissions from Beijing using differential column measurements, *Environmental Research Letters*, 17, 075 001, 2022.
- 493 Chen, H. W., Zhang, F., Lauvaux, T., Davis, K. J., Feng, S., Butler, M. P., and Alley, R. B.: Characterization of regional-scale CO₂ transport
494 uncertainties in an ensemble with flow-dependent transport errors, *Geophysical Research Letters*, 46, 4049–4058, 2019.
- 495 Chiquetto, J. B., Machado, P. G., Mouette, D., and Ribeiro, F. N. D.: Air quality improvements from a transport modal change in the São
496 Paulo megacity, *Science of The Total Environment*, 945, 173 968, 2024.
- 497 Connor, B. J., Boesch, H., Toon, G., Sen, B., Miller, C., and Crisp, D.: Orbiting Carbon Observatory: Inverse method and prospective error
498 analysis, *Journal of Geophysical Research: Atmospheres*, 113, 2008.
- 499 Crippa, M., Guizzardi, D., Muntean, M., Schaaf, E., Lo Vullo, E., Solazzo, E., Monforti-Ferrario, F., Olivier, J., and Vignati, E.: EDGAR v6.
500 0 greenhouse gas emissions. European Commission, Joint Research Centre (JRC), 2021.
- 501 Crisp, D.: Measuring atmospheric carbon dioxide from space with the Orbiting Carbon Observatory-2 (OCO-2), in: *Earth observing systems*
502 *xx*, vol. 9607, p. 960702, SPIE, 2015.
- 503 De Pue, J., Wieneke, S., Bastos, A., Barrios, J. M., Liu, L., Ciais, P., Arboleda, A., Hamdi, R., Maleki, M., Maignan, F., et al.: Temporal
504 variability of observed and simulated gross primary productivity, modulated by vegetation state and hydrometeorological drivers, *Biogeo-*
505 *sciences*, 20, 4795–4818, 2023.

506 Deng, A., Lauvaux, T., Davis, K. J., Gaudet, B. J., Miles, N., Richardson, S. J., Wu, K., Sarmiento, D. P., Hardesty, R. M., Bonin, T. A., et al.:
507 Toward reduced transport errors in a high resolution urban CO₂ inversion system, *Elem Sci Anth*, 5, 20, 2017.

508 Feng, S., Lauvaux, T., Newman, S., Rao, P., Ahmadov, R., Deng, A., Díaz-Isaac, L. I., Duren, R. M., Fischer, M. L., Gerbig, C., et al.: Los
509 Angeles megacity: a high-resolution land-atmosphere modelling system for urban CO₂ emissions, *Atmospheric Chemistry and Physics*,
510 16, 9019–9045, 2016.

511 Freitas, H. C. d.: A influência dos transportes advectivos na estimativa do balanço de CO₂ do ecossistema: Um estudo de caso para a mata
512 atlântica com uso de técnicas micrometeorológicas, Ph.D. thesis, Universidade de São Paulo, 2012.

513 Gatti, L. V., Miller, J., D’amelio, M. a., Martinewski, A., Basso, L., Gloor, M., Wofsy, S., and Tans, P.: Vertical profiles of CO₂ above eastern
514 Amazonia suggest a net carbon flux to the atmosphere and balanced biosphere between 2000 and 2009, *Tellus B: Chemical and Physical*
515 *Meteorology*, 62, 581–594, 2010.

516 Gavidia-Calderón, M., Schuch, D., Vara-Vela, A., Inoue, R., Freitas, E. D., Albuquerque, T. T. d. A., Zhang, Y., de Fatima Andrade, M., and
517 Bell, M. L.: Air quality modeling in the metropolitan area of São Paulo, Brazil: A review, *Atmospheric Environment*, p. 120301, 2023.

518 Gerbig, C., Körner, S., and Lin, J.: Vertical mixing in atmospheric tracer transport models: error characterization and propagation, *Atmo-*
519 *spheric Chemistry and Physics*, 8, 591–602, 2008.

520 Gourdji, S. M., Karion, A., Lopez-Coto, I., Ghosh, S., Mueller, K. L., Zhou, Y., Williams, C. A., Baker, I. T., Haynes, K. D., and Whetstone,
521 J. R.: A modified Vegetation Photosynthesis and Respiration Model (VPRM) for the eastern USA and Canada, evaluated with comparison
522 to atmospheric observations and other biospheric models, *Journal of Geophysical Research: Biogeosciences*, 127, e2021JG006 290, 2022.

523 He, J., Li, W., Zhao, Z., Zhu, L., Du, X., Xu, Y., Sun, M., Zhou, J., Ciais, P., Wigneron, J.-P., et al.: Recent advances and challenges in
524 monitoring and modeling of disturbances in tropical moist forests, *Frontiers in Remote Sensing*, 5, 1332 728, 2024.

525 Hersbach, H.: ERA5 reanalysis is in production, *ECMWF newsletter*, 147, 5, 2016.

526 Hong, S.-Y., Noh, Y., and Dudhia, J.: A new vertical diffusion package with an explicit treatment of entrainment processes, *Monthly weather*
527 *review*, 134, 2318–2341, 2006.

528 Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., and Collins, W. D.: Radiative forcing by long-lived greenhouse
529 gases: Calculations with the AER radiative transfer models, *Journal of Geophysical Research: Atmospheres*, 113, 2008.

530 Ibarra-Espinosa, S., Ynoue, R., O’Sullivan, S., Pebesma, E., Andrade, M. d. F., and Osses, M.: VEIN v0. 2.2: an R package for bottom-up
531 vehicular emissions inventories, *Geoscientific Model Development*, 11, 2209–2229, 2018.

532 IBGE: Instituto Brasileiro de Geografia e Estatística (IBGE), 2021.

533 Kaiser, W., Zhuravlev, R., Ganshin, A., Valsala, V. K., Andrews, A., Chmura, L., Dlugokencky, E., Haszpra, L., Langenfelds, R. L.,
534 Machida, T., et al.: A high-resolution inverse modelling technique for estimating surface CO₂ fluxes based on the NIES-TM-FLEXPART
535 coupled transport model and its adjoint.

536 Lian, J., Bréon, F.-M., Broquet, G., Lauvaux, T., Zheng, B., Ramonet, M., Xueref-Remy, I., Kotthaus, S., Haefelin, M., and Ciais, P.: Sensi-
537 tivity to the sources of uncertainties in the modeling of atmospheric CO₂ concentration within and in the vicinity of Paris, *Atmospheric*
538 *Chemistry and Physics*, 21, 10 707–10 726, 2021.

539 Mahadevan, P., Wofsy, S. C., Matross, D. M., Xiao, X., Dunn, A. L., Lin, J. C., Gerbig, C., Munger, J. W., Chow, V. Y., and Gottlieb, E. W.: A
540 satellite-based biosphere parameterization for net ecosystem CO₂ exchange: Vegetation Photosynthesis and Respiration Model (VPRM),
541 *Global Biogeochemical Cycles*, 22, 2008.

542 Morrison, H., Thompson, G., and Tatarskii, V.: Impact of cloud microphysics on the development of trailing stratiform precipitation in a
543 simulated squall line: Comparison of one-and two-moment schemes, *Monthly weather review*, 137, 991–1007, 2009.

544 Nerobellov, G., Timofeyev, Y., Smyshlyaev, S., Foka, S., Mammarella, I., and Virolainen, Y.: Validation of WRF-Chem model and CAMS
545 performance in estimating near-surface atmospheric CO₂ mixing ratio in the area of Saint Petersburg (Russia), *Atmosphere*, 12, 387,
546 2021.

547 Nogueira, T., Kamigauti, L. Y., Pereira, G. M., Gavidia-Calderon, M. E., Ibarra-Espinosa, S., Oliveira, G. L. d., Miranda, R. M. d., Vascon-
548 cellos, P. d. C., Freitas, E. D. d., and Andrade, M. d. F.: Evolution of vehicle emission factors in a megacity affected by extensive biofuel
549 use: results of tunnel measurements in São Paulo, Brazil, *Environmental Science & Technology*, 55, 6677–6687, 2021.

550 O'Dell, C., Connor, B., Bösch, H., O'Brien, D., Frankenberg, C., Castano, R., Christi, M., Eldering, D., Fisher, B., Gunson, M., et al.: The
551 ACOS CO₂ retrieval algorithm–Part 1: Description and validation against synthetic observations, *Atmospheric Measurement Techniques*,
552 5, 99–121, 2012.

553 Osterman, G., Eldering, A., Avis, C., Chafin, B., O'Dell, C., Frankenberg, C., Fisher, B., Mandrake, L., Wunch, D., Granat, R., et al.: Orbiting
554 Carbon Observatory–2 (OCO-2) Data Product User's Guide, Operational L1 and L2 Data Versions 8 and Lite File Version 9, Version 1,
555 Revision J., October 10, 2018, 2018.

556 Peiro, H., Crowell, S., Schuh, A., Baker, D. F., O'Dell, C., Jacobson, A. R., Chevallier, F., Liu, J., Eldering, A., Crisp, D., et al.: Four years
557 of global carbon cycle observed from the Orbiting Carbon Observatory 2 (OCO-2) version 9 and in situ data and comparison to OCO-2
558 version 7, *Atmospheric Chemistry and Physics*, 22, 1097–1130, 2022.

559 Raju, A., Sijikumar, S., Burman, P. K. D., Valsala, V., Tiwari, Y. K., Mukherjee, S., Lohani, P., and Kumar, K.: Very high-resolution Net
560 Ecosystem Exchange over India using Vegetation Photosynthesis and Respiration Model (VPRM) simulations, *Ecological Modelling*,
561 481, 110 340, 2023.

562 Ramonet, M., Ciais, P., Apadula, F., Bartyzel, J., Bastos, A., Bergamaschi, P., Blanc, P., Brunner, D., Caracciolo di Torchiareolo, L., Calzo-
563 lari, F., et al.: The fingerprint of the summer 2018 drought in Europe on ground-based atmospheric CO₂ measurements, *Philosophical
564 Transactions of the Royal Society B*, 375, 20190 513, 2020.

565 Rocha, H. R. d., Freitas, H. C., Rosolem, R., Juárez, R. I., Tannus, R. N., Ligo, M. A., Cabral, O. M., and Dias, M. A.: Measurements of CO₂
566 exchange over a woodland savanna (Cerrado *Sensu stricto*) in southeast Brasil, *Biota Neotropica*, 2, 1–11, 2002.

567 SEEG: SEEG – Greenhouse Gas Emissions and Removals Estimation System, Climate Observatory, Tech. rep., seeg.eco.br, 2019.

568 Seo, M.-G., Kim, H. M., and Kim, D.-H.: Effect of atmospheric conditions and VPRM parameters on high-resolution regional CO₂ simula-
569 tions over East Asia, *Theoretical and Applied Climatology*, 155, 859–877, 2024.

570 Seto, K. C., Güneralp, B., and Hutyrá, L. R.: Global forecasts of urban expansion to 2030 and direct impacts on biodiversity and carbon
571 pools, *Proceedings of the National Academy of Sciences*, 109, 16 083–16 088, 2012.

572 Seto, K. C., Dhakal, S., Bigio, A., Blanco, H., Carlo Delgado, G., Dewar, D., Huang, L., Inaba, A., Kansal, A., Lwasa, S., et al.: Human
573 settlements, infrastructure, and spatial planning, 2014.

574 Shimada, S., Ohsawa, T., Chikaoka, S., and Kozai, K.: Accuracy of the wind speed profile in the lower PBL as simulated by the WRF model,
575 *Sola*, 7, 109–112, 2011.

576 Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Liu, Z., Berner, J., Wang, W., Powers, J. G., Duda, M. G., Barker, D. M., et al.: A
577 description of the advanced research WRF version 4, NCAR tech. note ncar/tn-556+ str, 145, 2019.

578 Souto-Oliveira, C. E., Marques, M. T., Nogueira, T., Lopes, F. J., Medeiros, J. A., Medeiros, I. M., Moreira, G. A., da Silva Dias, P. L.,
579 Landulfo, E., and Andrade, M. d. F.: Impact of extreme wildfires from the Brazilian Forests and sugarcane burning on the air quality of
580 the biggest megacity on South America, *Science of the Total Environment*, 888, 163 439, 2023.

581 Souza Jr, C. M., Z. Shimbo, J., Rosa, M. R., Parente, L. L., A. Alencar, A., Rudorff, B. F., Hasenack, H., Matsumoto, M., G. Ferreira, L.,
582 Souza-Filho, P. W., et al.: Reconstructing three decades of land use and land cover changes in brazilian biomes with landsat archive and
583 earth engine, *Remote Sensing*, 12, 2735, 2020.

584 Tewari, M., Chen, F., Kusaka, H., and Miao, S.: Coupled WRF/Unified Noah/urban-canopy modeling system, Near WRF Documentation,
585 NCAR, Boulder, 122, 1–22, 2007.

586 Vara-Vela, A., Andrade, M. F., Kumar, P., Ynoue, R. Y., and Munoz, A. G.: Impact of vehicular emissions on the formation of fine particles
587 in the Sao Paulo Metropolitan Area: a numerical study with the WRF-Chem model, *Atmospheric Chemistry and Physics*, 16, 777–797,
588 2016.

589 Vara-Vela, A., de Fátima Andrade, M., Zhang, Y., Kumar, P., Ynoue, R. Y., Souto-Oliveira, C. E., da Silva Lopes, F. J., and Landulfo,
590 E.: Modeling of atmospheric aerosol properties in the São Paulo metropolitan area: impact of biomass burning, *Journal of Geophysical*
591 *Research: Atmospheres*, 123, 9935–9956, 2018.

592 Vara-Vela, A. L., Herdies, D. L., Alvim, D. S., Vendrasco, É. P., Figueroa, S. N., Pendharkar, J., and Reyes Fernandez, J. P.: A new predictive
593 framework for Amazon forest fire smoke dispersion over South America, *Bulletin of the American Meteorological Society*, 102, E1700–
594 E1713, 2021.

595 Vermote, E.: MODIS/Terra Surface Reflectance 8-Day L3 Global 500m SIN Grid V061, NASA EOSDIS Land Processes DAAC: Missoula,
596 MT, USA, 2021.

597 Wilmot, T. Y., Lin, J. C., Wu, D., Oda, T., and Kort, E. A.: Toward a satellite-based monitoring system for urban CO2 emissions in support
598 of global collective climate mitigation actions, *Environmental Research Letters*, 19, 084 029, 2024.

599 Zhang, L., Zhang, H., Li, Q., Wu, B., Cai, X., Song, Y., and Zhang, X.: Complexity of carbon dioxide flux in urban areas: A comparison with
600 natural surfaces, *Science of the Total Environment*, 895, 165 115, 2023.

601 Zhang, Y., Dubey, M. K., Olsen, S., Zheng, J., and Zhang, R.: Comparisons of WRF/Chem simulations in Mexico City with ground-based
602 RAMA measurements during the 2006-MILAGRO, *Atmospheric Chemistry and Physics*, 9, 3777–3798, 2009.