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

Rafaela Cruz Alves Alberti[1], Thomas Lauvaux[3], Angel Liduvino Vara-Vela[6,7,8], Ricard Segura Barrero[2], Christoffer Karoff[6,7,8], Maria de Fátima Andrade[1], Márcia Talita Amorim Marques[1], Noelia Rojas Benavente[5], Osvaldo Machado Rodrigues Cabral[4], Humberto Ribeiro da Rocha[1], and Rita Yuri Ynoue[1]

[1]Department of Atmospheric Sciences, University of São Paulo, Brazil
[2]Institute of Environmental Sciences and Technology, Universitat Autònoma de Barcelona, Spain.
[3]Universite´ de Reims Champagne-Ardenne, CNRS, GSMA, Reims, France
[4]Brazilian Agricultural Research Corporation, Embrapa Environment, Brazil
[5]Physics Institute, University of São Paulo, Brazil
[6]Department of Geoscience, Aarhus University, Denmark
[7]Department of Physics and Astronomy, Aarhus University, Aarhus, Denmark
[8]iCLIMATE Aarhus University Interdisciplinary Centre for Climate Change, Aarhus, Denmark

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

1   **Abstract.** Atmospheric CO_2 concentrations over urban areas indirectly reflect local fossil fuel emissions and biogenic fluxes,
2   offering a potential approach to assess city climate policies. However, atmospheric models used to simulate urban CO_2 plumes
3   face significant uncertainties, particularly in complex urban environments with dense populations and vegetation. This study
4   aims to address these challenges and fill the research gap regarding such vegetated and urbanized areas by conducting a com-
5   prehensive analysis of atmospheric CO_2 dynamics in the Metropolitan Area of São Paulo, Brazil, and its surroundings, using
6   the WRF-GHG atmospheric model. The simulations are evaluated using observations from ground stations collected across
7   the METROCLIMA GHG network, the first greenhouse gas monitoring network in South America, and column concentrations
8   (XCO_2) from the OCO-2 satellite spanning February to August 2019. We also assess and improve the performances of the bio-
9   spheric model Vegetation Photosynthesis and Respiration Model (VPRM) by optimizing the model parameters of the dominant
10 vegetation types (Atlantic forest, cerrado, sugarcane) using flux measurements from multiple eddy-covariance flux towers. We
11 evaluate the atmospheric model's ability to replicate seasonal variations in CO_2 concentrations by comparing the simulations
12 with measurements from two sites part of the GHG network in Sao Paulo. We conclude here that atmospheric concentrations
13 over metropolitan areas located in tropical areas largely depend on our ability to represent the biogenic contribution from the
14 surrounding vegetation, the large-scale contribution in global models, and the model's ability to represent the local atmospheric
15 dynamics.

16 ## 1   Introduction

17 Urban areas, although occupying only a small fraction of the Earth's surface, exert an outsized influence on global carbon
18 emissions. Accounting for a staggering 70% of CO_2 emissions from fossil fuel burning while covering just 2% of the planet's

landmass (Seto et al., 2014; Change et al., 2014), cities have become focal points for climate action. The relentless pace of urbanization has further exacerbated this phenomenon, driving up energy consumption and emissions levels (Seto et al., 2012). Consequently, combating climate change necessitates a targeted approach, with policies increasingly tailored to address urban emissions. In response to the growing need for climate action, initiatives like the International Council for Local Environmental Initiatives (ICLEI), the C40 Cities Climate Leadership Group (C40), and the Covenant of Mayors (CoM) have emerged to coordinate global efforts and share best practices among cities. These initiatives highlight the crucial role cities play in the fight against climate change and the importance of localized mitigation strategies. São Paulo, Brazil's largest municipality (IBGE, 2021), is a member of C40 and focuses on reducing greenhouse gas emissions, with transportation accounting for 58% of its total emissions (SEEG, 2019). The city is working towards carbon neutrality through projects in green infrastructure, urban planning, public transportation improvements, energy efficiency, and waste management (Caetano et al., 2021). These efforts aim to reduce emissions and enhance São Paulo's resilience, fostering a more sustainable urban environment. Central to these efforts is the need for accurate data and robust modeling frameworks to inform policy decisions effectively. Urban atmospheric networks, such as Metroclima in the Sao Paulo Metropolitan Area (MASP), in Brazil, provide vital insights into greenhouse gas concentrations and emission patterns. By leveraging these datasets alongside sophisticated atmospheric transport models and statistical techniques, policymakers gain tools for designing targeted interventions and monitoring their efficacy. However, the complexity of urban CO_2 dynamics presents significant challenges for modeling and analysis. Process-driven biosphere models and inverse modeling techniques offer complementary approaches for capturing the intricate spatio-temporal variabilities inherent in urban environments (Kaiser et al.; Che et al., 2022; Zhang et al., 2023; Wilmot et al., 2024). Despite advancements in modeling capabilities, gaps remain in our understanding of CO_2 dynamics, particularly at regional and national scales. South America, in particular, suffers from limited data availability, and research focusing on this region is scarce. Additionally, vegetation models in tropical regions often exhibit poor performance due to inaccuracies in simulating seasonality, oversimplified representations of biodiversity, and errors in carbon and water cycle interactions. These models struggle to capture the complex dynamics of tropical ecosystems, leading to underestimations of productivity and poor predictions of vegetation responses to climate variability (De Pue et al., 2023; He et al., 2024). This study aims to address these gaps by conducting a comprehensive analysis of anthropogenic and biospheric CO_2 dynamics near the MASP. By utilizing the Weather Research and Forecasting model coupled with chemistry (WRF-GHG) coupled to the Vegetation Photosynthesis and Respiration Model (VPRM) (Mahadevan et al., 2008) integrated with vehicular emissions from VEIN model (Ibarra-Espinosa et al., 2018) and industrial sector emissions (EDGAR), we seek to elucidate the underlying drivers of CO_2 variability. In addition, we utilized data from the OCO-2 satellite to cover the study domain, comparing smoothed XCO2 concentrations derived from WRF-GHG (considering biogenic and anthropogenic emissions). Through a combination of model simulations, field observations, and satellite data analysis, this study seeks to provide an understanding of CO_2 dynamics in urban environments. This is the first study in this field conducted in any city in the Global South, making it an innovative effort with significant implications. By setting a precedent, this research paves the way for future studies, contributing to a more comprehensive global picture of CO_2 dynamics in urban environments.

2 WRF-GHG

2.1 Model set-up

A set of high-resolution simulations of atmospheric Greenhouse Gas concentrations were performed with the Weather Research and Forecasting model coupled with Chemistry (WRF-Chem V4.0) and a modified version of the greenhouse gas chemistry module (WRF-GHG) (Beck et al., 2011). The WRF-GHG was used to simulate the transport of the mole fraction of CO_2, and no chemical processes or reactions have been used in this modified version (Beck et al., 2013). The period that has been simulated was from 1 February to 31 August, 2019. This period was selected due to available data from monitoring stations from the Metroclima network for CO_2. The simulations were made for each month. For each run, the simulation was initiated 5 days before and these 5 days were discarded as spin-up time. The single modeling domain was centered at 23.5°S and 46.3°W with a horizontal grid 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, 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 previous studies for this same area (Andrade et al., 2015; Vara-Vela et al., 2016; Gavidia-Calderón et al., 2023; Benavente et al., 2023). The meteorological initial and boundary conditions to drive the simulations were obtained from the European Centre for Medium-Range Weather Forecasts (ECMWF) global ERA5 reanalysis data set with 0.25º × 0.25º horizontal resolution and 6-hourly intervals (Hersbach, 2016), while chemical initial and boundary conditions were obtained from Carbon Tracker and have a horizontal resolution of 3° in longitude and 2° in latitude, with 25 vertical layers (http://carbontracker.noaa.gov). This global dataset was interpolated to provide the lateral boundary conditions for CO_2 simulations and ensure consistency with the input data for WRF-Chem. The main physics and chemistry options used in this study are listed in Table 1.

2.1.1 Anthropogenic Emissions

In MASP, the vehicular fleet represents the main source of CO_2 emissions (CETESB, 2019). Therefore, in this study, we used the Vehicle Emission Inventory model (VEIN), developed to identify the emissions contribution from mobile sources, considering exhaust, and evaporative, performing speciation, and incorporating functions to generate and spatially allocate emissions databases (Ibarra-Espinosa et al., 2018). It allows the users to load their emission factors, which we used derived from experimental campaigns conducted in different traffic tunnels in the megacity of São Paulo (Nogueira et al., 2021). The VEIN model reads and distributes by age of use, extrapolates hourly traffic data, and estimates emissions hourly and spatially. We also incorporated other emissions from the energy and industrial sectors obtained from EDGAR (Crippa et al., 2021).

2.1.2 Biogenic Fluxes

Biogenic CO_2 fluxes were simulated offline using the Vegetation Photosynthesis and Respiration Model (VPRM) (Mahadevan et al., 2008), implemented as a module within the WRF-GHG. This model estimates net ecosystem exchange (NEE) by calcu-
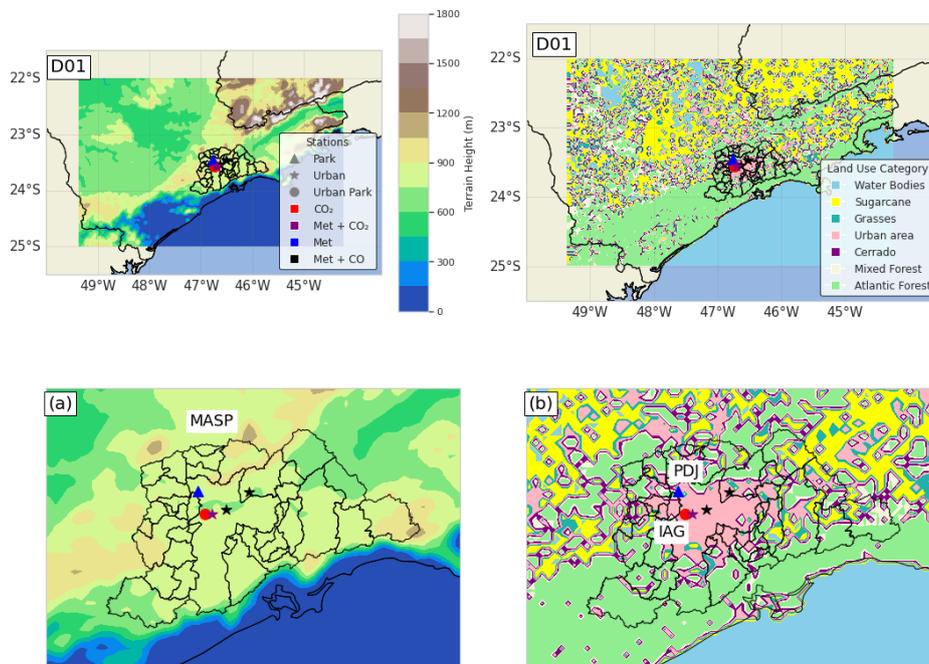
Figure 1. (a) The terrain height and highlight of the MASP area and (b) land use category map of the model domain (D01) considered in the WRF-GHG model and the locations of the CETESB and METROCLIMA monitoring stations.

Table 1. WRF-GHG 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)

4

83 lating the difference between gross ecosystem exchange (GEE) and ecosystem respiration (R), where negative fluxes indicate

84 CO_2 absorption by ecosystems (Equation 1).

85 $$NEE = GEE - R \tag{1}$$

86 The meteorological variables 2m air temperature (T) and downward shortwave radiation (PAR) from WRF model simula-

87 tions were used to calculate the GEE (Equation 2) and Respiration (Equation 3) fluxes. Additionally, factors such as the light

88 use efficiency (λ), PAR saturation (PAR0), and the Enhanced Vegetation Index (EVI), which refer to the fraction of shortwave

89 radiation absorbed by leaves were used to calculate GEE. The temperature sensitivity of the photosynthesis parameter (Tscale)

90 and the effects of leaf age on canopy photosynthesis parameter (Pscale) were both calculated as functions of the land surface

91 water index (LSWI) to identify the green-up (leaf expansion) and senescence phases (Mahadevan et al., 2008). These vegeta-

92 tion indices were derived from Moderate Resolution Imaging Spectroradiometer (MODIS) reflectance data from MOD09A1

93 Version 6 (Vermote, 2021).

94 $$GEE = \lambda \times T_{\text{scale}} \times P_{\text{scale}} \times W_{\text{scale}} \times EVI \times \frac{1}{1 + \frac{PAR}{PAR_0}} \times PAR \tag{2}$$

95 Respiratory fluxes (R) were estimated using a linear model based on air temperature and two parameters that represent the

96 linear sensitivity of respiration to air temperature (α) and the baseline respiration (β), as defined in Mahadevan et al. (2008).

97 $$R = \alpha \times T + \beta \tag{3}$$

98 The land cover data used by the VPRM were derived from the MapBiomas data (Souza Jr et al., 2020). The VPRM param-

99 eters (λ, PAR0, α, β) were optimized against flux tower NEE for the main land cover type over the study domain.

2.1.3 Meteorological data

101 Meteorological data from the São Paulo State Environmental Protection Agency (CETESB) air quality network were used to

102 evaluate the model's performance in simulating meteorological fields. CETESB manages automatic and manual air quality

103 stations over São Paulo state. These stations provide hourly information on meteorological and pollutant parameters, such

104 as air temperature, wind speed, and wind direction (Table 2), as well as the concentration of air pollutants. Monitoring fol-

105 lows instrumentation standards and directives from the Environmental Protection Agency (US EPA) and the World Health

106 Organization (WHO) respectively for air pollutants, and from the World Meteorological Organization (WMO) for meteoro-

107 logical variables (CETESB, 2019). The air quality and meteorological data are continuously published on the Qualar website

108 (https://qualar.cetesb.sp.gov.br/qualar/). This study used data from four stations located in the MASP (Figure 1): Parque D.

109 Pedro II, Pico do Jaraguá, Guarulhos, and Pinheiros. Table 2 provides the location of the sites, the classification type of the

110 stations, the 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	T2, Wd, Ws	CETESB
Pico do Jaraguá (PDJ)	23.45S, 46.76W	Park	T2, Wd, Ws and CO_2	CETESB/ METROCLIMA
Guarulhos	23.46S, 46.52W	Urban	T2, Wd, Ws	CETESB
Pinheiros	23.46S, 46.70W	Urban	T2, Wd, Ws and CO	CETESB
IAG	23.55S, 46.73W	Urban Park	CO_2	METROCLIMA

2.2 CO_2 observational data

2.2.1 Ground-based observations

For the surface model evaluation, we used CO_2 data from the METROCLIMA network in São Paulo (see Table 2), the first conventional in situ greenhouse gas measurement network established in South America (www.metroclima.iag.usp.br). The network comprises four continuously operating monitoring stations, all located within the MASP and equipped with cavity ring-down spectroscopy instruments (Picarro) that measure the concentrations of CO_2 following the directives from WMO. The monitoring stations are located at various locations within MASP: in a vegetated area at the extreme west (Pico do Jaraguá, 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 building (ICESP); and in an urban area in the east zone characterized by heavy traffic in the neighborhood (UNICID). However, we only used data from the IAG and PDJ sites, the only two that monitored CO_2 for the period chosen in this study, before the Covid-19 pandemic period (Souto-Oliveira et al., 2023).

2.2.2 CO_2 fluxes data

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 PTF 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', 45°03' 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

134   (Rocha et al., 2002). For sugarcane we used data from the municipality of Pirassununga, in São Paulo State, Brazil (21° 57'S,

135   47° 20'W at 655 m altitude) for the period from November 2016 to August 2017 (Cabral et al., 2020). The VPRM parameters

136   were optimized separately for each PFT using half-hourly observed fluxes from the flux towers over the entire observation

137   periods. We optimized the parameters for the GEE and R simultaneously, and for the default VPRM parameters we used

138   non-linear least squares minimization, between the modeled NEE and the flux tower estimation of the observed NEE. In the

139   optimization, the VPRM model is driven by the meteorological measurements of the sites and their specific land covers. The

140   vegetation indices (EVI and LSWI) were derived from the product MOD09A1 of MODIS at 500 m resolution and 8-daily

141   frequency using Google Earth Engine.

142   2.2.3   XCO_2 satellite observations

143   Satellite-based XCO_2 observations were utilized in addition to surface CO_2 measurements over the study domain. OCO-2,

144   NASA's inaugural Earth remote sensing satellite dedicated to atmospheric CO_2 observations, was launched in 2014 (Crisp,

145   2015). Operating on a solar synchronous orbit, OCO-2 conducts global measurements of CO_2 absorption and emission at 13:30

146   Local Solar Time. The OCO-2 observation data utilized were ACOS L2 Lite Output Filtered with oco2-lite_fle_prefilter_b9,

147   which were converted from Level 1 radiance to Level 2 data using the ACOS retrieval algorithm developed by O'Dell et

148   al. (2012). Data quality assessment for OCO-2 observations can be performed using the xco2_quality_flag and warn_level

149   parameters, as detailed in the OCO-2 Data Product User's Guide (Osterman et al.). In this study, we considered only OCO-2

150   data with a '0' xco2_quality_flag value that indicates "good" quality. Initially, simulated CO_2 concentrations were interpolated

151   to match the latitude, longitude, horizontal resolution, and vertical levels of OCO-2 data. Due to the difference in data types

152   and units between the simulated CO_2 concentrations and observed XCO_2 from satellites, a conversion was necessary prior to

153   comparison. Consequently, CO_2 concentrations simulated at each pressure level in WRF-GHG were transformed into XCO_2

154   concentrations following the methods by Connor et al. (2008) and O'Dell et al. (2012), as follows:

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

156   where X$CO_{2\mathrm{a}}$ 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

157   interpolated simulated CO_2 concentrations of WRF-GHG, and $CO_{2\mathrm{a}}$ is a priori CO_2.

158   2.3   Evaluation metrics

159   Several statistical metrics are available for assessing the effectiveness of atmospheric models. These include mean bias er-

160   ror (bias, Equation 5), indicating the average difference between the simulation and the observation; root-mean-square error

161   (RMSE, Equation 6), which quantifies the square root of the average squared deviation between simulation and observation;

162   and the correlation coefficient (R^2, Equation 7), representing the degree and direction of the linear connection between the

163   simulation and the observation. To evaluate the model results, bias, root mean square error (RMSE), and correlation (R^2) were

164   utilized and computed as follows:

165 $$Bias = \frac{\sum_{i=1}^{N}(pred_i - obs_i)}{N} \tag{5}$$

166 $$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(pred_i - obs_i)^2}{N}} \tag{6}$$

167 $$R^2 = \frac{\sum_{i=1}^{N}(pred_i - \overline{pred_i})(obs_i - \overline{obs_i})}{\sqrt{\sum_{i=1}^{N}(pred_i - \overline{pred_i})^2 \sum_{i=1}^{N}(obs_i - \overline{obs_i})^2}} \tag{7}$$

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

3 Results

170 Hourly simulations were conducted from 1 February to 31 August 2019, with each month simulation including a five-day

171 spin-up period. In the following sections, the performance of meteorological drivers will be first presented, followed by the

172 terrestrial surface CO_2 fluxes and atmospheric CO_2 concentrations from the IAG station and Pico do Jaraguá (PDJ) stations.

173 These measurements were used to evaluate the model performances and to assess the local impacts of the main CO_2 sources

174 and sinks on atmospheric CO_2 concentrations.

3.1 Model performance for meteorological drivers

176 The assessment of the meteorological model performances is essential for accurately simulating greenhouse gas concentrations.

177 In this study, the model represented the temporal variability and trends of 2-meter temperature (T_{2m}), 10-meter wind speed

178 (WS), and direction (WD) throughout the simulation period, as illustrated in Fig. 2 and the supplementary material. The WRF-

179 GHG model effectively captured significant changes in the observed variables, although it failed to accurately represent the

180 maximum and minimum peaks, particularly for wind speed. The simulated 2-meter temperature tended to overestimate values

181 at specific sites, such as Parque D. Pedro II (bias = 0.5°C), Guarulhos (bias = 0.1°C) (see figure A1a and A2a in Appendix), and

182 PDJ (bias = 0.7°C) (see Figure 2a). However, at the Pinheiros station, the simulated surface temperature was underestimated

183 (bias = -0.7°C) (Figure A3a in Appendix).

184 In terms of biases, the model overestimated the wind speed at all sites (bias < 1.5 ms^{-1}), with Pico do Jaraguá exhibiting the

185 highest mean bias (1.4 ms^{-1}). This overestimation could be attributed to the model's misrepresentation of land use, leading

186 to elevated wind speeds in areas classified as urban rather than vegetated. Notably, numerical models tend to lack sensitivity

187 in simulating very low-velocity speeds due to imperfections in land surface processes and the model's ability to accurately

188 resolve topographical features (Shimada et al., 2011; Zhang et al., 2009; Vara-Vela et al., 2018, 2021). The model's wind

189 directions showed sufficient sensitivity, aligning accurately with observed values. Both the model and observations indicated
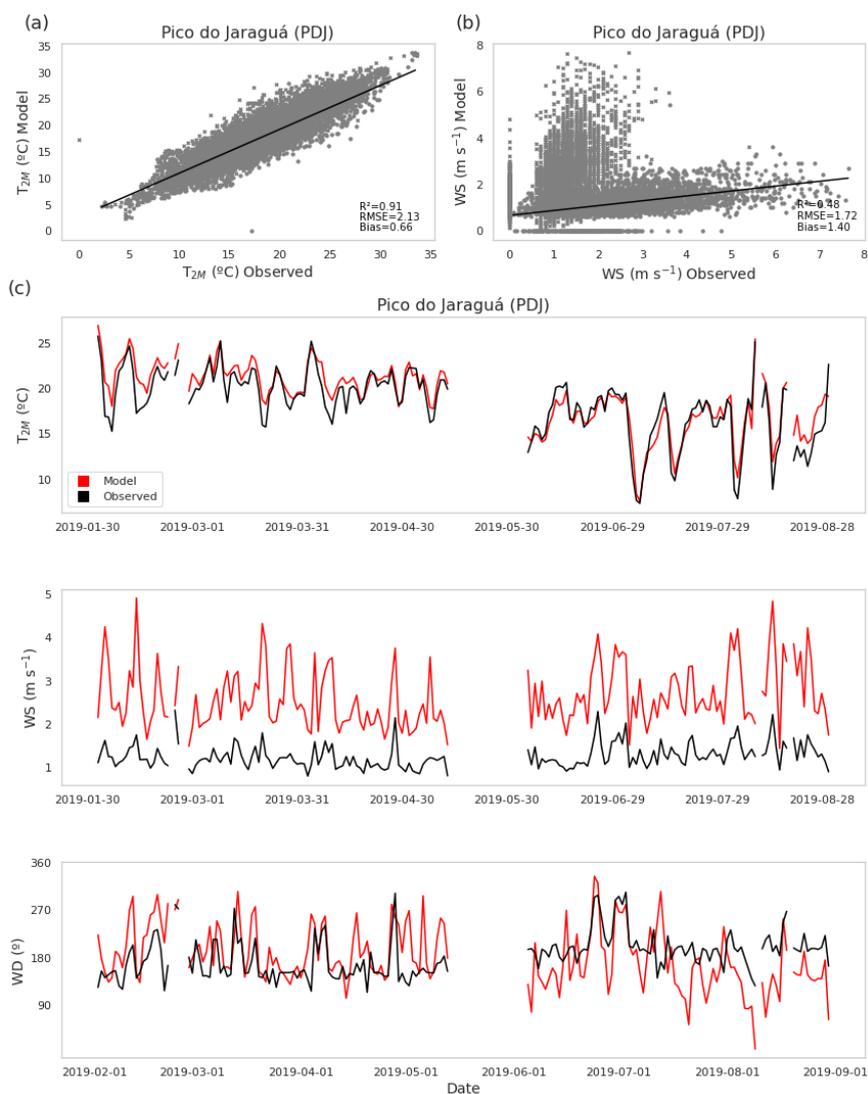
Figure 2. The panels in a) show the scatter plots of hourly measurements of 2 m air temperature (T_{2m}) and b) show 10 m wind speed (WS) compared to observed data from the PDJ station. The figure illustrates the relationship between modeled and observed data. The panels in c) show the daily averages 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.

190 that prevailing winds were predominantly from the southeast. In summary, the WRF model showed proficiency in reproducing
191 atmospheric conditions in the study area, particularly concerning air temperature and wind Direction, with similar performances
192 as previous studies (Feng et al., 2016; Deng et al., 2017).

3.2 The VPRM Model: Evaluation with Flux Tower Data

194 The optimization results are shown in Table 3. Substituting alpha and beta back into the respiration equation led to a better
195 model representation of NEE compared to NEE values simulated with default parameters (Mahadevan et al., 2008) for the
main PFT across the domain (Figure 3).

Table 3. 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 (PTFs)	Default				Optimized & Botía et al. (2022)			
	PARo	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

196

197 The optimized VPRM parameters for the Atlantic Forest exhibited the greatest discrepancies compared to other vegetation
198 classes. The geomorphological characteristics of the Atlantic forest differ from those of the evergreen forest studied by (Ma-
199 hadevan et al., 2008), where the default parameters (VPRM_default, represented by the red curve in Figure 3) were used. The
200 optimized VPRM parameters (VPRM_opt, shown as the purple curve in Figure 3) more accurately captured the seasonal cycle
201 in the daily average NEE for the three PFTs optimized in this study. The model was particularly successful in capturing the
202 seasonal profile for the agricultural ecosystem, which can be attributed to the more pronounced seasonal transitions of sugar-
203 cane (as indicated by the EVI), even though the low-resolution satellite indices do not fully capture the onset of the growing
204 season. However, this allowed the model to better represent the GEE equation for this ecosystem. For the cerrado, the model
205 smoothed the NEE peaks, and the GEE and respiration equations were also smoothed with the optimization. Optimizing the
206 VPRM parameters improved the representation of the growing season, especially for the Atlantic Forest and sugarcane, while
207 using either optimized or default parameters for the cerrado resulted in similar NEE simulation.

208 The first panel in Figure 4 shows the monthly net CO_2 flux simulated by the VPRM model for 2019. February represents a
209 summer month, while August represents a winter month. The second panel shows the monthly daily net CO_2 flux simulated
210 at the three flux tower sites used to optimize the VPRM model parameters. In February, negative NEE values are found in the
211 northern part of the MASP, while the southern part exhibits positive NEE fluxes in the coastal region. This gradient reflects
212 the distribution of vegetation types, their phenology, and productivity, as well as the impact of urbanization, with null fluxes
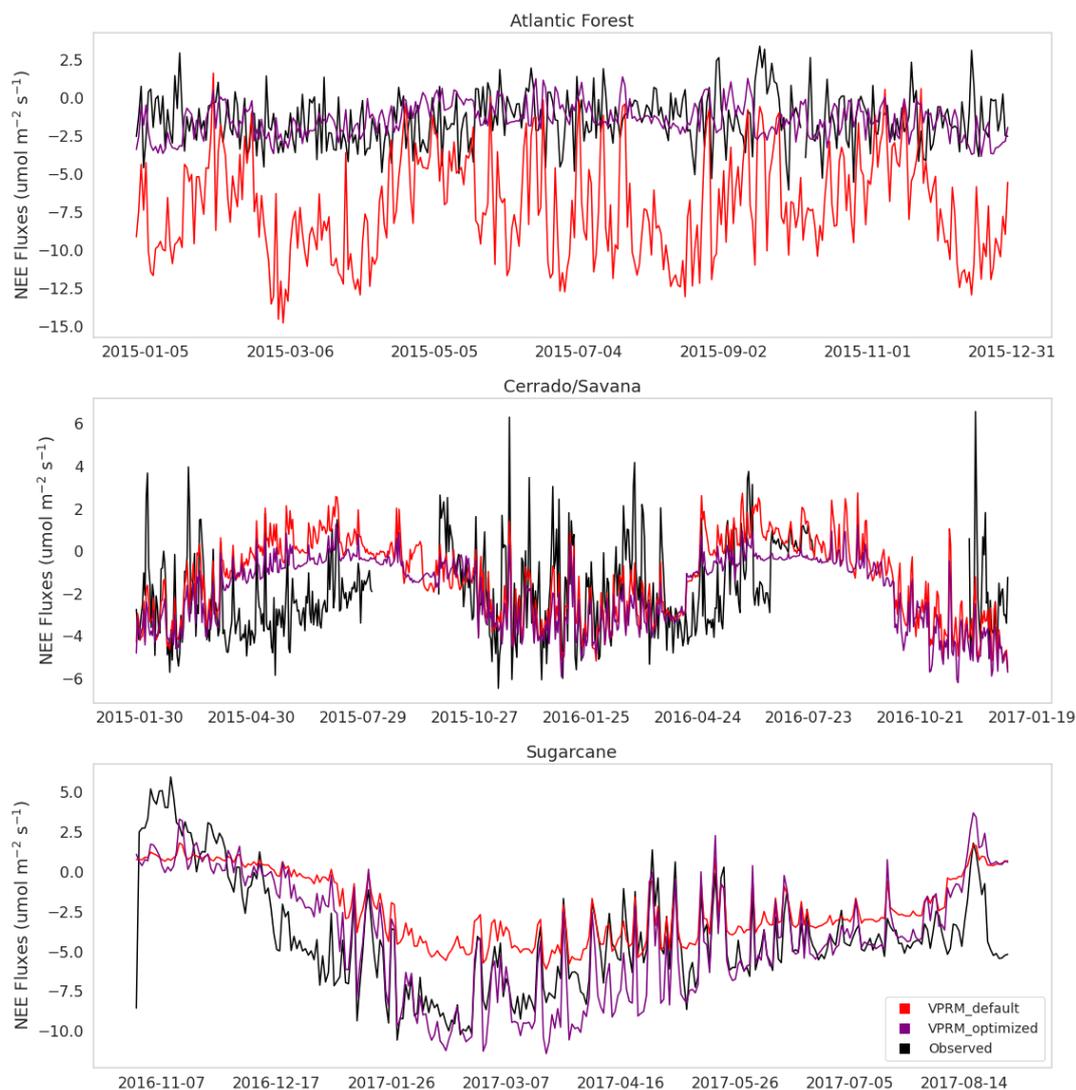
Figure 3. Daily variability of NEE fluxes from the flux tower (black line), alongside NEE fluxes simulated by the VPRM model using the default (red line) and optimized (purple line) parameters.

213 observed in highly urbanized regions such as downtown MASP. In the summer season in the Southern Hemisphere, vegetation

214 productivity reaches its peak across all land cover classes, leading to more negative NEE values (Figure 4a) specially for the

215 Cerrado, and sugarcane (within the state of São Paulo), but less intense in the Atlantic Forest in the southern coastal region.

216 This results in negative NEE fluxes (dark green across the domain), indicating that these areas acted as a CO_2 sink.
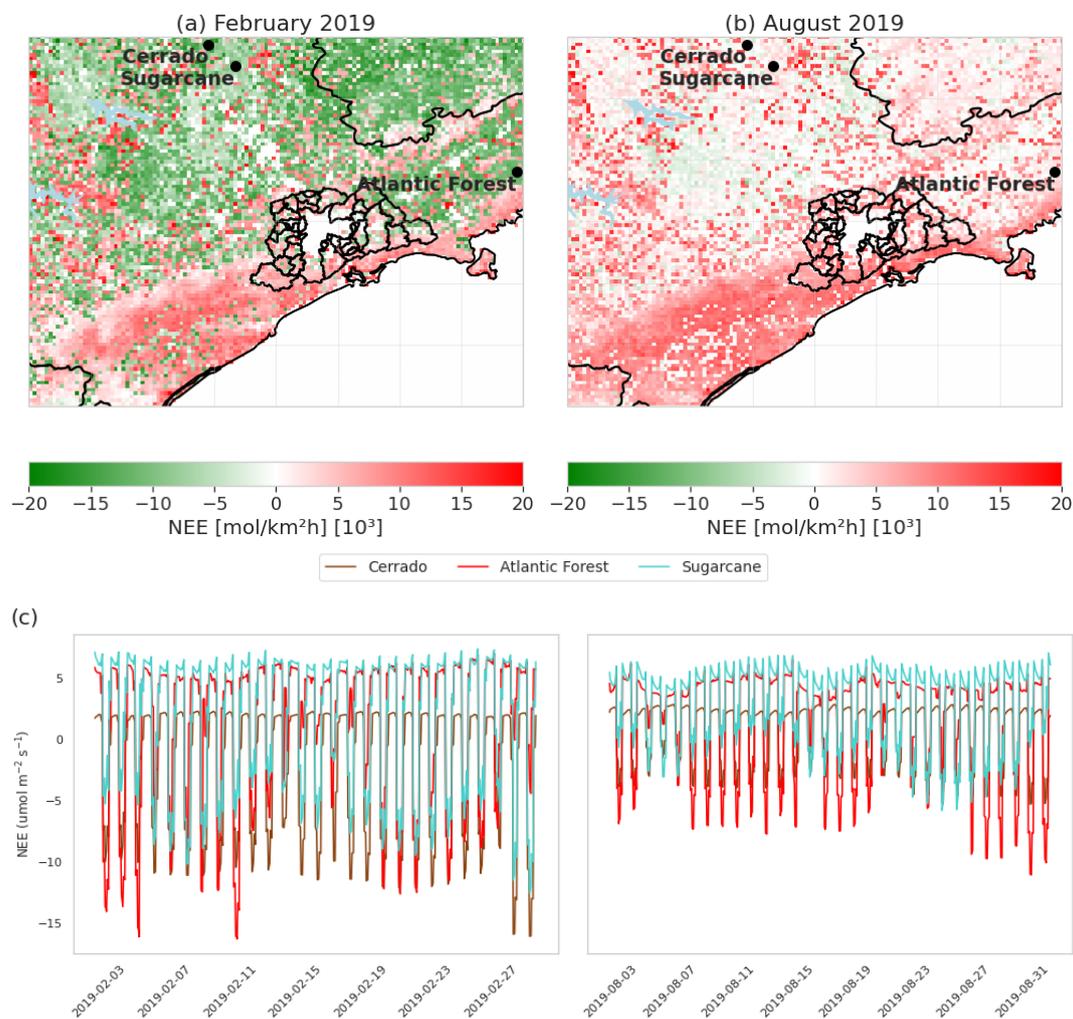
Figure 4. First panel shows the monthly hourly average of net ecosystem exchange (NEE) for February (a) and August (b) 2019. The second panel (c) shows the daily profile for the same months and for three different PTFs sites.

217 In August, the cold and dry conditions, due to reduced solar radiation and a lower leaf area index, resulted in positive fluxes

218 across most of the domain and low negative fluxes in only a few areas (Figure 4b). The highest positive NEE values are found in

219 the southern coastal region. Generally, larger areas with negative CO_2 fluxes are observed in February compared to August for

220 the same dominant land cover classes. This indicates greater CO_2 absorption by agriculture in February compared to forested

12

221 regions. Conversely, in August, CO_2 fluxes are predominantly lower and negative across most of the domain, with higher

222 positive values in the coastal area, especially in the south. Overall, the domain acts as a net CO_2 sink during summer, while

223 vegetation becomes a CO_2 source in winter, except for the Atlantic Forest in the southern part of the study area. The second

224 panel also shows simulated fluxes for the same flux tower sites, with negative net fluxes in February, particularly in the Atlantic

225 forest, sugarcane, and cerrado. This underscores the reduction in negative fluxes during winter, as seen in the August data for

226 all three vegetation types. Unfortunately, observed data from these flux towers for this period were not available for statistical

227 model evaluation. However, Figure 4 illustrates the significant influence of climatic drivers on reduced flux trends, consistent

228 with findings by Raju et al. (2023) for a tropical region. Note that the respiration equation in Mahadevan et al. (2008) is a

229 simple linear function of temperature and does not account for seasonal or spatial variability in biomass and litter inputs to soil

230 carbon pools Gourdji et al. (2022), which is particularly relevant for forest ecosystems like the Atlantic Forest.

3.3 Seasonal variations in observed and modeled CO_2 mixing ratios

232 Figure 5 and Table 4 depict the monthly mean, standard deviation, bias and RMSE of CO_2 concentrations at two sites in the

233 MASP. In 2019, the IAG station recorded CO_2 values ranging from 406 to 464 ppm. The seasonal variation peaked during

234 autumn (435.7 ± 29.1 ppm), followed by winter (434.6 ± 30.8 ppm), with the lowest values observed in summer (434.0 ± 26.1

235 ppm). This variation in CO_2 levels is primarily influenced by factors such as the latitude of the observation site, meteorological

236 conditions including wind speed and atmospheric stability, as well as seasonal patterns of photosynthesis and vehicular traffic.

237 The maximum and minimum monthly CO_2 concentrations at IAG were recorded in June (442.8 ± 32.8 ppm), during the

238 winter season, and March (430.2 ± 24.9 ppm), during the summer season, respectively. During the summer months in MASP,

239 high humidity, wind speed, and circulation patterns typically lead to lower atmospheric stability and increased dispersion of

240 various gasses and particles. Meanwhile, at the Pico do Jaraguá station, CO_2 levels ranged from 407 ppm to 425 ppm. The

241 seasonal variation peaked during summer (416.5 ± 10.7 ppm), followed by autumn (416.4± 9.1 ppm), with the lowest values

242 observed in winter (414.4 ± 6.7 ppm). The maximum monthly CO_2 mean at PDJ was identified in May (417.3 ± 9.1 ppm),

243 corresponding to the autumn season, while the minimum was recorded in July (414.0 ± 6.3 ppm), during the winter season.

244 Monthly values at PDJ exhibited less variability and a smaller standard deviation compared to the IAG site. This result was

245 expected, considering that the IAG site is significantly impacted by vehicular traffic in its vicinity. In contrast, PDJ is located at

246 a higher elevation in a more vegetated area, with less influence from local anthropogenic sources. Additionally, it was expected

247 that PDJ would show lower CO_2 concentrations during the summer due to the stronger vegetation signal at Pico do Jaraguá

248 compared to the IAG site. However, PDJ actually shows peak CO_2 levels in summer and the lowest values in winter, indicating

249 that additional ecological and ecosystem variables need to be considered for a better understanding of this location.

250 The simulated CO_2 concentrations for the IAG station ranged from 410 ppm to 436 ppm, with a seasonal variation peaking

251 in autumn (425.0 ± 15.0 ppm), followed by winter (422.8 ± 12.2 ppm), and the lowest values occurring in summer (416.8 ±

252 8.3 ppm), mirroring the observed data. Notably, the highest and lowest monthly CO_2 concentrations at IAG were identified

253 in June (431.7 ± 17.0 ppm) and February (414.6 ± 5.3 ppm), respectively. Although the maximum monthly value from the

254 model coincided with the observed data, the month with the minimum concentration was February, which may be attributed to
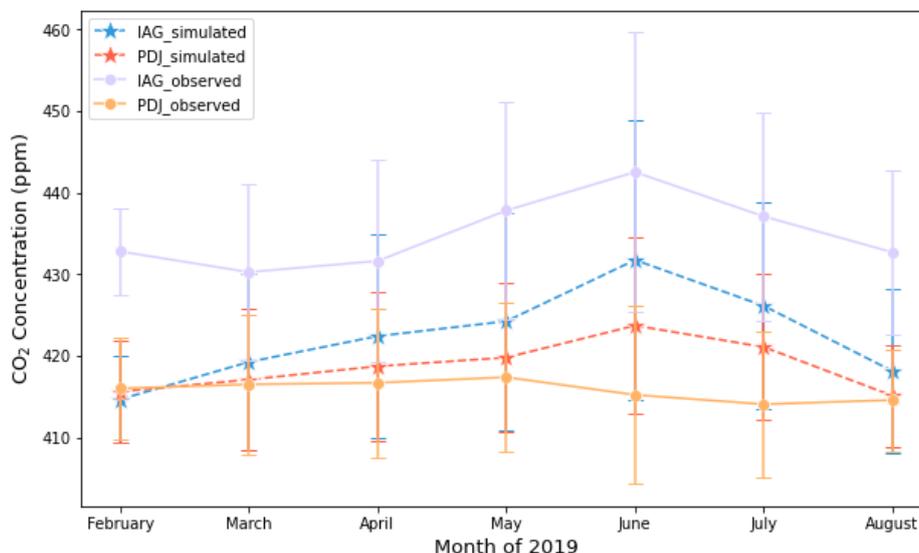
Figure 5. Monthly variations in CO2 concentration observed and simulated at IAG and Pico do Jaraguá stations in 2019. Error bars represent the monthly standard deviation.

gaps in measurement, which were not considered when calculating the mean, thereby influencing the observed monthly mean. The CO_2 concentrations at Pico do Jaraguá ranged from 415 ppm to 424 ppm, with seasonal variation peaking in autumn (419.9 ± 9.9 ppm), followed by winter (418.7 ± 8.7 ppm), and the lowest values occurring in summer (416.5 ± 7.5 ppm). The model data profile for Pico do Jaraguá more closely resembles the simulated IAG profile than the PDJ station's observed profile, likely due to the model's resolution, its limitations in representing land use, and underestimated vehicular emissions in these areas. However, negative biases were observed for all seasonal periods at IAG, indicating an underestimation of CO_2 concentrations and higher root mean square errors compared to the statistics for the PDJ station. The PDJ station exhibited low positive biases, indicating better agreement between the model and observations across all periods and lower errors between the model and observations.

Table 4. Monthly means and standard deviation of CO2 concentrations for IAG and Pico do Jaraguá (PDJ) stations.

Station	Season	CO_2 Observed (ppm)	CO_2 Simulated (ppm)	Bias (ppm)	RMSE (ppm)
	Summer (February to March)	434.0 ± 26.1	416.8 ± 8.3	-16.6	28.3
IAG	Autumn (March to June)	435.7 ± 29.1	425.0 ± 15.0	-10.9	26.7
	Winter (June to August)	434.6 ± 30.8	422.8 ± 12.2	-11.8	30.8
	Summer (February to March)	416.5 ± 10.7	416.5 ± 7.5	0.15	9.9
PDJ	Autumn (March to June)	416.4 ± 9.1	419.9 ± 9.9	3.50	11.4
	Winter (June to August)	414.4 ± 6.7	418.7 ± 8.7	4.37	10.2

14

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).
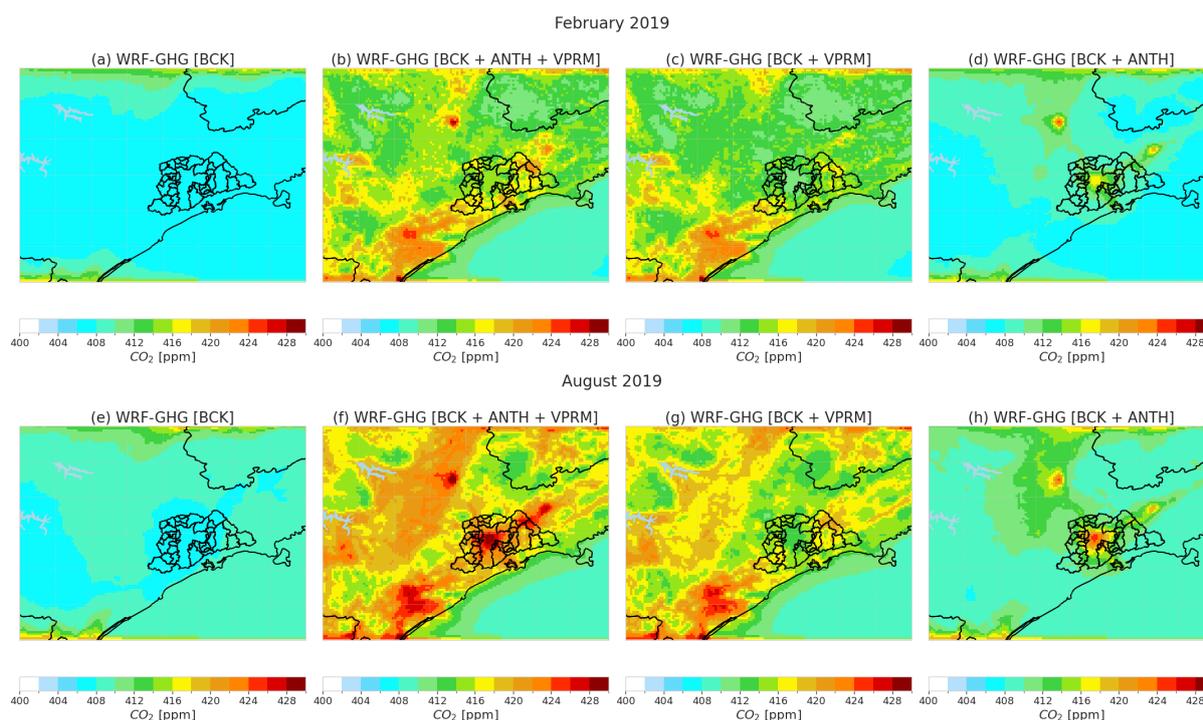
Figure 6. Atmospheric CO_2 concentrations under different emission scenarios (refer to the text). The panels in the first row represent the monthly mean concentration for February (a, b, c, d), while the panels in the second row represent the monthly mean concentration for the August period (e, f, g, 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.

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 does not calculate CO2 fluxes in the urban area, we can observe the influence of biogenic signals being transported to this area.

276 The coastal region southwest of MASP, characterized by a dense forest region, did exhibit the highest CO_2 concentrations,
277 between 420-424 ppm. The simulation with anthropogenic emissions (Figure 6d) stands out elevated CO_2 concentrations
278 over the center of the city of São Paulo, characterized by high vehicle emissions, as well as over other two urban areas in
279 the north and northeast of MASP. The monthly mean CO_2 concentration in these two urban areas was roughly 420 ppm,
280 attributed to emissions from refineries represented by the EDGAR datasets as well as vehicles. Figure 6b shows the simulated
281 CO_2 concentration considering both vegetation fluxes and anthropogenic emissions. As expected, this simulation combines
282 both contributions, resulting in high CO_2 concentrations over urban areas and along the coastal region. For August, it can be
283 observed that the background concentrations (Figure 6e) were slightly higher around MASP. Additionally, the monthly mean
284 CO_2 concentration for the scenario in August with only biogenic sources was 8 ppm higher than that in February, which can
285 be explained by the lower photosynthetic rates in this period, as observed in Figure 4. The Atlantic forest in the coastal region
286 exhibits more positive CO_2 fluxes and lower photosynthetic activities, characterized by lower amounts of rainfall in the region
287 that contribute to this reduced photosynthetic production by vegetation. The simulation with only anthropogenic emissions
288 (Figure 6h) shows higher CO_2 concentrations compared to those in February. This increase in CO_2 levels in August is attributed
289 to a lower planetary boundary layer height. However, it is important to point out that the EDGAR anthropogenic emission
290 inventory generally overestimates the emissions around local anthropogenic sources (e.g., urban areas) (Seo et al., 2024). The
291 higher simulated CO_2 concentration for August compared to February, in the scenario with both biogenic and anthropogenic
292 sources, is largely dependent on factors such as atmospheric stability and meteorological conditions. Atmospheric stability,
293 along with meteorological variables such as humidity, solar radiation, and temperature, plays a crucial role in determining
294 biogenic CO_2 concentrations. In addition, under stable atmospheric conditions, such as those often observed during winter
295 periods, CO_2 concentrations tend to accumulate near the surface, resulting in higher concentrations, especially in urban areas.
296 Therefore, the comparative analysis between simulations of CO_2 concentrations during summer and winter periods highlights
297 the importance of considering not only anthropogenic emissions but also biogenic fluxes from vegetation, along with local
298 atmospheric conditions.

3.3.2 Evaluation of sources contribution

300 In Figure 7, we applied a data selection scheme to all-time series to minimize the effects of local contributions and increase
301 the spatial representativeness of each record, it consists of retaining mid-afternoon (09–17 h local) data, when the air is well-
302 mixed, providing a large spatial representativeness with minimum influence from local sources (Gerbig et al., 2008; Ramonet
303 et al., 2020). Figure 7, shows the comparison of the daily mid-afternoon average CO2 concentrations simulated by the model
304 for February and August 2019, considering both biogenic and anthropogenic sources (see Figures 6b and 6f), at both IAG and
305 PDJ sites. The left panels (Figures 7a, 7c, 7e, and 7g) depict the simulated CO_2 concentration considering both anthropogenic
306 and biogenic sources (all_sources, in gray), alongside observed concentrations (observed, in purple) for both sites. Conversely,
307 the right panels (Figures 7b, 7d, 7f, and 7h) display the different simulations considering anthropogenic and biogenic sources
308 separately to the daily concentration. In Figure 7a, which represents the sole summer month with observed data in February
309 2019, the simulated values generally underestimated the observed concentrations. While the observed average CO_2 concentra-

tion stood at 424.0 ppm, this figure was somewhat compromised by missing data in the observed profile, whereas the simulated average was 412.4 ppm, indicating an approximate 11 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 22nd and 23rd, there was a peak in the CO_2 concentration of the observed data, but this was not represented in the simulations with all sources and in the anthropogenic simulation, but in the biogenic simulation, this same period stood out but with less intensity.

For the PDJ site, the observed average concentration was 414.3 ppm, with a simulated average of 412.2 ppm, marking a marginal difference of 2 ppm between observed and simulated concentrations. The model effectively captured peaks and profiles for this period and location (Figure 7c), with biogenic contributions emerging as more substantial (Figure 7d) compared to the IAG site. This difference arises because PDJ is located in a vegetated area away from traffic sources. In August, characterized by a drier, more stable boundary layer and lower wind speed, observed data for IAG showed an average of 426.2 ppm (Figure 7e), while with the model showed a monthly average of 413.2 ppm, resulting in a discrepancy of only 13.0 ppm, i.e. a closer approximation compared to February. In terms of the contributions of the sources (Figure 7f), simulations showed similar patterns, with a few days where CO_2 contributions from biogenic fluxes exceeded those from anthropogenic emissions. Oppositely, for PDJ (Figure 7g), the monthly average concentration stood at 412.6 ppm, slightly surpassing the simulated average of 412.0 ppm. While the model slightly underestimated the monthly average by 0.5 ppm, it exhibited a profile akin to the observed data, with a more pronounced biogenic signal compared to anthropogenic contributions in the CO_2 simulation (Fig. 7h), underscoring vegetation's substantial role as an important CO_2 source to consider in the simulations for both sites. Additionally, Figure 4 illustrates more positive CO_2 fluxes (representing CO_2 emissions to the atmosphere) by the VPRM model during this period. The bias and RMSE for each simulation at the IAG and PDJ sites for February and August 2019, are illustrated (see Figure A4 in Appendix). Overall, the bias tended to be negative across the board, indicating that the simulated surface CO_2 concentrations generally underestimated the observed values. Notably, with the exception of the ALL_PDJ simulation for August (Figure 8c), which displayed a small negative bias, CO_2 simulations consistently fell below-observed levels at this site. Among the six sets of simulations, PDJ exhibited the smallest bias, averaging at -3.0 ppm, while IAG displayed a larger average bias of -13.3 ppm. Further analysis revealed that simulations incorporating both biogenic and anthropogenic sources (ALL_*) consistently yielded the smallest biases. RMSE values at PDJ remained below 12 ppm, while those at IAG exceeded this threshold. Notably, simulations focusing solely on anthropogenic sources at PDJ exhibited the poorest RMSE for both February and August, highlighting the significance of vegetation fluxes at this site. On the other hand, at IAG, simulations relying solely on biogenic sources in February and on anthropogenic sources in August resulted in the highest RMSE values, highlighting the importance of anthropogenic emissions, especially traffic ones. In February, simulations ALL_IAG and ALL_PDJ displayed the lowest RMSEs at 25.25 ppm and 10.15 ppm, respectively (Figure A2b). In August, these figures stood at 26.24 ppm and 6.91 ppm (Figure A2d), respectively. Overall, simulations incorporating both biogenic and anthropogenic sources yielded better results in terms of RMSE and bias, indicating a closer alignment between simulated and observed surface CO_2 concentrations. Additionally, CO_2 simulations at PDJ demonstrated the closest resemblance to observed values among the six simulations.
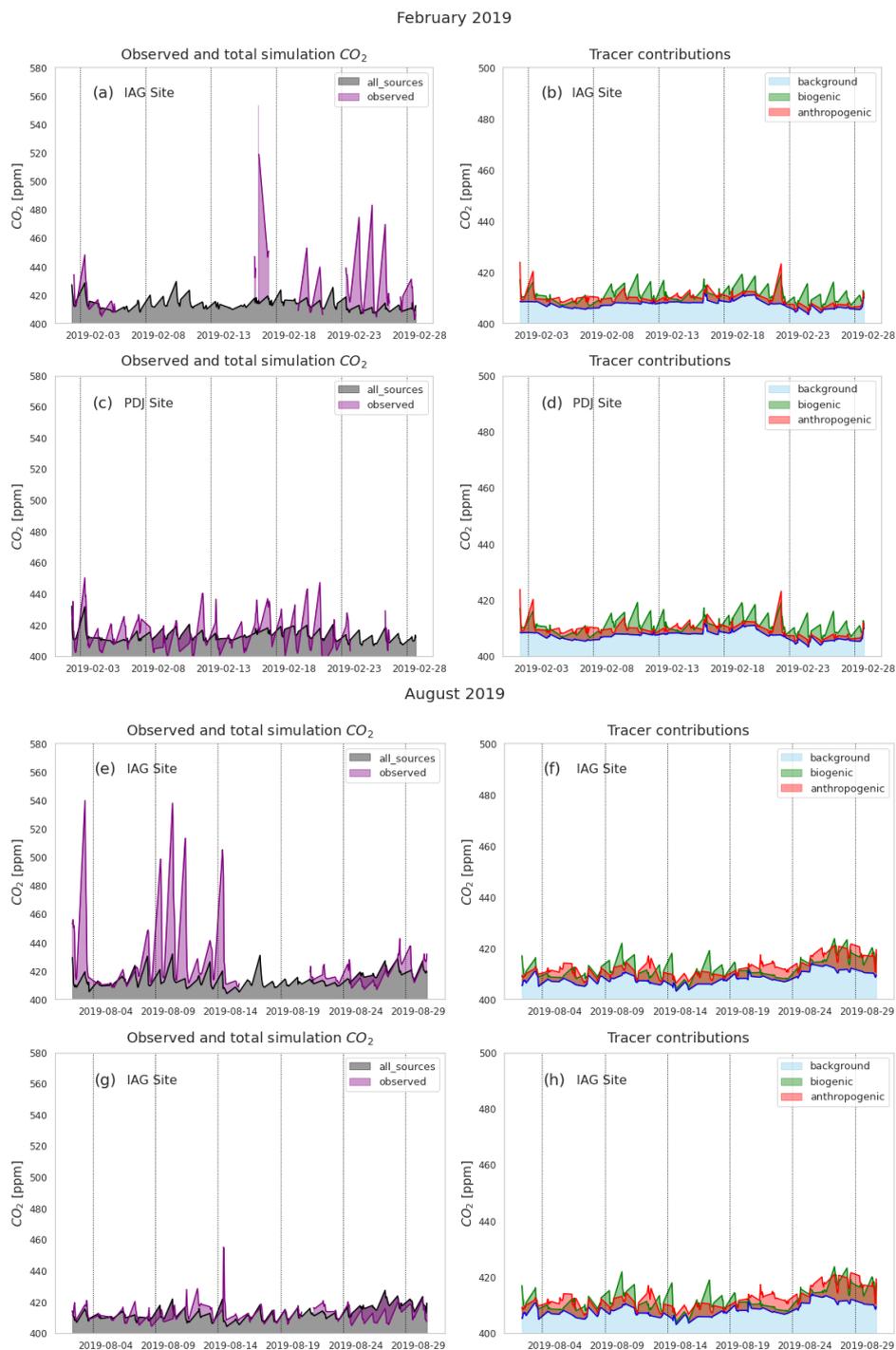
Figure 7. Daily mean CO_2 concentrations simulated and observed for the IAG site in February 2019 (a), for the PDJ site in February (c), for the IAG site in August (e), and for the PDJ site in August (g). And the daily simulated background concentrations, anthropogenic concentrations, and biogenic concentrations 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).
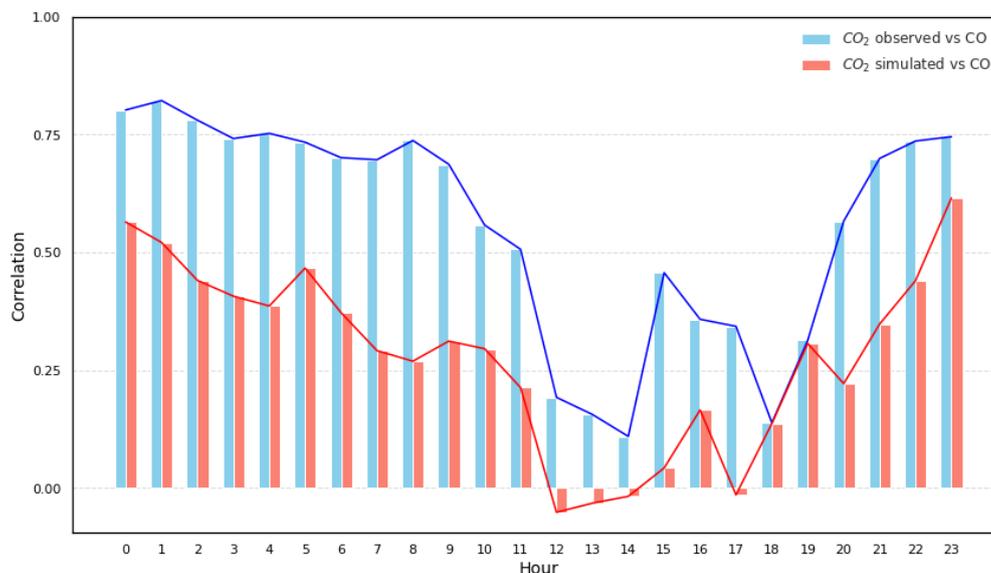
Figure 8. The hourly correlation between observed and simulated CO_2 concentrations at the IAG site and CO concentrations measured at the Pinheiros station.

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

In addition to the correlation between gases, Figure 9 indicates that both profiles (modeled and simulated CO_2) 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, 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 these high CO_2 concentrations since it assumes that emissions follow the same diurnal variation every day of the month. For the model to accurately simulate these high concentrations, the emissions must better represent the reality of urban emissions.
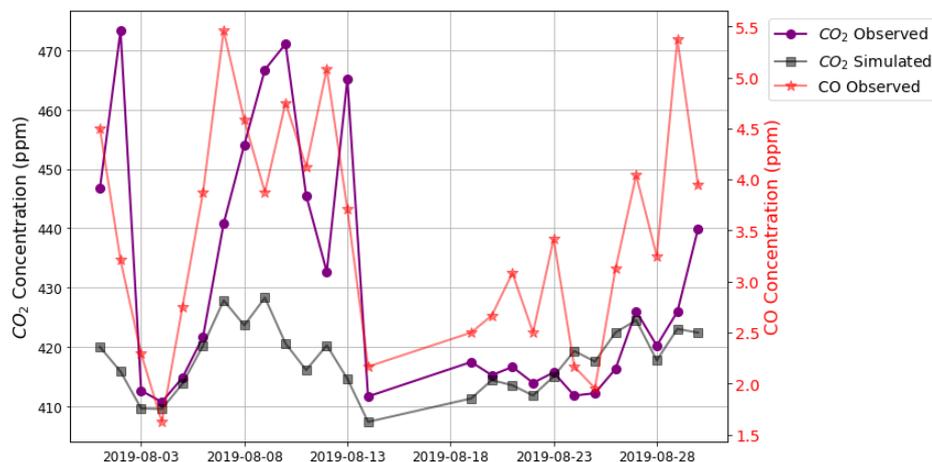
Figure 9. The hourly correlation between observed and simulated CO2 concentrations at the IAG site and CO concentrations measured at the Pinheiros station.

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 A5 in Appendix) 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 positive 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 the model's tendency to overestimate winds, 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 concentration provided by the Carbon Tracker.
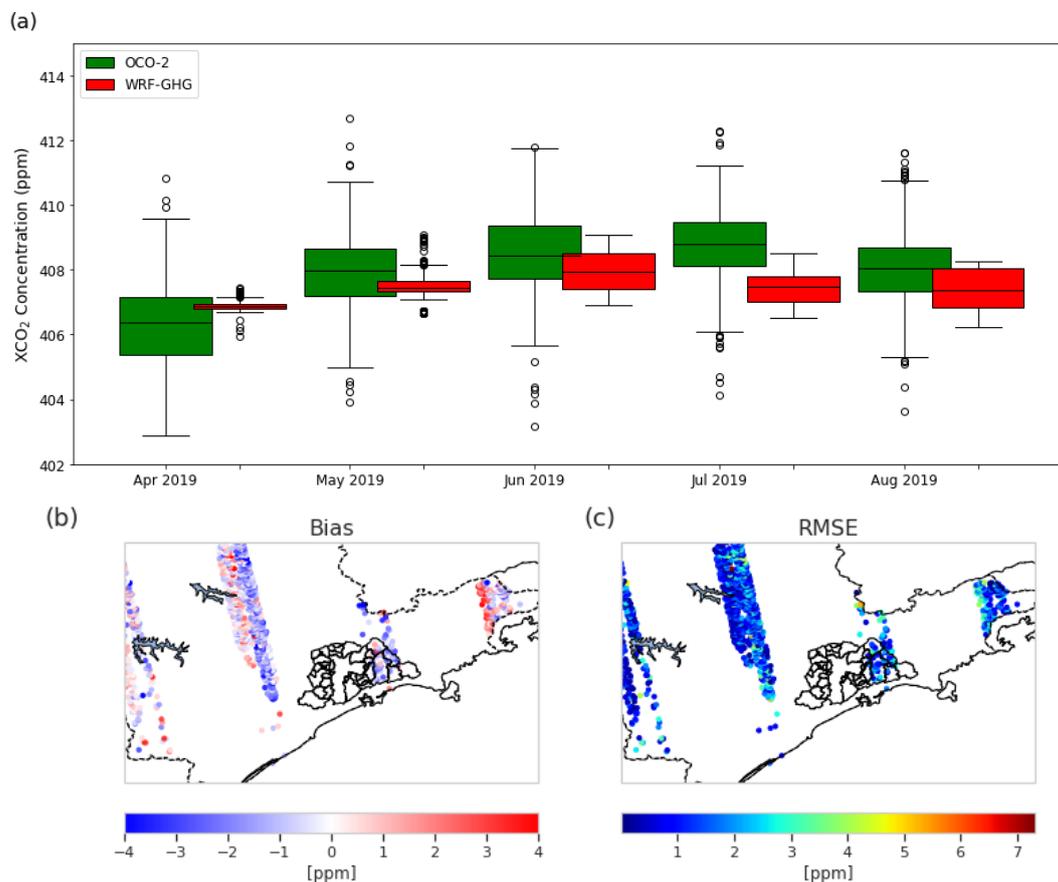
Figure 10. a) Monthly boxplots of observed and simulated XCO_2 concentrations for the period from 1 April 2019 to 31 August 2019, b) Bias and c) RMSE calculated by pixel over the study domain.

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. Anthropogenic emissions were curated from diverse models and products to accurately reflect real urban conditions. Boundary and initial conditions were scrutinized using global products. The spatial and temporal distribution of modeled CO_2 concentrations, stemming from anthropogenic, biogenic, and background emission processes, underwent comprehensive analysis. Wind dynamics emerged as a pivotal factor,

21

394 underscoring the importance of precise simulation of wind speed, wind direction, and planetary boundary layer dynamics.
395 The WRF-GHG model adeptly replicated meteorological variables such as temperature, however discrepancies in local wind
396 speed and direction persisted. This can be attributed to the intricate topography and the limited model resolution (3 km),
397 which impedes the capture of nuanced local dynamical processes. Surface CO_2 concentrations unveiled distinct diurnal cycles
398 shaped by local anthropogenic emissions, boundary layer dynamics, and vegetation respiration. Importantly, the modeled CO_2
399 concentrations exhibited high sensitivity not only to atmospheric vertical mixing near the surface but also to the prescribed
400 temporal profiles of anthropogenic and biogenic emissions, highlighting the underestimation of vehicular emissions. These
401 sources of error, particularly pronounced in winter, present challenges in accurately quantifying city emissions. In suburban
402 locations such as the PDJ site, distant from urban sources, anthropogenic emissions diminish, and the vertical gradient of
403 CO_2 concentration generated by city emissions attenuates through atmospheric convection and diffusion processes. However,
404 during the growing season, the contribution of biogenic flux to CO_2 concentration warrants attention, especially concerning the
405 simulation of nocturnal CO_2 concentrations and ecosystem respiration, improving the respiration equation in the VPRM model
406 (Gourdji et al., 2022). In general, the WRF-GHG model demonstrated proficiency in simulating seasonal variations, including
407 XCO_2, with profiles akin to OCO-2 data. This study underscores the imperative for further investigations and applications of
408 the WRF-GHG model in uncharted regions such as the MASP, showcasing its prowess in simulating meteorological fields and
409 CO_2 observations.

Appendix A: Scatter plots and Time series
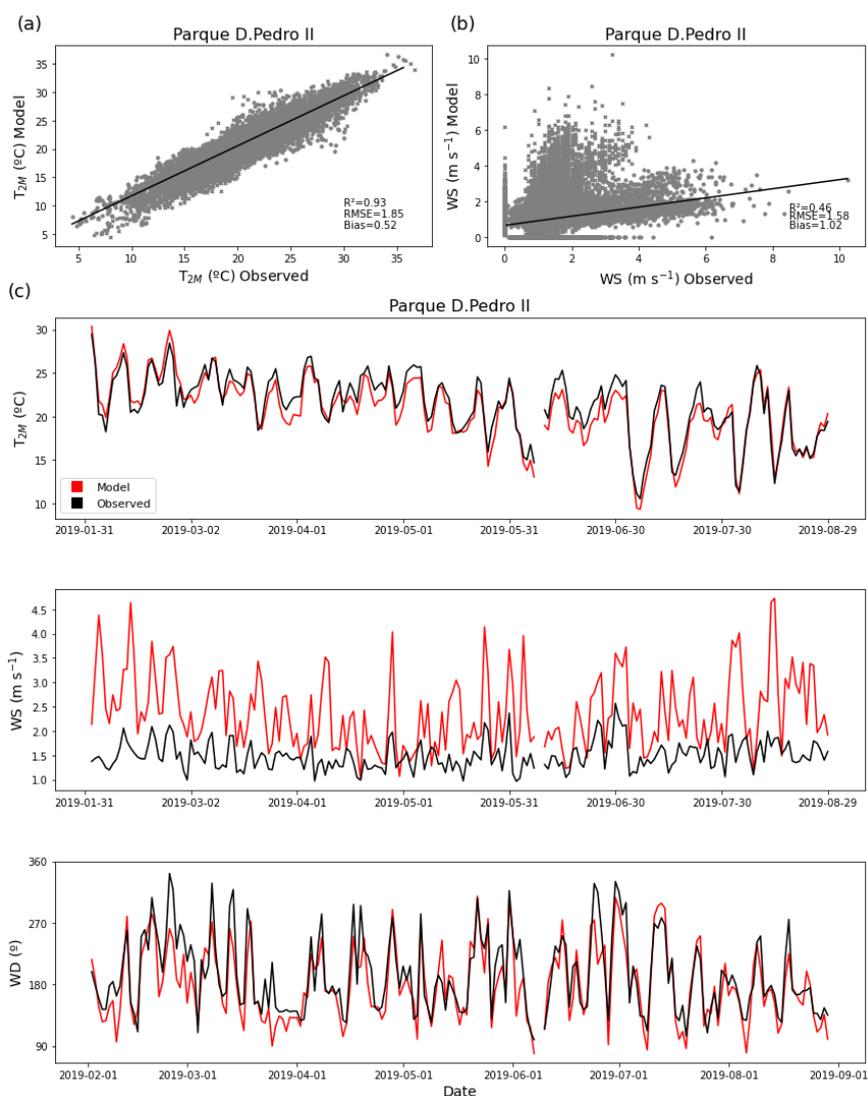
Figure A1. The panels in a) show the scatter plots of hourly measurements of 2 m air temperature (T2m) and 10 m wind speed (WS) compared to observed data from the Parque D.Pedro II. The figure illustrates the relationship between modeled and observed data. The panels in b) show the daily averages from February to August 2019 of 2 m air temperature (T2m), 10 m wind speed (WS), and wind direction (WD). Black line represents the observed data and red line represents the model simulation.
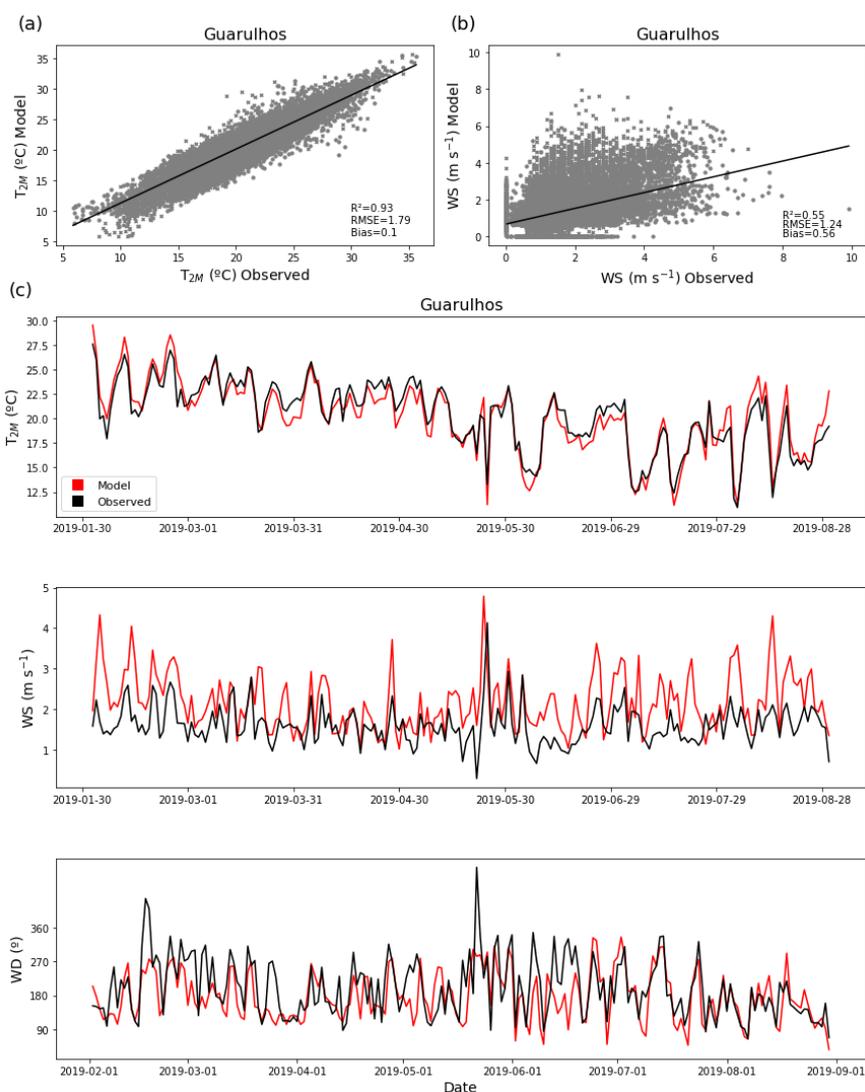
Figure A2. The panels in a) show the scatter plots of hourly measurements of 2 m air temperature (T2m) and 10 m wind speed (WS) compared to observed data from the Guarulhos. The figure illustrates the relationship between modeled and observed data. The panels in b) show the daily averages from February to August 2019 of 2 m air temperature (T2m), 10 m wind speed (WS), and wind direction (WD). Black line represents the observed data and red line represents the model simulation.
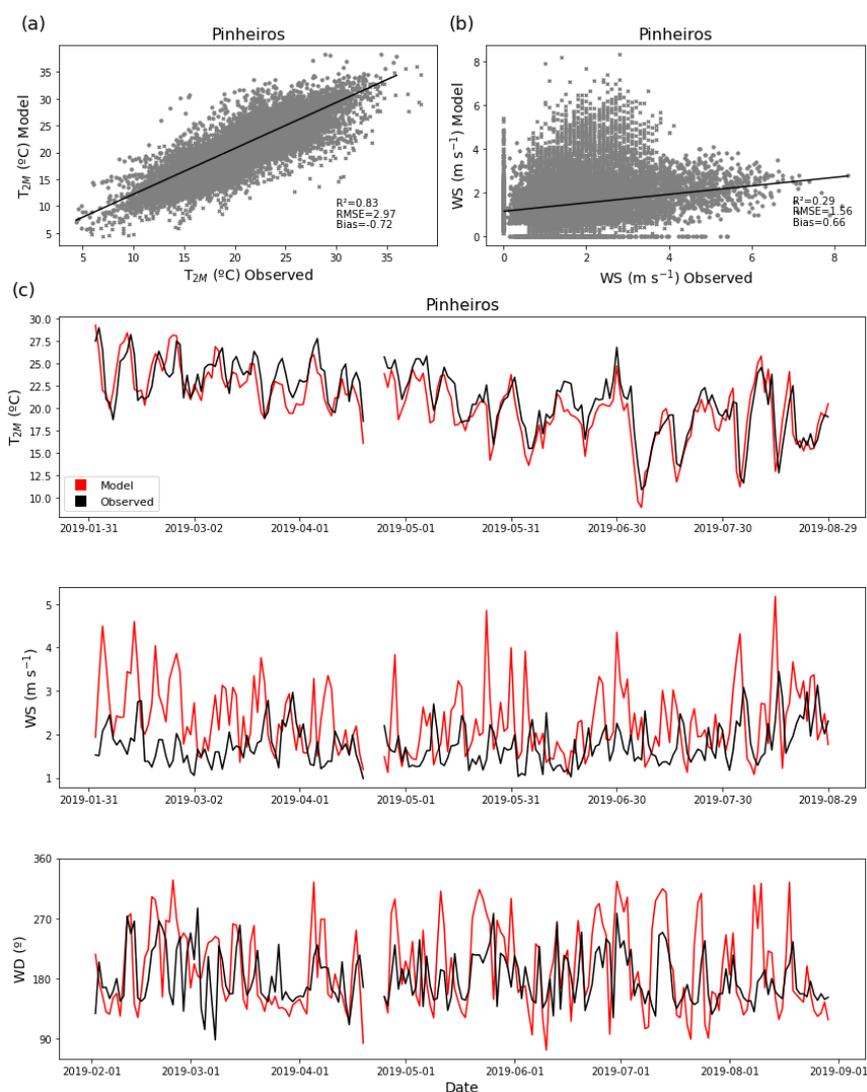
Figure A3. The panels in a) show the scatter plots of hourly measurements of 2 m air temperature (T2m) and 10 m wind speed (WS) compared to observed data from the Pinheiros. The figure illustrates the relationship between modeled and observed data. The panels in b) show the daily averages from February to August 2019 of 2 m air temperature (T2m), 10 m wind speed (WS), and wind direction (WD). Black line represents the observed data and red line represents the model simulation.
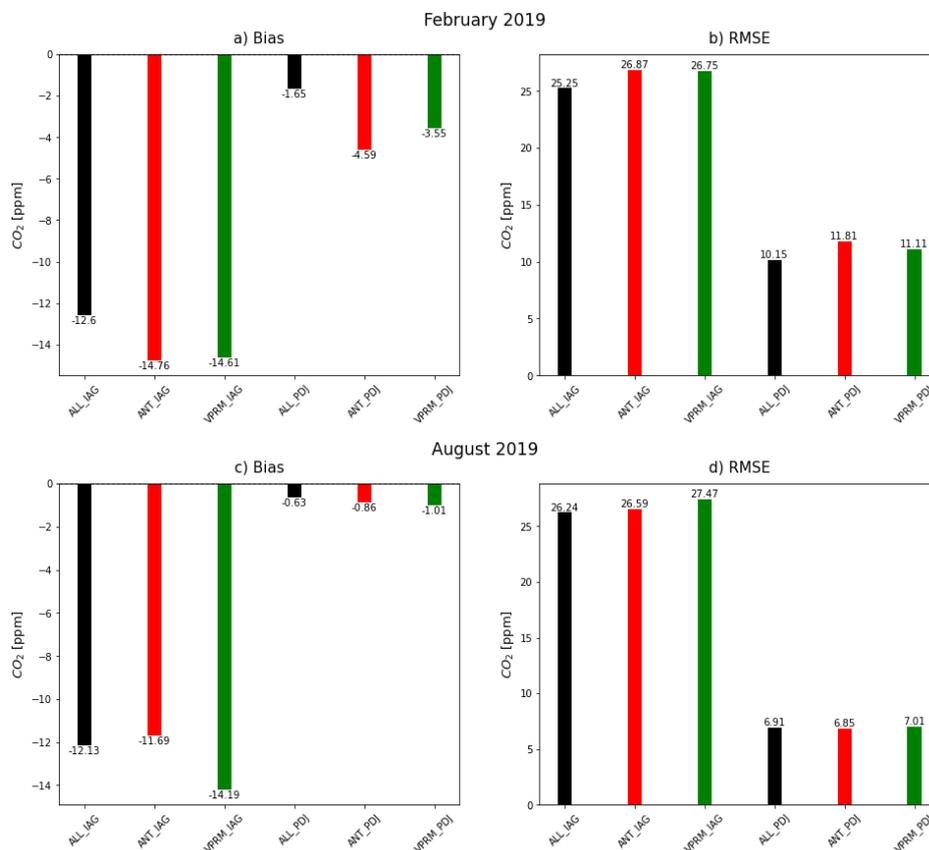
Figure A4. Bias (ppm) and RMSE (ppm) for each simulation at the surface CO_2 observation sites. Panels (a) and (b) represent the simulations for February, while panels (c) and (d) represent the simulations for August (ALL_*: black, ANT_*: red, VPRM_*: green) *Represents the observation sites, e.g. IAG and PDJ.
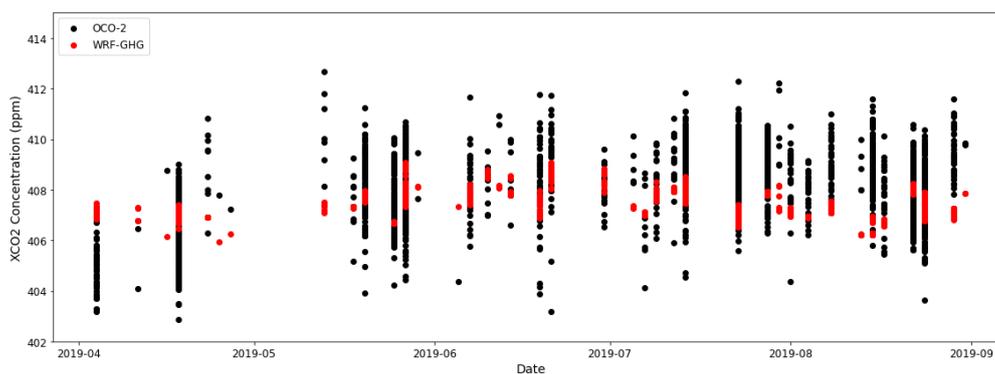
Figure A5. Time series of smoothed column concentrations observed (black) and modeled (red).

References

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.: Air quality forecasting system for Southeastern Brazil, Frontiers in environmental Science, 3, 9, 2015.

Beck, V., Koch, T., and Kretschmer, R.: WRF Greenhouse Gas Model (WRF-GHG) The, Technical Reports, 2011, 2011.

Beck, V., Gerbig, C., Koch, T., Bela, M. M., Longo, K. M., Freitas, S. R. d., Kaplan, J. O., Prigent, C., Bergamaschi, P., and Heimann, M.: WRF-Chem simulations in the Amazon region during wet and dry season transitions: evaluation of methane models and wetland inundation maps, Atmospheric Chemistry and Physics, 13, 7961–7982, 2013.

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 quality simulation with WRF-Chem over southeastern Brazil, part I: Model description and evaluation using ground-based and satellite data, Urban Climate, 52, 101 703, 2023.

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 CO_2 record at the Amazon Tall Tower Observatory: A new opportunity to study processes on seasonal and inter-annual scales, Global Change Biology, 28, 588–611, 2022.

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, H. R., and Rossi, P.: The sustainability of a sugarcane plantation in Brazil assessed by the eddy covariance fluxes of greenhouse gases, Agricultural and Forest Meteorology, 282, 107 864, 2020.

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: A Policy to Embrace Urban Environmental Services and Green Infrastructure Inequalities in the Global South, Frontiers in Sustainable Cities, 3, 685 875, 2021.

CETESB: Relatório de Qualidade do Ar no Estado de São Paulo, Tech. rep., CETESB, São Paulo, Brasil, 2019.

Change, I. C. et al.: Impacts, adaptation and vulnerability, Part A: global and sectoral aspects. Contribution of working group II to the fifth assessment report of the intergovernmental Panel on Climate Change, 1132, 2014.

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 CO_2 emissions from Beijing using differential column measurements, Environmental Research Letters, 17, 075 001, 2022.

Connor, B. J., Boesch, H., Toon, G., Sen, B., Miller, C., and Crisp, D.: Orbiting Carbon Observatory: Inverse method and prospective error analysis, Journal of Geophysical Research: Atmospheres, 113, 2008.

Crippa, M., Guizzardi, D., Muntean, M., Schaaf, E., Lo Vullo, E., Solazzo, E., Monforti-Ferrario, F., Olivier, J., and Vignati, E.: EDGAR v6. 0 greenhouse gas emissions. European Commission, Joint Research Centre (JRC), 2021.

Crisp, D.: Measuring atmospheric carbon dioxide from space with the Orbiting Carbon Observatory-2 (OCO-2), in: Earth observing systems xx, vol. 9607, p. 960702, SPIE, 2015.

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 variability of observed and simulated gross primary productivity, modulated by vegetation state and hydrometeorological drivers, Biogeosciences, 20, 4795–4818, 2023.

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.: Toward reduced transport errors in a high resolution urban CO_2 inversion system, Elem Sci Anth, 5, 20, 2017.

460   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
461   Angeles megacity: a high-resolution land–atmosphere modelling system for urban CO 2 emissions, Atmospheric Chemistry and Physics,
462   16, 9019–9045, 2016.

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

465   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
466   Bell, M. L.: Air quality modeling in the metropolitan area of São Paulo, Brazil: A review, Atmospheric Environment, p. 120301, 2023.

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

469   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,
470   J. R.: A modified Vegetation Photosynthesis and Respiration Model (VPRM) for the eastern USA and Canada, evaluated with comparison
471   to atmospheric observations and other biospheric models, Journal of Geophysical Research: Biogeosciences, 127, e2021JG006 290, 2022.

472   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
473   monitoring and modeling of disturbances in tropical moist forests, Frontiers in Remote Sensing, 5, 1332 728, 2024.

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

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

477   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
478   gases: Calculations with the AER radiative transfer models, Journal of Geophysical Research: Atmospheres, 113, 2008.

479   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
480   vehicular emissions inventories, Geoscientific Model Development, 11, 2209–2229, 2018.

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

482   Kaiser, W., Zhuravlev, R., Ganshin, A., Valsala, V. K., Andrews, A., Chmura, L., Dlugokencky, E., Haszpra, L., Langenfelds10, R. L.,
483   Machida, T., et al.: A high-resolution inverse modelling technique for estimating surface CO2 fluxes based on the NIES-TM-FLEXPART
484   coupled transport model and its adjoint.

485   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
486   satellite-based biosphere parameterization for net ecosystem CO2 exchange: Vegetation Photosynthesis and Respiration Model (VPRM),
487   Global Biogeochemical Cycles, 22, 2008.

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

490   Nerobelov, G., Timofeyev, Y., Smyshlyaev, S., Foka, S., Mammarella, I., and Virolainen, Y.: Validation of WRF-Chem model and CAMS
491   performance in estimating near-surface atmospheric CO2 mixing ratio in the area of Saint Petersburg (Russia), Atmosphere, 12, 387,
492   2021.

493   Nogueira, T., Kamigauti, L. Y., Pereira, G. M., Gavidia-Calderon, M. E., Ibarra-Espinosa, S., Oliveira, G. L. d., Miranda, R. M. d., Vascon-
494   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
495   use: results of tunnel measurements in São Paulo, Brazil, Environmental Science & Technology, 55, 6677–6687, 2021.

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 ACOS CO 2 retrieval algorithm–Part 1: Description and validation against synthetic observations, Atmospheric Measurement Techniques, 5, 99–121, 2012.

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

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

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

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 exchange over a woodland savanna (Cerrado Sensu stricto) in southeast Brasil, Biota Neotropica, 2, 1–11, 2002.

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

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

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

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 settlements, infrastructure, and spatial planning, 2014.

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, Sola, 7, 109–112, 2011.

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 description of the advanced research WRF version 4, NCAR tech. note ncar/tn-556+ str, 145, 2019.

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., Landulfo, E., and Andrade, M. d. F.: Impact of extreme wildfires from the Brazilian Forests and sugarcane burning on the air quality of the biggest megacity on South America, Science of the Total Environment, 888, 163 439, 2023.

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., Souza-Filho, P. W., et al.: Reconstructing three decades of land use and land cover changes in brazilian biomes with landsat archive and earth engine, Remote Sensing, 12, 2735, 2020.

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

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 in the Sao Paulo Metropolitan Area: a numerical study with the WRF-Chem model, Atmospheric Chemistry and Physics, 16, 777–797, 2016.

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, E.: Modeling of atmospheric aerosol properties in the São Paulo metropolitan area: impact of biomass burning, Journal of Geophysical Research: Atmospheres, 123, 9935–9956, 2018.

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 framework for Amazon forest fire smoke dispersion over South America, Bulletin of the American Meteorological Society, 102, E1700–E1713, 2021.

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

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 of global collective climate mitigation actions, Environmental Research Letters, 19, 084 029, 2024.

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 natural surfaces, Science of the Total Environment, 895, 165 115, 2023.

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