Universidade de São Paulo
Instituto de Astronomia, Geofísica e Ciências Atmosféricas
Departamento de Astronomia

Noelia Rojas Benavente

Analysis of the Spatial and Temporal Variability of Gases and Particulate Material Concentrations in the São Paulo Metropolitan Region and its Transport to other regions using Satellite Images and Numerical Modeling

São Paulo
2022

Noelia Rojas Benavente

Analysis of the Spatial and Temporal Variability of Gases and Particulate Material Concentrations in the São Paulo Metropolitan Region and its Transport to other regions using Satellite Images and Numerical Modeling

Thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Sciences in the Institute of Astronomy, Geophysics and Atmospheric Sciences

Major field: Atmospheric Sciences
Advisor: Prof. Dr. Marcia Akemi Yamasoe
Co-Advisor: Prof. Dr. Maria de Fatima Andrade

Corrected version. The original is available in the Unit

São Paulo

2022

This research I dedicate to my parents Joel and Rosa, who have given me their support and unconditional love, and to be the archetypes of my life, learning that everything that one sets out is achieved with effort and dedication. Brothers and brothers in-law, for theirs words of encouragement and support to complete my studies, to my grandmother Fara Alvarado for teaching me to have strength and courage. To my uncles, cousins and nephews for the great love they have for me and for understanding my priorities. And last but not least to my friends Angel and Janaina for the trust, motivation, unconditional support and for the teachings they gave me both in my life and in this research work.

Noelia Rojas Benavente

Acknowledgements

I thank Instituto de Astronomia, Geofísica e Ciências Atmosféricas da Universidade de São Paulo (IAG-USP) for the opportunity as a student in the graduate program in Meteorology. I am grateful for the financial support of Coordenacao de Aperfeicoamento de Pessoal de Nivel Superior (CAPES), for this doctoral project (Process: 05/56487-9).
A special thanks to my supervisors Marcia Akemi Yamasoe and Maria de Fatima Andrade, for giving me the opportunity to show my ability and for passing on much of their knowledge and enthusiasm. Also, thank you for your trust and respect. To all the professors of the Department of Atmospheric Sciences who contributed to the strengthening of my knowledge.
I also have special thanks to Janaina and Angel , colleagues, collab-orators and friends, with whom important part of this work was developed and who had the patience to have a look at this thesis. I am very grateful to all my colleagues at IAG-USP, with whom I had the opportunity to learn so much during these years.

Esta tese foi escrita em LaTeX com a classe IAGTESE, para teses e dissertações do IAG.

Resumo

A variabilidade espacial e temporal das concentrações de $PM_{2.5}$, NO_x (NO + NO_2), CO e O_3 e valores de AOD sobre o sudeste do Brasil foi analisada, utilizando dados medidos próximos da superfície e integrados na coluna atmosférica e simulações do modelo WRF-Chem versão 3.9.1. Sessenta estações de qualidade do ar, quatro estações AERONET, três produtos de satélite e quatro diferentes períodos de simulações (Junho de 2017, Julho de 2018, e Junho e Agosto de 2019) foram considerados para caracterizar a qualidade do ar nesta região. As concentrações dos poluentes atmosféricos de uma estação móvel foram também incluídas para avaliar o impacto do transporte regional dos poluentes atmosféricos gerados na Região Metropolitana de São Paulo (MASP) para áreas vizinhas. As emissões de entrada compreendem emissões veiculares derivadas de um modelo de emissão bottom-up e cálculos on-line de taxas de emissões biogênicas e de queimadas. Para determinar uma configuração do modelo que seja adequada para representar a fotoquímica atmosférica, primeiro avaliamos a capacidade do modelo de reproduzir as condições meteorológicas. Foi encontrada uma boa concordância entre as simulações do modelo e as observações para temperatura (T_{2m}) e umidade relativa (RH_{2m}) a dois metros de altura, com R $>$ 0.85 na maioria dos períodos. Subestimativas ocorreram para a precipitação acumulada diária devido às limitações do esquema de microfísica de nuvens ou parametrização de cumulus. O modelo apresentou deficiências para representar bem a precipitação local, embora tenha sido capaz de reproduzir padrões de precipitação em escala regional. As concentrações de $PM_{2.5}$, NO_x, CO e O_3 simuladas concordaram bem com as variações e tendências temporais observadas, com discrepâncias entre o modelo e as observações devido principalmente às incertezas nos inventários de emissões. Além disso, as simulações de NO_2, CO e AOD foram comparadas com os produtos de satélite do OMI, MOPPIT e

MODIS, respectivamente, obtendo, em geral, coeficientes de correlação entre moderados e altos ($0.56 < R < 0.82$). As diferenças entre os dados simulados e do satélite estão relacionadas, principalmente, às incertezas dos inventários de emissões e simplificações do modelo de aerossol (para a análise de AOD). Por outro lado, a evolução temporal de CO e $PM_{2.5}$ mostrou uma forte dependência da velocidade e direção do vento em superfície ao longo dos períodos de estudo. Os padrões diurnos característicos de O_3 observados sobre áreas urbanas, mas principalmente sobre áreas rurais, onde os níveis de O_3 foram ainda maiores devido às contribuições dos precursores de O_3 mediante o transporte regional da MASP para Botucatu, foram bem representadas pelo modelo. Tanto a estação da qualidade do ar como o modelo WRF-Chem mostraram altas concentrações de O_3 sobre Botucatu (> 115 μgm^{-3}), superior ao registrado na MASP (≈ 90 μgm^{-3}) e chegando a níveis de O_3 maiores do que o habitual. Essas altas concentrações de O_3 medidas na estação móvel indicaram a necessidade de um monitoramento mais ativo da qualidade do ar nas regiões do interior do sudeste do Brasil, onde vivem milhões de pessoas. Este trabalho representa um primeiro esforço combinando diferentes ferramentas: estações em superfície, sensoriamento remoto e modelagem numérica, para melhorar a compreensão de como as emissões de poluentes impactam a qualidade do ar na MASP e áreas adjacentes.

Abstract

We analyzed the spatial and temporal variability of $PM_{2.5}$, NO_x (NO + NO_2), CO and O_3 concentrations, and AOD values over southeastern Brazil, using ground-based and atmospheric column data, simulations of the WRF-Chem model version 3.9.1. Sixty air quality stations, four AERONET stations, three satellite products and four different simulation periods (June 2017, July 2018, and June and August 2019) were considered to characterize the air quality over this region. Air pollutant concentrations from a mobile station were also included to evaluate the impact of the regional transport of air pollutants generated at the Metropolitan Area of São Paulo (MASP) to neighboring areas. Input emissions comprise vehicular emissions derived from a bottom-up emission model, and on-line calculations of biogenic and fire emission rates. To determine a model configuration that is suitable for representing atmospheric photo-chemistry, first we analyzed the model's ability to reproduce the meteorological conditions. Good agreement between model simulations and observations for temperature (T_{2m}) and relative humidity (RH_{2m}) at two meters height was found, with R > 0.85 in most periods. Underestimations occurred for the daily accumulated precipitation due to the limitations of the cloud microphysics scheme or cumulus parameterization. The model showed deficiencies to represent well the local precipitation, although it was able to reproduce precipitation patterns on a regional scale. Model predictions of $PM_{2.5}$, NO_x, CO and O_3 agreed well with observed temporal variations and trends, with model-observation discrepancies, mainly due to uncertainties in the emission inventories. Also, simulations of NO_2, CO and AOD were compared with the satellite products of OMI, MOPITT and MODIS, respectively, obtaining, in general, moderate to high correlation coefficients ($0.56 < R < 0.82$). The differences between the model and the satellite are related to uncertainties of the emission inventories and aero-

sol model simplifications (for AOD analysis). On the other hand, the temporal evolution of CO and $PM_{2.5}$ showed a strong dependence on wind speed (WS_{10m}) and direction (WD_{10m}) measured at 10 meters height throughout the study periods. The characteristic O_3 diurnal patterns observed over urban areas, but mainly over rural, where the O_3 levels were even higher due to the contributions of O_3 precursors from regional transport from MASP and upwind regions in between, were well represented by the model. Both the air quality station and the WRF-Chem model showed high O_3 concentrations in Botucatu ($>$ 115 μgm^{-3}), higher than that registered in the MASP ($\approx$ 90 μgm^{-3}) and reaching higher O_3 levels than the habitual. These high O_3 concentrations measured at the mobile station indicate the need for more active air quality monitoring over inland regions in southeastern Brazil where millions of people live in. This work represents a first effort combining different tools: ground-based observations, remote sensing and numerical modeling,to improve understanding of how pollutants emissions impact the air quality in the MASP and surrounding areas.

List of Figures

2.1 Relative emissions by type of sources in the MASP (Adapted from (CETESB, 2021)). 31

2.2 Graphic depiction of radiation flows between the Earth's surface and the atmosphere. 37

2.3 Transmission of shortwave solar and long wavelength radiation from the Earth's surface through the atmosphere.(Adapted from (Godish, 2004)) . . 38

2.4 Observation geometry of nadir and backward modes 40

3.1 Default (left side) terrain height (m) and (right side) land use categories in the WRF-Chem model, both over the State of São Paulo ($20°S$ - $27°S$ latitude and $51°W$ - $38°W$ longitude). Colored dots in both images represent the CETESB air quality stations distributed over the State of São Paulo. (a) and (b) represent the zoom of the black frame within the study domain, showing the MASP. 43

3.2 Temporal distribution of NO, NO_2, CO and Aldehyde emission in $mol\ km^{-2}hr^{-1}$, representing light and heavy vehicular traffic. 46

3.3 Spatial distribution of (left side) NO and (right side) $PM_{2.5}$ emission in $mol\ km^{-2}hr^{-1}$ and $\mu gm^{-3}\ ms^{-1}$, respectively, representing the hourly vehicular traffic. 47

3.4 Spatial distribution of (Top-left) annual mean emissions of isoprene and monthly leaf area index in (Top-right) June, (Bottom-left) July and (Top-left) August, respectively, all for 2000, over the modeled domain and calculated by the MEGAN model. 48

3.5 Spatial distribution of CO emissions from cumulative burning events, estimated by the FINNv1 model, throughout (Top-left) June 2017, (Top-right) July 2018, (Bottom-left) June 2019 and (Bottom-right) August 2019 and over our study domain. 50
3.6 Passages of the Terra and Aqua satellites (from 2014 to 2019) close to the AERONET stations (distance less than 1km, 3km and 10km) distributed (numbers in red) over the study domain (D1). 55

4.1 Flowchart representing the basic steps for running WRF-Chem model. Regional (MEGAN and FINN) and local (LAPAt) emissions are considered to include the chemical and aerosol modules in the WRF modeling. 63
4.2 (a) AOD values from MODIS sensor at 10km spatial resolution (MOD-L2), (b)remapping of gridded MODIS data at 9km spatial resolution and (c) AOD values obtained by WRF-chem model at 9km spatial resolution. . . . 67
4.3 (a) NO_2 concentrations by the OMI instrument at 13x24 km^2 of spatial resolution, (b) remapping of satellite gridded data at 9x9 km^2 of spatial resolution and NO_2 concentrations obtained by the WRF-Chem model at 9x9 km^2 of spatial resolution. 69
4.4 The sketch of the interpolation-remaping method.(Adapted from (Tao et al., 2014)) 70

5.1 Scatter plots of NME versus NMB, representing the performance of the MODIS products (dots with different colors) for Terra satellite at four AERONET stations (different markers) for the six years of study. Each panel represents the MODIS performance at a specific time window: t $\leq \pm$ 15, 30 and 60 minutes, and a specific spatial window: d $\leq$ 1, 3 and 10 km. 72
5.2 Scatter plots of NME versus NMB, representing the performance of the MODIS products (dots with different colors) for Aqua satellite at four AERONET stations (different markers) for the six years of study. Each panel represents the MODIS performance at a specific time window: t $\leq \pm$ 15, 30 and 60 minutes, and a specific spatial window: d $\leq$ 1, 3 and 10 km. 72

5.3 Color diagram of the NMB and NME of the spatio-temporal windows for the four AERONET stations, for the five satellite aerosol products and for each season (DJF, MAM, JJA and SON). 74
5.4 (Left side) Temporal variability (2014-2019) and (Right side) the correlation of the AOD values at 550nm obtained by the AERONET station (black line) and MODIS instrument (red line), using MAIAC (yellow dots), DT-3K (blue dots), DB-L2 (gray dots), DT-DB-L2 (green dots) and DT-L2 (magenta dots) algorithms at the São Paulo station. 75
5.5 (Left panel) The observed and simulated T_{2m} temporal variability and (right panel) its correlation in the four categories of land use: Regional urban, urban park, urban and industrial for July 2018. The measured and simulated results are represented by the black and red line, respectively. 79
5.6 (Left panel) The observed and simulated T_{2m} temporal variability and (right panel) its correlation in the IAG-USP station for July 2018. The measured and simulated results are represented by the black and red line, respectively. 80
5.7 (Left panel) The observed and simulated RH_{2m} temporal variability and (right panel) its correlation in the four categories of land use: Regional urban, urban park, urban and industrial for July 2018. The measured and simulated results are represented by the black and red line, respectively. . . 82
5.8 (Left panel) The observed and simulated RH_{2m} temporal variability and (right panel) its correlation in the IAG-USP station for July 2018. The measured and simulated results are represented by the black and red line, respectively. 82
5.9 (Left panel) The observed and simulated WS_{10m} temporal variability and (right panel) its correlation in the four categories of land use: Regional urban, urban park, urban and industrial for July 2018. The measured and simulated results are represented by the black and red line, respectively. . . 85
5.10 (Left panel) The observed and simulated WS_{10m} temporal variability and (right panel) its correlation in the IAG-USP station for July 2018. The measured and simulated results are represented by the black and red line, respectively. 85

5.11 (Left panel) The observed and simulated WD_{10m} temporal variability in the four categories of land use: Regional urban, urban park, urban and industrial for July 2018. The measured and simulated results are represented by the black and red line, respectively. (Right panel) Wind rose plot of observed and simulated wind over the State of São Paulo. 87
5.12 Time series of accumulated precipitation (mm) averaged (left panel) over the Regional urban, urban Park, urban and industrial regions and (right panel) over the IAG-USP station for July 2018, for the WRF-Chem model and the IMERG Final Run V.06 product. 88
5.13 Spatial distribution of daily accumulated precipitation (mm) from WRF-Chem model (left panels) and IMERG Final Run V.06 product (middle panels) over the southeastern Brazil for July 2018. The black square represents the zoom of the MASP. The gray dots represent the AQS distributed over our study area and the magenta star represents the location of the IAG-USP station. 89
5.14 Time series and scatter plot of the observed and simulated NO_x concentrations in the four categories of land use: Regional urban, urban park, urban and industrial for July 2018. The measured and simulated results are represented by the black and red line, respectively. 91
5.15 Simulated and observed hourly mean NO_x and O_3 concentrations in the four categories of land use: Regional urban, urban Park, urban and industrial over the State of São Paulo for July 2018. The measured and simulated results are represented by the black and red line, respectively. 92
5.16 Time series and the correlation of the observed and simulated O_3 concentrations in the four categories of land use: Regional urban, urban park, urban and industrial, for July 2018. The measured and simulated results are represented by the black and red line, respectively. 93
5.17 Time series and scatter plot of the observed and simulated CO concentrations in the four categories of land use: Regional urban, urban park, urban and industrial, during July 2018. The measured and simulated results are represented by the black and red line, respectively. 95

5.18 Time series and scatter plot of the observed and simulated $PM_{2.5}$ concentrations in the four categories of land use: Regional urban, urban park, urban and industrial for July 2018. The measured and simulated results are represented by the black and red line, respectively. 97
5.19 Simulated and observed hourly mean CO and $PM_{2.5}$ concentrations in the four categories of land use: Regional urban, urban park, urban and industrial over the State of São Paulo for July 2018. The measured and simulated results are represented by the black and red line, respectively. 98
5.20 Simulated (first column) and observed (second column) mean spatial distribution of tropospheric NO_2 VCD, total CO VCD and AOD values over southeastern Brazil for June 2017. The third and fourth columns represent the relative difference and correlation between both tools. The blue lines, in the first three columns, represent Botucatu city and the MASP and the gray space indicates missing data. 100
5.21 Simulated (first column) and observed (second column) monthly mean spatial distribution of tropospheric NO_2 VCD, total CO VCD and AOD values over southeastern Brazil for July 2018. The third and fourth columns represent the relative difference and respective scatter plot. The blue lines, in the first three columns, represent Botucatu city and the MASP and the gray space indicates missing data. 101
5.22 Simulated (first column) and observed (second column) mean spatial distribution of tropospheric NO_2 VCD, total CO VCD and AOD values over southeastern Brazil for June 2019. The third and fourth columns represent the relative difference and correlation between both tools. The blue lines, in the first three columns, represent Botucatu city and the MASP and the gray space indicates missing data. 102
5.23 Simulated (first column) and observed (second column) monthly mean spatial distribution of tropospheric NO_2 VCD and AOD values over southeastern Brazil for August 2019. The third and fourth columns represent the relative difference and correlation between both tools. The blue lines, in the first three columns, represent Botucatu city and the MASP and the gray space indicates missing data. 103

5.24 Box plot of the WRF-Chem, MODIS, and AERONET AOD values at 550 nm over the São Paulo station for July 2018. 104
5.25 Overlay of observed and predicted O_3 (in μg m^{-3}) (left panels), NO_x (in ppb) (middle panels) and HCHO (in μg m^{-3}) (right panels) spatial distributions during daytime for July 17, 2018. The black arrows in each panel represent the wind direction at 10 m. The observed values are indicated by the colored star and circles. The blue lines represent Botucatu city and the MASP. 106
5.26 Overlay of observed and predicted O_3 (in μg m^{-3}) (left panels), NO_x (in ppb) (middle panels) and HCHO (in μg m^{-3}) (right panels) spatial distributions during nighttime for July 17, 2018. The black arrows in each panel represent the wind direction at 10 m. The observed values are indicated by the colored star and circles. The blue lines represent Botucatu city and the MASP. 107
5.27 Overlay of observed and predicted CO (in ppmv) (left panels), $PM_{2.5}$ (in μg m^{-3}) (middle panels) and AOD (right panels) spatial distributions during daytime for July 17, 2018. The black arrows in each panel represent the wind direction at 10 m. The observed values are indicated by the colored star and circles. The blue lines represent Botucatu city and the MASP. . . 108
5.28 Overlay of observed and predicted CO (in ppmv) (left panels), $PM_{2.5}$ (in μg m^{-3}) (middle panels) and AOD (right panels) spatial distributions during nighttime for July 17, 2018. The black arrows in each panel represent the wind direction at 10 m. The observed values are indicated by the colored star and circles. The blue lines represent Botucatu city and the MASP. . . 109
5.29 Spatial distribution of model-simulated and tropospheric NO_2 VCD, Terra- and Aqua-retrieved AOD at 550 nm integrated into the atmosphere for July 17, 2018. 110

5.30 Vertical cross-sections (altitude–downwind distances) of simulated O_3 concentrations over the downwind area at 9:00, 13:00, 15:00 19:00 e 21:00 LT of July 17, 2018. The yellow line drawn in the first panel (up-right side) of the Figure represents the downwind area and the dashed gray lines in the other panels represent the planetary boundary layer (PBL).The black background of each panel represents the topography of the surface. 112
5.31 Vertical cross-sections (altitude–downwind distances) of simulated O_3 concentrations over the downwind area at 9:00, 13:00, 15:00 19:00 e 21:00 LT of June 17, 2019. The yellow line drawn in the first panel (up-right side) of the Figure represents the downwind area and the dashed gray lines in the other panels represent the planetary boundary layer (PBL).The black background of each panel represents the topography of the surface. 113
5.32 Vertical cross-sections (altitude–downwind distances) of simulated O_3 concentrations over the downwind area at 9:00, 13:00, 15:00 19:00 e 21:00 LT of August 11, 2019. The yellow line drawn in the first panel (up-right side) of the Figure represents the downwind area and the dashed gray lines in the other panels represent the planetary boundary layer (PBL).The black background of each panel represents the topography of the surface. 114

A.1 Vertical coordinate system η of the WRF model.(Adapted from (Skamarock et al., 2008)) 145
A.2 Horizontal (left) and vertical (right) grid of Arakawa type C.(Adapted from (Skamarock et al., 2008)) 146
A.3 Various nest configurations for multiple grids. (a) Telescoping nests and (b) nests at the same level with respect to a parent grid. (Adapted from (Skamarock et al., 2008)) 147
A.4 Interactions between physical parameterizations in WRF model. 150

B.1 Overlay of observed and predicted O_3 (in μg m^{-3}) (left panels), NO_x (in ppb) (middle panels) and HCHO (in μg m^{-3}) (right panels) spatial distributions during daytime for June 17, 2019. The observed values are indicated by the colored star and circles. The blue lines represent Botucatu city and the MASP. 151

B.2 Overlay of observed and predicted O_3 (in μg m^{-3}) (left panels), NO_x (in ppb) (middle panels) and HCHO (in μg m^{-3}) (right panels) spatial distributions during nighttime for June 17, 2019. The observed values are indicated by the colored star and circles. The blue lines represent Botucatu city and the MASP. 152
B.3 Overlay of observed and predicted CO (in ppmv) (left panels), $PM_{2.5}$ (in μg m^{-3}) (middle panels) and AOD (right panels) spatial distributions during daytime for June 17, 2019. The observed values are indicated by the colored star and circles. The blue lines represent Botucatu city and the MASP. . . 153
B.4 Overlay of observed and predicted CO (in ppmv) (left panels), $PM_{2.5}$ (in μg m^{-3}) (middle panels) spatial distributions during nighttime for June 17, 2019. The observed values are indicated by the colored star and circles. The blue lines represent Botucatu city and the MASP. 154
B.5 Spatial distribution of model-simulated and OMI-retrieved tropospheric NO_2 VCD, Terra- and Aqua-retrieved AOD at 550 nm integrated into the atmosphere for June 17, 2019. 155
B.6 Overlay of observed and predicted O_3 (in μg m^{-3}) (left panels), NO_x (in ppb) (middle panels) and HCHO (in μg m^{-3}) (right panels) spatial distributions during daytime for August 11, 2019. The observed values are indicated by the colored star and circles. The blue lines represent Botucatu city and the MASP. 156
B.7 Overlay of observed and predicted O_3 (in μg m^{-3}) (left panels), NO_x (in ppb) (middle panels) and HCHO (in μg m^{-3}) (right panels) spatial distributions during nighttime for August 11, 2019. The observed values are indicated by the colored star and circles. The blue lines represent Botucatu city and the MASP. 157
B.8 Overlay of observed and predicted CO (in ppmv) (left panels), $PM_{2.5}$ (in μg m^{-3}) (middle panels) and AOD (right panels) spatial distributions during daytime for August 11, 2019. The observed values are indicated by the colored star and circles. The blue lines represent Botucatu city and the MASP. 158

B.9 Overlay of observed and predicted CO (in ppmv) (left panels), $PM_{2.5}$ (in μg m^{-3}) (middle panels) and AOD (right panels) spatial distributions during nighttime for August 11, 2019. The observed values are indicated by the colored star and circles. The blue lines represent Botucatu city and the MASP. 159
B.10 Spatial distribution of model-simulated and OMI-retrieved tropospheric NO_2 VCD, Terra- and Aqua-retrieved AOD at 550 nm integrated into the atmosphere for August 11, 2019. 160

List of Tables

3.1 WRF-Chem model configurations 44
3.2 Air Quality Stations distributed over the State of São Paulo 52
3.3 Additional stations on the State of Sao Paulo 53
3.4 AERONET stations over southeastern Brazil 54
3.5 Satellite Configurations 61

4.1 Meteorological parameter benchmarks 66
4.2 Recommended benchmarks for photochemical model performance statistics 66

5.1 General statistics of the validation between the AOD values measured by the AERONET stations and MODIS products, considering the most appropriate spatio-temporal window (t $\leq \pm$ 60 minutes and d $\leq$ 10 km) from 2014 to 2019 (n = Matchup Number) 77
5.2 Performance statistics of the validation between predicted Temperature at 2 meters (T_{2m} in $^{\circ}C$) and measurements by the AQS and the IAG-USP and Botucatu stations for the four periods. 80
5.3 Performance statistics of the validation between predicted Relative Humidity at 2 meters (RH_{2m} in %) and measurements by the AQS and the IAG-USP station for the four periods. 83
5.4 Performance statistics of the validation between predicted Wind Speed at 10 meters (WS_{10m} in ms^{-1}) and measurements by the AQS and the IAG-USP and Botucatu stations for the four periods. 86

5.5 Performance statistics of the validation between predicted Nitrogen Oxides (NO_x in ppb) and measurements by the AQS and Botucatu station for the four periods. 92
5.6 Performance statistics of the validation between predicted Ozone (O_3 in μg m^{-3}) and measurements by the AQS and the IAG-USP and Botucatu stations for the four periods. 94
5.7 Performance statistics of the validation between predicted Carbon Monoxides (CO in ppmv) and measurements by the AQS and the IAG-USP and Botucatu stations for the four periods. 96
5.8 Performance statistics of the validation between predicted daily Particulate Material ($PM_{2.5}$ in μg m^{-3}) and measurements by the AQS and the IAG-USP and Botucatu stations for the four periods. 99
5.9 Performance statistics of the validation between model-simulated and satellite-retrieved data for the four periods. 104

Contents

1. Introduction 23

1.1 Motivation 23

1.2 Objectives 27

1.2.1 Specific Objectives 27

1.3 Chapter organization 27

2. Literature Review 29

2.1 Air pollution in State of São Paulo 29

2.1.1 Emission and transport 30

2.1.1.1 Local emission 30

2.1.1.2 Transport 34

2.1.2 Urban air pollution formation 35

2.2 Interaction between pollutants and radiation 36

2.2.1 Satellite measuring techniques for Air pollutants retrievals 40

3. Modeling framework and datasets 42

3.1 Model 42

3.1.1 WRF-Chem model 42

3.1.2 WRF-Chem model configurations 42

3.1.3 Gas-phase and Aerosol chemical mechanism 45

3.1.4 Emissions inventories 45

3.1.4.1 Vehicular emissions 46

3.1.4.2 Biogenic emissions 47

3.1.4.3 Biomass Burning emissions 49

3.2 Ground-based data . . . 51
3.2.1 CETESB air quality stations . . . 51
3.2.2 IAG-USP meteorological station . . . 53
3.2.3 Botucatu station . . . 53
3.2.4 AERONET Surface Observation Network . . . 54
3.3 Satellite data . . . 55
3.3.1 MODIS sensor . . . 55
3.3.1.1 Aerosol Products . . . 56
3.3.2 OMI . . . 59
3.3.3 MOPITT . . . 59
3.3.4 GPM . . . 60

4. Methodology . . . 62
4.1 Metrics to assess photochemical model performance . . . 62
4.2 Observational data and WRF-Chem model . . . 65
4.3 AERONET network - MODIS and WRF-Chem model . . . 66
4.4 OMI, MOPITT and GPM with WRF-Chem model . . . 68
4.5 Images Remapping . . . 70

5. Results . . . 71
5.1 Comparison between MODIS products and AERONET stations . . . 71
5.2 Validation of WRF-Chem against ground-level and column observations . . . 78
5.2.1 Meteorological parameters . . . 78
5.2.1.1 Temperature . . . 78
5.2.1.2 Humidity . . . 82
5.2.1.3 Wind Direction and Speed . . . 84
5.2.1.4 Precipitation . . . 87
5.2.2 Chemical species . . . 90
5.2.2.1 Comparison with ground-based observations . . . 90
5.2.2.2 Comparison with Satellite retrievals . . . 99
5.3 Pollutant transport from the MASP to other regions . . . 105

6. Conclusions and Future work 115
6.1 General conclusions 115
6.2 Future Works 120

Bibliography 121

Appendix 142

A. A brief description of the atmospheric WRF model components 144
A.1 General mathematical formulation 144
A.2 Geographic allocation, initial and boundary conditions 146
A.3 Physical parameterizations 147

B. Pollutant transport from the MASP to other regions 151
B.1 Episode - 17 June 2019 151
B.2 Episode - 11 August 2019 156

Chapter 1

Introduction

1.1 Motivation

Air pollution continues to be a big problem, with a tendency to worsen in the future, as reported by WHO (2021). The air pollution growth is basically related to the population increase, which mainly occurs in urban and industrial areas (Junyan et al., 2020) through transportation and industry, as the main sources of air pollution (Andrade et al., 2012; Rojas Benavente, 2014; Angelevska et al., 2021). One of the consequences of this air pollution is the gases and particles emitted to the troposphere as also the oxidation products and formation of new particles that represent a direct human health risk for the respiratory system (Saldiva et al., 2002; Arbex et al., 2012; Liu et al., 2019). In developing countries, as Brazil, with 212 million inhabitants, the problem of air pollution can be aggravated by a combination of some key elements, such as intensive use of fossil fuels both in the industry and transport sector and the aging of the vehicle fleet (Alonso et al., 2010; Andrade et al., 2015). The Metropolitan Area of São Paulo (MASP), located in southeastern Brazil, is the largest urban center in the country and has a fleet of over 7 million vehicles (CETESB, 2019). The vehicle fleet is a type of emission source that contributes to the increase in air pollution, mainly in urban areas (Pérez-Martínez et al., 2015).

Since 1960, the MASP and other Brazilian cities began to grow more rapidly than before, generating a disorderly urbanization and an increase in sources of air pollution (Oyama et al., 2016; Andrade et al., 2017). Urban air quality traditionally has been monitored with networks of ground monitoring stations worldwide. In this context, for the purpose of monitoring the air quality in the State of São Paulo, the Environmental Agency of the State of

São Paulo (CETESB) was created in the late 1960s. Over a decade later, an air pollution control plan was designed with the aim of reducing the concentrations of particulates and sulfur dioxide (SO_2). The first action to be carried out was the relocation of industrial plants located within urban centers to the interior of the State of São Paulo or to other Brazilian States (Andrade et al., 2012). Cleaner technologies and production practices were then implemented in the industries to reduce their emissions and keep the related air pollutant concentrations within the primary air quality standards (Andrade et al., 2015). Later, in 1986, the National Council for the Environment (CONAMA) established the Programa de Controle de Poluição do Ar por Veículos Automotores (PRONCOVE), a vehicle emission control strategy for the reduction of emissions from new vehicles, using air quality standards for four main pollutants (suspended particles, carbon monoxide (CO), SO_2 and ozone (O_3)) with levels similar to the primary standards adopted by the US Environmental Protection Agency (https://www.epa.gov/criteria-air-pollutants/naaqs-table). In 1975, the National Ethanol Program (PROALCOOL) was created in order to reduce CO emissions from the vehicle fleet, stimulating the use of ethanol as fuel (Andrade et al., 2017). Despite the efforts made to create several programs with the purpose of reducing both industrial and vehicular emissions, secondary pollutants such as O_3 (Schuch et al., 2019) and fine particulate matter ($PM_{2.5}$) have not yet been controlled (Andrade et al., 2012; Pérez-Martínez et al., 2015).

Even though CETESB air quality monitoring network continuously operates throughout the State of São Paulo, the measurement sites are not spatially well distributed over the region. In fact, there are several urban areas in the State of São Paulo lacking any air pollution monitoring system. In consequence of these facts, the use of numerical atmospheric chemical-transport models has been considered to assess emissions and improve the analysis on the investigation of changes in air quality at discrete points. In addition, satellite remote-sensing technologies also have possible applications in monitoring air quality in rural or remote regions where no ground-based monitoring network exists (Chudnovsky et al., 2014). Although they provide limited quantitative information on air quality compared with ground-based monitoring (Valin et al., 2011), they are useful for air quality research on identifying the location of peak concentrations and determining concentration gradients between surface monitoring stations (Rojas Benavente, 2014; Roy et al., 2017).

Several studies have applied chemical transport models as, for example, the Weather Research and Forecasting with Chemistry (WRF-Chem) model to simulate the air quality over MASP. The large percentage of these studies was focused on gaseous species, such as O_3, nitrogen oxides (NO_x), CO, among others (Silva and Andrade, 2013; Hoshyaripour et al., 2016; Schuch et al., 2018; Albuquerque et al., 2018; Gavidia-Calderón et al., 2018), while there are few studies focused on aerosols (Vara-Vela et al., 2016, 2018) with the aim to validate the efficiency of the WRF-Chem model at a regional and/or local scale. Other studies have shown that regional and local air quality may also strongly depend on emissions at regional and local scales (Schneider et al., 1997; Ponche and Vinuesa, 2005). Therefore, the fine scale emission inventories are very useful for air pollution prediction, improving the primary and secondary pollutants estimation (Taghavi et al., 2005; Vara-Vela et al., 2016; Gavidia-Calderón et al., 2018). The spatial and temporal resolutions of the global inventories are generally low and, consequently, they do not capture the specific dynamic characteristics of each region, mainly with respect to the representation of urban centers (Hoshyaripour et al., 2016; Gavidia-Calderón et al., 2021). Andreão et al. (2020) showed the importance using local anthropogenic emission inventories. The local inventories distinguish the point emission sources, such as highways, industries, etc. On that line, Vara-Vela et al. (2016) evaluated the impact of vehicular emissions on the formation of fine particles ($PM_{2.5}$) over MASP, using the WRF-Chem model with local emissions inventory. Vara-Vela et al. (2016) obtained a good performance of the model and showed that the vehicle emission of primary gases led to the formation of secondary particles. More than 40 % of the $PM_{2.5}$ formation was attributed to the emission of hydrocarbons (mainly aromatic). Gavidia-Calderón et al. (2018) also considered the local emissions inventory in the WRF-Chem model simulations to examine tropospheric ozone formation to and from urban areas. Gavidia-Calderón et al. (2018) evaluated the impact of using static and time-dependent chemical boundary conditions in the ozone simulations over the MASP, finding that the model is more efficient when time-dependent chemical boundary conditions were considered.

Regarding the use of remote sensing to study air quality over the MASP, there are several global observations available for a wide range of atmospheric pollutants. For instance, the AErosol RObotic NETwork station network (AERONET) (Holben et al., 1998) and the MODerate Resolution Imaging Spectroradiometer (MODIS) products (Anderson et al.,

2003) are the most used to study the properties of aerosols. Ozone Monitoring Instrument (OMI) Levelt et al. (2006), Tropospheric Monitoring Instrument (TROPOMI), Measurements of Pollution in the Troposphere (MOPPIT), Atmospheric Infrared Sounder (AIRS) and Orbiting Carbon Observatory 2 (OCO2) are used to study gaseous species (such as O_3, NO_x, CO, $HCHO$) (Fu et al., 2018; Martínez-Alonso et al., 2022; Sekiya et al., 2022). Overall, satellite sensors provide a broad view of urban pollution and help determine the impact on urban air quality by vehicular transport, dust storms and local fires, or transport of pollutants from more distant sources, including those crossing national or regional boundaries. In addition, satellite aerosol products have been widely used to retrieve surface $PM_{2.5}$ concentration due to their coverage and continuous spatial distribution. Finding a good relationship between AOD and PM2.5 will help us understand the $PM_{2.5}$ pollution over a particular region (Wang and Christopher, 2003; Van Donkelaar et al., 2006; Kumar et al., 2007) and over the MASP, only few studies analyzed this relationship. Damascena et al. (2020) explored the relationship between MAIAC AOD values at high resolution and ground based measurements of $PM_{2.5}$ over MASP. Their work showed a weak correlation and according to them, one of the reasons is the aerosol optical properties model used by the MAIAC inversion algorithm. The types of aerosol particles in the MASP are different from what is considered in the algorithm. AERONET stations and the MODIS products present limitations in spatial and temporal coverage, respectively. In cloudy conditions, these platforms exclude all information in the atmospheric column and for that reason, aerosols present below, inside or above clouds are not retrieved (Ichoku et al., 2002; Anderson et al., 2003; Huffman et al., 2015). There are only few studies that combine remote sensing products with model simulations to analyze aerosol formation and transport over the southeastern Brazil, mainly over the MASP (Vara-Vela et al., 2018, 2021). The integrated use of these two tools, and ground stations measurements, should strengthen the scientific understanding of air quality in urban and rural areas. Nevertheless, air quality management is hampered by uncertainties in traditional "bottom-up"and "top-down"emission inventories based on the application of emission factors, biomass combustion coefficient, the diurnal cycle of the total mass of emitted species, total dry matter consumption, combustion completeness, fuel loading, etc, as the formation of O_3 and secondary aerosols depends on a complicated manner upon the sources of their precursors. Satellite remote sensing has been shown as an important tool for help on the reducing uncertainty in the

spatial distribution of these harmful species and the processes affecting them (Chance and Martin, 2017).

1.2 Objectives

The main objective of the present work is to analyze the spatial and temporal variability of gases and particulate material concentrations in the Metropolitan Area of São Paulo (MASP) and its transportation to other regions using satellite images and a numerical atmospheric chemical-transport model.

1.2.1 Specific Objectives

- Evaluate the air pollutants concentrations simulated by the Weather Research Forecasting coupled with Chemistry (WRF-Chem) model in the State of São Paulo.
- Analyze Particulate Material (PM) data from Botucatu and CETESB air quality monitoring stations available in the state of São Paulo.
- Analyze the secondary gases vertical column densities (VCDs) data by using satellite images and numerical modeling.
- Analyze the transport of air pollutant concentrations from the Metropolitan area of São Paulo to other cities around by using modeling, satellite and ground-based data.
- Select transport events to discuss possible impacts of urban pollutants upwind MASP to inland areas.

1.3 Chapter organization

The following is a brief description of the thesis content:

Chapter 1 presents an introduction in which the importance and relevance of the proposed study are discussed and some of the contextualization of the problem to be addressed, as well as the objectives and chapter organization of the work.

Chapter 2 makes a literature review on the subject of the work, showing the theoretical assumptions necessary for a good understanding of the studies that will follow.

Chapter 3 presents the data used, including a more detailed description of the WRF-Chem model (input data and chemical and aerosol mechanism used), and observational and satellite data considered in this work.

Chapter 4 shows the methodology used, a description of the processing of data from the model, satellite and ground based data. In addition, it presents the statistical indices used to analyze the results.

Chapter 5 describes the results obtained and discussions. This chapter has been divided into three parts. The first part presents a spatial and temporal analysis of the AERONET and MODIS data to choose the best product of the Terra and Aqua satellites and then be compared with the WRF-Chem model in the study of aerosols. The second part brings an analysis of the performance of the WRF-Chem model using the physical and chemical parameters, detailed in chapter 3. And finally, the third part shows specific events of pollutant transport that occurred during the studied periods.

Chapter 6 presents the conclusions and future directions.

Chapter 7 shows the bibliographic references used as the basis for the work.

Chapter 2

Literature Review

2.1 Air pollution in State of São Paulo

For 22 years (2000-2021) the air pollution in State of São Paulo has also been analyzed by the Institute for Energy and the Environment (IEMA). Despite that air quality data published by CETESB showed improvements, the air quality indices in the State of São Paulo still do not reach the minimum recommendations of the World Health Organization (WHO) for particulate material (PM), ozone (O_3) and nitrogen dioxide (NO_2); the latter being a problem for the city based on current WHO recommendations (WHO, 2021; IEMA, 2022). Particulate material consists of solid and/or liquid particles ($PM_{2.5}$ fine and PM_{10} coarse). Both PM, O_3 and nitrogen oxides (NO_x - the sum of nitric oxide (NO) and NO_2) are the main atmospheric pollutants due to the historical importance, concentration and overall effects on plants, animals and people (Wang et al., 2004). They are generally believed to be the most hazardous of ambient pollutants, and it has been estimated that reducing ambient air concentrations of PM_{10} from 70 to 20 μgm^{-3} would lower the number of air quality related deaths by approximately 15 % (Liu et al., 2019). However, air pollution has been increasing in major cities, especially those found in developed and developing countries (such as in: Brazil, United states, Russia, India, Indonesia and China) as a result of rapid urbanization (Hertel et al., 2009). The latest International Gas Union (IGU) report shows that 1.6 million deaths in China throughout 2015 was directly caused by $PM_{2.5}$ pollution. Despite the decrease in pollutant levels in Europe since 1950 and the implementation of the first European Commission Directive on ambient air quality in 1980, important disparities in exposure to air pollution still persist today between and within European countries (Pascal et al., 2013). Urban air pollution in Asia is responsible for an

estimated 537,000 premature deaths annually (Gouveia and Fletcher, 2000; WHO, 2002; Martins et al., 2004; Schwela et al., 2012). Overall, exposure to PM2.5 caused about 4.5 million premature deaths worldwide annually (Cohen et al., 2017). Since 2000, in many major urban areas around the world, air pollution is characterized more by the formation of O_3 and other oxidants rather than by oxides of sulfur, particles and sulfuric acid (Finlayson-Pitts and Pitts, 2000). The atmospheric composition is under the influence of the man kind evolution, both in terms of populational growth, as in industrial and agricultural and household development. The energy consumption and industrial evolution are main causes of the man made pollution, causing modification of the air quality (Popescu and Ionel, 2010). In addition, although the nature and characteristics of urban emissions can be variable in time and space, changes in pollutant dispersion pathways both locally and regionally often determine the temporal and spatial patterns of atmospheric composition in urban areas. Also, severe atmospheric particulate material pollution episodes, from anthropogenic origin, are usually presented in cold seasons, not only in urban and industrial areas but also in regional and rural areas (Pey et al., 2010).

2.1.1 Emission and transport

Urban air pollution is the product between emission processes which increase pollutant concentrations, and dispersion, advection and deposition processes that reduce and remove them. The impact on pollution levels of the dispersion and removal processes are governed by the local meteorological conditions, which also vary heavily with the physical location of the city. Air pollution concentrations in an urban environment are naturally the result of local emissions as well as contributions from pollution transport from more remote sources. The location of the city is very important for local dispersion conditions, which are governed by the meteorology and affected by topographical conditions (Hertel et al., 2009).

2.1.1.1 Local emission

By their nature, the air pollution sources can be classified, mainly, by physical, chemical and biological sources, and secondly by natural and anthropogenic sources (Popescu and Ionel, 2010). Natural air pollution are generated during volcanic eruptions, sandstorms (Saharan sand deposits), forest fires, processes related to air-chemistry in thunderstorms,

and also emissions of volatile organic compounds (VOCs) from plants. However, the anthropogenic air pollution is of major concern as there are no borders and, no matter where the pollutants are released into the atmosphere, it will have an impact over the global environment (Schwela et al., 2012). The anthropogenic pollution sources are exhausting diverse species to the atmosphere and are pushing over the limit the natural possibility of the ecosystem to adapt to the caused concentration augmentation (Popescu and Ionel, 2010). These sources, dominated by the combustion of fossil fuels, can be conveniently grouped into a few major sectors (energy, industry, agriculture and waste). The transportation accounts for most of the total pollutants emitted, mainly, in urban areas such as São Paulo. Fuel combustion emissions from electrical utilities, industries, and other categories are other major sources of air pollution emissions

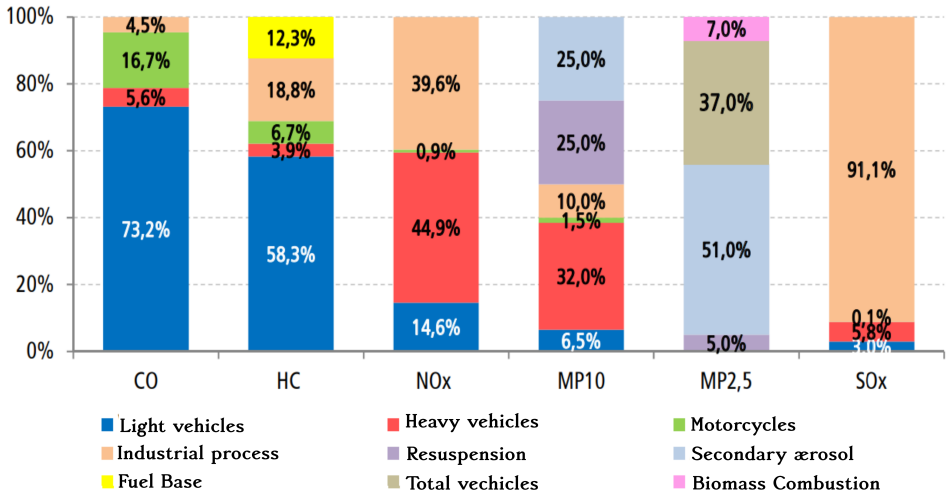

Figure 2.1: Relative emissions by type of sources in the MASP (Adapted from (CETESB, 2021)).

Figure 2.1 shows the contribution of the main sources to the emissions of these main pollutants in the MASP for the year 2021 (CETESB, 2021). Local NO_x emissions in the atmosphere have been decreasing steadily since pre-industrial times (from ∼238 to ∼153 10^3t), motivated by the incorporation of vehicles with new technologies to replace older and more polluting vehicles. ∼44.9 % and ∼14.6 % of the annual NO_x emissions to the atmosphere are produced by heavy and light vehicles that circulate on the MASP tracks. ∼39.6 % is associated with the industrial processes. All NO_x is emitted as NO which is then rapidly oxidized to the more toxic NO_2. Moreover, the main natural and anthropogenic sources of NO_x are lightning, forest fires, bacterial activity in soils, road transport

and public electricity and heat sector (Finlayson-Pitts and Pitts, 2000; Koppmann, 2008; Stocker et al., 2013; CETESB, 2021). In addition to NO_x being a precursor to tropospheric O_3 formation, it also contributes to the atmospheric fine particulate loading as a result of oxidation to form nitrate aerosols (Holman, 1999).

CO is a colorless, tasteless and odorless gas, flammable and dangerous to human health due to its toxicity and for promoting chemical asphyxiation. CO is caused by the incomplete combustion in major urban areas. CETESB (2021) reported that 73.2 % of CO emissions are caused by light vehicles and 16.7 % by motorcycles (Figure 2.1). However, combustion processes related to energy, heating, biomass burning, the oxidation of methane and volatile organic compounds (VOCs) also are main sources to emit CO (Rozante et al., 2017). Highest CO concentrations occur along major traffic lanes during morning and evening rush hours and decrease relatively rapidly with distance from roadways (Schwela et al., 2012). Also, elevated levels of CO tend to pollute cities which are located at relatively high altitudes (Godish, 2004).

VOCs comprise a wide range of chemical compounds including hydrocarbons (alkanes, alkenes, aromatics), oxygenates (alcohols, aldehydes, ketones, and ethers) and halogen-containing species. Historically, organic compounds in the troposphere have been measured as non-methane VOCs (NMVOCs) since methane oxidizes relatively slowly in the troposphere and its short-range does not contribute to O_3 formation compared to other organic compounds, but all act as O_3 precursors. Isoprene is the most reactive and plays a significant role in O_3 formation (Alvim et al., 2017, 2018; Santos et al., 2022). The main natural source of NMVOCs is vegetation when temperatures are high. These biogenic emissions consist of a wide variety of species including isoprene, terpenes and oxygenated compounds. Anthropogenic fraction which emits NMVOCs is related to the use of fossil fuels for transportation (Brito et al., 2015; Dominutti et al., 2020), the production of consumer goods, and various industrial processes. Emissions of alkanes and alkenes are dominated by anthropogenic sources, but are also produced by soils, wetlands and oceans (Koppmann, 2008).

Particulate material is a collective term used to describe small solid and liquid particles

that are present in the atmosphere over brief (minutes) to extended periods of time (days to weeks) (Popescu and Ionel, 2010). Individual particles vary in size, geometry, mass, concentration, chemical composition, and physical properties. Particles in ambient air typically appear in three rather distinct size classes (or modes) usually called ultrafine (diameter: 0.01–0.1 μm or 10 - 100 nm), fine (diameter: 0.1 - 2.5 μm) and coarse (diameter: $>$ 2.5 μm) mode particles (Hertel et al., 2009). However, air quality standards for particulate material worldwide is expressed in particulate material less than 10 μm in size, commonly called PM_{10}, and particulate material less than 2.5 μm in diameter, as $PM_{2.5}$ (Finlayson-Pitts and Pitts, 2000; Stocker et al., 2013). Such suspended particles characterize atmospheric aerosol. They can be described as being primary or secondary based on their origin and processes by which they form. Primary particles are emitted directly into the atmosphere from a variety of natural and anthropogenic (as a direct or indirect result of human activities) sources (Popescu and Ionel, 2010). The natural sources include volcanoes, forest fires, dust, ocean spray, biologic sources and meteoric debris. The anthropogenic source includes transportation, fuel combustion in stationary sources, a variety of industrial processes, solid waste disposal, and miscellaneous sources such as agricultural activities and fugitive emissions from roadways. In the MASP, the heavy vehicles and secondary aerosols are the main sources that produce PM (PM_{10}= 32 %, $PM_{2.5}$ = 51 %, respectively) (see Figure 2.1). The particles that appear in traffic exhaust are found mainly in the ultrafine fraction and include elemental carbon (EC) and organic carbon (OC) (Hertel et al., 2009). Secondary particles are produced from both naturally emitted gas-phase substances and anthropogenic sources as a result of chemical processes involving gas-phase pollutants, aerosol particles, and water vapor (H_2O) (Brito et al., 2018; Miranda et al., 2018). Secondary aerosol particles include sulfates, nitrates, and oxyhydrocarbons produced by direct, catalytic, and photochemical oxidation of sulfur (S), nitrogen (N), and volatile NMVOCs. The formation and increase of atmospheric aerosol is a major air quality concern because aerosol particles may scatter light, reducing visibility, pose an inhalation hazard to humans and animals, affect climate on a regional and global scale (Finlayson-Pitts and Pitts, 2000; Popescu and Ionel, 2010).

2.1.1.2 Transport

In many parts of the world, urban areas are downwind of other large conurbations. Regional scale transport of pollutants into urban areas can make a significant contribution to ambient levels (Chaithanya et al., 2021). However, conditions which promote large scale regional transport of pollutants are also associated with periods when dispersion of local pollutants is poor. One of the biggest challenges facing air quality researchers is determining how much of the ambient air pollution measured in urban areas is transported into other urban or rural areas compared to emissions from local pollutant sources. For instance, highest concentrations of particulate material usually occur during stable atmospheric conditions and light winds. However, long-range transport also contributes significantly to precursor gas-phase molecules and mainly to $PM_{2.5}$. O_3, the major photochemical pollutant, can be transported across national boundaries (Simpson and Malik, 1996). For that reason, regional air quality models are used to evaluate the relative contribution of local pollutants and estimate the transport of pollutants to and from cities (Hertel et al., 2009), considering that the models require the parameterization of urban areas to provide the boundary conditions.

Short- and long-range transport to and from urban areas such as the MASP is an important and considerable source. This can contribute to the increase and formation of new pollutants both in the MASP and in nearby cities, which can be urban or rural areas (Sánchez-Ccoyllo et al., 2006; Yadav et al., 2017; Chiquetto et al., 2018; Chaithanya et al., 2021; Squizzato et al., 2021). For instance, about short-range transport, Chiquetto et al. (2018) showed that pollutants emitted in closer municipalities to MASP, such as Cubatão (located halfway between the MASP and the Atlantic ocean) are transported to the MASP by the sea breeze circulation, showing a strong influence over air pollutants concentration. Sánchez-Ccoyllo et al. (2006) studied the pollution transport in the MASP to regions further inland and found out that, ocasionally, transport from the Northeast part of the state of São Paulo to the center of the city (MASP) can generate high $PM_{2.5}$ concentrations. Squizzato et al. (2021) carried out an air quality study using a mobile station installed near the municipality of Botucatu (200 km southwest of the MASP), focusing on ozone concentrations and found that the region studied is affected by activities in the MASP and

by local activities. For long-range transport of pollutants, there are more studies showing the influence they have on air quality in several places (Cuspilici et al., 2017; Yadav et al., 2017; Martins et al., 2018). The long-range transport of biomass burning in the Amazon when it reaches the state of São Paulo generates high pollutants concentrations mainly during the dry season (Nascimento et al., 2022).

2.1.2 Urban air pollution formation

Once the pollutants are emitted into the atmosphere by various sources, they can react with gases and/or particles in the atmosphere, thus forming new atmospheric pollutants. Following this reasoning, the urban atmosphere is a giant chemical reactor in which polluting gases such as hydrocarbons (HC) and oxides of nitrogen and sulfur react under the influence of sunlight to create a variety of secondary oxidized products, including ozone and sub-micrometer aerosols (or particulate material). O_3, being a secondary pollutant, is formed locally and/or by transport of certain gases (NO_x, VOCs and CO), in addition to sunlight. During the day, tropospheric O_3 is produced by the photolysis of NO_2, while at night it is declined through the process of NO_x titration. NO_x emissions are known to be responsible for much of the O_3 formation that occurs in rural areas. However, in regions closer to cities, O_3 formation is enhanced by VOC emissions, that are released from road traffic and the use of products containing organic solvents. The O_3 production rates depend on the abundance of OH^-, which depends on sources and sinks of odd hydrogen radicals as a whole. The different contributions to the O_3 levels in the troposphere can be divided into background O_3, regional photochemical production and depletion of O_3 by deposition and by loss reactions. Background O_3 can be interpreted as stratospheric injection and photochemical production far away from local sources Finlayson-Pitts and Pitts (2000); Seinfeld and Pandis (2016).

Particulate material is comprised of a complex mixture of constituents, including inorganic components (such as sulfates, nitrate and ammonium), organic compounds (organic aerosols), elemental carbon (EC), crustal and trace elements, and water. Among the inorganic components, the term sulfate aerosols is used for a suspension of fine solid particles of a sulfate or tiny droplets of a solution of a sulfate or of sulfuric acid (hydrogen sulfate). They are primarily produced through the gas-phase oxidation of SO_2 by the OH^- radical (Wang

et al., 2006; Seinfeld and Pandis, 2016). The fine nitrate (NO_3) formation is dominated by the reactions of gaseous nitric acid (HNO_3) or between NO_2 and O_3 during the nighttime (Seinfeld and Pandis, 2016). Where HNO_3 is also produced in reactions between NO_3 and formaldehyde (HCHO). NO_3 plays a significant role in nighttime oxidation of biogenically produced VOCs such as isoprene and the pinenes (compounds that have a major role in tropospheric O_3 chemistry) (Feng and Penner, 2007; Hertel et al., 2009). For the organic compound (carbon-containing), they contribute ∼20-50% to the total $PM_{2.5}$ mass at continental mid-latitudes and as much as 90% in tropical forested areas. They can occur in two forms: primary and secondary, while their sources are poorly understood (Turpin et al., 2000; Kanakidou et al., 2005). Primary organic aerosol (POA) is emitted directly from exhaust pipe and is considered a precursor for secondary organic aerosol (SOA). In contrast to POA, SOA is formed via homogeneous nucleation through the successive oxidation of gas-phase organic compounds, or through condensation on pre-existing particles. These gas-phase species exert high vapor pressures, meaning they are volatile and stable in the gas-phase. Following oxidation, the reduction in the volatility of the molecules results in a reduction in vapor pressure. With sufficiently low pressure, the gas-phase compound partitions into the solid-phase, producing secondary organic material (the particle phase of secondary organic aerosol). The oxidation of aromatic molecules (such as benzene, toluene and xylene) in urban areas lead to the incorporation of substantial quantities of SOA (in the fine particle mode). Both POA and the precursors of SOA can be from urban, biogenic, and biomass burning sources and the major hydrocarbon precursors of SOA are isoprene, monoterpenes, sesquiterpenes and hydrocarbons.

2.2 Interaction between pollutants and radiation

Both gases and particles in the atmosphere can interact with solar and terrestrial radiation through the absorption, emission and scattering, affecting thus the atmospheric radiation transfer. Particles also have a strong effect upon global climate through their role as cloud condensation nuclei (CCN), directly affecting the propensity of clouds to form and through indirect effects of the induced droplet size distribution affecting cloud brightness and lifetime.

Radiation is the flow of electromagnetic energy between the sun and the Earth's surface and it includes both solar (shortwave) and thermal (long-wave) radiation. Solar radiation is emitted by the sun and it corresponds to ultra-violet (UV), visible (0.4 to 0.7 μm) and near infra-red wavelengths (0.1 to 4 μm). Thermal radiation is emitted by the atmosphere or the Earth's surface and it corresponds mainly to thermal infra-red and far infra-red wavelengths (4 to 100 μm) (Andrews, 2010; Sportisse, 2010). The solar (incoming) radiation interacting with the atmosphere is influenced by cloud droplets, atmospheric gases and aerosol particles, and they are the cause of the attenuation process (absorption and scattering) of radiation. A percentage of this radiation goes back to space (outgoing radiation) either by reflection or by emission from the Earth's surface and/or atmospheric elements (Godish, 2004). Both the incoming and outgoing radiation of the Sun-Earth system are parts of components of the Earth's radiation budget (Figure 2.2).

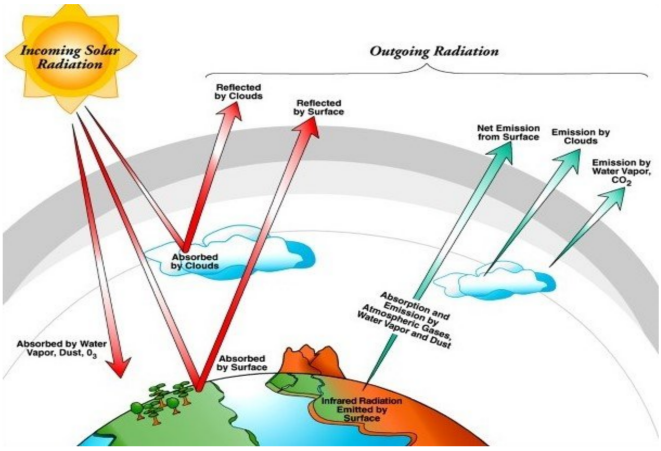

Figure 2.2: Graphic depiction of radiation flows between the Earth's surface and the atmosphere.

The Earth's atmosphere is a thin gaseous envelope comprised mostly of N_2 and O_2 and other gases and particles of small amounts, such as H_2O, CO_2, O_3, CH_4, dust and Chlorofluorocarbons (CFCs) (Godish, 2004; Ahrens, 2011). Although the Earth's atmosphere may appear transparent to radiation, it plays a very important role in the Earth's energy balance, since it controls the amount of solar radiation, both that reaches the surface and that escapes into space (Godish, 2004). When solar radiation interacts with the atmosphere, a part is absorbed by certain atmospheric gases, implying that not all solar radiation reaches the Earth's surface (see Figure 2.3). Panel (a) of Figure 2.3 shows the

solar spectral irradiance at the top of the atmosphere and that at sea level, where shaded regions refer to absorption by molecules. While the other panels (b - d) show the percentage of spectral absorption of the main atmospheric gases on solar and terrestrial radiation (0.1 - 30 μm).

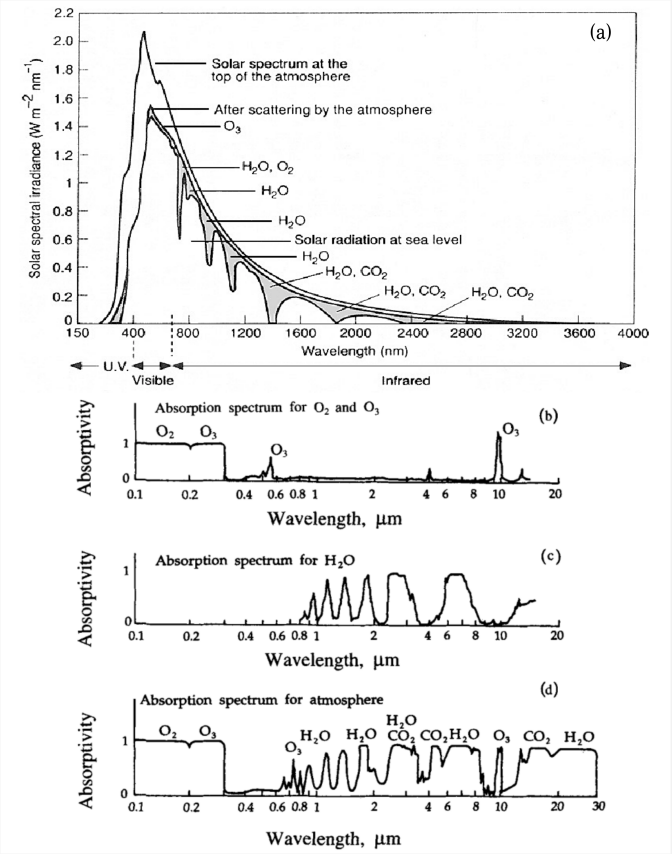

Figure 2.3: Transmission of shortwave solar and long wavelength radiation from the Earth's surface through the atmosphere.(Adapted from (Godish, 2004))

The spectral absorption indicates that absorption is so strong in some spectral regions that no solar energy in those regions reaches the Earth's surface. The absorption by O_2 and O_3 is responsible for the elimination of practically all the incident radiation $\lambda < 290$ nm (Figure 2.3(b)). Between 300 and 800 nm, the atmospheric absorption is not strong, forming an atmospheric window (region where the atmosphere is transparent). H_2O absorbs in a complicated way, and mostly in the region where the solar and terrestrial radiation overlap (from 0.8 to 2 μm). Thus, we note that not all molecules absorb over the same spectral ranges, their geometrical structure explains the reason for these differences (Godish, 2004; Sportisse, 2010). For example, H_2O, CO_2 and O_3 strongly interact with radiation above 4 μm, while N_2 and O_2 do not. H_2O and O_3 behave as molecules with an electric dipole

moment while N_2 and O_2 are symmetric. Molecules with dipole moments interact strongly with electromagnetic radiation because the wave's electric field causes forces in opposite directions and thus accelerations in the electrons and nuclei at one end of the molecule compared to the other (Godish, 2004). The CO_2 molecule is linear but can be easily bent, resulting in an induced dipole moment. A transverse vibrational mode exists for CO_2 at 15 μm, where the Earth emits most of its infrared radiation. Consequently, CO_2, H_2O and O_3 are known as greenhouse gases since they also absorb infra-red radiation from the Earth (outgoing radiation). NO_2 (being an important molecule in the troposphere) absorbs throughout the UV and visible range ($\lambda > 290$ nm) of the solar spectrum in the lower atmosphere (Godish, 2004). In the case of aerosols, these and cloud reflect about a quarter of the incoming radiation back to space. Atmospheric aerosols play an important role with electromagnetic radiation, known as direct effect (Sportisse, 2010; Xu and Penner, 2012; Myhre et al., 2013). Most of the atmospheric aerosols scatter incoming solar radiation, while other few absorb it, depending on the composition of the particles (Xu and Penner, 2012; Myhre et al., 2013). For instance, sulfates and nitrates reflect nearly all radiation they encounter, cooling the atmosphere (Xu and Penner, 2012). Black carbon, in contrast, absorbs radiation readily, warming the atmosphere but also shading the surface. Organic carbon has a warming influence on the atmosphere depending on the brightness of the underlying ground. Dust impacts radiation to varying degrees, depending on the composition of the minerals that comprise the dust grains, and whether they are coated with black or brown carbon. Salt particles tend to reflect all the solar radiation they encounter (Zhang et al., 2016). The absorption and loss of energy by the Earth system are almost responsible for the Earth's weather on both global and local scales (Seinfeld and Pandis, 2016).

All these processes that occur in the Earth system are studied through the radiative transfer equation (RTE), which describes these interactions mathematically, and radiometric observations made from the ground, the air, and space (e.g. satellites). Satellites have sensors on board with the ability to measure electromagnetic radiation at a certain distance, where the radiation is characterized by a specific wavelength that is sensitive to some physical aspect of the medium (Roy et al., 2017). This mechanism is known as remote sensing. Particularly, satellite makes use of four spectral regions: UV, visible, infra-red (near and thermal infra-red) and microwave; especially where radiation passes through at-

mospheric windows (Liou, 2002). Next, we will discuss about the characteristics and types of satellites, and the techniques they used to estimate air pollutants (gases and aerosols) volume column density (VCD).

2.2.1 Satellite measuring techniques for Air pollutants retrievals

Satellite measurements of trace species began in the 1970s with the objective to study their distribution and influence on a global scale. The primary satellites providing information on lower tropospheric trace constituents fly in near-polar, Sun-synchronous, low Earth orbits (Lenoble, 1993). Orbit altitude is 705 km, yielding 100 min per orbit and approximately 14 complete orbits per day. The satellite crosses the equator twice an orbit, once in the southward or descending direction, and once in the northward or ascending direction. At low and mid latitudes, the resultant observations occur at a near-constant local time. The point directly beneath the satellite is called the nadir (Figure 2.4) and the nadir-viewing geometry observes the entire atmospheric column and is the only geometry that probes the boundary layer. Near-global coverage for a Sun-synchronous orbit is achieved in more than 12 hours for daytime and nighttime measurements of thermal emission, and more than 24 hours for daytime measurements of solar backscatter.

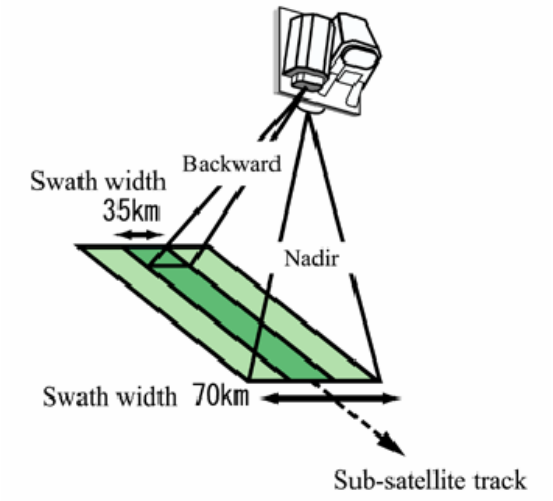

Figure 2.4: Observation geometry of nadir and backward modes

Remote sensing is divided into two broad categories: passive and active. Passive remote sensors (such as photographic camera, electro-optical and microwave sensor) detect only

electromagnetic radiation naturally emanating from the Earth's surface. The majority of instruments employ passive techniques, observing either solar backscatter (e.g. AIRS, MOPPIT, IMG and IASI; $\lambda < 4$ μm) or thermal infra-red emission (e.g. TOMS, GOME, SCIAMACHY, MODIS, MISR and OMI; $4 < \lambda < 50$ μm). Active sensors, such as laser radar, illuminate the object of study with their own artificially produced radiation and record the reflected component. It is worth emphasizing that remote sensing instruments do not directly measure atmospheric composition. Rather, a retrieval is conducted by calculating the atmospheric composition that best reproduces the observed radiation. Such retrievals often require external information on geophysical fields. The development of a variety of algorithms to extract physical parameters by accounting for atmospheric radiative transfer has been integral to the success of modern remote sensing. For example, MOPPIT CO retrieval uses a nonlinear optimal estimation algorithm and a fast radiative transfer model to invert the measured A- and D-signals. CO retrievals use up to six channels (twelve measured signals) in two bands: two channels in a solar relfectance band (≈ 2.3 μm) e four channels in a thermal emission band (≈ 4.3 μm), to measure carbon monoxide and methane in the troposphere over the entire globe. OMI uses 390 and 780 channels (from different spectral bands) to determinate Ozone profile and total column of other species (SO_2, NO_2, HCHO and O_3), and finally MODIS uses 8 channels to determine the aerosol properties.

Chapter 3

Modeling framework and datasets

3.1 Model

3.1.1 WRF-Chem model

The Weather Research Forecasting (WRF) model is widely used to simulate the flow of the atmosphere in a 3D cube with grid spacing from tens of meters to thousands of kilometers (Skamarock et al., 2008). WRF is supported by the National Center for Atmospheric Research (NCAR) and National Oceanic and Atmospheric Administration (NOAA) with partnerships and collaborations with universities and other government agencies in the United States and overseas. The model uses two dynamic core options, such as the Advanced Research WRF (ARW) and the Nonhydrostatic Mesoscale Model (NMM), to solve the Eulerian and non-hydrostatic equations. The ARW is the most used for research purposes and the NMM for operational demands (Skamarock et al., 2008). The model is employed in diverse applications that include meteorological research, operational numerical weather forecasting, climate and air quality simulations, prediction of forest fire behavior, studies of tropical storms, among others (Skamarock et al., 2008). WRF-Chem model is a fully coupled "online"meteorology-chemistry transport model and the complete description of the model is given by Grell et al. (2005).

3.1.2 WRF-Chem model configurations

The WRF-Chem model version 3.9.1 was used to perform simulations during four different months. These simulations were considered using a single domain with a horizontal resolution of 9 km. The modeling domain was centered at $23.5°S$ and $46.3°W$, projected on a Lambert plane. It consists of 150 points in the west-east direction, 100 points in the

north-south direction, and 35 vertical levels that extend from the surface up to 50 hPa (20 km) (see Figure 3.1). The time periods were selected according to the availability of observational data and favorable meteorological conditions for pollutant accumulation: a) from 00:00 UTC May 31 to 18:00 UTC June 30, 2017; b) from 00:00 UTC June 30 to 18:00 UTC July 31, 2018; c) from 00:00 UTC May 31 to 18:00 UTC June 30, 2019 and d) from 00:00 UTC July 31 to 18:00 UTC August 31, 2019. The first 24 hours in each run are discarded as spin-up time. The main physical parameterizations included the Morrison two-moment scheme (Morrison et al., 2009) for cloud microphysics, the Rapid Radiative Transfer Model for general circulation models (GCMs) (RRTMG) scheme (Mlawer et al., 1997) for shorwave and longwave radiation, the Yonsei University (YSU) scheme (Hong et al., 2006) for turbulence, the Grell 3D Ensemble (Grell and Devenyi, 2002) for cumulus clouds, the unified Noah Land Surface model (Tewari et al., 2004) for land surface and the Revised Atmospheric Research Mesoscale Model (MM5) for surface layer.The main physics and chemistry options used in this study are listed in Table 3.1. In order to analyze the atmospheric chemical composition through atmospheric models, it is important to properly represent the meteorological conditions, background concentration of chemical species and emission sources in the study area.

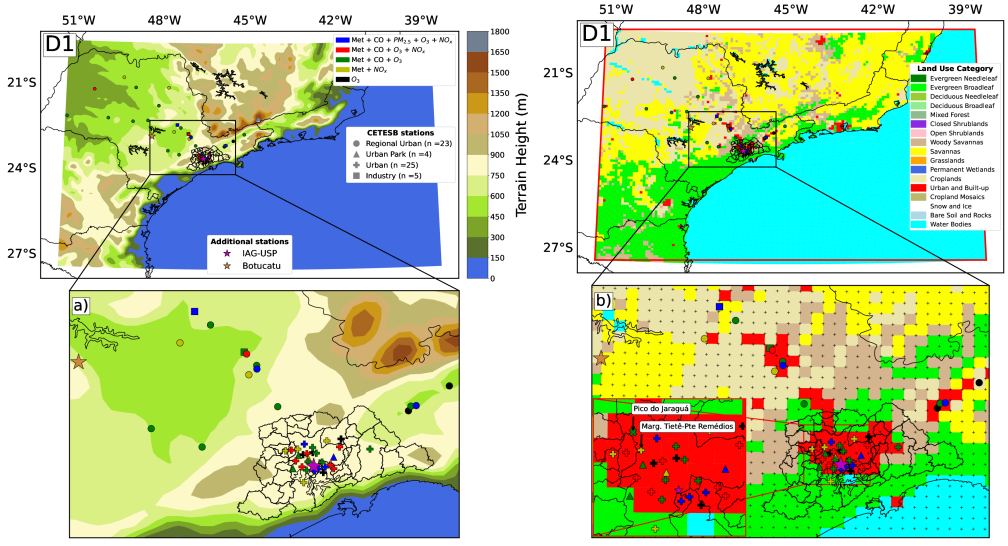

Figure 3.1: Default (left side) terrain height (m) and (right side) land use categories in the WRF-Chem model, both over the State of São Paulo ($20°S$ - $27°S$ latitude and $51°W$ - $38°W$ longitude). Colored dots in both images represent the CETESB air quality stations distributed over the State of São Paulo. (a) and (b) represent the zoom of the black frame within the study domain, showing the MASP.

The meteorological initial and boundary conditions were obtained from the NCEP global

forecasting system (GFS) model, which are available every 6 hours. The GFS data have global latitude and longitude grids of 0.25° horizontal resolution and generate data for dozens of atmospheric and land-soil variables, including temperatures, winds, precipitation, soil moisture, and atmospheric ozone concentration, necessary to run the WRF-Chem model. GFS data is freely available on the website *https://rda.ucar.edu/datasets/ds084.1/*. The chemical initial and boundary conditions were obtained from the Community Atmosphere Model with Chemistry (CAM-Chem) model (Lamarque et al., 2012) at 0.9° x 1.5° horizontal resolution. The CAM-Chem is a component of the NCAR Community Earth System Model (CESM) and has demonstrated the ability to represent tropospheric and stratospheric conditions including temperature structure and dynamics (Lamarque et al., 2012). The CAM-Chem uses the Model for Ozone and Related chemical Tracers, version 4 (MOZART-4) chemical mechanism (Emmons et al., 2010).

Table 3.1 - WRF-Chem model configurations

Attribute	Configurations
Model Version	3.9.1
Horizontal grid	D1 = (150x100) $\Delta x = 9km$
Vertical layer	35 (top = 20km)
Time step	45 s
Initial and boundary conditions	NCEP GFS 0.25° horizontal resolution
Simulation period	00 UTC on May 31th - 18 UTC on June 30th, 2017
	00 UTC on June 30th - 18 UTC on July 31th, 2018
	00 UTC on May 31th - 18 UTC on June 30th, 2019
	00 UTC on July 31th - 18 UTC on August 31th, 2019
	the first twenty-four hours were discarded as spin-up time
Physics Options	Scheme
Micro physics	Morrison 2-Moment (mp_physics = 10)
Planetary Boundary Layer (PBL)	Yonsei University (bl_pbl_physics = 1)
Cumulus	Grell 3D Ensemble (cu_physics = 5)
Radiation (Long- and Short-wave)	RRTMG (ra_lw_physics = 4)
Land Surface	Unified Noah Land Surface Model (sf_surface_physics = 2)
Surface Layer	Revised MM5 (sf_sfclay_physics = 1)
Chemistry Options	Mechanism
Gas phase	MOZART
Aerosol module	GOCART (chem_opt = 112)
Photolysis rate	Madronich F-TUV photolysis (phot_opt = 3)
Emissions	Model
Anthropogenic	LAPAt
Biogenic	MEGAN (bio_emiss_opt = 3)
Fire	FINN (biomass_burn_opt = 2)
Sea Salt	GOCART sea salt emissions (seas_opt = 1)

Aerosol Optical Properties	Based upon volume approximation (aer_op_opt = 1)

3.1.3 Gas-phase and Aerosol chemical mechanism

For the modeling framework proposed here, the MOZART-4 (Emmons et al., 2010) for the gas-phase chemistry coupled with the Goddard Chemistry Aerosol Radiation and Transport (GOCART, (Chin et al., 2002)) for aerosol chemistry, is used. The MOZART-4 mechanism includes 85 gas-phase species, 12 bulk aerosol compounds, 39 photolysis rates and 157 gas-phase reactions. It also includes an updated isoprene oxidation scheme and a better treatment of volatile organic compounds to represent alkanes and alkenes (Emmons et al., 2010). GOCART is a bulk aerosol mechanism, used within a model to approximate particulate matter processes, for the major tropospheric aerosol types (sulfate, dust, organic carbon (OC), black carbon (BC) and sea salt aerosols). Despite the GOCART scheme does not simulate nitrate and secondary organic aerosols, it is potentially the most suitable tool to link satellite and in-situ observations (Chin et al., 2000, 2002; Ginoux et al., 2001), due to its ability to simulate aerosol optical depth (AOD) with better results, compared to the other aerosol modules. GOCART model is driven by the assimilated meteorological fields, which are generated in the Goddard Earth Observing System Data Assimilation System (GEOS DAS). The GOCART aerosol model module includes organic carbon (OC1 and OC2), black carbon (BC1 and BC2) both emitted in a combination of hydrophilic and hydrophobic forms, respectively, as well as sulfate and fine unspeciated aerosol contribution (P_{25}). Five type of dust in different sizes representing effective radius of 0.5, 1.4, 2.4, 4.5 and 8 μm; and four type of sea salt (SEAS_1, SEAS_2, SEAS_3 and SEAS_4) with effective radii of 0.3, 1.0, 3.25, and 5 μm, are assumed in GOCART.

3.1.4 Emissions inventories

An emissions inventory is a database that lists, by source, the amount of air pollutants emitted to the atmosphere during a time period. In this work, three types of sources (vehicular, biogenic and biomas burning emissions) were considered.

3.1.4.1 *Vehicular emissions*

As vehicular emissions (see Table 14 from CETESB, (2019)) over southeastern Brazil are the largest sources of pollution (Andrade et al., 2015; Gavidia-Calderón et al., 2018), emissions from sectors other than ground transport (such as industrial and residential) are not taken into account in the simulations. In addition, because the emissions from the industrial and residential sector from EDGAR-HTAP global emission dataset are not suitable for our study domain. The ground transport emissions, specifically on-road vehicles, were derived from the bottom-up transport emission model described by Andrade et al. (2015). This model uses information on number of vehicles, emission factors and intensity of use (mileage) for different vehicle types (motorcycles, light-duty vehicles, and heavy-duty vehicles) and different fuel types (gasohol, ethanol, combination of any proportion of gasohol and ethanol, and diesel) from different agencies, such as DENATRAN, CETESB and the IAG-USP Laboratory of Atmospheric Processes (LAPAt). Volatile Organic Compounds (VOCs) speciation and pollutant temporal profiles were derived from in-tunnel field campaigns located in the MASP during 2004 and 2011 (Martins et al., 2006; Nogueira et al., 2021).

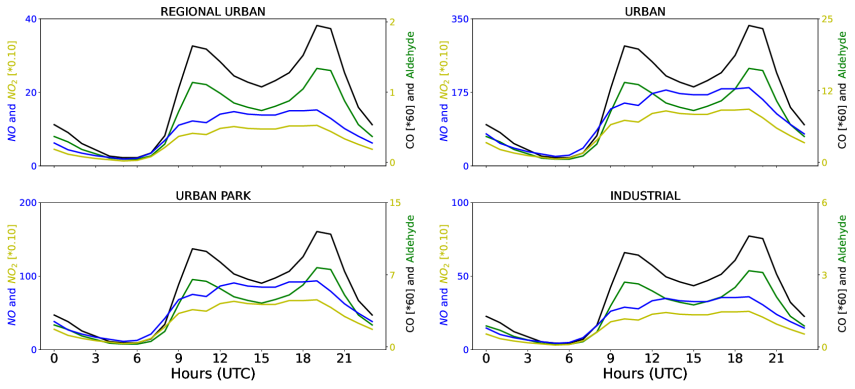

Figure 3.2: Temporal distribution of NO, NO_2, CO and Aldehyde emission in $mol\ km^{-2}hr^{-1}$, representing light and heavy vehicular traffic.

For the temporal distribution, the profile for light and heavy vehicles was considered, generating emissions throughout the day. Figure 3.2 shows the hourly profile of the emissions of atmospheric pollutants emitted by light vehicles (CO and Aldehyde) and heavy vehicles (NO and NO_2), considering four different regions (regional urban, urban park, urban and industrial) within the state of São Paulo (these regions will be explained in section 4.1). The maximum emission of light and heavy vehicles start at 10:00 UTC, while the

emissions of heavy vehicles remain constant throughout the day, the emissions of light vehicles have another maximum peak at 19:00 UTC, coinciding with the higher traffic of vehicles in those hours. For spatial distribution, vehicles within the modeled domain were distributed proportionally to the road length in each grid cell. Road length was calculated as the sum of various road types found within each grid cell. For this study, the total vehicle emissions, spatially and temporally distributed over our study domain, were obtained from the CETESB report for the year 2015 (CETESB, 2015). Figure 3.3 illustrates the spatial distribution of NO and $PM_{2.5}$ emissions at 20:00 UTC over southeastern Brazil. High emissions of both pollutants can be observed above the MASP and low emissions are presented outside the MASP. This contrast of emissions reflects the high flow of vehicles that circulate through MASP.

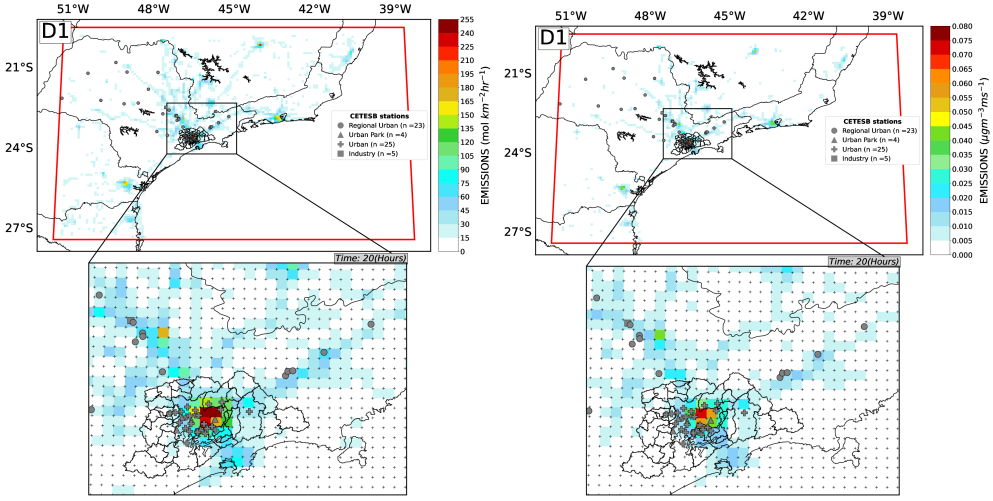

Figure 3.3: Spatial distribution of (left side) NO and (right side) $PM_{2.5}$ emission in $mol\ km^{-2}hr^{-1}$ and $\mu gm^{-3}\ ms^{-1}$, respectively, representing the hourly vehicular traffic.

3.1.4.2 Biogenic emissions

Biogenic emission data were calculated online based on the global Model of Emissions of Gases and Aerosols from Nature version 2.1 (MEGAN2.1), designed for both global and regional emissions modeling. This model estimates the net monthly or annual emission of 20 compound classes generated by urban, rural and agricultural ecosystems, except by biomass burning, and that can influence the atmosphere. This component then, breaks down into 150 individual species such as isoprene, monoterpenes, sesquiterpenes, carbon monoxide, nitrogen monoxide, alkanes, alkenes, aldehydes, acids, ketones, and other oxy-

genated VOCs. The MEGAN2.1 model includes additional compounds, emission types, and controlling processes, different from previous versions; and has simple mechanistic algorithms. MEGAN has a spatial resolution of approximately 1km and it can be used for both regional air quality modeling and global earth system modeling studies (Guenther et al., 2006; Arneth et al., 2011).

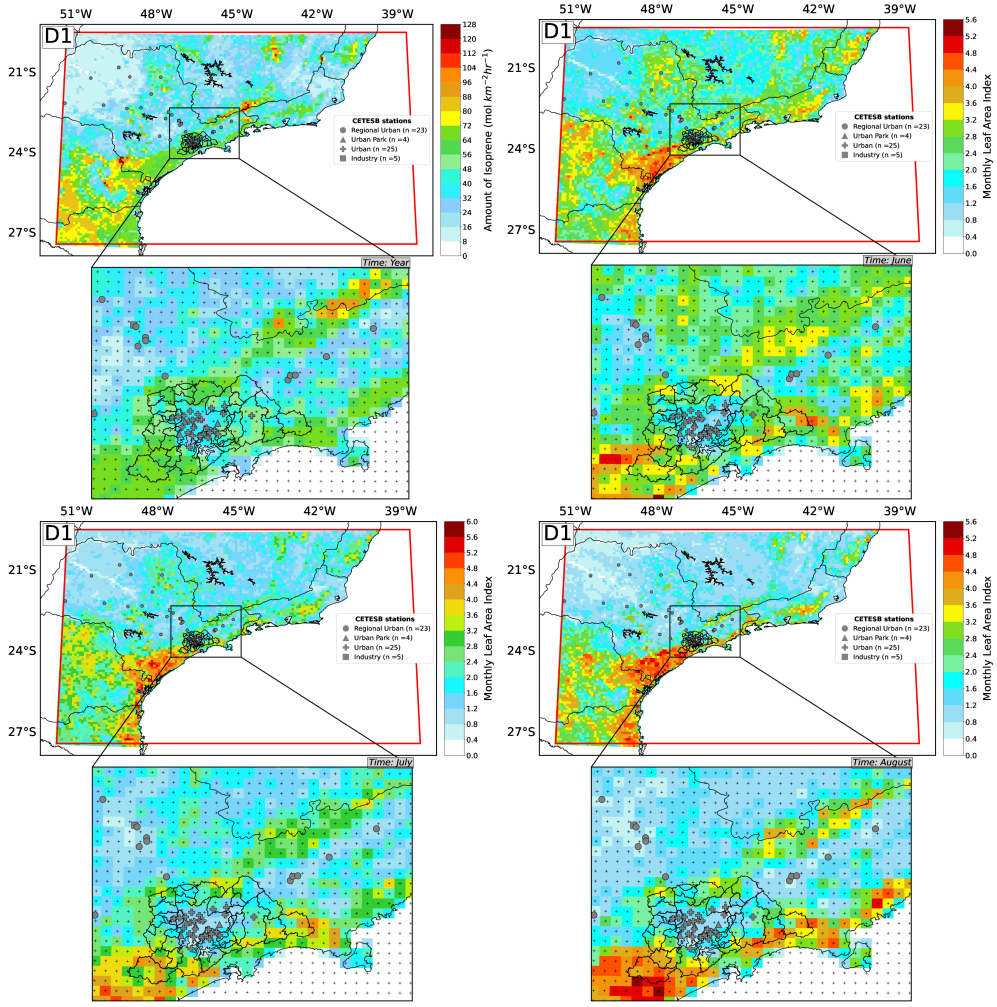

Figure 3.4: Spatial distribution of (Top-left) annual mean emissions of isoprene and monthly leaf area index in (Top-right) June, (Bottom-left) July and (Top-left) August, respectively, all for 2000, over the modeled domain and calculated by the MEGAN model.

To take into account this biogenic-emissions option, we downloaded and compiled the utility from NCAR web site https://www.acom.ucar.edu/wrf-chem/download.shtml. The utility prepares the MEGAN input data files for the WRF-Chem model (named wrfbiochemi_d0x files) as shown in Figure 3.4. The term "d0x"means the number of domains considered. MEGAN provides some suitable variables for the WRF-Chem model such as: amount of isoprene (MSEBIO_ISOP), percentage of broad leaf (PFTP_BT), percentage of

needle leaf (PFTP_NT), percentage of shrubs (PFTP_SB), percentage of herbaceous biota (PFTP_HB), monthly leaf area index (MLAI), monthly air temperature (MTSA, in K) and monthly downward short wave radiation (MSWDOWN, Wm^{-2}). The upper left panel in Figure 3.4 represents the spatial distribution of the annual mean emissions of ISOP over southeastern Brazil, with focus on the MASP. The other panels represent the spatial distribution of MLAI for the months of June, July and August. We can seen high ISOP and LAI values on the MASP boundaries and over the state of Parana and Santa Catarina for all periods, although for August the LAI values were higher compared to the other months. The LAI is an important indicator of radiation and precipitation interception, energy conversion, and water balance. It is a reliable parameter for plant growth (Jang et al., 2019).

3.1.4.3 Biomass Burning emissions

Biomass burning is a significant air pollution source, with global, regional and local impacts on air quality, public health and climate. The State of São Paulo often experiences high air pollution levels due to emissions of particulate pollutants from local sources and long-range transport of air masses originated from biomass burning. This pollution source is being monitored annually using satellite images and numerical simulation. For this work, we have considered the Fire INventory from NCAR version 1.0 (FINNv1) model to obtain fire emission data over our study domain. This inventory was chosen because it differs from other inventories by providing a unique combination of high temporal and spatial resolution, global coverage, and estimates for a large number of chemical species.

FINNv1 has been designed to use any fire detection data available. However, here we have used satellite observations processed by MODIS sensor onboard the Terra and Aqua satellites, to detect active fires and land cover. These data are needed to estimate a high spatial resolution (1km) daily data of global emissions of particulates and trace gases from different types of vegetation fires (wildfires, agricultural burns, and prescribed burns). On the other hand, this inventory does not include biofuels use and burning (Wiedinmyer et al., 2011). The daily data estimated by FINNv1 have been specifically developed to model atmospheric chemistry and air quality in a consistent framework at local to global scales (Freitas et al., 2007; Grell et al., 2011). Emission factors have been updated, particularly, for Non-methane Organic Compounds (NMOC). The list of chemical speciation

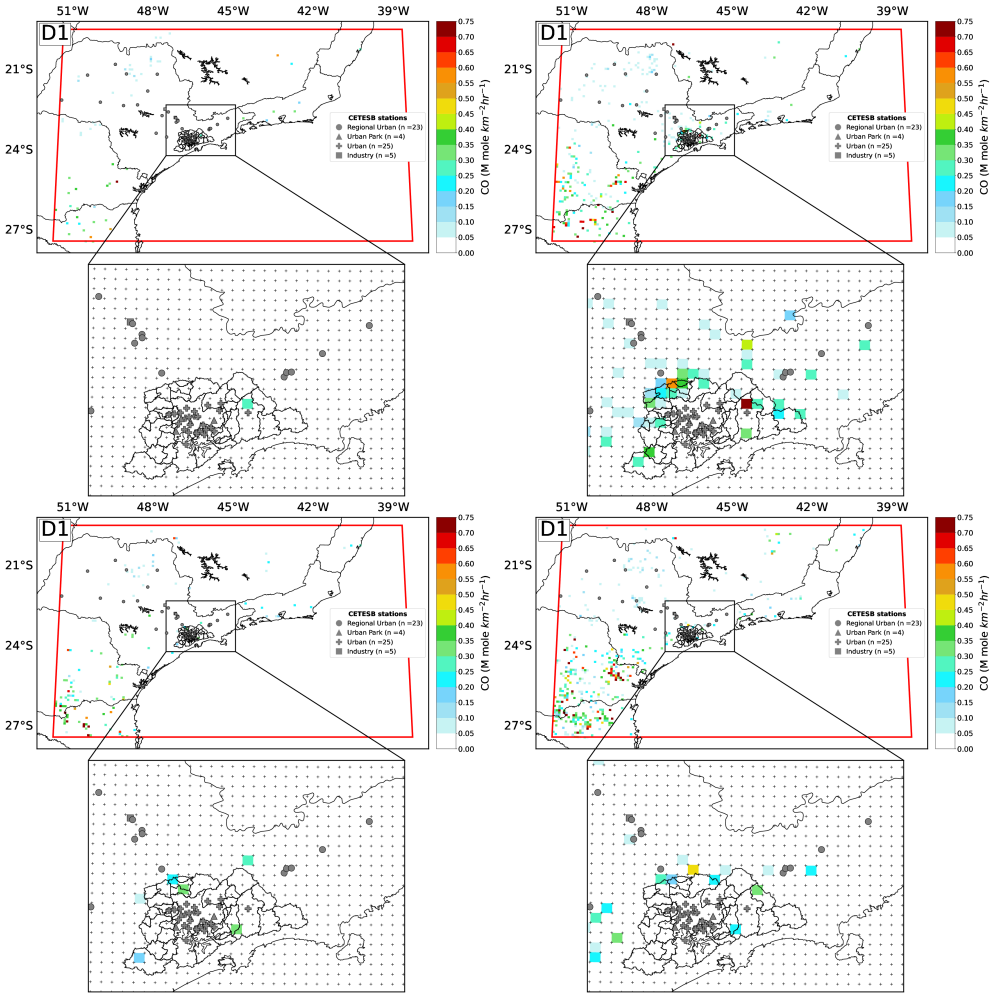

Figure 3.5: Spatial distribution of CO emissions from cumulative burning events, estimated by the FINNv1 model, throughout (Top-left) June 2017, (Top-right) July 2018, (Bottom-left) June 2019 and (Bottom-right) August 2019 and over our study domain.

needed to assign total NMOC emission estimates to grouped species for use in chemical transport models is provided for three widely used chemical mechanisms: SAPRC99, GEOS-CHEM, and MOZART-4. Also, FINN provides estimates of organic compounds, including formaldehyde and methanol (Wiedinmyer et al., 2011). To use this fire emissions option, we downloaded the near-real-time fire emissions from the FINN, based on MODIS Rapid Response fire count (FIRMS) and the utility, both available in the NCAR website https://www.acom.ucar.edu/Data/fire/. The utility prepares FINN input data files for use in the WRF-Chem model (called wrffirechemi_d0x files). Figure 3.5 represents the spatial distribution of CO emissions from monthly cumulative burning events over our study domain and for each period analyzed (June 2017, July 2018, June and August 2019). We can see that during July 2018, there was a higher emission of CO, compared to the other

three periods, this will be in the results chapter.

3.2 Ground-based data

3.2.1 CETESB air quality stations

The Companhia Ambiental do Estado de São Paulo (CETESB) is an environmental agency that since the late 1970s is responsible for the control and monitoring of air pollutants in the State of São Paulo (CETESB, 2017). Given the seriousness of the pollution generated by on-road vehicles, during the 1980s, the CETESB laid the foundation for the development of the technical bases of CONAMA resolution No. 18/86 (BRASIL, 1986) that established the Program for the Control of Air Pollution Emissions by Motor Vehicles (Programa de Controle da Poluiçao do Ar por Veículos Automotores, PROCONVE). PROCONVE required that new vehicles comply with the maximum emission limits, in standardized tests and with reference fuels. This program became the main driver for reducing air quality violations in Brazilian cities (Andrade et al., 2017); however, its emission reduction protocols are now outdated and do not satisfy the current limits, due in part to urbanization and demographic changes. And also due to the fact São Paulo has different characteristics, in territorial area and population, compared to other states inside and outside of Brazil.

CETESB manages a monitoring network with 63 automatic air quality stations (61 fixed and 2 mobile) and 22 manual stations. The air quality stations (hereinafter referred to as AQS) provide hourly information on meteorological and pollutant parameters, such as air temperature at 2m (T_{2m}), relative humidity at 2m (RH_{2m}), wind speed and wind direction at 10m (WS_{10m} and WD_{10m}, respectively), nitrogen monoxide (NO), nitrogen dioxide (NO_2), tropospheric ozone (O_3), carbon monoxide (CO), fine particulate material ($PM_{2.5}$), among others (CETESB, 2019). These air quality and meteorological data are continuously published on the Qualar website (https://cetesb.sp.gov.br/ar/qualar/). In this work, 60 AQS spread out across the state of São Paulo, mainly over the MASP (see Figure 3.1 and Table 3.2), were selected for the analysis.

Table 3.2 - Air Quality Stations distributed over the State of São Paulo

Stations	Type	Code	Latitude	Longitude	Meteorological Variables				Pollutants			
					T_{2m}	RH_{2m}	WS_{10m}	WD_{10m}	NO_x	O_3	CO	$PM_{2.5}$
Capão Redondo	Urban Park	269	-23.67	-46.78	x[1]	x	x	x	x	x		
Cid.Universitária-USP-Ipen	Urban Park	95	-23.57	-46.74					x	x		x
Ibirapuera	Urban Park	83	-23.59	-46.66					x	x	x	x
Itaquera	Urban Park	97	-23.58	-46.47						x		
Carapicuíba	Urban	263	-23.53	-46.84	x	x	x	x	x	x	x	
Cerqueira César	Urban	91	-23.55	-46.67					x		x	
Congonhas	Urban	73	-23.62	-46.66					x		x	x
Diadema	Urban	92	-23.69	-46.61						x		
Grajaú-Parelheiros	Urban	98	-23.78	-46.70					x	x	x	x
Guarulhos-Paço Municipal	Urban	264	-23.46	-46.52	x	x	x	x	x	x		x
Guarulhos-Pimentas	Urban	279	-23.44	-46.41	x	x	x	x	x	x	x	x
Itaim Paulista	Urban	266	-23.50	-46.42						x		x
Interlagos	Urban	262	-23.68	-46.68	x	x	x	x	x	x		
Marg.Tietê-Pte Remédios	Urban	270	-23.52	-46.74	x	x	x	x	x		x	x
Mauá	Urban	65	-23.67	-46.47					x	x		
Mogi das Cruzes	Urban	287	-23.52	-46.19	x	x	x	x	x	x		
Mooca	Urban	85	-23.55	-46.60	x	x	x	x		x	x	
N.Senhora do Ó	Urban	96	-23.48	-46.69						x		
Osasco	Urban	120	-23.53	-46.79					x		x	
Parque D.Pedro II	Urban	72	-23.54	-46.63	x	x	x	x	x	x	x	x
Pinheiros	Urban	99	-23.56	-46.70	x	x	x	x	x	x	x	x
Santo Amaro	Urban	64	-23.65	-46.71						x	x	
São Caetano do Sul	Urban	86	-23.62	-46.56					x	x	x	
Santana	Urban	63	-23.51	-46.63	x	x	x	x		x		x
S.André Capuava	Urban	100	-23.64	-46.49	x	x	x	x		x		
S.André-Paço Municipal	Urban	254	-23.66	-46.53							x	
S.Bernardo-Centro	Urban	272	-23.70	-46.55	x	x	x	x	x	x	x	x
S.Bernardo-Paulicéia	Urban	102	-23.67	-46.58	x	x	x	x				
Taboão da Serra	Urban	103	-23.61	-46.76					x		x	
Araçatuba	Regional Urban	107	-21.19	-50.44	x	x	x	x		x		
Araraquara	Regional Urban	106	-21.78	-48.19	x	x	x	x	x	x		
Bauru	Regional Urban	108	-22.33	-49.09	x	x	x	x	x	x		
Campinas-Centro	Regional Urban	89	-22.90	-47.06							x	
Campinas-Taquaral	Regional Urban	276	-22.87	-47.06	x	x	x	x	x	x		
Campinas-V.União	Regional Urban	275	-22.95	-47.12	x	x	x	x	x	x		x
Catanduva	Regional Urban	248	-21.14	-48.98	x	x	x	x	x	x		
Guaratinguetá	Regional Urban	289	-22.80	-45.19	x	x	x	x	x	x		
Jaú	Regional Urban	110	-22.30	-48.57	x	x	x	x	x	x		
Jundiaí	Regional Urban	109	-23.19	-46.90	x	x	x	x	x	x		
Limeira	Regional Urban	281	-22.56	-47.41	x	x	x	x	x	x		
Marília	Regional Urban	111	-22.20	-49.96	x	x	x	x	x	x		
Paulínia Sul	Regional Urban	112	-22.79	-47.14					x	x		
Piracicaba	Regional Urban	113	-22.70	-47.65	x	x	x	x	x	x		x
Presidente Prudente	Regional Urban	114	-22.12	-51.41	x	x	x	x	x	x		
Ribeirão Preto	Regional Urban	288	-21.15	-47.83	x	x	x	x	x		x	x
São José Do Rio Preto	Regional Urban	116	-20.78	-49.40	x	x	x	x	x	x		x
Sorocaba	Regional Urban	67	-23.50	-47.48	x	x	x	x	x	x		
S.José Campos	Regional Urban	88	-23.19	-45.87	x	x	x	x	x	x		
S.José Campos-Jd.Satélite	Regional Urban	277	-23.22	-45.89	x	x	x	x	x	x	x	x

[1] The value "x"represents the existence of meteorological variables and chemical species available for each station.

S.José Campos-Vista Verde	Regional Urban	278	-23.18	-45.83	x	x	x	x				
Taubaté	Regional Urban	280	-23.03	-45.58	x	x	x	x	x	x	x	x
Tatuí	Regional Urban	256	-23.36	-47.87	x	x	x	x	x	x		
Paulínia	Industrial	117	-22.77	-47.15	x	x	x	x	x	x		
Santa Gertrudes	Industrial	273	-22.46	-47.54	x	x	x	x				

3.2.2 IAG-USP meteorological station

The meteorological station of the Institute of Astronomy, Geophysics and Atmospheric Sciences of the University of São Paulo (IAG-USP) is located in the Parque do Estado, Bairro da Água Funda, in the south-east side of the MASP (see Table 3.3). The station is located far from important roads with a high traffic density of light and heavy vehicles. The meteorological station provides hourly data of different meteorological parameters such as T_{2m}, WD_{10m}, WS_{10m}, WD_{10m} and precipitation.

3.2.3 Botucatu station

The mobile Boucatu station was located 200 km northwest of the MASP, on the outskirts of Botucatu. Botucatu is a small city with a population of 148.130 inhabitants, and is representative of urban areas in inland São Paulo State that are often affected by pollutant transport from the MASP region (Squizzato et al., 2021). The sampling took place during winter time of 2019 (from June 07 to August 25) in an open field area, far from the city center (north of Botucatu) and with low influence from local sources. Among the ambient parameters that were monitored are CO, NO, NO_2, O_3, and particulate matter ($PM_{2.5}$ and PM_{10}), in addition to meteorological conditions such as T_{2m}, WS_{10m}, WD_{10m} and precipitation.

Table 3.3 - Additional stations on the State of Sao Paulo

Station	City	Geo location	Latitude (°)	Longitude (°)	Period
IAG-USP	São Paulo	Urban Park	-23.65	-46.62	June 2017, July 2018, June and August 2019
Botucatu	São Paulo	Regional Urban	-22.85	-48.43	June and August 2019

3.2.4 AERONET Surface Observation Network

The AERONET (Aerosol Robotic Network) program has an extensive network of sunphotometers on a global scale that provides solar and sky spectral radiance measurements in different wavelengths (340, 380, 440, 500, 675, 870, 940, and 1020 nm) at 15-min sampling frequencies (Holben et al., 2001). Solar direct measurements are used to calculate the aerosol optical depth at 340, 380, 440, 500, 675, 870 and 1020 nm (based on Beer-Lambert-Bouguer law) (Holben et al., 1998). The 940 nm channel is used to retrieve the total precipitable water in centimeters. The sky radiance measurements along with the solar direct measurements, at 440, 675, 870 and 1020 nm, are used to retrieve the aerosol size distribution, complex refractive index, single scattering albedo and asymmetry parameter using the methodology of Dubovik et al. (2000). AERONET provides three quality level of data (level 1.0, 1.5 and 2.0) and here we selected level 1.5 (cloud screened) and 2.0 (cloud screened and quality-assured) data to validate the satellite and model data. These measurements are freely available on the AERONET website (*https://aeronet.gsfc.nasa.gov/*). For this study, over our study area (Figure 3.6), we have considered four AERONET stations (see Table 3.4) and a study period of six years (from 2014 to 2019). For the sake of comparison, new AERONET AOD values at 550 nm were calculated based on the Angstrom expression (Angstrom, 1929; Eck et al., 1999):

$$\alpha = -\frac{ln(\frac{\tau_{a_2}}{\tau_{a_1}})}{ln(\frac{\lambda_2}{\lambda_1})} \tag{3.1}$$

where λ is the wavelegth (in nm) and τ_a is the AOD. The subscripts 1 and 2 refer to two different wavelengths (440nm and 675nm).

Table 3.4 - AERONET stations over southeastern Brazil

Station	City	Metropolitan Region	Geo location	Population[2]	Latitude (°)	Longitude (°)	Period	Number of data	Level
São Paulo	São Paulo	São Paulo	Urban	21,571,281	-23.561	-46.735	2014-2019	16259	2.0
SP-EACH	São Paulo	São Paulo	Urban		-23.482	-46.500	2016-2019	12114	2.0
Cachoeira Paulista	São Paulo	Vale do Paraíba	Rural	33,581	-22.689	-45.006	2016-2019	9127	1.5
Itajuba	Minas Gerais	South of Minas	Semi-Urban	97,334	-22.413	-45.452	2014-2019	13074	2.0

[2] https://www.ibge.gov.br/explica/codigos-dos-municipios.php.

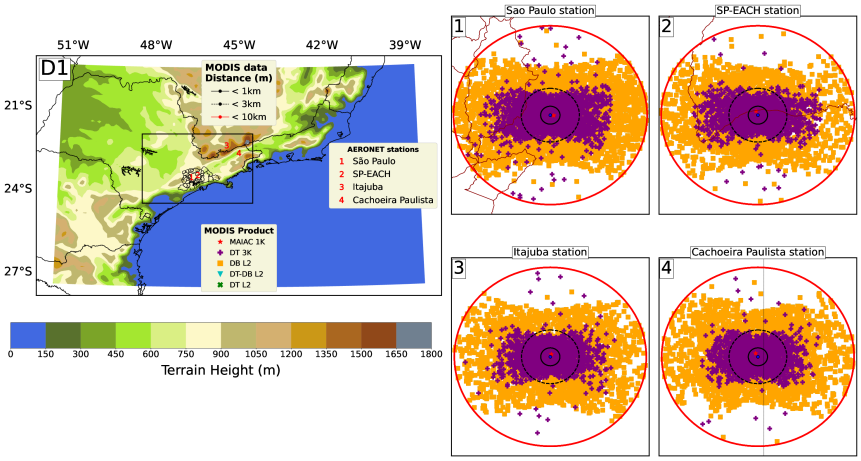

Figure 3.6: Passages of the Terra and Aqua satellites (from 2014 to 2019) close to the AERONET stations (distance less than 1km, 3km and 10km) distributed (numbers in red) over the study domain (D1).

3.3 Satellite data

3.3.1 MODIS sensor

The Moderate Resolution Imaging Spectroradiometer (MODIS) sensor, on board the Terra (passing over the Equator at 10:30 local time - LT) and Aqua (passing over the Equator at 13:30 LT) satellites, was designed to monitor the atmosphere, ocean and land surface with its 36 spectral bands (visible-thermal infrared). This instrument measures spectral radiance at a spatial resolution of 250m (bands 1-2), 500m (bands 3-7) and 1000m (bands 8-36), every 1-2 times per day (Levy et al., 2013). Through these measurements, the MODIS sensor uses different algorithms to derive aerosol properties, including aerosol optical depth (AOD), such as: Dark Target (DT), Deep Blue (DB) and MAIAC. These algorithms are the most commonly used MODIS products to conduct air quality studies. All of these products have level 2.0 data processing and can be downloaded freely from the LAADS DAAC website (*https://ladsweb.modaps.eosdis.nasa.gov/search/*). All MODIS data products are provided in Hierarchal Data Format Files (HDF), and are labeled MODXX for Terra and MYDXX for Aqua. Each HDF file provides metadata and scientific data sets (SDSs). SDSs may be multi-dimensional (e.g., across-track×along-track×bands). The L2 aerosol product files are known as MOD04 (Terra) and MYD04 (Aqua), or MxD04 (in general).

3.3.1.1 Aerosol Products

a) **DT Products**

The DT algorithm estimates the AOD (in a clear-sky scene) based on the assumption that the surface reflectance in two visible bands (440 and 644nm) presents a good linear relationship with the surface reflectance in the shortwave infrared band (2119 nm) in dense vegetated areas, given that the measurement in the shortwave infrared band is transparent with the aerosol particle (Kaufman et al., 1997; Levy et al., 2013). The DT algorithm groups and selects pixels, including cloud masking and filtering out bright and dark areas. It separates the surface from the atmosphere, considers distinct regional aerosol models based on climatological studies that combine observed TOA spectral reflectance with look-up tables (LUTs), and finally infers the ambient conditions of the aerosol (model weights and total AOD). The DT algorithm produces two aerosol resolution products in Collection 6.0 and 6.1, 10km x 10km (known as $MxD04_L2$) and 3km x 3km (known as $MxD04_3K$). Initially, the $MxD04_L2$ was sufficient to characterize global and regional aerosol loading for climate studies (Kaufman et al., 1997; Remer et al., 2005; Levy et al., 007a,b, 2013). However, due to the expansion of MODIS aerosol product to a wide variety of purposes, it is recommended to generate a product with a finer nominal resolution, $3km$ at nadir (Levy et al., 2013; Remer et al., 2013). These two resolution products are similar in the AOD retrieval processing, except in the pixels group. The product $MxD04_L2$ organizes the pixels in groups of 20x20, whereas the product $MxD04_3K$ organizes the pixels in groups of 6x6. Both consider the visible and shortwave infrared bands at a spatial resolution of 500m (Remer et al., 2013). On the other hand, as the DT algorithm is not suitable for urban, arid, semi-arid and desert areas, the DB algorithm is required to retrieve the properties of aerosols over bright surfaces, using the blue bands (Hsu et al., 2004). In this study, we consider the $MxD04_L2$ and $MxD04_3K$ products from the Terra and Aqua latest Collection version 6.1. In DT products, the scientific datasets (SDS), named the "*Optical_Depth_Land_And_Ocean*", were considered with a QA = 3, indicating high confidence data (Levy et al., 2013).

b) **DB Products**

The DB algorithm estimates the AOD based on the hypothesis that, on bright sur-

faces (urban and desert regions), the reflectance in the Deep Blue band (412nm) is much lower than in the longer wavelength bands (Hsu et al., 2004). The 412 nm spectral band was chosen due to the contrast that exists between the cloudy and non-cloudy pixels. Over land, the surface reflectance in 412 nm band is darker compared to the reflectance at other longer spectral bands. This algorithm organizes groups of 3x3 pixels (in the 412 nm band) to minimize the problem of cloud contamination. Then, to retrieve the properties of the aerosols, in cloud-free pixels, the reflectance of the apparent surface is calculated (especially in arid and semi-arid regions) using the LUTs as a function of solar/sensor geometry, AOD and aerosol scattering/absorption properties. These calculations are achieved to find the best match with the observed TOA reflectance at 420, 470 and 650 nm. Another procedure carried out in this algorithm was to use aerosol models, since there are critical values in the apparent reflectance where the satellite does not detect the aerosol loading. Aerosol models use values that represent most of the range of representative properties of each aerosol type in different parts of the world.

The algorithm retrieves the AOD at a spatial resolution of 1km and then averages it on a scale of 10km (known as $MxD04_L2$) "retrieval product pixel"(hereafter, "retrieval") (Hsu et al., 2013). In Collection 6.0, the surface reflectance database is enhanced using knowledge of the normalized difference vegetation index, scattering angle, and season (Hsu et al., 2013). In collection 6.1, the main improvements were in radiometric calibration, dense smoke detection, artifact reduction in heterogeneous terrain, high terrain surface model, and regional/seasonal aerosol optical models (Hsu et al., 2017). In this study, we consider the $MxD04_L2$ product from the Terra and Aqua latest Collection version 6.1. The same as the DT products, SDSs named "$Deep_Blue_Aerosol_Optical_Depth_550_Land$" and "$Optical_Depth_Land_And_Ocean$", were considered with a QA = 3, indicating high confidence data (Hsu et al., 2017).

c) **MAIAC Products** The MAIAC algorithm estimates the AOD based on a dynamic time series analysis that allows separating the contribution of aerosol and clouds from the surface properties, and considers the effects of bidirectional surface reflectance (Lyapustin et al., 2018). MAIAC uses MODIS level 1B land-surface radiances in a fixed grid of 1km of spatial resolution. The gridded data are split into 1200 km tiles

and placed in a queue containing from 5 (over the poles) to 16 (over the equator) days to retrieve the Ross-Thick Li-Sparse (RTLS) bi-directional surface reflectance distribution function using measurements in the SWIR band (Lucht et al., 2000). MAIAC was devised to be more effective technique for the detection of absorbing aerosols (smoke or dust) without the need to mask them as clouds. The methodology improves cloud and snow cover and enhances aerosol retrieval on bright surface (Lyapustin et al., 012c). For aerosol retrievals, it is essential to calculate the linear spectral region coefficient (SRC). This coefficient relates the bidirectional reflectance factors (BRF) at 470 nm with the short wave infrared spectral band, 2119 nm (being this band transparent to the atmosphere, except in dusty conditions). Unlike DT and DB products, MAIAC uses eight geographically prescribed aerosol models based on aerosol climatology obtained from AERONET. Current MAIAC aerosol models are static and do not take into account seasonal variations in aerosol properties, which is one of the limitations of the MAIAC aerosol product. Finally, the AOD at 470 nm is retrieved by searching the minimum spectral residual between the theoretical TOA reflectance of the LUT and the measurements in the red and SWIR bands (Lyapustin et al., 2018). The AOD is originally retrieved at 470 nm, then it is calculated at 550 nm considering the AOD at 470 nm as a function of the spectral properties, expressed by the regional aerosol model (Lyapustin et al., 2018). The high spatial resolution of MAIAC AOD retrieval improves the ability to determine the characteristics of fine aerosols and distinguish aerosol sources (Mhawish et al., 2018). In this study, we consider the $MDC19A2$ or $MAIAC_1km$ product from the Terra and Aqua latest Collection version 6.1. Same as the DT and DB products, SDSs named "$Optical_Depth_055$", were considered with a QA = 3, indicating high confidence data.

Overall, five algorithms were considered (MAIAC-1km, DT-3K, DT-L2, DB-L2 and DT-DB-L2) over our study domain for a period of six years (2014-2019). The DT-DB-L2 algorithm is the combination of DT and DB, and its validation, statistically, is similar to the DT algorithm (Ali et al., 2019; Ibrahim and Halounova, 2019; Li et al., 2020).

3.3.2 OMI

In 2004 the Aura spacecraft was launched with the purpose of studying the earth's ozone layer, air quality and climate. This satellite is a polar sun-synchronous orbit with an equator-crossing time of 13:45 LT (ascending node). In addition, this satellite has four instruments to study atmospheric chemistry and dynamics, one of them being the OMI (Aura, 2021). OMI (Ozone Monitoring Instrument) is an imaging spectrometer that measures solar light back-scattered by the Earth's atmosphere and surface by covering ultraviolet (270 nm-380 nm) and visible (350 nm-500 nm) spectrum (Joiner and Bhartia, 1995). This instrument is based on the TOMS (Total Ozone Mapping Spectrometer) and GOME (Global Ozone Monitoring Experiment) instruments of NASA and ESA (European Space Agency), respectively. The diference is that OMI can measure more atmospheric constituents than TOMS and its spatial resolution (13km x 24km) is better than of GOME (40km x 320km). On the other hand, OMI uses different algorithms to measure O_3 and atmospheric trace gases (such as NO_2, SO_2, HCHO, ClO_2, Br_2O and others). Some of them in the entire atmospheric column and others in the stratosphere and troposphere. For this work, we have considered retrieval of geolocated tropospheric VCD (vertical column density) NO_2 (in units of molecules cm^{-2}) extracted from the DOMINIO (Dutch OMI NO_2) v2.0 algorithm through the EU FP7 QA4ECV project. This project has led the development of a new OMI NO_2 data product (Boersma et al., 2011). The DOMINIO data constitute in Level 2 products provide geophysical information for each and every ground pixel observed by the instrument. The product also contains intermediate results such as: a priori NO_2 profile, averaging kernel, among others, which will be explained in the next chapter. Finally, these data are available at www.temis.nl. DOMINIO v2.0 data are stored in version 5 of the EOS Hierarchical Data Format (HDF-EOS). Each file contains data for the diurnal portion of the orbit ($\approx$ 14 orbits).

3.3.3 MOPITT

The MOPITT (Measurements of Pollution In The Troposphere) instrument, launched in 1999 on board the NASA Terra satellite, provides the longest global record of tropospheric CO available to date (2000-present), with a spatial resolution of 22x22 km^2 at nadir (Deeter et al., 2003). This instrument is a cross-track scanning gas correlation radiometer with sun-

synchronous orbit at 705 km of altitude and with satellite overpass time at approximately 10:30 LT (ascending node) and 22:30 LT (descending node). MOPITT measures radiance at both TIR band near 4.7 μm and the NIR band near 2.3 μm and uses these measurements to determine the CO concentration profile using an optimal estimation-based retrieval algorithm and a fast radiative transfer model (Deeter et al., 2003; Edwards et al., 1999). This study uses version 9.0 Level 3 MOPPITT data products from the TIR band which provide CO mixing ratios at 10 pressure levels between the surface and 100 hPa, with a difference of 100 hPa between the levels and with a spatial resolution of 1 degree. The data were downloaded from the EARTHDATA Atmospheric Science Data Center (ASDC) website[3] for three of the four study periods (June 2017, July 2018 and June 2019). There are no data available for August 2019.

3.3.4 GPM

The GPM (Global Precipitation Measurement) mission is an international network of satellites launched in 2014 by NASA and JAXA space agencies. It provide the global observations of rain and snow to understand Earth's water and energy cycles. The Integrated Multi-satellite Retrievals for GPM (IMERG) Final Run is an algorithm prepared to calibrate, mix and interpolate precipitation estimates from all passive-microwave instruments in the GPM Constellation (Huffman et al., 2015). IMERG covers the globe from 60° S to 60° N with a spatial resolution of 10km and a temporal resolution of 30 minutes and has three runs: early, late, and final, with different latency and accuracy. The early run (delay of 6 hours) employs morphing techniques that use the information from geostationary infrared data at a fine time scale to complete the gaps in coverage of microwave overpasses (Joyce et al., 2004). The late run (delay of 18 hours) allows additional microwave observations to be used in the morphing techniques. The final run (delay of 4 months), is adjusted to the monthly gauge measurements. This study uses version 6 GPM Level 3 IMERG Final Run (GPM_3IMERGDF) product with a daily temporal resolution and a spatial resolution of 10km. This product represents the final estimate of daily accumulated precipitation in mm. Data, over the entire domain for the four periods (June 2017, July 2018, June and August 2019), were downloaded from the GES DISC website[4].

[3] https://asdc.larc.nasa.gov/data/MOPITT/MOP03T.009/

[4] https://disc.gsfc.nasa.gov/datasets/GPM_3IMERGDF_06/summary

Table 3.5 - Satellite Configurations

Characteristics \ Satellite	Terra		Aqua	Aura	GPM
Sensor	MODIS	MOPITT	MODIS	OMI	IMERG
Product	MODxx	MOP03T	MYDxx	DOMINIO v2.0	$GPM_3IMERGDF$
Level	Level 2	Level 3	Level 2	Level 2	Level 3
Local equator crossing time	10:30	10:30/22:30	13:30	13:45($\pm$ 15 minutes)	-
Spatial Resolution	1, 3 and 10 km	1 degree	1, 3 and 10 km	13x24 km^2	10x10 km^2
Temporal Resolution	1 day	2 day	1 day	1 day	1 day
Repeat Cycle	16 days	16 days	16 days	16 days	-
Pollutants	AOD	CO	AOD	NO_2	Precipitation

Chapter 4

Methodology

This section will show both the preparation of ground-based (CETESB, Botucatu, IAG-USP and AERONET network) and column (MODIS sensor, MOPPIT, OMI and GPM) observational data, and the post-processing performed on the model outputs for the comparison model-observation. A set of meaningful statistical indices will be used to evaluate the performance of the WRF-Chem model. Furthermore, a specific methodology describing how the AOD from AERONET and MODIS was compared is also presented.

A schematic representation of the modeling system, including the pre-processing tools, is shown in Figure 4.1. The modeling framework is composed of three principal components: pre-processing, processing, and post-processing. The pre-processing, also known as WRF Preprocessing System (WPS), defines the model horizontal domain, horizontally interpolates static geographical data and masked surface fields to the model grid, and prepares the meteorological, background concentrations, and emission fields to be read during the numerical integration. The processing is a program that correctly incorporate the external fields prepared during the pre-processing, and then accurately simulate the state and chemical composition of the atmosphere. Finally, a procedure designed to post-process the model outputs is applied.

4.1 Metrics to assess photochemical model performance

Being confident that the model is capable of effectively producing a proper spatial and temporal distribution of meteorological variables and chemical species is not an easy task. That is the reason why it has been considered to use statistical evaluation metrics. It has

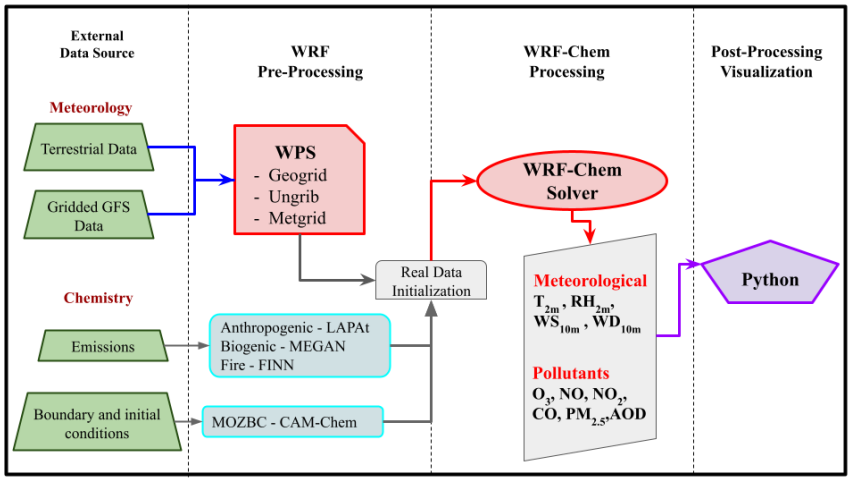

Figure 4.1: Flowchart representing the basic steps for running WRF-Chem model. Regional (MEGAN and FINN) and local (LAPAt) emissions are considered to include the chemical and aerosol modules in the WRF modeling.

also been necessary to consider periods with certain characteristics in order to fulfill the main objective of the work: negligible external source to the study domain, low presence of clouds, atmospheric stability and high contaminant concentration index. Our chosen metrics include mean bias (MB), mean error (ME), and Pearson correlation coefficient (R). These statistical indicators were calculated using the following equations (Emery et al., 2017):

a) **Pearson correlation coefficient (R)**

The Pearson correlation coefficient is also known as the bi-variate correlation simply as the correlation coefficient. It is a measure of linear correlation between two sets of data, and ranges between -1 and 1. An absolute value of exactly 1 implies that a linear equation describes the relationship between two sets of data perfectly. A value of 0 implies that there is no linear dependency between the variables.

$$R = \frac{\sum_{i=1}^{N}(pred_i - \overline{pred_i})(obs_i - \overline{obs_i})}{\sqrt{\sum_{i=1}^{N}(pred_i - \overline{pred_i})^2 \sum_{i=1}^{N}(obs_i - \overline{obs_i})^2}}, \quad Unitless \quad -1 \leq R \leq +1 \tag{4.1}$$

b) **Root Mean Square Error (RMSE)** evaluates the error of model simulation compared to observed data with an emphasis on outliers. RMSE is a way of measuring how good our predictive model is, the smaller the RMSE, the better the model will perform.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(pred_i - obs_i)^2}{N}}, \qquad 0 \leq RMSE \leq \infty \tag{4.2}$$

c) **Mean Bias (MB)**

Evaluates the average difference between the simulated and observed data. It provides information about the tendency of the model to overestimate or underestimate a variable, it quantifies the systematic error of the model.

$$MB = \frac{\sum_{i=1}^{N}(pred_i - obs_i)}{N}, \qquad -\infty \leq MB \leq \infty \tag{4.3}$$

The mean bias is normalized to make it dimensionless.

$$NMB = \sum_{i=1}^{N}(\frac{pred_i - obs_i}{obs_i})x100, \qquad -100\% \leq NMB \leq +\infty \tag{4.4}$$

d) **Mean Error (ME)**

Another statistic parameter that describes information similar to the RSME is the mean error defined as:

$$ME = \frac{\sum_{i=1}^{N}|pred_i - obs_i|}{N}, \qquad 0 \leq ME \leq \infty \tag{4.5}$$

To take into account the weight of the error with respect to the value of the measured variable, the absolute error is normalized, taking the normalized mean error (NME):

$$NME = \sum_{i=1}^{N}\frac{|pred_i - obs_i|}{obs_i}x100, \qquad 0\% \leq NME \leq \infty \tag{4.6}$$

e) **Index Of Agreement (IOA)**

To complete the information provided by these statistics, an index that provides information on the behavior of the model when comparing the results with observations, has also been calculated. The IOA is calculated as:

$$IOA = 1 - \frac{\sum_{i=1}^{N}(obs_i - pred_i)^2}{\sum_{i=1}^{N}(|pred_i - \overline{obs}_i| + |obs_i - \overline{obs}_i|)^2}, \quad Unitless \quad 0 \leq IOA \leq +1 \tag{4.7}$$

The $pred_i$ and obs_i are the simulated and observed data, respectively. The $\overline{pred}_i$ and $\overline{obs}_i$ are the corresponding mean values. "N"is the number of paired samples. R = 1 is perfect correlation; R = 0 is totally uncorrelated.
In this work three tools are considered: observational, satellite and model data. Therefore, three types of comparisons were considered to evaluate the performance of the model:

- The first evaluation is made between the observational from ground based measurements ("obs") versus satellite ("pred") data.
- The second evaluation is made between the observational from surface measurements ("obs") versus model ("pred") data, and
- The third evaluation is made between the satellite ("obs") versus model ("pred") data.

4.2 Observational data and WRF-Chem model

Air pollution and meteorological parameters inferred from the WRF-Chem simulations were extracted for the closest grid points to the AQS, IAG-USP and Botucatu stations (see Tables 3.2 and 3.3). The AQS are distributed throughout the State of São Paulo. Due to the large number of AQS, both observed and simulated parameters were grouped into and then averaged assuming four categories of land use: regional urban, urban park, urban and industrial (see Table 3.2).

To evaluate the performance of the WRF-Chem model when reproducing the pollutant concentrations, we first evaluate the simulated meteorological variables with observed data (CETESB, IAG-USP and Botucatu), highlighting the fact that local meteorological variables strongly affect the transport and mixing of trace gases and aerosols (Banta et al., 2005; Darby, 2005). This work follows the recommendations of Monk et al. (2019) (Table 4.1) and Emery et al. (2017) (Table 4.2), to evaluate the photo-chemical model considering

a simple and complex terrain. The benchmarks for simple terrain were mostly developed for flat terrain and simple meteorological conditions. Instead, the benchmarks for complex terrain were developed for more complex conditions, considering mountainous terrain. The purpose of the benchmarks is to understand how good or bad our results are relative to other model applications run for Brazil. We have considered the most representative statistical indices (see section 4.1) to test them with the benchmarks.

Table 4.1 - Meteorological parameter benchmarks

Variable	Terrain Type	MB	ME	RMSE	IOA
T_{2m} ($^{\circ}C$)	Simple Terrain	$\leq \pm$ 0.5	$\leq$ 2		$\geq$ 0.8
	Complex Terrain	$\leq \pm$ 1.0	$\leq$ 3		
RH_{2m} (%)	Simple Terrain	$\leq \pm$ 10	$\leq$ 20		$\geq$ 0.6
WS_{10m} (ms^{-1})	Simple Terrain	$\leq \pm$ 0.5		$\leq$ 2	$\geq$ 0.6
	Complex Terrain	$\leq \pm$ 1.5		$\leq$ 2.5	
WD_{10m} (degree)	Simple Terrain	$\leq \pm$ 10	$\leq$ 30		
	Complex Terrain	$\leq \pm$ 10	$\leq$ 55		

Table 4.2 - Recommended benchmarks for photochemical model performance statistics

	NMB		**NME**		**R**	
Species	Goal	Criteria	Goal	Criteria	Goal	Criteria
1-hr O_3 (μg m^{-3})	$< \pm$ 5 %	$< \pm$ 15 %	$< \pm$ 15 %	$< \pm$ 25 %	$>$ 0.75	$>$ 0.50
24-hr $PM_{2.5}$ (μg m^{-3})	$< \pm$ 10 %	$< \pm$ 30 %	$< \pm$ 35 %	$< \pm$ 50 %	$>$ 0.70	$>$ 0.40

4.3 AERONET network - MODIS and WRF-Chem model

For decades, MODIS data validation has been made using AOD data from AERONET stations, using the matching method. This type of method uses the closest pixel (granule of each satellite) to the location of an AERONET station and the AERONET data in an approximate satellite overpass time. Due to the lack of satellite and AERONET data and to increase the matchup data number, a spatiotemporal window has been adopted, following

the approach suggested by Ichoku et al. (2002). This methodology averages the satellite AOD data in a spatial window around the AERONET stations (1 km, 3 km and 10 km) so that the AERONET data is averaged in the temporal window centered on the satellite overpass time (30, 60 and 120 minutes). The reason for choosing these spatial windows is because we used satellite data with a resolution of 1km up to 10km (Figure 3.6), allowing us to include as much information as possible. The simulated AOD is obtained at 300-, 400-,600- and 1000-nm wavelengths, at each layer of the atmosphere (35 levels) and every hour (from 00:00 UTC to 23:00 UTC). A number of previous analyses were necessary to make the simulated AOD comparable to that one provided by the AERONET and MODIS estimates. The first step was to calculate the AOD at a wavelength of 550nm considering the same method used with the AERONET data (equation 3.1, using AOD at 400 and 600 nm); the second step was to integrate the AOD at the wavelength of 550nm in the entire atmospheric column (0-20km); and the third step was divided into two parts: (a) For comparison with AERONET AOD at 550nm, the simulated AOD values were extracted for the closest pixel (d< 2km) for each AERONET station (see Table 3.4), and the observed mean AOD data for the closer times (t< ±60 minutes) to the modeled AOD data was considered. (b) For comparison with the MODIS AOD, it was necessary to perform a remapping of the gridded data (see Figure 4.2). The resolution of the satellite data has been adjusted to have the same grid as the model. This procedure was performed using a package (Earth System Modeling Framework - xESMF) in python (see section 4.5 for more details). Only the spatial distribution of the simulated AOD values with "1 hour"after to the time passage of each satellite (Terra and Aqua) were selected.

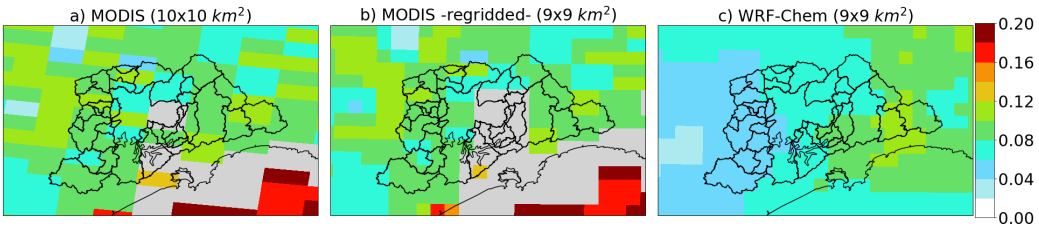

Figure 4.2: (a) AOD values from MODIS sensor at 10km spatial resolution (MOD-L2), (b)remapping of gridded MODIS data at 9km spatial resolution and (c) AOD values obtained by WRF-chem model at 9km spatial resolution.

4.4 OMI, MOPITT and GPM with WRF-Chem model

As gases and aerosols behave differently near the surface and far from it, the usefulness of satellite data to evaluate the model is relevant. The model results were estimated in 35 levels, from the surface to an altitude of 20 km, covering the troposphere and a percentage of the stratosphere. Unlike ground-based observations, satellite retrievals should not be compared directly with the model output (Kumar et al., 2012). The retrievals of trace gases depend on the relative sensitivity in different altitudes of the atmosphere, and on the a priori information about the retrieved trace gas amounts. Therefore, any modeled profile must explicitly take into account the a priori information and sensitivity of the retrieved profiles to true retrievals before its comparison with the satellite retrieval. A number of steps were employed to compare model results with the satellite data: (i) Due to the spatial resolution difference between the satellite and simulated data, it was necessary to perform a remapping of the gridded data (see section 4.5 for more details). For NO_2, the satellite data were adjusted to the spatial resolution of the model (Figure 4.3), while for CO, the simulated data were adjusted to the spatial resolution of the MOPPIT. (ii) Since that the model outputs and the satellite data have a different temporal resolution, simulated data closer to passage time of each satellite has been considered to performed the comparison. (iii) Then, the model data were mapped on to the pressure grids of the different instruments (OMI and MOPPIT). (iv) Finally, the spatially and temporally matched model results were transformed using the averaging kernel (A) and a priori profiles used in the satellite retrievals to obtain a model profile. This last procedure was performed using the following equations:

$$Y_{trop} = A_{trop}.X_{trop}, \qquad where, \;\; A_{trop} = A.\frac{AMF}{AMF_{trop}} \tag{4.8}$$

$$X_{rtv-model} = X_a + \mathbf{A}(X_{true-model} - X_a) \tag{4.9}$$

Equation 4.8 is applied for adjusting the WRF-Chem simulated tropospheric column NO_2 abundances for comparison to OMI retrievals, while the equation 4.9 is used to transform the WRF-Chem simulated column CO abundances for comparison with MOPPIT (Dekker et al., 2017). A_{trop} is the total column averaging kernel, AMF and AMF_{trop} are the air mass

factors for the total columns and tropospheric columns, respectively, Y_{trop} is the transformed model profile and X_{trop} is the tropospheric WRF-Chem NO_2 profile interpolated to the OMI pressure grid. All these variables are extracted from the satellite data of the DOMINIO v2.0 product[1]. $X_{rtv-model}$ is the retrieved state vector and X_a and $X_{true-model}$ are the a priori state vector and the true state vector, respectively. **A** (in the equation 4.9) is the retrieval averaging kernel matrix and has a size of 10x10. It represents the sensitivity of retrieved profiles to actual profiles (Kopacz et al., 2010; Kumar et al., 2012; Nandi et al., 2020; Tang et al., 2020). More information about the MOPITT products can be found in the user guide[2].

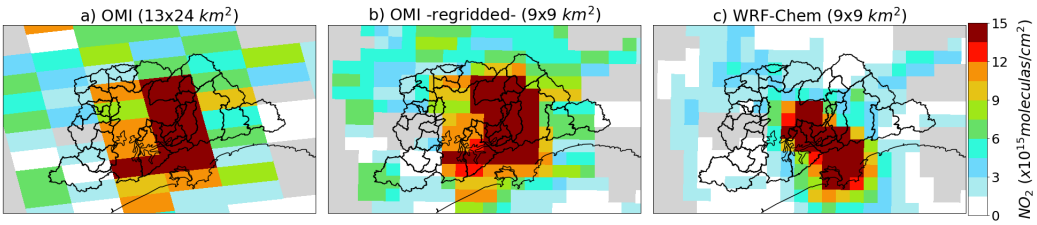

Figure 4.3: (a) NO_2 concentrations by the OMI instrument at 13x24 km^2 of spatial resolution, (b) remapping of satellite gridded data at 9x9 km^2 of spatial resolution and NO_2 concentrations obtained by the WRF-Chem model at 9x9 km^2 of spatial resolution.

In the case of the precipitation data, both the model and the GPM provide data at the surface level. The simulated precipitation data was calculated by adding the "Accumulated total cumulus precipitation (RAIN)"and "Accumulated total grid scale precipitation (RAINNC)"outputs. These fields contain the total accumulated precipitation over the simulated period. Thus, the total accumulated precipitation is simply the sum of these fields (RAINC + RAINNC). Here, the model also was adjusted to the spatial resolution of the satellite by doing the remapping of the gridded data. In relation to the temporal resolution between the model and the satellite, the daily accumulated simulated precipitation was considered as the sum between 00:00 LT to 23:00 LT each day.

[1] https://d37onar3vnbj2y.cloudfront.net/static/docs/OMI_NO2_HE5_2.0_2011.pdf
[2] https://www2.acom.ucar.edu/sites/default/files/documents/v9_users_guide_20220203.pdf

4.5 Images Remapping

Regridding is a process of interpolating from one grid resolution to a different grid resolution. There are numerous grid interpolation methods, here we have considered the "nearest_s2d"and "conservative"methods to regrid the simulated or satellite data (see Figure 4.4). These methods preserve the original coarse grid structure (although the data is now defined on a finer grid). Using an inappropriate interpolation scheme may lead to misleading results. On the other hand, most climate model grids are georeferenced on a sphere where pole singularities and longitude meridians can converge, and thus, special procedures need to be addressed. Grids used in this work are in the rectilinear and curvilinear categories. Grids with one-dimensional coordinates (latitude and longitude - 1D) are called rectilinear and grids with two-dimensional coordinates (latitude and longitude - 2D) are called curvilinear.

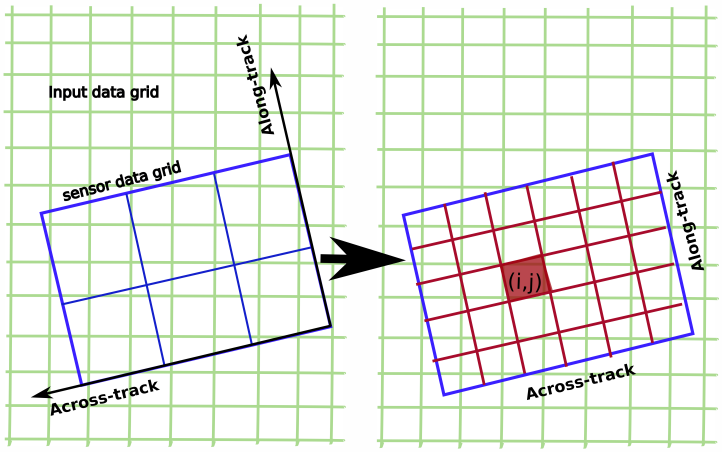

Figure 4.4: The sketch of the interpolation-remaping method.(Adapted from (Tao et al., 2014))

The considered interpolations here, uses a multi-step approach. First, we specify the type of interpolation to be performed (ex. nearest_2d and conservative) and the source and destination grid spatial coordinates. The software then generates a netCDF file which contains an array of weights. And finally, we can read the generated weights and perform a "sparse matrix multiply"to perform the regridding. This is a very fast operation and is ideally suited to processing a large number of grids.

Chapter 5

Results

The results are organized into three sections. In section 5.1, we will show the choice of the best MODIS product that represent the AOD values over the four chosen AERONET stations. This choice will be made by comparing the AOD values from MODIS products and AERONET stations by using the matching methodology among themselves. In the section 5.2 we will compare observed and simulated meteorological and chemical data, along with satellite data to assess the performance of the WRF-Chem model. Finally, in the section 5.3 we will analyze specific regional transport events of air pollutant that occurred during our study periods.

5.1 Comparison between MODIS products and AERONET stations

In order to find periods with availability of AOD data, a broad 6-year analysis (2014-2019) of both MODIS and AERONET products has been carried out over the southeastern Brazil. This analysis was performed using a matching methodology, that is, considering spatio-temporal windows (see Figure 3.6) for coincident observations between satellite and ground-truth (Ichoku et al., 2002; Martins et al., 2017). Figures 5.1 and 5.2 show the performance of MODIS sensor (for Terra and Aqua satellites, respectively), in terms of NMB and NME for the period of 6 years. Each panel represents the MODIS performance for a specific spatio-temporal window. For example, the top-left panel in Figure 5.1 represents the performance statistics for coincident observations (MODIS, AERONET) distant less than or equal 1 km and with a time span of up to ±15 minutes (the time refers to the satellite pass over the AERONET station). The temporal window changes for each column (t $\leq$ ± 15, 30 and 60 minutes) and the spatial window changes for each row (d $\leq$ 1, 3 and 10

km).

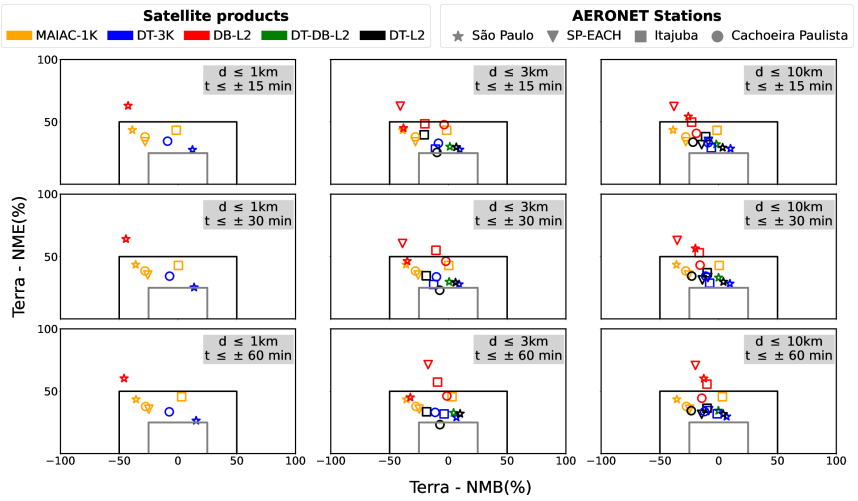

Figure 5.1: Scatter plots of NME versus NMB, representing the performance of the MODIS products (dots with different colors) for Terra satellite at four AERONET stations (different markers) for the six years of study. Each panel represents the MODIS performance at a specific time window: t $\leq \pm$ 15, 30 and 60 minutes, and a specific spatial window: d $\leq$ 1, 3 and 10 km.

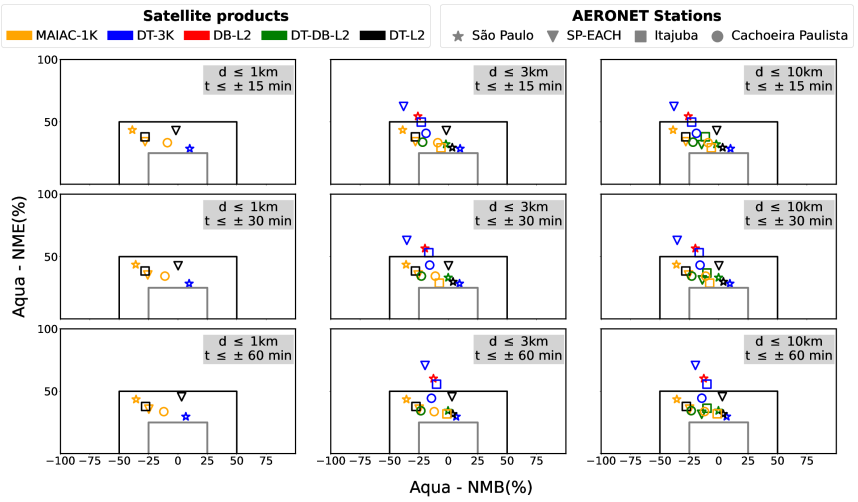

Figure 5.2: Scatter plots of NME versus NMB, representing the performance of the MODIS products (dots with different colors) for Aqua satellite at four AERONET stations (different markers) for the six years of study. Each panel represents the MODIS performance at a specific time window: t $\leq \pm$ 15, 30 and 60 minutes, and a specific spatial window: d $\leq$ 1, 3 and 10 km.

This same analysis is also considered for the Aqua satellite (see Figure 5.2). As observed in the Figures 5.1 and 5.2, the most appropriate matchup period for both Terra and Aqua satellites have an interval of $\pm$ 15 minutes using a spatial window of 10 x 10 km^2. Remembering that the most appropriate matchup is the one with NMB and NME closest to zero.

With this spatio-temporal window, a greater number of valid pairs MODIS-AERONET is found, for the y-axis, close or inside the second rectangle of each panel ($NME < 50\%$) and close to zero ($NMB \approx 0$) for the x-axis. However, DB-L2 and DT-3K products do not represent well the AOD values on the SP-EACH station (for Terra and Aqua satellites, respectively).

Considering the four seasons of the year, Summer (DJF: December, January and February), Autumn (MAM: March, April and May), Winter (JJA: June, July and August) and Spring (SON: September, October and November), and using the matching methodology, we have found different temporal and spatial windows matchups compared to the annual analysis. Figure 5.3 shows the performance of MODIS products in terms of NMB and NME for each season during the 6 years of study. We can see that the months with the highest number of coincident observations are those between March and November, mainly when considering a spatial window of 3 and 10 km (for both the Terra and Aqua satellites). Therefore, we concluded that for the autumn season (MAM), the most suitable spatio-temporal window is ± 15 minutes and 10 km^2 for Terra satellite, while a spatio-temporal window of ± 60 minutes and 3 km best suited for Aqua satellite. For the winter season (JJA), the most suitable matchup period is ± 60 minutes using a spatial window of 10 km^2 for both satellites. Finally, for the months with the greatest influence of forest fires (SON), the most appropriate spatio-temporal window is ± 60 minutes and 3 km for both satellites. Overall, the DT-3K and DT-L2 products have performed better in winter season over most of the AERONET stations in southeast Brazil, while MAIAC and DB-L2 products presented the most discrepant results. The gray boxes at the Figure 5.3 indicate that there are no coincident observations of AOD data (MODIS and AERONET) to analyze. The lack of data is found within 1 km spatial window from each AERONET station, for most of the products (except for MAIAC). We concluded that during the four seasons, the presence of rain and/or clouds over our study area was responsible for the lack of AOD values between MODIS and AERONET (summer and early fall seasons). Also, the validation approach being applied to the average MODIS retrievals within 10x10 km^2 centered at each site as well as to the average of AERONET measurements within ± 60 minutes around the time of the satellite overpass. Aerosols are highly variable and this variability increases when the atmosphere is affected by the transport from other external sources. Therefore,

during the dry season in Brazil, when biomass burning activities take place, a smaller time difference must be considered (t < ±15 minutes). By contrast, in winter the atmosphere is usually stable and has low biomass burning transport , so it is convenient to use a time limit of up to ± 60 minutes.

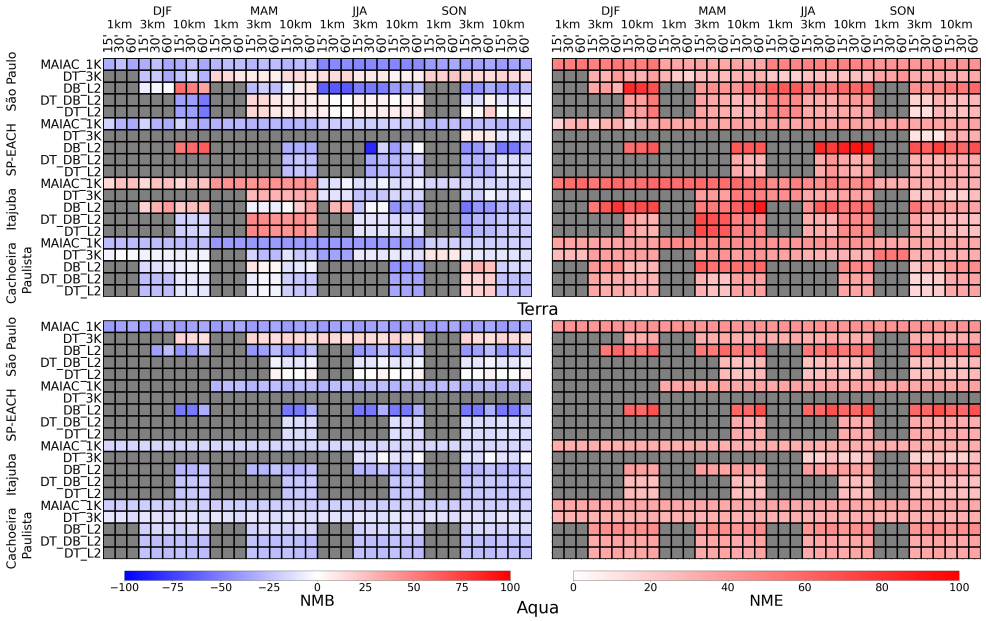

Figure 5.3: Color diagram of the NMB and NME of the spatio-temporal windows for the four AERONET stations, for the five satellite aerosol products and for each season (DJF, MAM, JJA and SON).

Once the appropriate spatio-temporal windows to validate MODIS AOD data against AERONET estimates were determined, a more detailed analysis has been carried out. Figure 5.4 shows the temporal variability (left side) and scatter plot (right side) of the AOD values measured at the São Paulo station and retrieved by the MODIS products (MAIAC, DT-3K, DB-L2, DT-DB-L2 and DT-L2) for both Terra and Aqua satellites, from 2014 to 2019. The choice to show the assessment at the São Paulo station was based on the larger amount of data available compared to the other stations. Both the AERONET and MODIS data represent average values considering the most suitable spatio-temporal window (d ≤ 10 km and t ≤ ± 60 minutes). The largest number of matchup AOD values is found between March and October, particularly from August to October when forest fires represent a dominant source of aerosols over the central Amazon (Hoelzemann et al., 2009). The emission of aerosols to the atmosphere caused by the fire events in the Amazon can be effectively transported to urban areas in southeastern Brazil, such as the MASP, causing environmental contamination over those regions (Martins et al., 2018; Vara-Vela

et al., 2018, 2021). Analyzing the temporal variability (Figure 5.4 - left side), we can observe that the MAIAC product underestimates AOD values between September and October (burning months) compared to the other products, while DB product estimates higher AOD values throughout the study period.

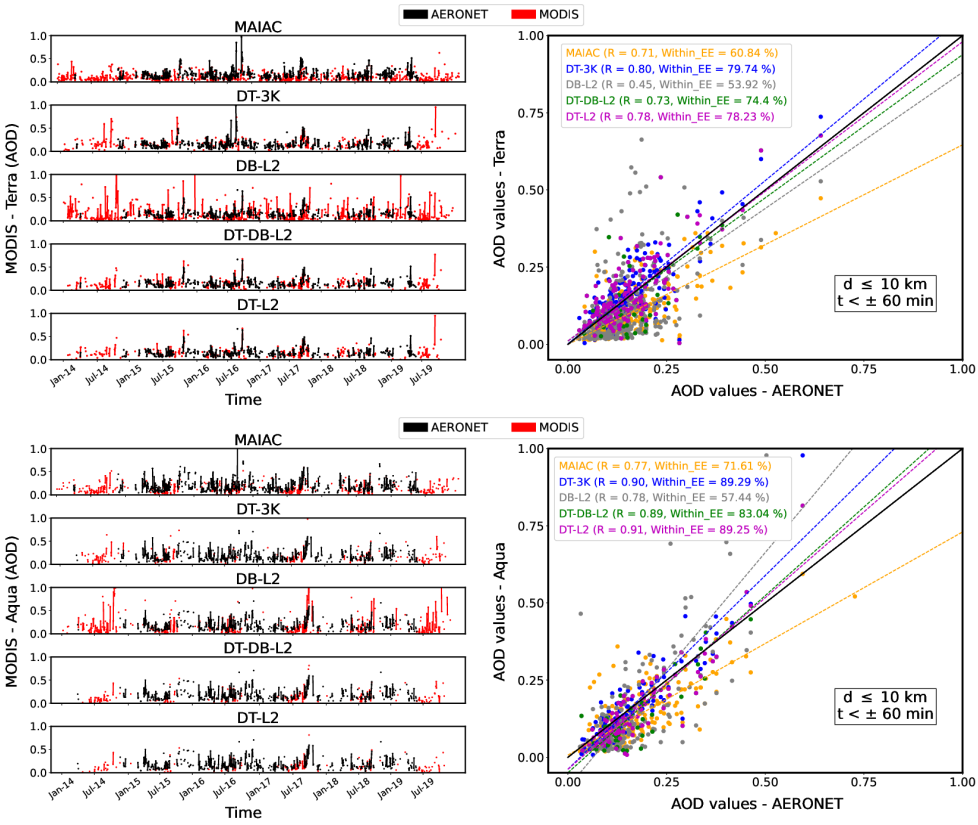

Figure 5.4: (Left side) Temporal variability (2014-2019) and (Right side) the correlation of the AOD values at 550nm obtained by the AERONET station (black line) and MODIS instrument (red line), using MAIAC (yellow dots), DT-3K (blue dots), DB-L2 (gray dots), DT-DB-L2 (green dots) and DT-L2 (magenta dots) algorithms at the São Paulo station.

Based on the correlation coefficient, for the São Paulo station (Figure 5.4 - right side), we can see that the DT (3K/L2) product represents well the AOD values (Terra: R = 0.80/78 and Aqua: R = 0.90/91), indicating the best correlation with the AERONET AOD data, while DB product presented the lowest correlation coefficient with AERONET values in the region (Terra: R = 0.45 and Aqua: R = 0.78). Despite the higher resolution of MAIAC product, it is not as efficient as the DT product (Terra: R = 0.71 and Aqua: R = 0.77) in representing the AOD values over the São Paulo station. One justification for this result is related to the aerosol model used by the MAIAC algorithm which fails over urban areas with land use characteristics different from those found in United States

cities. The MAIAC aerosol model considered over the study area is the same for the United States, dominated by non-absorbing aerosols, while in the MASP a more absorbing aerosol prevails, with lower single scattering albedo values (Damascena et al., 2020).

The basic statistics between the AOD values measured by the four AERONET stations and MODIS products (MAIAC-1km, DT-3K, DT-L2, DB-L2 and DT-DB-L2), considering the most appropriate spatio-temporal window ($t \leq \pm$ 60 minutes and $d \leq 10$ km) from 2014 to 2019, are showed in the Table 5.1. Analyzing the other AERONET stations, we can observe that the DB-L2 product performed worst in estimating AOD values at SP-EACH ($R = 0.48$) and Cachoeira Paulista ($R = 0.56$) while the MAIAC product have a worse performance at the Itajuba station ($R = 0.57$) in the afternoon (Aqua satellite). For bias, MAIAC product underestimated the observed AOD values at São Paulo, SP-EACH and Cachoeira Paulista stations, while it overestimated the observed AOD values at Itajuba station. Overall, the product that performed the best in most AERONET stations in the studied area is DT (3K and L2) with a correlation coefficient greater than 0.84 and 0.87, respectively. This analysis supports the fact that the DT-L2 product is the most suitable MODIS AOD product for studying aerosols over our study area. The DT method was developed to perform better over dark vegetated targets and does not work appropriately over bright land surfaces. In this case, most of the AERONET stations are located in green areas. MODIS Aerosol product reaches a good accuracy when more than 66% of the retrievals fall within the EE envelope (Remer et al., 2005). Using this condition, 79.7 % (Terra) , 89.3 % (Aqua) and 78.2 % (Terra), 89.3 % (Aqua) of the retrievals from DT-3K and DT-L2, respectively, fall within the EE envelope, indicating a good accuracy at the São Paulo station. Comparing the MAIAC and DB products, only 60.8 % (Terra), 71.6 % (Aqua) and 53.9 % (Terra), 57.4 % (Aqua) of the retrievals fall within the EE envelope, indicating a bad accuracy for the MAIAC and DB products. For SP-EACH, Cachoeira Paulista and Itajuba stations, the aerosol products that reach good accuracy are: DT-3K (80 %), DT-L2 (78.3 %) and DT-L2 (90.7 %), respectively, for Terra satellite. For Aqua satellite are: MAIAC (88.7 %), DB-L2 (80.3 %) and DT-3K (93.1 %), respectively. Taking into account this result, the WRF-Chem model outputs are compared with the DT-L2 product (next section).

Table 5.1 - General statistics of the validation between the AOD values measured by the AERONET stations and MODIS products, considering the most appropriate spatio-temporal window (t $\leq \pm$ 60 minutes and d $\leq$ 10 km) from 2014 to 2019 (n = Matchup Number)

Station	**Algorithm/**	**Terra**								**Aqua**							
	Resolution	n	R	MB	NMB	ME	NME	RMSE	EE (%)	n	R	MB	NMB	ME	NME	RMSE	EE (%)
Sao Paulo	MAIAC-1km	332	0.71	-0.05	-35.64	0.06	43.52	0.08	60.84	236	0.77	-0.04	-23.35	0.06	35.34	0.07	71.61
	DT-3K	153	0.80	0.01	6.70	0.05	29.68	0.07	79.74	112	0.90	0.0	2.27	0.04	24.47	0.06	89.29
	DT-L2	147	0.78	0.0	3.90	0.05	31.66	0.07	78.23	93	0.91	-0.02	-11.26	0.04	23.56	0.06	89.25
	DB-L2	293	0.45	-0.02	-12.66	0.09	60.18	0.13	53.92	195	0.78	-0.0	-3.06	0.09	52.14	0.14	57.44
	DT-DB-L2	168	0.73	-0.0	-0.23	0.05	34.38	0.07	74.40	112	0.89	-0.02	-14.22	0.05	28.71	0.07	83.04
SP-EACH	MAIAC-1km	88	0.70	-0.04	-24.78	0.05	35.68	0.07	68.18	62	0.93	-0.03	17.58	0.04	22.94	0.05	88.71
	DT-3K	15	0.49	-0.01	-8.60	0.05	34.47	0.07	80.0	3	-	-	-	-	-	-	-
	DT-L2	33	0.85	-0.03	-14.48	0.06	31.51	0.07	72.73	17	0.94	0.03	18.82	0.04	26.72	0.06	76.47
	DB-L2	98	0.48	-0.03	-19.80	0.12	70.68	0.16	35.71	56	0.60	-0.0	-2.87	0.11	61.16	0.16	42.86
	DT-DB-L2	33	0.85	-0.03	-14.48	0.06	31.51	0.07	72.73	19	0.92	0.03	20.53	0.05	30.72	0.06	73.68
Itajuba	MAIAC-1km	201	0.74	0.0	3.11	0.04	45.59	0.06	83.58	161	0.57	0.03	44.72	0.05	67.50	0.08	77.64
	DT-3K	61	0.91	-0.0	-1.26	0.04	31.90	0.08	83.61	29	0.97	-0.02	-22.16	0.04	33.48	0.04	93.10
	DT-L2	108	0.90	-0.01	-9.96	0.04	36.31	0.04	90.74	37	0.95	-0.04	-27.90	0.05	35.34	0.06	75.68
	DB-L2	171	0.73	-0.01	-10.03	0.05	55.69	0.08	82.46	150	0.81	-0.01	-15.81	0.04	46.38	0.05	82.0
	DT-DB-L2	108	0.90	-0.01	-9.96	0.04	36.31	0.04	90.74	37	0.95	-0.04	-27.90	0.05	35.34	0.06	75.68
Cachoeira Paulista	MAIAC-1km	183	0.82	-0.04	-27.53	0.05	37.82	0.07	77.05	157	0.83	-0.03	-26.39	0.05	39.89	0.06	78.98
	DT-3K	197	0.93	-0.02	-13.09	0.05	33.51	0.07	77.21	152	0.91	-0.03	-19.65	0.05	38.01	0.06	77.36
	DT-L2	201	0.78	-0.03	-24.20	0.05	35.34	0.06	78.32	140	0.87	-0.03	-22.40	0.05	39.33	0.07	74.75
	DB-L2	263	0.56	-0.02	-17.85	0.06	43.33	0.08	75.0	238	0.75	-0.02	-13.60	0.05	39.74	0.07	80.25
	DT-DB-L2	201	0.78	-0.03	-24.20	0.05	35.34	0.06	78.32	140	0.87	-0.03	-22.40	0.05	39.33	0.07	74.75

5.2 Validation of WRF-Chem against ground-level and column observations

For this study, four months were considered: June 2017, July 2018, June and August 2019, however in this section the meteorological and chemical validation for July 2018 will be shown, since the analysis was similar in the other periods. At the end of each meteorological and chemical variable analysis, a table with a statistical summary will be presented, including the validation for the four study periods.

5.2.1 Meteorological parameters

To quantitatively evaluate the performance of the WRF-Chem model for representing the concentrations of atmospheric pollutants, a model evaluation of the simulated meteorological parameters against observed data is first necessary. Considering that local meteorological variables strongly affect the transport and mixing of trace gases and aerosols, the meteorological variables analyzed in this work are: temperature, humidity, wind speed and direction and precipitation. These variables are extracted from the air quality stations and model simulations. Precipitation measurements are available only at the IAG and Botucatu stations, for this reason, model precipitation results will be compared only with data from those stations and from IMERG product. Here the analysis follows the recommendation of Monk et al. (2019), to evaluate the photo-chemical model over a simple and complex terrain. After calculating the statistical indices (discussed in section 4.1), these are compared with the benchmarks (see Table 4.1).

5.2.1.1 Temperature

Temperature plays an important role in air pollution due to the temperature dependence of photo-chemical and aerosol processes (affecting air movement), as well as vertical dispersion from buoyancy due to the surface heating. With the aim of better understanding how well the WRF-Chem model represents the meteorological variables over southeast Brazil, the analysis assesses its performance based on the comparison between observational and simulated data. Figures 5.5 and 5.6 show the temporal evolution and scatter plot of the hourly T_{2m} extracted from the WRF-Chem model and AQS and IAG-USP station for July 2018. We can see that the model reproduced much better the temporal variability of the

T_{2m} (R = 0.95), despite still having difficulties in representing the maximum and minimum peaks. In addition, the model was able to represent the sharp drops in temperature during the days 10-11, 21-22 and 23-24 July 2018. The last two temperature falls were only observed over urban and urban park (including the IAG-USP station) regions. The temperature decrease was caused by frontal systems that acted over southern and southeastern Brazil. On 10 July, this system acted over the State of São Paulo, causing colder temperatures than usual. Other system began to act between the north of Bolivia and Paraná on 20 July, although with weak convection activity, causing only an increase in cloudiness and then decrease in temperature. Right away, this front rapidly moved towards the ocean, due to its weakening and the presence of a trough in 500 hPa on 21 July, although continued to favor the increase in cloudiness and the temperature fall on the east of São Paulo and the south coast of Rio de Janeiro. Finally, on 24 July, a cold front reached the east and coast of São Paulo, leaving cold air masses and causing a temperature fall. The model was able to represent the fall in temperature registered in the urban and urban park regions.

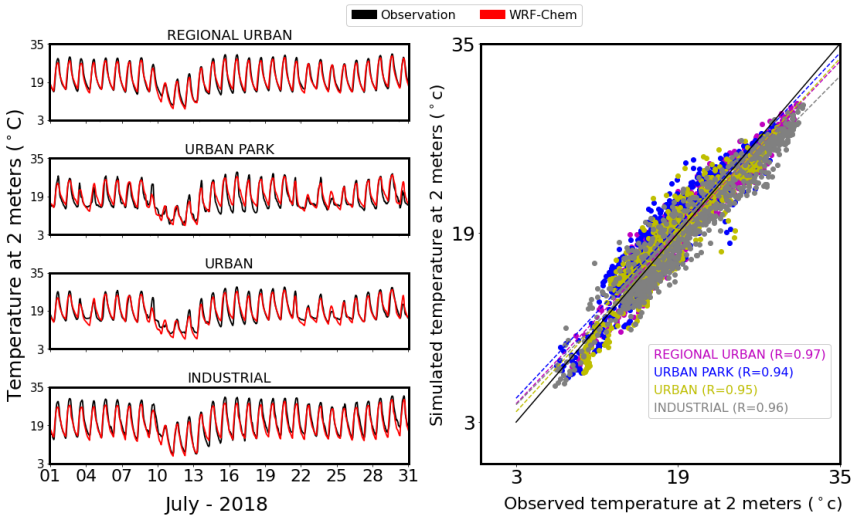

Figure 5.5: (Left panel) The observed and simulated T_{2m} temporal variability and (right panel) its correlation in the four categories of land use: Regional urban, urban park, urban and industrial for July 2018. The measured and simulated results are represented by the black and red line, respectively.

In terms of statistics (see Table 5.2), the model underestimated the data measured by the AQS on the regional urban (MB = -0.10 $°C$), urban (MB = - 0.16 $°C$) and industrial (MB = -0.81 $°C$) areas. However, the model overestimated the observed data over urban park region (MB = 0.73 $°C$), being more noticeable over the IAG-USP station (MB =1.46 $°C$). Moreover, high correlations are found for the urban region (R = 0.97) and a lower correla-

tion on the IAG-USP station (R = 0.84), but still fulfilling with the expected benchmarks recommended by (Monk et al., 2019).

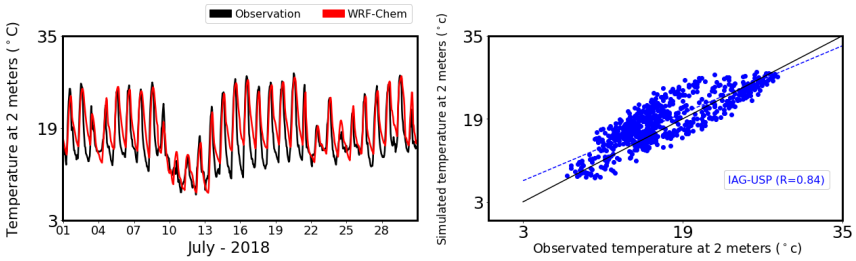

Figure 5.6: (Left panel) The observed and simulated T_{2m} temporal variability and (right panel) its correlation in the IAG-USP station for July 2018. The measured and simulated results are represented by the black and red line, respectively.

Table 5.2 - Performance statistics of the validation between predicted Temperature at 2 meters (T_{2m} in $°C$) and measurements by the AQS and the IAG-USP and Botucatu stations for the four periods.

Station	Type of Region	June 2017								July 2018							
		n	R	MB	NMB	ME	NME	RMSE	IOA	n	R	MB	NMB	ME	NME	RMSE	IOA
AQS	Regional Urban	697	0.95	-0.20	-1.07	1.19	6.25	1.49	0.97	721	0.97	-0.10	-0.48	1.06	5.19	1.32	0.98
	Urban Park	692	0.91	0.08	0.47	1.56	9.12	2.09	0.94	721	0.94	0.73	4.15	1.47	8.32	1.82	0.96
	Urban	697	0.91	-0.33	-1.88	1.53	8.70	2.08	0.93	721	0.95	-0.16	-0.85	1.21	6.50	1.52	0.97
	Industry	697	0.91	-0.74	-3.81	1.75	9.04	2.14	0.94	721	0.96	-0.81	-4.00	1.55	7.61	1.84	0.97
IAG-USP	Urban Park	696	0.84	0.41	2.43	2.04	12.13	2.61	0.90	720	0.84	1.46	8.58	2.43	14.32	3.03	0.89
		June 2019								August 2019							
AQS	Regional Urban	649	0.96	-0.44	-2.17	1.06	5.19	1.28	0.98	718	0.96	0.35	1.73	1.12	5.53	1.50	0.98
	Urban Park	629	0.94	0.26	1.45	1.04	5.67	1.35	0.97	713	0.95	0.98	5.87	1.40	8.37	1.79	0.97
	Urban	649	0.94	-0.52	-2.75	1.13	5.94	1.37	0.96	718	0.95	0.11	0.60	1.20	6.80	1.52	0.98
	Industry	649	0.94	-0.90	-4.46	1.49	7.35	1.84	0.96	718	0.95	0.23	1.13	1.35	6.69	1.78	0.97
IAG-USP	Urban Park	648	0.75	0.99	5.60	2.22	12.62	2.84	0.84	718	0.88	1.11	6.78	1.96	11.92	2.58	0.93
Botucatu	Regional Urban	505	0.88	-1.17	-5.85	1.64	8.16	2.11	0.90	598	0.88	0.48	2.62	1.98	10.91	2.57	0.94

Temperature estimated with the WRF-Chem model in the four periods (see Table 5.2) presented biases at the surface, mostly within the recommended values (± 1°C). Overall, the model adequately represented the temporal variability of the observed T_{2m} and its trend for all the periods, although the model failed to represent the maximum and minimum peaks on some days of June 2017. The model underestimated the observed T_{2m}

measured by the AQS on regional urban, urban and industrial areas, while overestimated the observed T_{2m} over urban park region for almost all periods with the exception of the last period, August 2019 (see Table 5.2). The stations classified as the urban park group are Ibirapuera, Capão Redondo, Cid. Universitária-USP-Ipen and Itaquera (see Table 3.2), located near a green area (between 158 - 300 km^2). Over urban park region, the land use is poorly represented by the model, which leads to higher temperatures than those measured by these AQS. This underrepresentation is also applied to the IAG-USP station, located near a green area ($\approx$ 13 km^2), where the model presented overestimated results when compared with measurements. For IAG-USP station, the model simulated hourly temperatures with a mean bias of 1.46°C for July 2018, the highest mean bias compared to all the stations and all the study periods. For Botucatu station, on the other hand, the model underestimated the observed temperature, with a bias of -1.17°C for June 2019. The hourly values of most of the simulated temperatures were close to the observed values, with better model performance over AQS located on regional urban areas.

Compared to previous studies conducted for the State of Sao Paulo, these results presented higher correlation coefficients (Vara-Vela et al., 2016, 2018; Gavidia-Calderón et al., 2018). Vara-Vela et al. (2016) reported a correlation coefficient of 0.71 between the temperature simulated by the WRF-chem model and the IAG-USP and Interlagos stations, both located within the MASP and classified as urban park and urban areas, respectively. That work was carried out between 7 August and 6 September, 2012, a period that would resemble our fourth period (August 2019). According to the Table 5.2, we can see that the model better represented the temperature measured at the AQS (R $\geq$ 95) and the IAG-USP and Botucatu stations (both, R=0.88), considering a better configuration in the model (see Table 3.1). On the other hand, (Gavidia-Calderón et al., 2018) also evaluated the WRF-Chem model by comparing the simulated T_{2m} with 68 AQS distributed throughout the State of São Paulo. The statistical performance between simulated and observed data remained within the benchmarks recommended by Monk et al. (2019) (R = 0.95), being their results closest to this work.

5.2.1.2 Humidity

Moisture in the atmosphere influences both photo-chemistry and aerosol formation. High humidity increases the rate of harmful or toxic chemicals in the air (Health, 2006). In this subsection, the WRF-Chem model ability to simulate the observed diurnal cycle of near-surface relative humidity is analyzed in the four periods. The temporal evolution and scatter plot of the hourly RH_{2m} extracted from the WRF-Chem model and AQS and IAG-USP station for July 2018 are shown in Figures 5.7 and 5.8, respectively.

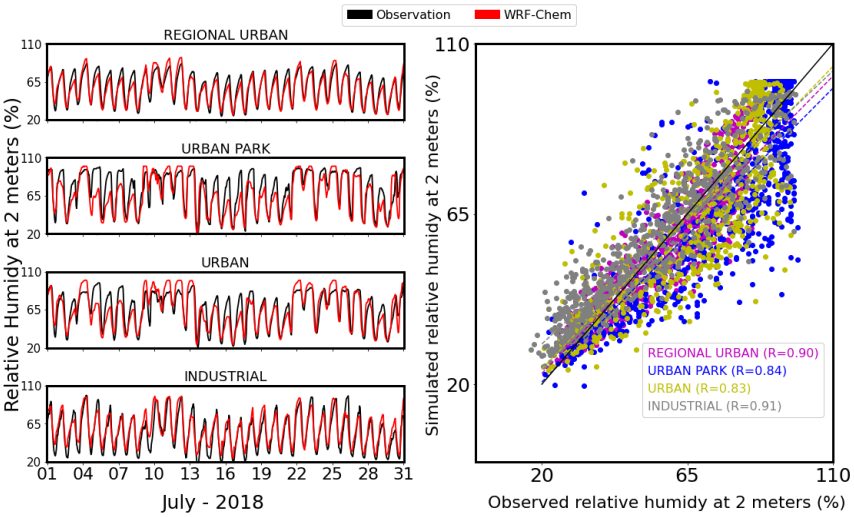

Figure 5.7: (Left panel) The observed and simulated RH_{2m} temporal variability and (right panel) its correlation in the four categories of land use: Regional urban, urban park, urban and industrial for July 2018. The measured and simulated results are represented by the black and red line, respectively.

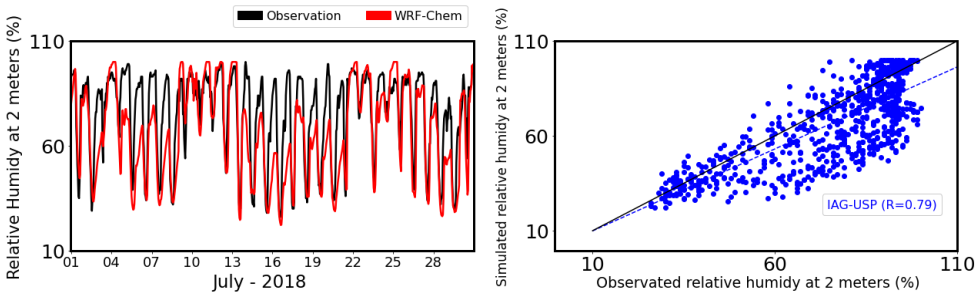

Figure 5.8: (Left panel) The observed and simulated RH_{2m} temporal variability and (right panel) its correlation in the IAG-USP station for July 2018. The measured and simulated results are represented by the black and red line, respectively.

From the time series, we can see that the model was able to consistently represent the temporal variability of the RH_{2m} registered by all the stations. The model managed to detect the strong changes of RH_{2m} in the presence of frontal systems that occurred in July

2018 (between 10-11, 21-22 and 23-24). Over the regional urban and industrial areas, the model was able to estimate the minimum and maximum values of RH_{2m}. However, over urban and urban park (including the IAG-USP station) areas, the model failed to reach the maximum RH_{2m} values as registered by the AQS.

Regarding the performance statistics (see Table 5.3), an overall good correlation between model simulations and observations was found for both AQS (R = 0.91) and IAG-USP (R = 0.79). However, the model overestimated the observed values at AQS classified as industrial stations (MB = 3.68 %) and underestimated the observed values at IAG-USP station (MB = -8.95 %), classified as urban park station. Overall, the simulated RH_{2m} in the State of São Paulo correlated quite well with the observed values (R > 0.87). The model represented the small variations in RH_{2m} in the presence of the cold front that acted on the State of São Paulo during June 2017, while during June 2019, the model had a greater deficiency to represent these RH_{2m} variations. From the results, one can conclude that the model underestimated RH_{2m} values in almost all the study periods and overestimated in July 2018 and June and August 2019 over industrial and urban areas (see Table 5.3). Using the nearest neighbor estimation technique (pixel-station), the model was deficient in representing the RH_{2m} for all the study periods (R = 0.62). This confirms that a horizontal grid resolution of 9 km does not favor the representation of surface relative humidity at individual points such as the IAG-USP station. The overall performance was better at the AQS over regional urban areas (lower error and bias) with R, RMSE and IOA of 0.91, 7.75 and 0.95, respectively.

Table 5.3 - Performance statistics of the validation between predicted Relative Humidity at 2 meters (RH_{2m} in %) and measurements by the AQS and the IAG-USP station for the four periods.

Station	**Type of Region**	**June 2017**								**July 2018**							
		n	R	MB	NMB	ME	NME	RMSE	IOA	**n**	R	MB	NMB	ME	NME	RMSE	IOA
AQS	Regional Urban	697	0.89	-2.21	-2.98	5.98	8.08	7.83	0.94	721	0.90	-2.03	-3.48	6.15	10.54	7.63	0.95
	Urban Park	692	0.84	-4.77	-5.85	8.75	10.73	12.13	0.90	721	0.84	-6.66	-9.16	10.84	14.90	14.07	0.90
	Urban	697	0.83	-1.21	-1.55	8.35	10.72	10.90	0.90	721	0.83	-1.17	-1.76	9.72	14.62	12.02	0.91
	Industry	697	0.85	-0.59	-0.82	7.81	10.78	10.38	0.92	721	0.91	3.68	6.68	7.53	13.69	9.49	0.94
IAG-USP	Urban Park	696	0.77	-4.13	-5.05	9.75	11.94	13.59	0.86	720	0.79	-8.95	-11.90	12.42	16.51	16.53	0.85
		June 2019								**August 2019**							

AQS	Regional Urban	649	0.91	0.48	0.71	5.56	8.12	7.00	0.95	718	0.92	-4.39	-6.93	6.78	10.69	8.53	0.94
	Urban Park	649	0.86	-4.85	-6.15	8.24	10.44	11.04	0.91	713	0.89	-7.36	-9.25	9.82	12.34	12.55	0.91
	Urban	649	0.85	2.05	2.84	7.89	10.92	9.81	0.92	718	0.87	0.32	0.44	8.60	12.04	10.64	0.93
	Industry	648	0.88	5.48	8.26	9.08	13.68	11.12	0.91	718	0.90	-2.44	-3.93	8.10	13.02	10.14	0.94
IAG-USP	Urban Park	648	0.57	-5.46	-6.83	13.46	16.84	17.76	0.74	718	0.36	-4.47	-5.77	18.93	24.47	23.73	0.61

Similar to temperature, the WRF-Chem model fulfilled the expected benchmarks for relative humidity for simple terrain conditions. In this work, the statistical performance of RH_{2m} was better compared to the works of Vara-Vela et al. (2016) and Gavidia-Calderón et al. (2018) (similar to the temperature). Even so, the results of Vara-Vela et al. (2016) and Gavidia-Calderón et al. (2018) fulfilled the expected benchmarks (MB $\leq \pm 10$, ME $\leq \pm 20$). The Grell 3D Ensemble and Morrison 2-Moment schemes, cumulus and microphysics parameterizations, respectively, generated less errors in the estimation of RH_{2m}.

5.2.1.3 Wind Direction and Speed

Winds influence the spatial distribution and concentrations of pollutants through processes of dispersion and advection. It is, therefore, critical to accurately characterize the air flow speed and direction for input into air quality models. According to Jiménez and Dudhia (2013), there is a relationship between model skill and the wind speed when considering wind direction. They also found large differences between simulated and observed winds at low wind speeds and over complex terrain, while the opposite was found for higher wind speeds and flatter terrain. Wind direction statistics will not be examined in this assessment, however, it is interesting to investigate how well the models capture the strong sea breeze over southeastern Brazil.

The WS_{10m} time series and the respective scatter plot, together with the observed data from the AQS and IAG-USP station for July 2018, are depicted in Figures 5.9 and 5.10, respectively. We can observe that the model simulated higher values of WS_{10m} compared to the values of WS_{10m} measured by the AQS over the four regions. The simulated WS_{10m} ranged from 0.005 to 7 $\mathrm{m}s^{-1}$, while the observed WS_{10m} ranged from 0.005 to 3 $\mathrm{m}s^{-1}$ for AQS, indicating that the winds were not strong for July 2018. Despite these differences, the model was able to represent the temporal variability of the WS_{10m} measured by the

AQS, at the same time that it accurately reproduced the local regime of winds at IAG-USP (Figure 5.10). There is an overestimation (MB = 0.93 ms^{-1}) of the wind speed in all the stations considered in this analysis.

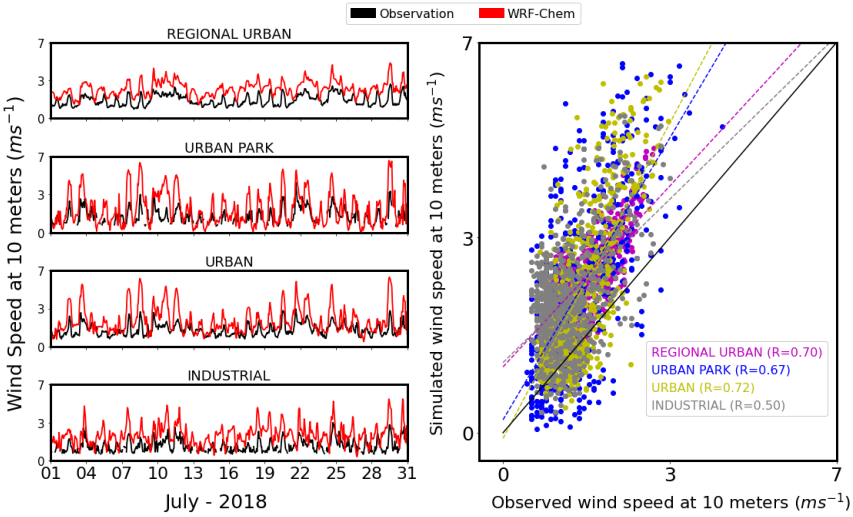

Figure 5.9: (Left panel) The observed and simulated WS_{10m} temporal variability and (right panel) its correlation in the four categories of land use: Regional urban, urban park, urban and industrial for July 2018. The measured and simulated results are represented by the black and red line, respectively.

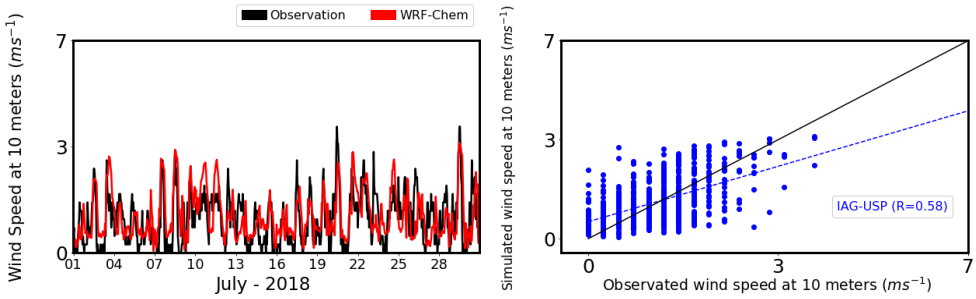

Figure 5.10: (Left panel) The observed and simulated WS_{10m} temporal variability and (right panel) its correlation in the IAG-USP station for July 2018. The measured and simulated results are represented by the black and red line, respectively.

The performance statistics for WS_{10m}, presented in Table 5.3, show a poor index of agreement (above 0.53). However, regarding MB and RMSE, they fell into the acceptable values (MB $\leq$ 1.5 ms^{-1} and RMSE $\leq$ 2.5 ms^{-1}) for complex terrain conditions. All periods show positive MB for AQS and IAG-USP stations, indicating a model overestimation, although over Botucatu it can be seen a significantly underestimation for June and August 2019.

Considering these results, the first fact that becomes clear is that the model is somewhat sensitive to whether it is simulating warm or cold periods. Also, the statistical evaluation parameters showed that the model had its best site-specific performance at IAG-USP, although better correlation coefficients were found for AQS. Based on what was just mentioned, we can conclude that the use of the AQS classification (averaged data) generates greater uncertainty than considering pixel-station (e.g. IAG-USP station). However, this problem is generated by the poor land use representation in the model.

Overall, the MM5–YSU–Noah parameterization set presented better results compared to other works executed in the State of São Paulo, which used the Monin-YSU-Noah parameterization set (Vara-Vela et al., 2016, 2018; Gavidia-Calderón et al., 2018). Some studies reported general tendency of WRF model to overestimate WS_{10m}, mainly in the plains and valleys (Cheng and Steenburgh, 2005; Randall et al., 2007; Roux et al., 2009). This shows that the model is unable to capture sub-grid scale local fluctuations. It can also be due to poor representation of surface drag or roughness length exerted by unresolved topography (Mass and Ovens, 2010, 2011).

Table 5.4 - Performance statistics of the validation between predicted Wind Speed at 10 meters (WS_{10m} in ms^{-1}) and measurements by the AQS and the IAG-USP and Botucatu stations for the four periods.

Station	**Type of Region**	**June 2017**								**July 2018**							
		n	R	MB	NMB	ME	NME	RMSE	IOA	**n**	R	MB	NMB	ME	NME	RMSE	IOA
AQS	Regional Urban	697	0.71	0.95	54.81	0.98	56.64	1.10	0.60	721	0.65	1.20	78.13	1.20	78.25	1.31	0.53
	Urban Park	693	0.67	1.36	110.37	1.53	123.47	2.01	0.58	721	0.67	1.03	91.27	1.20	106.05	1.53	0.65
	Urban	697	0.75	1.17	76.17	1.23	79.94	1.56	0.60	721	0.69	1.02	73.77	1.11	79.84	1.37	0.61
	Industry	697	0.59	1.10	96.38	1.15	100.85	1.32	0.57	721	0.49	1.28	115.88	1.34	121.51	1.54	0.52
IAG-USP	Urban Park	696	0.66	-0.01	-1.06	0.63	45.93	0.81	0.81	720	0.58	0.11	9.87	0.59	53.20	0.74	0.76
		June 2019								**August 2019**							
AQS	Regional Urban	649	0.77	1.09	66.47	1.10	66.91	1.21	0.60	718	0.77	1.13	49.25	1.14	49.85	1.27	0.64
	Urban Park	633	0.72	0.99	69.29	1.20	83.80	1.52	0.74	717	0.82	0.84	36.20	1.06	45.31	1.35	0.83
	Urban	649	0.75	1.19	83.20	1.25	87.32	1.52	0.62	718	0.79	1.32	66.75	1.35	68.16	1.62	0.60
	Industry	649	0.59	1.17	94.77	1.25	101.16	1.45	0.56	718	0.65	1.08	60.82	1.18	66.32	1.41	0.64
IAG-USP	Urban Park	648	0.38	0.19	15.95	0.81	69.36	1.02	0.62	718	0.53	0.04	2.18	0.70	42.47	0.88	0.73
Botucatu	Regional Urban	505	0.61	-1.57	-51.22	1.59	51.73	1.72	0.46	598	0.76	-1.84	-45.77	1.86	46.19	2.04	0.57

Figure 5.11 shows the temporal variability and wind roses to analyze the observed and simulated wind direction over the State of São Paulo for July 2018. The wind directions switched from east to west throughout this month, mainly in stations inside the MASP. The model was able to estimate the variations of the wind direction observed by the AQS, underestimating over regional urban (MB = -8.91°), urban (MB = -2.49°) and industrial (MB = -10.31°) areas, and overestimating over urban park (MB = 5.30°) area. The dominant observed and simulated wind direction were west-northwest wind (around 292.5°).

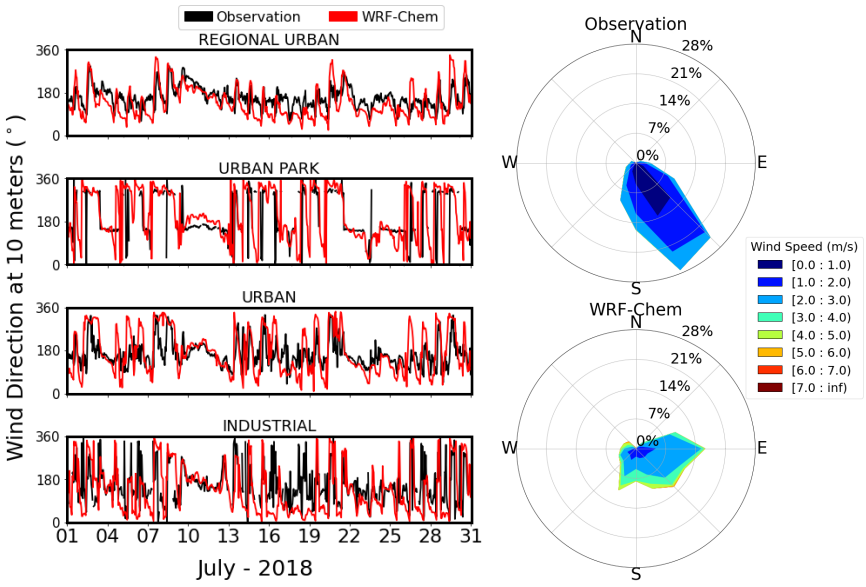

Figure 5.11: (Left panel) The observed and simulated WD_{10m} temporal variability in the four categories of land use: Regional urban, urban park, urban and industrial for July 2018. The measured and simulated results are represented by the black and red line, respectively. (Right panel) Wind rose plot of observed and simulated wind over the State of São Paulo.

5.2.1.4 Precipitation

Precipitation impacts air quality predominantly through wet deposition and aerosol scavenging. The precipitation over southeastern Brazil is often affected by the South American Monsoon System (SAMS) during the austral summer (December–February) (Carvalho et al., 2011). The South Atlantic Convergence Zone (SACZ), one of the main features of SAMS, is characterized by a northwest-southeast water vapor band that extends from the Amazon through southeastern Brazil to the subtropical South Atlantic causing extreme precipitation events in southeastern Brazil (Carvalho et al., 2002; Vera et al., 2006). However, in the dry season (June - August) the low precipitation is affected by the few cold

fronts reaching the region and by the easterly winds that transport water vapor from the Atlantic Ocean to the southeastern coast of Brazil (Carvalho et al., 2011). To evaluate daily accumulated precipitation (mm), we compared rain-gauge observations with their corresponding model and satellite grid points. Figure 5.12 shows the temporal evolution of the daily accumulated precipitation (mm) estimated by the WRF-Chem model and the IMERG Final Run V.06 product at AQS (left panel) and IAG-USP (right panel) stations for July 2018. The simulated daily accumulated precipitation values were higher than their corresponding values from IMERG product and AQS and IAG-USP stations. An overall mean bias of 0.98 mm was found, indicating that the model overestimated the observed precipitation. The model had deficiencies to represent well the local precipitation, although it was able to recognize rainy days on a regional scale.

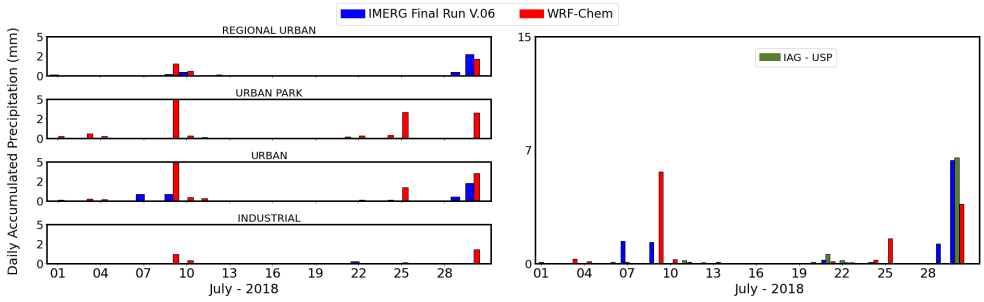

Figure 5.12: Time series of accumulated precipitation (mm) averaged (left panel) over the Regional urban, urban Park, urban and industrial regions and (right panel) over the IAG-USP station for July 2018, for the WRF-Chem model and the IMERG Final Run V.06 product.

Figure 5.13 shows the spatial distribution of accumulated precipitation over the southeastern Brazil for July 10 and 11, 2018, using the WRF-Chem model and IMERG Final Run V.06 product. On 10 July, there was an intense convective system over the Atlantic Ocean (23-26°S, 44-38°W), and that was well represented by the model. This precipitation was caused by a front that acted on the State of São Paulo and moved towards the Atlantic Ocean, remaining even on 11 July. The precipitation on 11 July was even lower than on 10 July (bottom-right panel of Figure 5.13). Overall, the displacement of short wave instabilities embedded in the baroclinic flow, which left the climate unstable, caused a strong influence on the magnitude of daily precipitation. The model had the same behavior for all the study periods, it was able to reproduce days with rain events, although large differences in accumulated precipitation exist between observations and model simulations. The spatial distribution of daily accumulated precipitation from all data sources (model-based

and satellite-based) showed higher magnitudes of precipitation over the Atlantic Ocean (latitude: 24 °S and longitude: 40 °W). The magnitude of the daily accumulated precipitation was lower over the center and coast of the State of São Paulo and Rio de Janeiro. The model was able to capture the spatial distribution of daily accumulated precipitation (mm), although it showed considerable underestimation when compared with the IMERG Final Run V.06 product over southeastern Brazil, particularly for June 2017 (MB = -0.48 mm, ME = 3.18 mm). Modeling clouds remains a challenging task implying that current models are not capable of quantitatively represent precipitation events. For June 2019, due to the start of data collection, which began on 7 June, the Botucatu station did not register data during the passage of the frontal system. In general, both the simulations and satellite estimates tend to overestimate the total accumulated precipitation observed at site-specific locations.

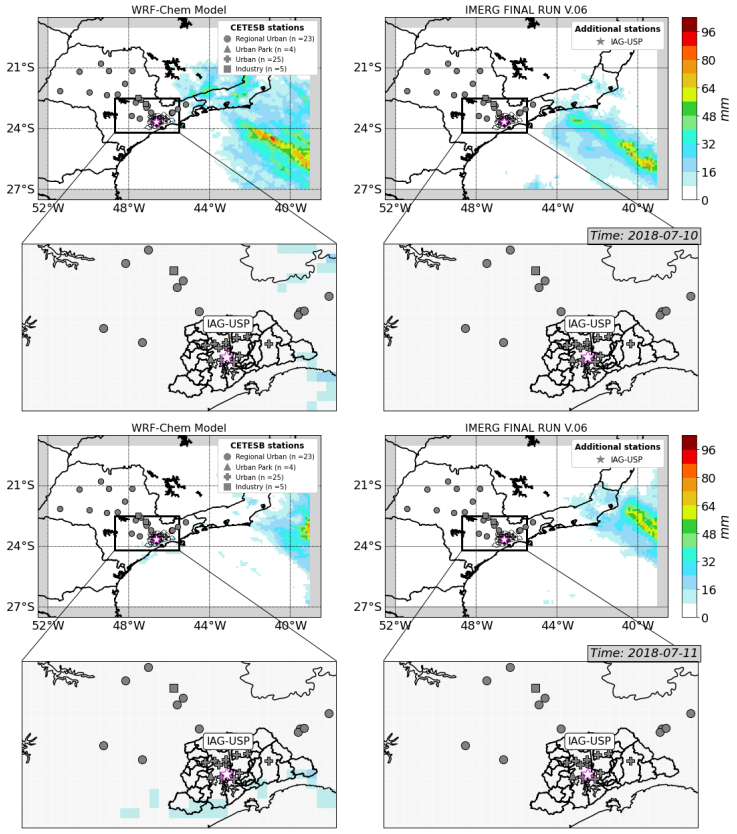

Figure 5.13: Spatial distribution of daily accumulated precipitation (mm) from WRF-Chem model (left panels) and IMERG Final Run V.06 product (middle panels) over the southeastern Brazil for July 2018. The black square represents the zoom of the MASP. The gray dots represent the AQS distributed over our study area and the magenta star represents the location of the IAG-USP station.

5.2.2 Chemical species

After evaluating the meteorological parameters and having a good agreement between the simulated and observed values, we performed WRF-Chem simulations including chemistry and aerosols. These simulations generated concentrations of several pollutants , including nitrogen oxides ($NO_x = NO + NO_2$), ozone (O_3), fine particulate material ($PM_{2.5}$) and carbon monoxide (CO), and aerosol optical depth (AOD) to be analyzed with observed (from the AQS, Botucatu and AERONET stations) and satellite (from OMI, MOPITT and MODIS instruments) data.

5.2.2.1 Comparison with ground-based observations

NO_x is a traffic-related pollutant, as it is emitted from automobile motor engines (Richmond-Bryant et al., 2017). It is a pollutant that causes irritation to the respiratory system as it penetrates deep into the lung, inducing respiratory illnesses, cough, wheezing, dyspnea, bronchospasm and even pulmonary edema when inhaled at high levels. O_3 is a strong oxidant, 52 % stronger than chlorine. It is formed in the stratosphere, but it could also be formed following chain reactions of photo-chemical smog in the troposphere. O_3 can travel to distant areas from its source of origin, moving with air masses (Villányi et al., 2010). O_3 levels over urban areas are low, in contrast to the increasing amounts occurring in rural areas which could become harmful to cultures, forests and vegetation as well as contribute to reduce carbon assimilation (Lorenzini and Saitanis, 2003; Fares et al., 2013). The temporal variability and scatter plot of NO_x concentrations simulated and measured by AQS over the State of São Paulo in July 2018 are depicted in Figure 5.14. The model was able to represent the different magnitudes of NO_x over the four types of land use regions (urban regional, urban park, urban and industrial area). High NO_x concentrations are observed over urban and urban park regions, while low concentrations of this pollutant are shown over regional urban and industrial regions. The NO_x removal by precipitation was captured by the model on July 10 and 11. Simulated and observed NO_x concentrations reached peaks of up to 250 ppb, with NO_x concentration peaks being also observed over the following nine days (between July 13-21). These high NO_x concentrations, both simulated and observed, were only registered at urban and urban park areas.

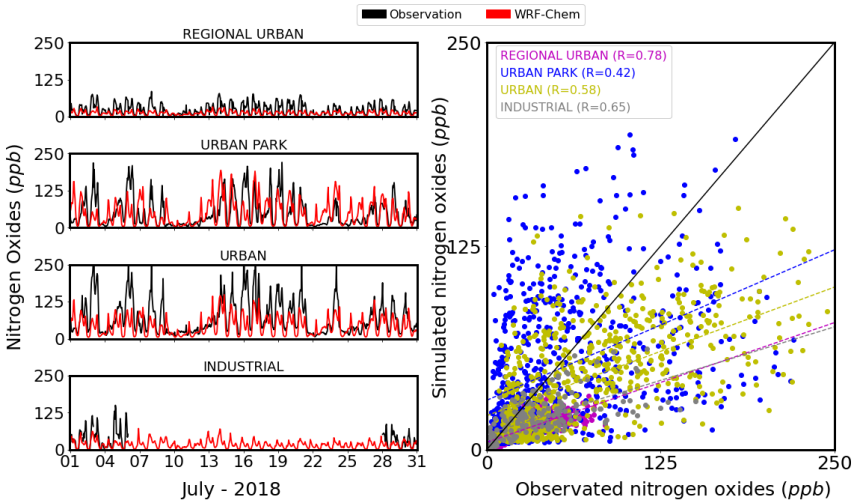

Figure 5.14: Time series and scatter plot of the observed and simulated NO_x concentrations in the four categories of land use: Regional urban, urban park, urban and industrial for July 2018. The measured and simulated results are represented by the black and red line, respectively.

Analyzing the diurnal cycle of NO_x in Figure 5.15, we can notice the morning and night peaks around 06:00-09:00 LT and 18:00-21:00 LT, respectively, generating high concentrations of up to 150 ppb over urban and urban park regions, while lower concentrations of less than 75 ppb were observed over regional urban and industrial regions. In terms of performance statistics for July 2018, the model underestimated the NO_x concentrations over all regions (see Table 5.5), although correlated well with the concentrations observed at regional urban stations (R = 0.78).

Looking at the performance statistics for all study periods in the Table 5.5, we can conclude that the simulated NO_x in the State of São Paulo had a correlation coefficient greater than 0.57, mainly in the industrial and regional urban areas, while over urban park areas the correlation coefficient was lower. The poor representation of the land use category and NO_x emissions by the inventories considered in the model somewhat justifies the low correlation between the model and the AQS on urban park areas. For almost all periods, with the exception of July 2018, the model overestimated the values observed on the urban park areas. This overestimation is related to either inaccurate representations of emissions or meteorology, while an underestimation of NO_x is related to a lack of emission sources or a combination of both. On the other hand, Botucatu is located far from any significant sources and the model was able to fairly represent the low NO_x concentrations observed at this station, showing better statistics compared to the urban and urban park areas for

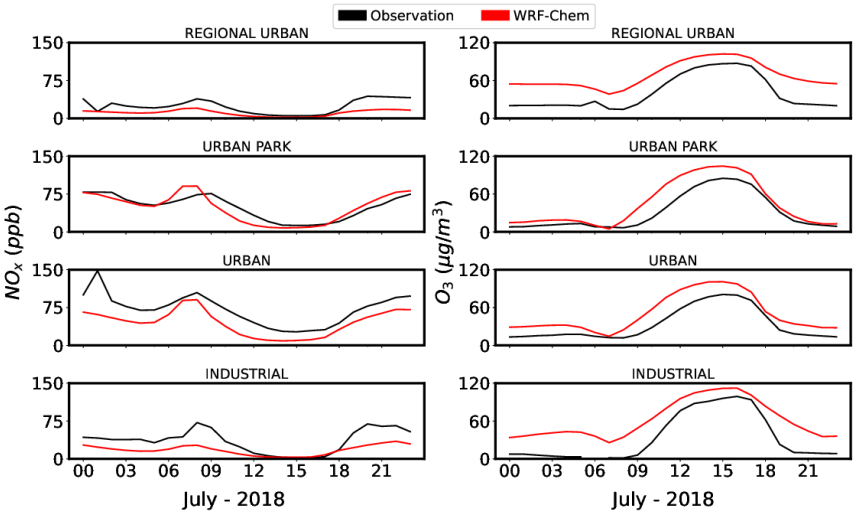

Figure 5.15: Simulated and observed hourly mean NO_x and O_3 concentrations in the four categories of land use: Regional urban, urban Park, urban and industrial over the State of São Paulo for July 2018. The measured and simulated results are represented by the black and red line, respectively.

June and August, 2019.

Table 5.5 - Performance statistics of the validation between predicted Nitrogen Oxides (NO_x in *ppb*) and measurements by the AQS and Botucatu station for the four periods.

Station	Type of Region	June 2017								July 2018							
		n	R	MB	NMB	ME	NME	RMSE	IOA	n	R	MB	NMB	ME	NME	RMSE	IOA
AQS	Regional Urban	697	0.69	-5.33	-26.63	8.47	42.32	11.70	0.76	721	0.78	-12.35	-51.74	13.06	54.74	18.18	0.62
	Urban Park	668	0.40	19.92	52.77	36.65	97.09	52.84	0.58	719	0.42	-0.41	-0.84	33.69	69.19	48.13	0.65
	Urban	697	0.48	-3.76	-6.26	34.02	56.59	45.38	0.70	721	0.58	-27.89	-39.16	37.88	53.19	55.44	0.64
	Industry	668	0.57	0.17	0.68	14.73	58.08	21.06	0.74	188	0.65	-16.77	-46.86	19.61	54.81	30.43	0.62
		June 2019								August 2019							
AQS	Regional Urban	649	0.72	-9.24	-44.41	9.91	47.63	14.64	0.63	718	0.81	-5.68	-36.50	6.82	43.82	10.69	0.74
	Urban Park	609	0.27	8.01	23.92	30.20	90.13	45.98	0.51	665	0.42	7.19	25.23	25.25	88.62	44.48	0.61
	Urban	649	0.48	-27.70	-41.94	36.98	55.98	53.51	0.57	718	0.65	-13.22	-27.49	26.12	54.31	38.87	0.76
	Industry	588	0.61	-8.72	-33.46	13.32	51.12	23.39	0.64	655	0.67	-3.06	-15.24	10.36	51.58	17.23	0.77
Botucatu	Regional Urban	308	0.59	0.36	8.02	2.26	49.64	3.13	0.76	578	0.66	1.40	45.33	2.34	75.82	3.36	0.75

Unlike CO and NO_x, which are directly emitted from surface sources, O_3 is photochemically produced and its chemical formation has a complicated non-linear relationship with O_3 precursors (CO, NO, and VOCs). Analyzing the behavior of O_3 over the four classified regions in the State of São Paulo, the model closely simulated the temporal variation

of surface O_3 at the AQS for July 2018 (Figure 5.16). The model was able to capture ozone peaks ($\geq$ 100 $\mu g\ m^{-3}$) throughout the month. It is interesting to note that the July episode is more exceptional, there are several episodes of similar magnitude between simulated and observed data. Scatter plots of O_3 concentrations over the different regions for July 2018 are shown in Figure 5.16. The model performed the best at regional urban (R = 0.89), urban (R = 0.85) regions and slightly worse at urban park (R = 0.82) region. Nevertheless, the model accurately simulated many of the high O_3 days during the warm days in comparison to cooler days with lower O_3, and the typical diurnal variation of O_3 at the four regions (see Figure 5.15). Over regional urban and industrial areas, the discrepancy between observed and simulated surface O_3 was greater at night.

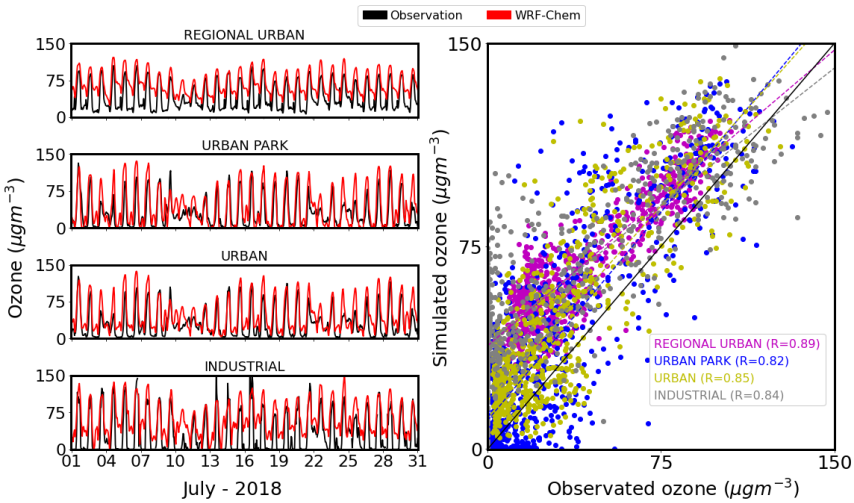

Figure 5.16: Time series and the correlation of the observed and simulated O_3 concentrations in the four categories of land use: Regional urban, urban park, urban and industrial, for July 2018. The measured and simulated results are represented by the black and red line, respectively.

For the other periods, the model also correlated better with the observed values, obtaining a higher correlation coefficient for all the stations, compared to the other chemical species (see Table 5.6). Over AQS, O_3 simulation reached the recommended goal benchmarks for the correlation coefficient ($>$ 0.75), although not fulfilled with the expected benchmarks for the bias (Table 4.2). However over Botucatu station, WRF-Chem model fulfilled with the expected benchmarks both the correlation coefficient and bias. In terms of correlation coefficient, this work showed better results compared to Vara-Vela et al. (2016, 2018); Gavidia-Calderón et al. (2018), although in terms of bias, this work shows higher biases. (Vara-Vela et al., 2016) considered two chemical-aerosol mechanisms:

RADM2-MADE/SORGAM and CB05-MADE/VBS, for two periods: August 7 to September 6, 2012 and August 19 to September 3, 2014, obtaining a mean correlation of 0.63 and 0.71, respectively. (Gavidia-Calderón et al., 2018) also evaluated the WRF-Chem model by comparing the simulated O_3 with 68 AQS distributed throughout the State of São Paulo. He used the CBMZ as a chemical mechanism but without considering the aerosol model, obtaining a mean correlation of 0.8, result within the benchmarks recommended by Monk et al. (2019). The chemical-aerosol mechanisms used were different from those considered in this work. Therefore, we can conclude that MOZART inhibits higher net O_3 photochemical production rates than RADM2 and CBMZ (Mar et al., 2016).

Table 5.6 - Performance statistics of the validation between predicted Ozone (O_3 in $\mu g\ m^{-3}$) and measurements by the AQS and the IAG-USP and Botucatu stations for the four periods.

Station	**Type of Region**	**June 2017**								**July 2018**							
		n	R	MB	NMB	ME	NME	RMSE	IOA	**n**	R	MB	NMB	ME	NME	RMSE	IOA
AQS	Regional Urban	669	0.84	21.01	74.30	21.08	74.56	24.34	0.75	721	0.65	1.20	78.13	1.20	78.25	1.31	0.53
	Urban Park	696	0.69	10.52	54.28	17.93	92.48	25.56	0.75	721	0.82	11.45	37.51	18.76	61.45	24.96	0.87
	Urban	697	0.72	10.74	49.25	16.16	74.11	22.95	0.77	721	0.85	16.61	53.08	19.51	62.34	24.00	0.86
	Industry	565	0.79	22.61	146.53	23.70	153.58	30.60	0.72	689	0.84	28.64	84.22	31.04	91.28	36.22	0.80
		June 2019								**August 2019**							
AQS	Regional Urban	649	0.90	26.83	81.50	26.83	81.52	28.85	0.73	718	0.77	16.63	37.83	17.03	38.74	20.52	0.84
	Urban Park	649	0.84	10.22	37.61	16.98	62.51	22.32	0.86	718	0.79	4.03	10.32	16.04	41.09	21.05	0.87
	Urban	649	0.86	16.79	66.67	19.24	76.38	23.60	0.82	718	0.80	10.70	30.25	16.59	46.89	21.46	0.84
	Industry	561	0.81	31.82	138.68	31.95	139.25	37.90	0.72	686	0.81	30.98	122.17	31.41	123.87	37.10	0.71
Botucatu	Regional Urban	501	0.74	7.46	10.76	12.01	17.33	15.29	0.82	587	0.68	-0.52	-0.78	11.28	16.73	15.29	0.82

Carbon monoxide is produced by fossil fuel when combustion is incomplete. The affinity of carbon monoxide to hemoglobin is much greater than that of oxygen. As a result, serious poisoning may occur in people exposed to high levels of carbon monoxide for a long period of time. Carbon monoxide affects the greenhouse gases which are tightly connected to global warming and climate. Figure 5.17 shows the temporal evolution and scatter plot of the hourly *CO* simulated and measured by the AQS for July 2018. We can observe that the model was able to represent the temporal variations and trends in the CO concentrations, mainly over the urban and urban park areas, although it did not reach the maximum peaks over urban areas. The CO removal by precipitation in July 2018, was captured by both

the model and the AQS over all the stations. The model estimated higher CO values over urban and urban park areas, with CO values up to 2 ppmv, and low CO concentrations estimated by the model were observed over regional urban areas, reaching CO values up to 0.3 ppmv, while the AQS registered CO values up to 1 ppmv. This result implied that the model's performance is found to be moderate over all AQS with correlation coefficient value from 0.45 to 0.64. In terms of bias, the model slightly overestimated the observed values over the urban park (MB = 0.07 ppmv) and greatly underestimated the urban areas (MB = -0.56 ppmv). On the other hand, over regional urban areas, the model was not able to represent the temporal variations of observed CO concentrations, underestimating them (MB = -0.32 ppmv), with a moderate correlation coefficient of 0.62.

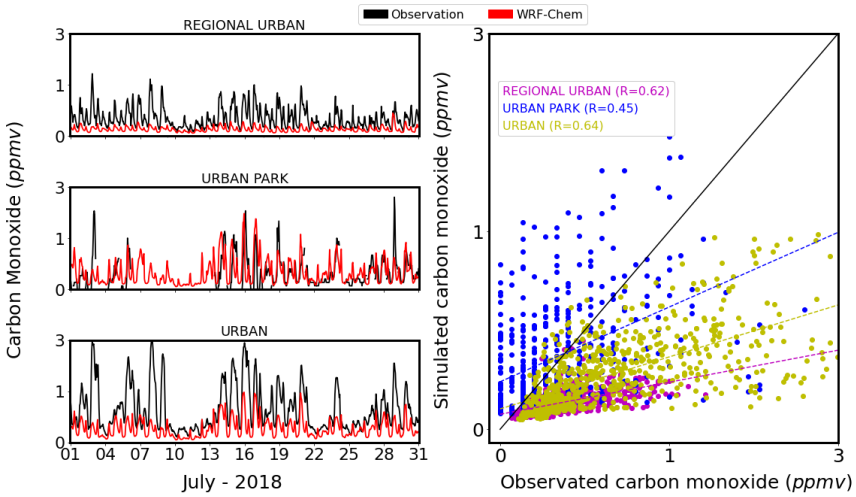

Figure 5.17: Time series and scatter plot of the observed and simulated CO concentrations in the four categories of land use: Regional urban, urban park, urban and industrial, during July 2018. The measured and simulated results are represented by the black and red line, respectively.

Analyzing the Table 5.7, for all periods, the model was able to represent the temporal variability of observed CO, mainly over urban areas, although it was not able to represent the maximum peaks. The model estimated CO values closer to the observed values over urban park areas, while in regional urban areas, the model greatly underestimated the observed values. For June and August 2019, the low correlation between the model and the data observed at the Botucatu station was due to the absence of realistic sources over that region. Possible reasons for the discrepancies between simulated and observed CO are: the anthropogenic emission in the model is represented using the local LAPAt inventory which is representative of 2011, while the biogenic emission is represented using the regional

MEGAN inventory which is representative of 2010. These inventories are six and seven years old for our study period. Due to rapid changes in the human activities in southeast Brazil, that time can contribute to large error observed over this areas. Considering which, motor vehicle exhaust currently accounts for the majority of CO emissions nationwide, and as much as 96.7 % of CO emissions in cities with high traffic congestion (CETESB,2019).

Table 5.7 - Performance statistics of the validation between predicted Carbon Monoxides (*CO* in ppmv) and measurements by the AQS and the IAG-USP and Botucatu stations for the four periods.

Station	Type of Region	June 2017								July 2018							
		n	R	MB	NMB	ME	NME	RMSE	IOA	n	R	MB	NMB	ME	NME	RMSE	IOA
AQS	Regional Urban	685	0.57	-0.45	-69.15	0.45	69.15	0.51	0.45	721	0.62	-0.32	-61.95	0.32	62.12	0.41	0.51
	Urban Park	564	0.32	0.28	108.48	0.36	138.86	0.53	0.49	506	0.45	0.07	14.54	0.32	69.07	0.46	0.66
	Urban	697	0.60	-0.36	-45.87	0.40	51.63	0.51	0.64	721	0.64	-0.56	-57.71	0.57	58.73	0.77	0.58
		June 2019								August 2019							
AQS	Regional Urban	649	0.61	-0.27	-59.81	0.27	59.92	0.35	0.49	708	0.73	-0.26	-61.11	0.26	61.27	0.33	0.49
	Urban Park	612	0.36	-0.10	-19.63	0.27	54.06	0.46	0.55	484	0.50	-0.13	-25.93	0.27	54.99	0.53	0.59
	Urban	649	0.60	-0.53	-61.34	0.54	61.84	0.71	0.53	718	0.76	-0.34	-54.18	0.35	55.74	0.51	0.61
Botucatu	Regional Urban	412	0.16	-0.45	-75.85	0.45	75.85	0.48	0.35	599	0.45	-0.55	-82.09	0.55	82.09	0.57	0.34

Fine Particulate matter ($PM_{2.5}$) contains tiny liquid or solid droplets that can be inhaled and cause serious health effects, including morbidity and mortality. In Figure 5.18, we show the time series and scatter plot of simulated and observed daily $PM_{2.5}$ concentrations for July 2018. This temporal evolution indicates that the magnitudes of simulated and observed daily concentrations are in the range of 0.05 - 30 μgm^{-3} and 0.05 - 60 μgm^{-3}, respectively, with high concentrations over urban and urban park areas and low concentrations over regional urban areas.

The hourly variability of simulated $PM_{2.5}$ showed in Figure 5.19, follows the behavior of the observed values over urban park areas, mainly from noon, showing low concentrations between 12:00 and 17:00 LT and high concentrations between 18:00 to 00:00 local time. Between 06:00 and 09:00 local time, the model showed a maximum peak while the AQS registered maximum values between 09:00 and 12:00 local time. This simulated hourly variability has the same behavior as over urban areas due to the poor representation of

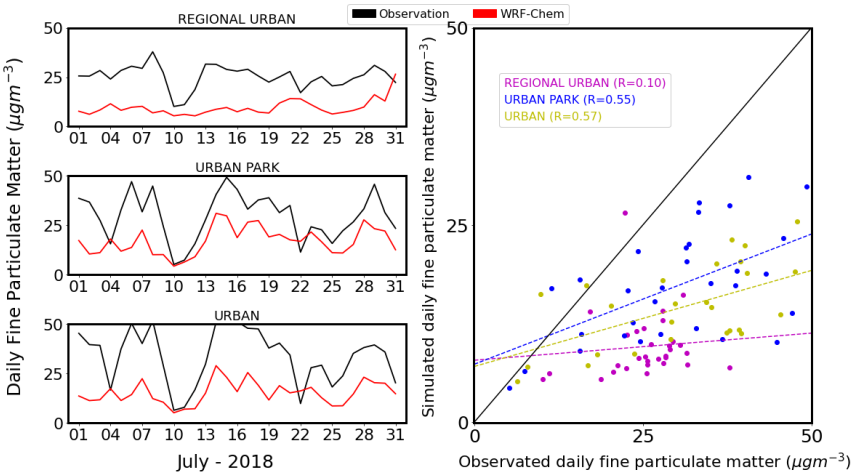

Figure 5.18: Time series and scatter plot of the observed and simulated $PM_{2.5}$ concentrations in the four categories of land use: Regional urban, urban park, urban and industrial for July 2018. The measured and simulated results are represented by the black and red line, respectively.

land use in the model. Thus, the AQS registered high $PM_{2.5}$ values from 18:00 to 12:00 LT and low values from 12:00 to 18:00 LT. Over regional urban areas, the model did not keep the same behavior as over the other regions, although showed an increase from 18:00 local time, similar to the AQS. Thus, the model performed well at different magnitudes over urban areas, indicating again that the anthropogenic emissions were well represented for urban areas. In general, the model showed a better correlation (R = 0.57) with the observed $PM_{2.5}$ values over urban areas than over regional urban areas (R = 0.10). On the other hand, higher simulated $PM_{2.5}$ concentrations, compared to the other periods were generated for July 2018. These high concentrations can be partly attributed to the absence of precipitation (see section 5.2.1.4) and the enhanced photo-chemical conditions which lead to secondary aerosol formation. Furthermore, the high biomass burning emissions (see section 3.1.4.3) for July 2018, mainly over the MASP, had an important contribution to PM levels over our modeling domain. As can be seen from Table 5.8, the WRF-Chem model underestimates the observed $PM_{2.5}$ concentrations throughout the period. The model generated lower bias over urban (MB = -13.13 μgm^{-3}) and urban park (MB = -7.82 μgm^{-3}) areas than over regional urban areas (MB = -15.41 μgm^{-3}), due to the increased $PM_{2.5}$ concentrations during various periods of July 2018 which appear at urban and urban park (Figure 5.18) areas.

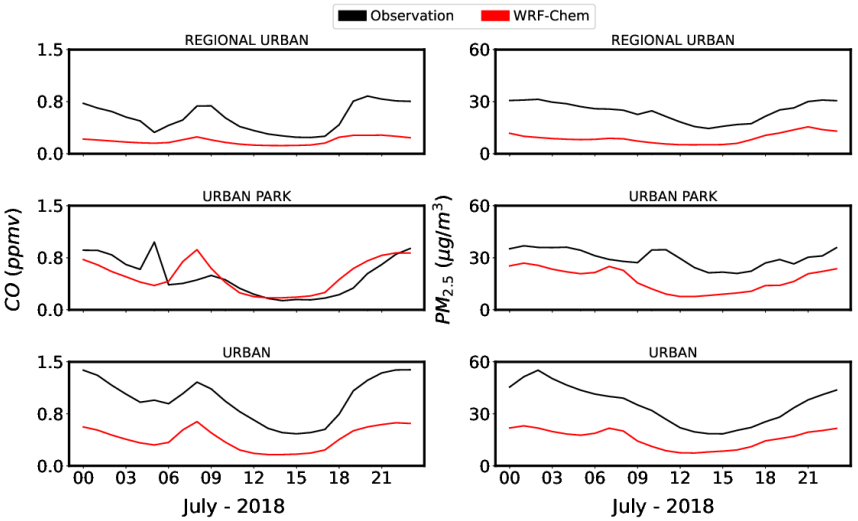

Figure 5.19: Simulated and observed hourly mean CO and $PM_{2.5}$ concentrations in the four categories of land use: Regional urban, urban park, urban and industrial over the State of São Paulo for July 2018. The measured and simulated results are represented by the black and red line, respectively.

Table 5.8 shows a summary of the performance of the WRF-Chem model when analyzing the simulated daily mean $PM_{2.5}$ concentrations with the daily mean concentrations measured at the AQS and Botucatu stations for June 2017, July 2018, June and August 2019. From the results of daily $PM_{2.5}$ at the AQS, we can observe that the simulated values reached the recommended criteria benchmarks for the mean correlation coefficient (R $>$ 0.40) for all periods, although over some types of areas such as regional urban (July 2018 and August 2019) and urban (June 2019), the simulated $PM_{2.5}$ did not fulfilled with the recommended benchmarks. Over Botucatu station, the simulated $PM_{2.5}$ values reached the recommended criteria benchmarks only for August 2019 (R = 0.42) while not fulfilled with the expected benchmarks for June 2019. In terms of NMB, the simulated $PM_{2.5}$ values reached the expected criteria benchmarks only over urban park areas for July 2018 and over Botucatu station for August 2019 ($\leq \pm$ 30%). Instead, in relation to the NME, the model fulfilled with the recommended criteria benchmarks over urban park areas and Botucatu station for all simulations ($\leq \pm$ 50%).

Besides emissions and meteorology, the choice of chemical mechanisms plays an important role in the prediction of air pollutants. Model bias in simulation of $PM_{2.5}$ depends on the photo-chemistry and VOCs speciation assumptions in the chemical mechanisms. Most of the chemical mechanisms make use of lumping approaches to reduce the number of che-

mical reactions, thus avoiding the treatment of a huge number of chemical species (Zaveri and Peters, 1999; Sarkar et al., 2016). The use of a more complex, or a more simplified gas-phase chemistry, photolysis and dry deposition schemes in chemical transport models does not guarantee an improvement in the model predictions.

Table 5.8 - Performance statistics of the validation between predicted daily Particulate Material ($PM_{2.5}$ in $\mu g\ m^{-3}$) and measurements by the AQS and the IAG-USP and Botucatu stations for the four periods.

Station	Type of Region	June 2017								July 2018							
		n	R	MB	NMB	ME	NME	RMSE	IOA	n	R	MB	NMB	ME	NME	RMSE	IOA
AQS	Regional Urban	30	0.43	-9.90	-64.38	9.90	64.38	10.44	0.37	31	0.11	-15.41	-60.79	15.68	61.85	16.86	0.33
	Urban Park	30	0.71	-5.36	-30.40	6.30	35.77	8.66	0.66	31	0.52	-7.82	-26.26	12.74	42.78	15.65	0.68
	Urban	30	0.70	-10.24	-49.74	10.74	52.18	12.88	0.52	31	0.58	-13.13	-38.76	14.97	44.20	18.25	0.66
		June 2019								August 2019							
AQS	Regional Urban	29	0.45	-12.03	-65.26	12.03	65.26	13.14	0.38	30	0.10	-9.67	-57.98	9.79	58.70	11.36	0.43
	Urban Park	29	0.46	-5.90	-31.55	8.18	43.73	10.20	0.63	30	0.63	-4.06	-27.83	5.84	40.03	7.56	0.63
	Urban	29	0.39	-15.17	-59.60	15.17	59.60	19.36	0.48	30	0.61	-7.93	-45.52	9.02	51.76	11.17	0.58
Botucatu	Regional Urban	23	0.06	-2.44	-30.52	3.14	39.38	3.51	0.39	24	0.42	0.21	3.42	2.49	40.54	3.11	0.63

5.2.2.2 Comparison with Satellite retrievals

In order to reinforce the WRF-Chem model evaluation over our study domain, satellites retrievals have been considered to analyze some simulated atmospheric pollutants, such as: tropospheric NO_2 VCD, CO VCD and AOD values. As detailed in the previous chapter, the simulated NO_2 and CO concentrations went through a small data treatment since direct comparison between model and satellite is not recommended (Kumar et al., 2012). Figures 5.20 to 5.23 show the comparison between the simulated and observed column mass abundances of NO_2 and CO, and AOD values over southeastern Brazil for June 2017, July 2018 and June and August 2019, respectively. Furthermore, we observe the relative difference and scatter-plot between the two tools. Both the model and satellite showed, during all study periods, higher NO_2 values in regions of industrial activities and/or dense population (10 x 10^{15} molecules cm^{-2}), within the MASP, and lower values over the center of the State of São Paulo and less urbanized regions (by about 2 x 10^{15} molecules cm^{-2}). The

model overall captured the spatial features of retrieved NO_2. The comparison between re-gridded OMI-retrieved and WRF-Chem-modeled tropospheric NO_2 VCD showed that the model agreed well with satellite retrievals. The spatial correlation averaged over the entire domain ranged between a 0.64 in June and August 2019 and 0.77 in July 2018, obtaining from moderate to high correlation coefficients.

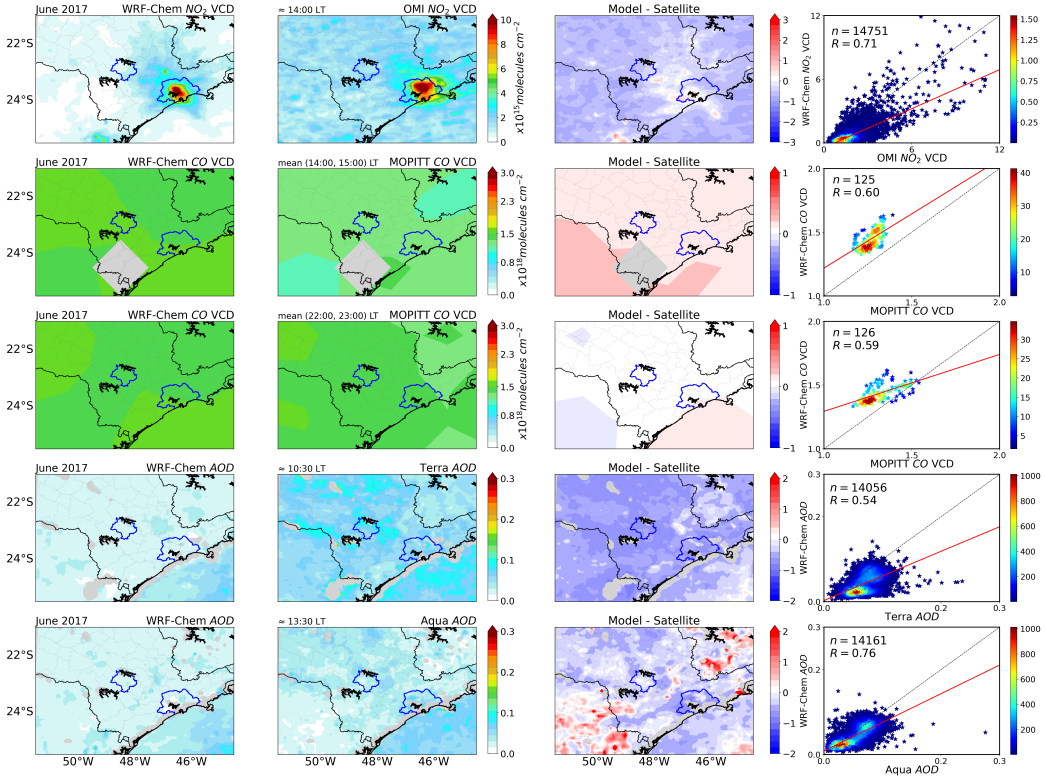

Figure 5.20: Simulated (first column) and observed (second column) mean spatial distribution of tropospheric NO_2 VCD, total CO VCD and AOD values over southeastern Brazil for June 2017. The third and fourth columns represent the relative difference and correlation between both tools. The blue lines, in the first three columns, represent Botucatu city and the MASP and the gray space indicates missing data.

According to the performance statistics for the model and OMI estimates, we can observe that the WRF-Chem underestimated the magnitude of retrieved tropospheric NO_2 column over most of the study domain for four study periods (see Table 5.9). Highest NMB (-68.09 %) and NME (68.92 %) values were found for August 2019, indicating that the model was more deficient than the values retrieved by the OMI. These large differences between the simulated and observed NO_2 tropospheric-column are associated with the poor representation of the contribution of long-range transport of biomass burning from the Amazon and that affected the air quality over the State of São Paulo, mainly during 19

and 20 August 2019. In addition, the underestimation of the model is mostly related to the underestimation of NO_x emissions over southeastern Brazil where modeled values are significantly lower than observed values. Another possible source of difference can arise from errors in simulating tropospheric NO_x chemistry (Martin et al., 2003; Valin et al., 2011) and non-anthropogenic emission sources including soil NO_x, lightning, or biomass burning.

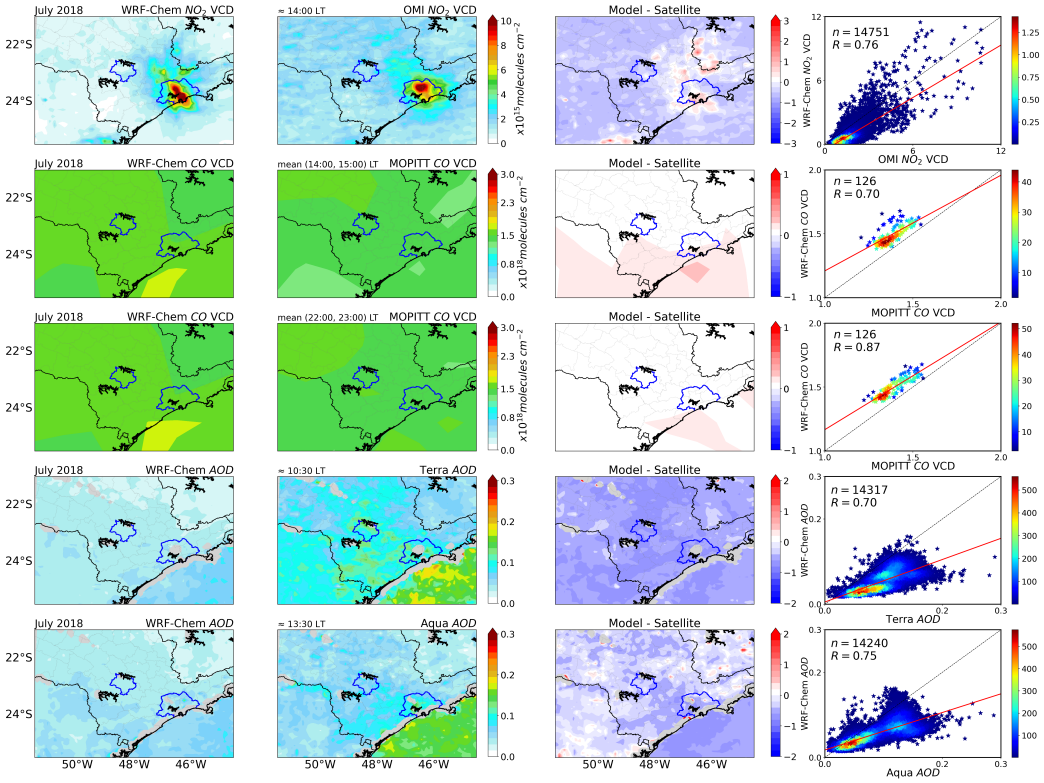

Figure 5.21: Simulated (first column) and observed (second column) monthly mean spatial distribution of tropospheric NO_2 VCD, total CO VCD and AOD values over southeastern Brazil for July 2018. The third and fourth columns represent the relative difference and respective scatter plot. The blue lines, in the first three columns, represent Botucatu city and the MASP and the gray space indicates missing data.

In relation to the comparison between re-gridded WRF-Chem-modeled and MOPITT-retrieved CO VCD, we can observe that both the model and satellite show a homogeneous spatial distribution with values of CO from 1.1 up to 1.8 x 10^{18} molecules cm^{-2} both in the afternoon (around at 14:00 LT) and evening (around at 22:00 LT) periods. The model slightly overestimated the satellite values for all study periods, with averaged MB of 0.14 x 10^{18} molecules cm^{-2}. The model integrated in the atmospheric column showed a good spatial correlation with the MOPITT, showing a high correlation coefficient averaged of 0.71 (see Table 5.9). The model performed well for simulating CO with a NMB $< 10\%$

of MOPITT. The agreement between WRF-Chem and MOPITT is higher in July 2018 (NMB = 8.17 %) than June 2017 (NMB = 14.05 %). For August 2019 (Figure 5.23), MOPITT CO data was not found.

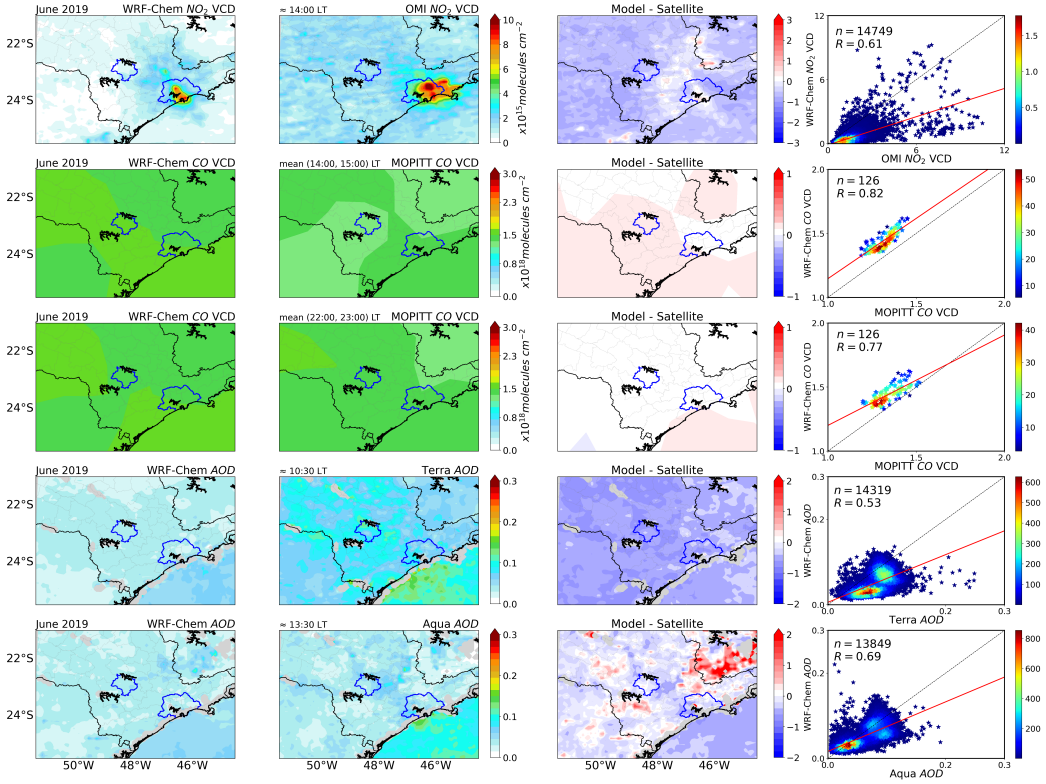

Figure 5.22: Simulated (first column) and observed (second column) mean spatial distribution of tropospheric NO_2 VCD, total CO VCD and AOD values over southeastern Brazil for June 2019. The third and fourth columns represent the relative difference and correlation between both tools. The blue lines, in the first three columns, represent Botucatu city and the MASP and the gray space indicates missing data.

Regarding the AOD values, we can see that the model was able to represent low AOD values similar to retrieval by the Terra and Aqua satellites, while it was deficient in estimating high AOD values. The model does not consider the option of formation of secondary organic aerosols (SOA) due to the type of aerosol module used in the simulations and as CETESB (2019) showed that 52% of $PM_{2.5}$ comes from secondary aerosols. This justifies the model's deficiency in estimating low AOD values where both the Terra and Aqua satellites retrieval high AOD values. The spatial correlation ranged between 0.44 (Terra satellite) and 0.65 (Aqua satellite) in August 2019, and 0.70 (Terra satellite) and 0.75 (Aqua satellite) in July 2018 (see Table 5.9). As discussed above, the model did not represent the contribution of long-range transport of biomass burning in August 2019. This reflects the

low correlation between the model and MODIS. The model-simulated AOD values were lower than Terra- and Aqua-retrieved AOD values with an averaged NMB of -39.15% and -24.25%, respectively, for our study period. In relation to the Terra and Aqua-retrieved AOD values against WRF-Chem model, we can see that the simulated AOD values are closer to the Aqua-retrieved AOD value, showing better statistical results (see Table 5.9). The differences between the model and the satellite are related to uncertainties of the emission inventories and aerosol model simplifications. The temporal profile used to distribute the emissions over the modeling domain is the same in all grid points with any emission value. This limitation can lead to possible mismatches between model predictions and satellite observations in regard to the timing of peak of AOD values at any particular time and location.

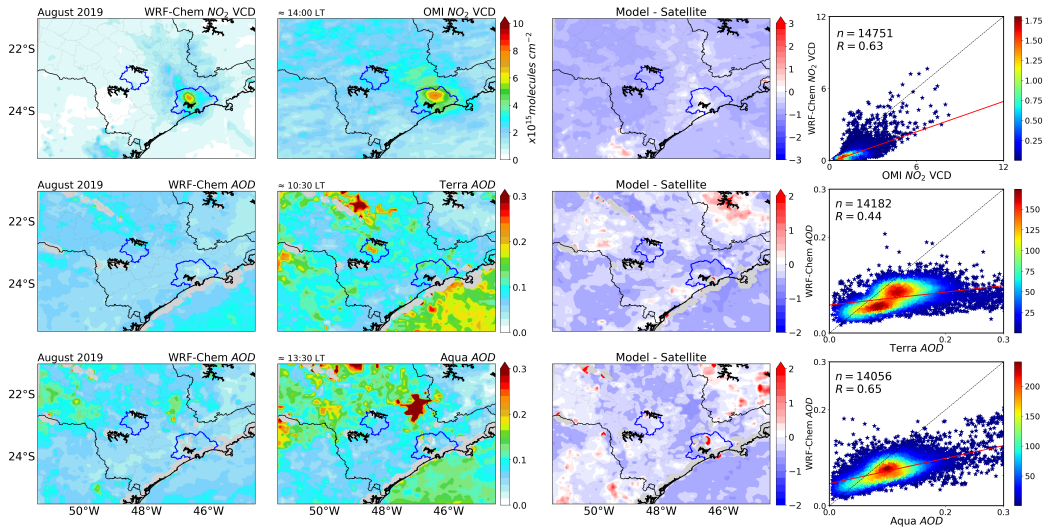

Figure 5.23: Simulated (first column) and observed (second column) monthly mean spatial distribution of tropospheric NO_2 VCD and AOD values over southeastern Brazil for August 2019. The third and fourth columns represent the relative difference and correlation between both tools. The blue lines, in the first three columns, represent Botucatu city and the MASP and the gray space indicates missing data.

Unfortunately, due to the difficulty of processing MODIS data under cloudy conditions and the scarcity of data over our study domain, the comparison between WRF-Chem model results and satellite retrievals does not reach 100% confidence for the study of aerosols. For this reason, the AOD values retrieved by the AERONET network are also compared to the model results. Figure 5.24 shows box plots of the WRF-Chem, Terra-Aqua and AERONET AOD values for July 2018, where it can be observed that WRF-Chem model

systematically underestimated AOD values. For Terra satellite, i.e. during morning overpasses, AOD at Sao Paulo station presented higher variability than AERONET retrievals, while the opposite was observed during Aqua overpasses.

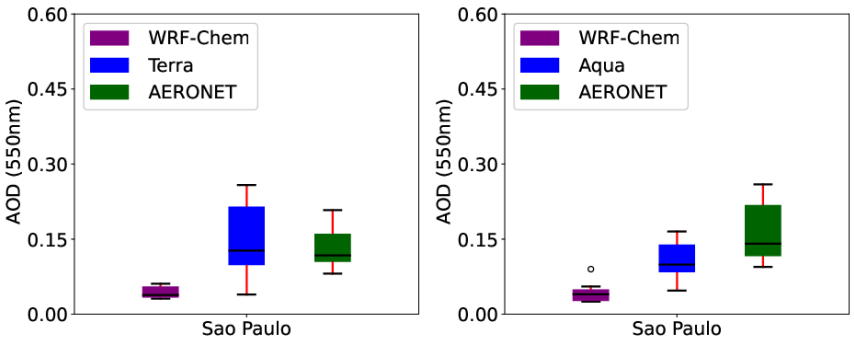

Figure 5.24: Box plot of the WRF-Chem, MODIS, and AERONET AOD values at 550 nm over the São Paulo station for July 2018.

Among all the air pollutants here analyzed, the $PM_{2.5}$ and AOD are the most difficult to simulate, as aerosol formation depends on multiple atmospheric processes such as secondary organic aerosol (SOA) gas-particle conversion. Knowledge gaps in aerosol modeling in terms of formation and aging processes, including aerosol feedback effects on meteorology, represent other sources of uncertainty (Vara-Vela et al., 2021).

Table 5.9 - Performance statistics of the validation between model-simulated and satellite-retrieved data for the four periods.

Variable	**Periods**	**n**	R	MB	NMB	ME	NME	RMSE	IOA
	June 2017	14751	0.71	-1.07	-66.57	1.09	68.08	1.21	0.59
NO_2	July 2018	14751	0.77	-0.80	-53.46	0.86	57.21	0.96	0.71
	June 2019	14751	0.64	-0.95	-64.32	0.97	65.85	1.14	0.57
	August 2019	14751	0.64	-1.07	-68.09	1.08	68.92	1.20	0.53

Variable	**Periods**	Daytime								Nighttime							
		n	R	MB	NMB	ME	NME	RMSE	IOA	**n**	R	MB	NMB	ME	NME	RMSE	IOA
	June 2017	125	0.60	0.18	14.05	0.18	14.05	0.19	0.38	126	0.59	0.11	8.30	0.12	8.68	0.13	0.58
CO	July 2018	126	0.70	0.11	8.17	0.11	8.17	0.13	0.57	126	0.87	0.11	7.55	0.11	7.55	0.11	0.67
	June 2019	126	0.82	0.13	9.69	0.13	9.69	0.13	0.55	126	0.77	0.10	7.36	0.10	7.40	0.11	0.65

Variable	**Periods**	Terra								Aqua							
		n	R	MB	NMB	ME	NME	RMSE	IOA	**n**	R	MB	NMB	ME	NME	RMSE	IOA
	June 2017	14056	0.54	-0.03	-37.91	0.03	39.92	0.03	0.56	14161	0.76	-0.01	-15.64	0.01	25.87	0.02	0.84
AOD	July 2018	14317	0.70	-0.04	-44.66	0.04	45.0	0.05	0.60	14240	0.75	-0.03	-35.34	0.03	37.54	0.04	0.71
	June 2019	14319	0.53	-0.03	-36.94	0.03	39.69	0.04	0.57	13849	0.69	-0.01	-15.38	0.02	27.82	0.02	0.80
	August 2019	14182	0.44	-0.04	37.09	0.05	40.06	0.08	0.47	14056	0.65	-0.03	-29.89	0.04	36.08	0.06	0.60

5.3 Pollutant transport from the MASP to other regions

After obtaining the reliability of the photo-chemical WRF-Chem model, by validating it with observational data at surface and column, few periods of regional air pollutants transport from urban to rural areas have been found within our study periods. Squizzato et al. (2021) conducted an air quality study over a rural area, placing a mobile station near the municipality of Botucatu ($22.85°S$; $48.43°W$, 200 km from the MASP) from June 7 to August 25, 2019. During that period, the collected samples presented higher O_3 concentrations than at the Pico do Jaraguá (located within the MASP) as a consequence from local activity and due to the advection of pollutants emitted in MASP. Based on that study, three similar episodes were analyzed to characterize the regional transport from the MASP to the inland regions of southeastern Brazil: 17 July 2018, 17 June and 11 August 2019, these last two dates being within the period of work of Squizzato et al. (2021).

In order to analyze the O_3 formation and transport from the MASP to other cities, the sensitivity of O_3 photochemical production to its main precursors (VOCs and NO_x) in the southeast Brazil was analyzed. Some works such as Orlando et al. (2010); Alvim et al. (2018) and Dominutti et al. (2020) suggested that the O_3 formation in the MASP is limited by the VOCs concentrations (VOC-limited). Spatial distributions of O_3, NO_x, HCHO, CO and $PM_{2.5}$ concentrations, and AOD values during the transport event between July 17 and 18, 2018 are shown in Figures 5.25 to 5.28, while the other transport events are shown in the Appendix - B (Figures B.1 to B.9). The colored circles represent the AQS locations, considering only available data on way of the pollutants transport. For July 17, 2018 (Figure 5.25), we can observe high simulated and observed NO_x and HCHO concentrations at 9:00 LT over downtown and northwest of the MASP. These simulated high concentrations reached values of up to 50 and 9 μgm^{-3} for NO_x and HCHO, respectively, while observed values (colored circles) reached higher values ($>$ 60 μgm^{-3} for NO_x). Over the hours, the increase in sunlight and the NO_x and HCHO degradation led to O_3 formation during its transport downwind, reaching concentrations of approximately 110 to 140 μgm^{-3}. The maximum O_3 concentration took place at 15:00 LT over almost the entire study region, including the Botucatu and Londrina (located in the State of Parana) cities, and even the MASP.

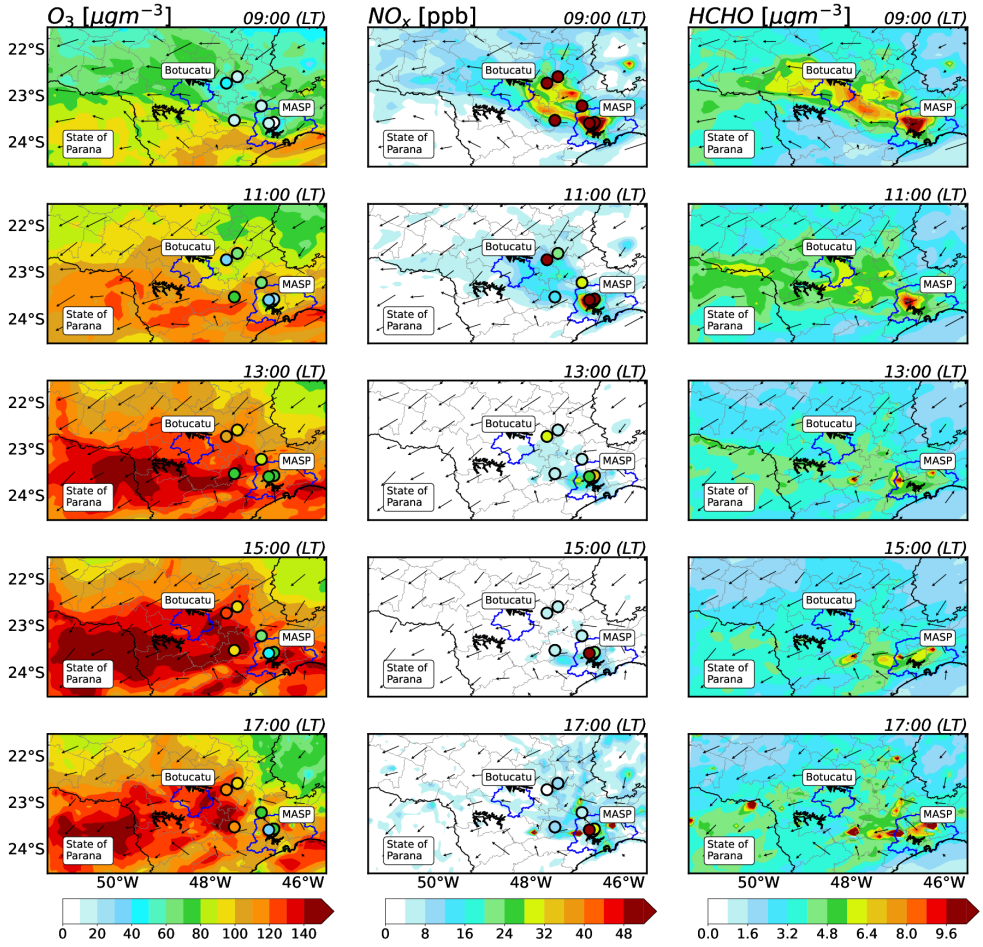

Figure 5.25: Overlay of observed and predicted O_3 (in $\mu g\ m^{-3}$) (left panels), NO_x (in ppb) (middle panels) and HCHO (in $\mu g\ m^{-3}$) (right panels) spatial distributions during daytime for July 17, 2018. The black arrows in each panel represent the wind direction at 10 m. The observed values are indicated by the colored star and circles. The blue lines represent Botucatu city and the MASP.

Vehicle and biogenic emissions contributed to the permanence of this pollutant over several hours, only decreasing after 18:00 LT (Figure 5.26). After that time, both the model and the AQS show an increase in NO_x and HCHO concentrations between the Botucatu municipality and the MASP, reaching higher values than 48 and 10 μ gm^{-3}, respectively. That increase could be attributed to the start of anthropogenic activities, such as vehicular traffic and biomass burning from 17:00 LT (Figures 5.25 and 5.27). Compared to the other events (Figures B.1 and B.6 in the Appendix - B), this period shows a higher HCHO concentration during daytime and nighttime, mainly between the MASP and Botucatu region. Greater number of burning sources, compared to the other periods, were registered by the MODIS sensor, considered in the WRF-Chem model simulations, which contributes part

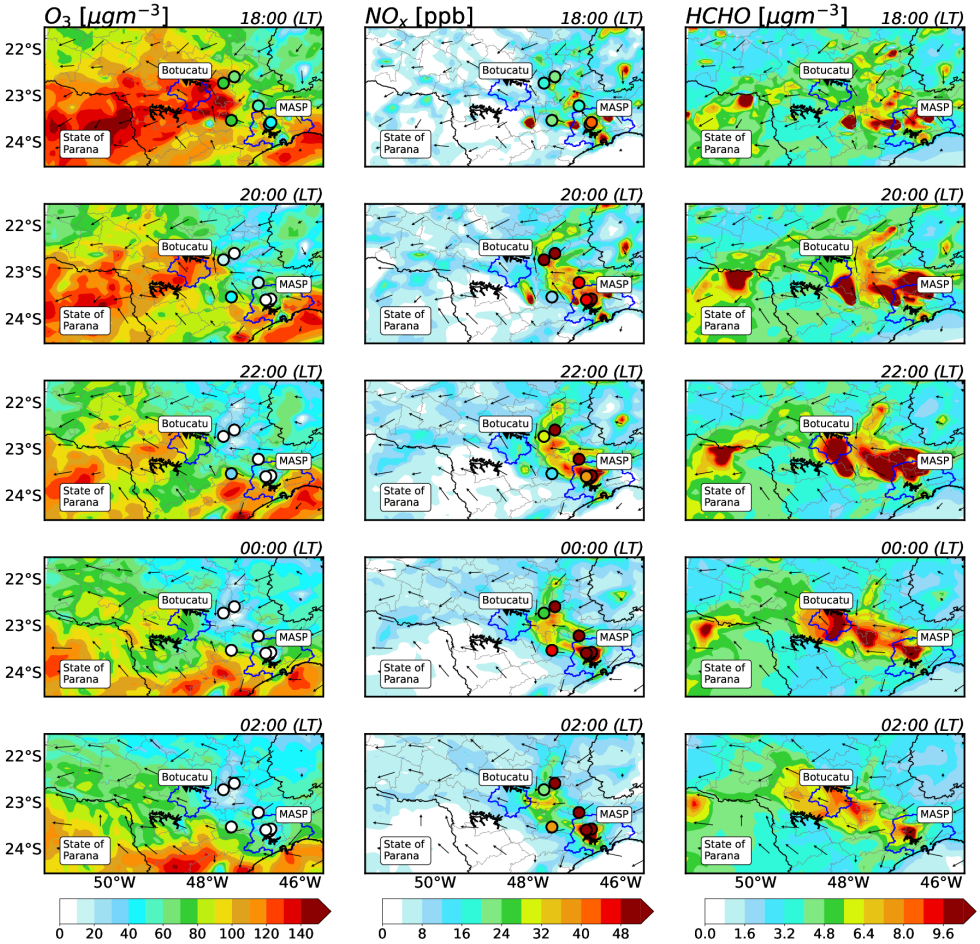

Figure 5.26: Overlay of observed and predicted O_3 (in μg m^{-3}) (left panels), NO_x (in ppb) (middle panels) and HCHO (in μg m^{-3}) (right panels) spatial distributions during nighttime for July 17, 2018. The black arrows in each panel represent the wind direction at 10 m. The observed values are indicated by the colored star and circles. The blue lines represent Botucatu city and the MASP.

to the high concentrations of analyzed pollutants from 18:00 LT (see Figures 5.26 and 5.28).

Looking at the CO and $PM_{2.5}$ concentrations from Figures 5.27 and 5.28, we can see that both the observed and simulated values showed high concentrations at 9:00 LT between MASP and Botucatu region, while the simulated AOD values were low over the MASP and high over Botucatu. Over time, both CO and aerosol particles were easily advected downwind from the MASP toward southwesterly direction close to the surface. During the daytime period, the northeasterly winds prevailed over southeast Brazil, driving the regional transport of these pollutants emitted at MAPS and Botucatu to the state of Parana. At nighttime, the prevailing winds over southeast Brazil were the southeasterly winds, transporting these pollutants emitted at MASP toward the Botucatu municipality.

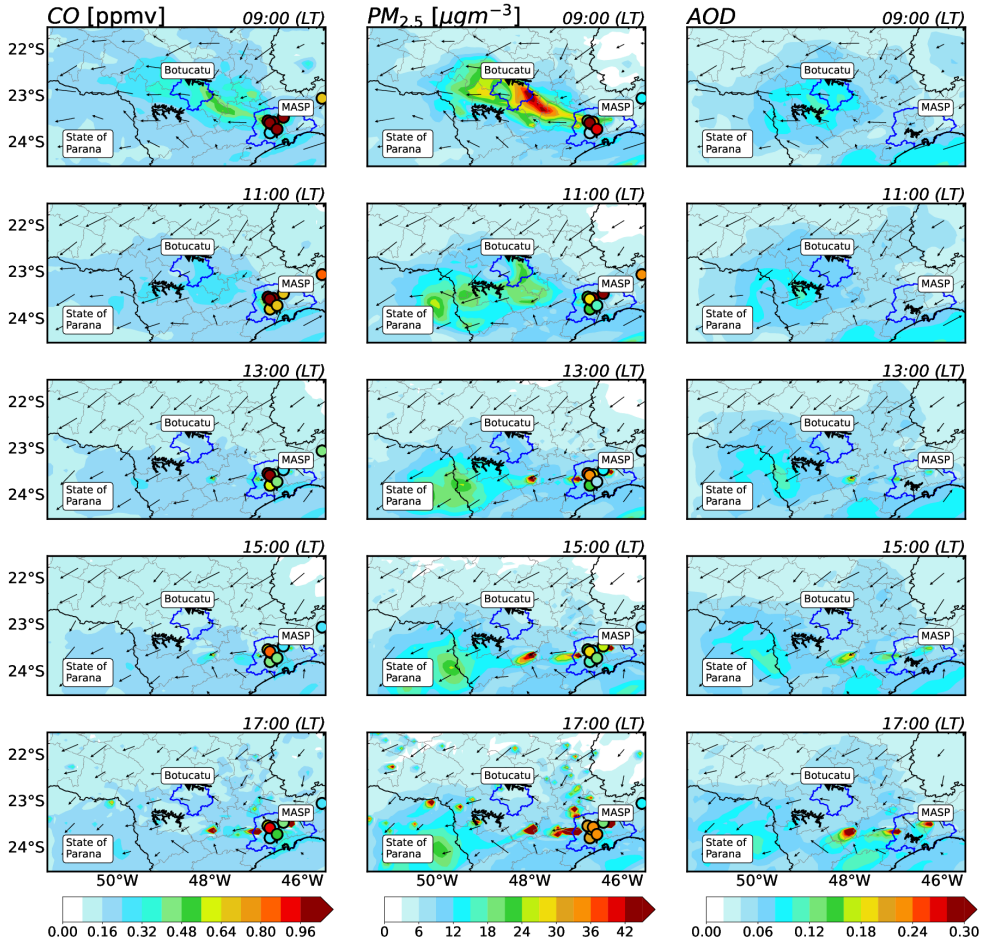

Figure 5.27: Overlay of observed and predicted CO (in ppmv) (left panels), $PM_{2.5}$ (in $\mu g\ m^{-3}$) (middle panels) and AOD (right panels) spatial distributions during daytime for July 17, 2018. The black arrows in each panel represent the wind direction at 10 m. The observed values are indicated by the colored star and circles. The blue lines represent Botucatu city and the MASP.

This result suggests that the wind direction plays an important role on the dispersion of primary and secondary pollutants, such as O_3. Biomass burning events started from 17:00 LT over the analyzed region and possibly contributed to the increase in CO and $PM_{2.5}$ concentrations and AOD values between 18:00 and 2:00 local time, as shown in Figure 5.28. In addition, the underestimation of observed CO concentrations over Botucatu was associated with an underestimation of CO transport coming from the southeast MASP.

Similar to the simulated NO_x and HCHO concentrations for the other events, CO and aerosol particles showed lower values than in this period (Figures B.3 and B.9 in the Appendix - B.1). Thus, we can conclude that the burning events contributed greatly to the increase of the concentrations of the main pollutants over our region. In this study, the

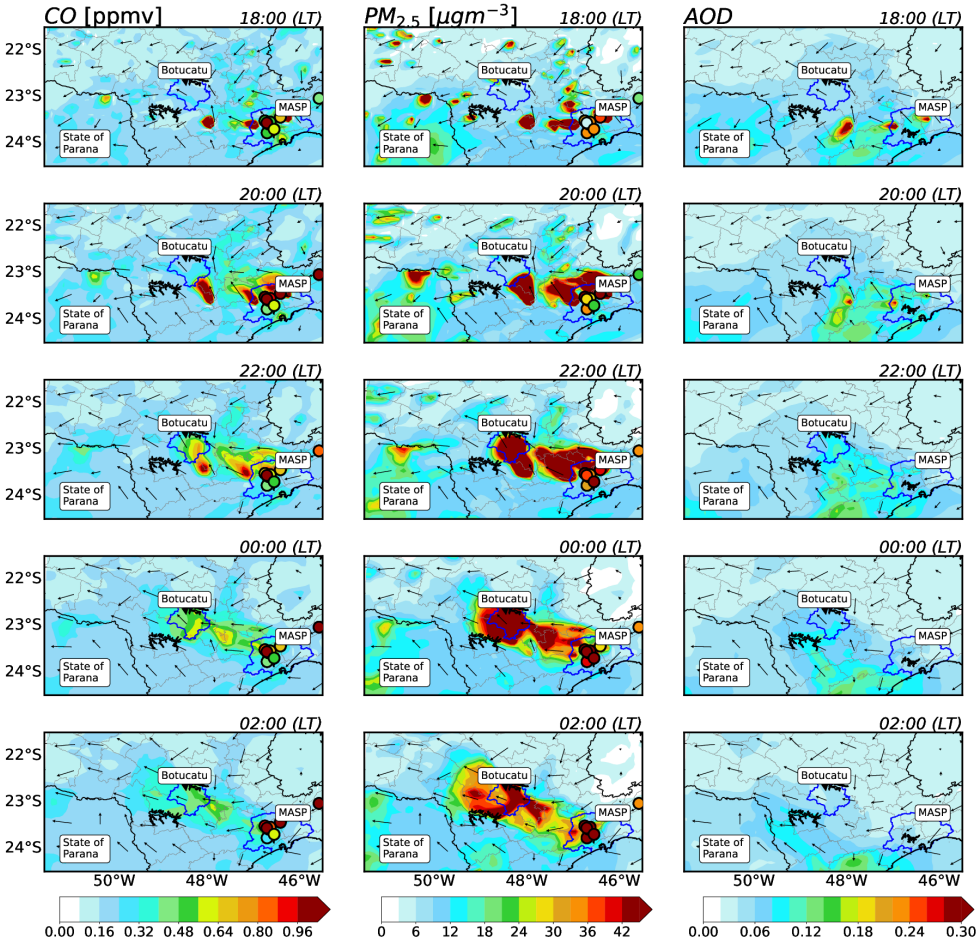

Figure 5.28: Overlay of observed and predicted CO (in ppmv) (left panels), $PM_{2.5}$ (in $\mu g\ m^{-3}$) (middle panels) and AOD (right panels) spatial distributions during nighttime for July 17, 2018. The black arrows in each panel represent the wind direction at 10 m. The observed values are indicated by the colored star and circles. The blue lines represent Botucatu city and the MASP.

scientific focus is not only O_3 formation in the MASP, but also O_3 formation on a regional scale, such as the rural and surrounding areas of MASP. To strengthen the analysis of these transport events, Figures 5.29, and B.5 and B.10 in the Appendix - B, show the spatial distribution of tropospheric NO_2 VCD and AOD values at 550 nm, integrated in the entire atmospheric column estimated by the WRF-Chem model and retrieved by OMI and MODIS, respectively, during July 17, 2018 and June 17 and August 11, 2019. From Figure 5.29, We can see that both the OMI and model present high NO_2 values over the MASP towards Botucatu during the beginning of the afternoon, showing that NO_2 contributed to the high O_3 levels registered at Botucatu.

To characterize the vertical structures of O_3 transport regionally, Figures 5.30 - 5.32 illus-

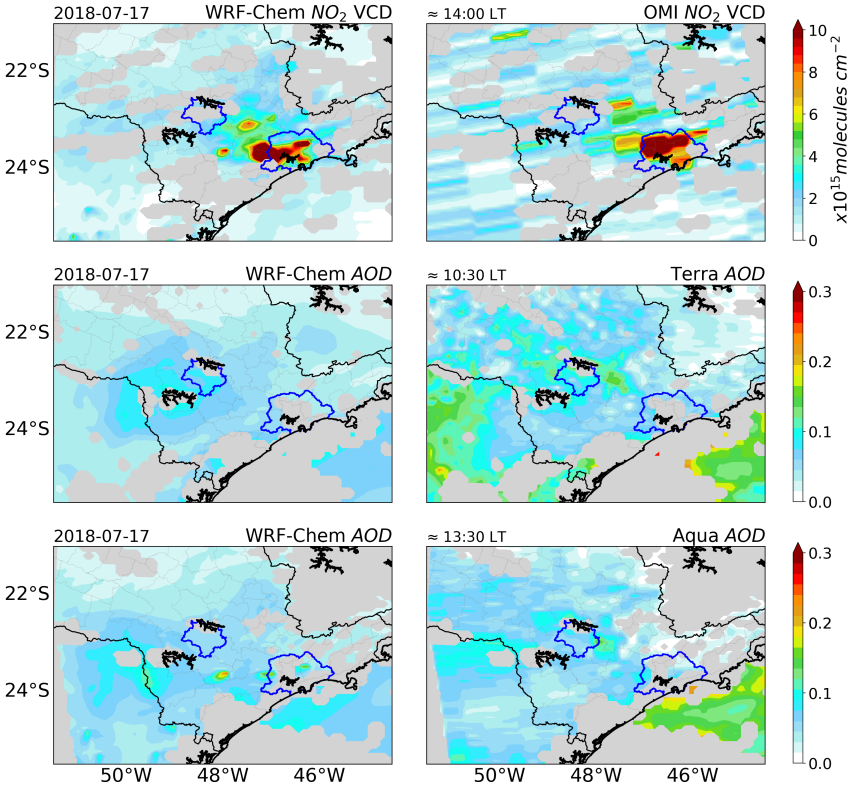

Figure 5.29: Spatial distribution of model-simulated and tropospheric NO_2 VCD, Terra- and Aqua-retrieved AOD at 550 nm integrated into the atmosphere for July 17, 2018.

trate the vertical section of the simulated O_3 transported from MASP to Botucatu municipality at six different times (from 11:00 to 21:00 LT) together with concentration ranges of O_3 (showed in violin plots) for the three events. Since the O_3 evolution is similar for all three events, we will explain at least one of them, July 17, 2018. From Figure 5.30 (left panels), we can observe an increase in O_3 at 13:00 LT between 55 and 111 km downwind from the MASP, reaching values between 120-140 μgm^{-3}, below the PBL (about 2.0 km) and where southeasterly winds predominate in the vertical. Anthropogenic sources such as vehicles as well as VOCs emissions from vegetation (BVOCs) often contribute to the photochemical pollution over this region. At 15:00 local time, O_3 concentrations reached their maximum value (around of 155 μgm^{-3}), being limited below 2.4 km in the lower troposphere while at 17:00 LT, the high O_3 concentrations exceeded the PBL (above 1.5 km). Between 15:00 and 17:00 LT, we can observe high O_3 concentrations around 56 km to 168 km downwind from the MASP. These higth concentrations arrived at Botucatu station at about 19:00 LT. Thus, the violin plots in Figure 5.30 quantitatively corroborates the displacement of O_3 values from the MASP to the Botucatu station. In relation to Figures

5.31 and 5.32 for June and August 2019, we can observe a similar behavior to the event just discussed. The maximum O_3 concentrations for both events took place at about 17:00 LT over a region 111 km far away from the MASP and the high concentrations arrived at the Botucatu station at 21:00 local time, indicating regional O_3 transport between MASP and Botucatu. Previous studies indicate that the O_3 concentrations over rural areas depend on the meso- and micro- meteorological conditions, and resulting wind flow (Thielmann et al., 2002; Zaveri et al., 2003). The major contributing factors to the high simulated O_3 concentration in Botucatu were related not only to local sources, but mainly to the regional transport of pollutants from MASP and upwind regions in between. Similar results were also found by (Squizzato et al., 2021), however here it was demonstrated with numerical modeling and satellite images.

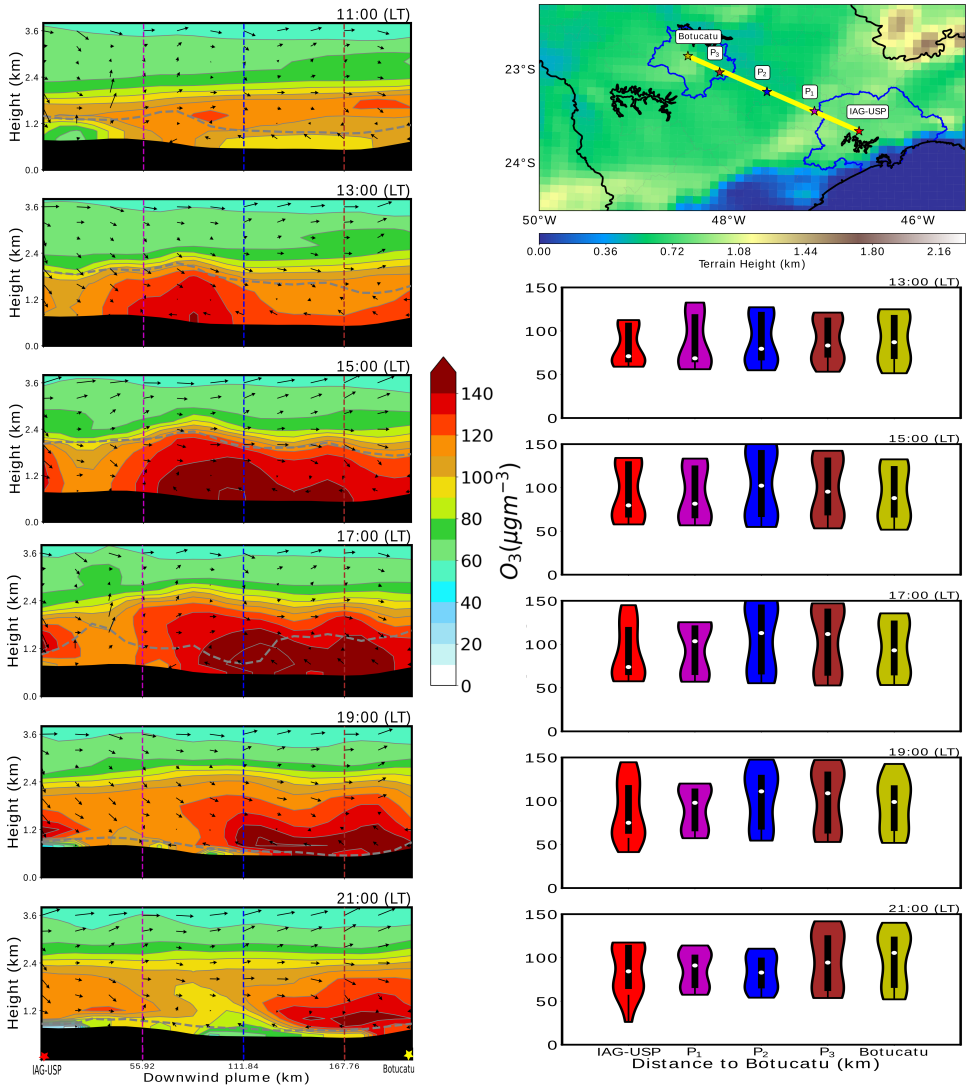

Figure 5.30: Vertical cross-sections (altitude–downwind distances) of simulated O_3 concentrations over the downwind area at 9:00, 13:00, 15:00 19:00 e 21:00 LT of July 17, 2018. The yellow line drawn in the first panel (up-right side) of the Figure represents the downwind area and the dashed gray lines in the other panels represent the planetary boundary layer (PBL).The black background of each panel represents the topography of the surface.

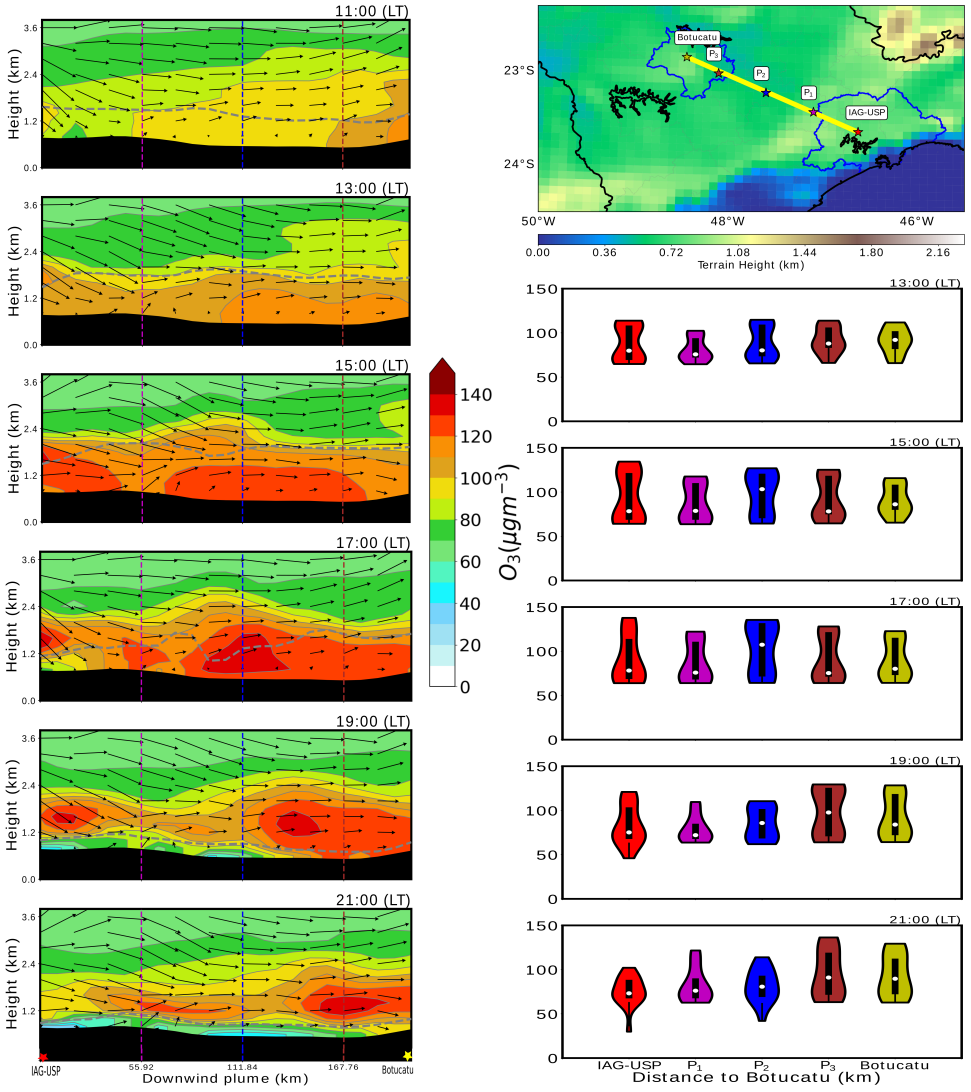

Figure 5.31: Vertical cross-sections (altitude–downwind distances) of simulated O_3 concentrations over the downwind area at 9:00, 13:00, 15:00 19:00 e 21:00 LT of June 17, 2019. The yellow line drawn in the first panel (up-right side) of the Figure represents the downwind area and the dashed gray lines in the other panels represent the planetary boundary layer (PBL).The black background of each panel represents the topography of the surface.

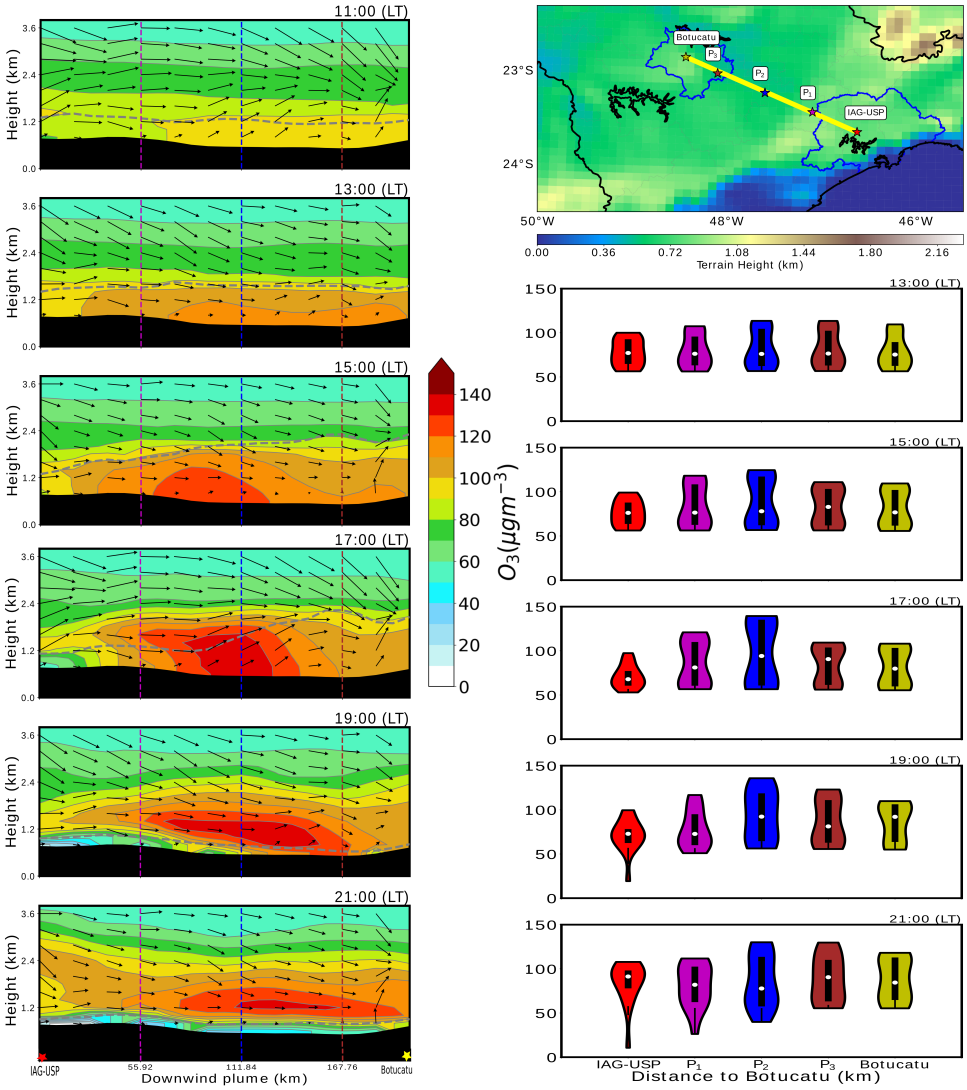

Figure 5.32: Vertical cross-sections (altitude–downwind distances) of simulated O_3 concentrations over the downwind area at 9:00, 13:00, 15:00 19:00 e 21:00 LT of August 11, 2019. The yellow line drawn in the first panel (up-right side) of the Figure represents the downwind area and the dashed gray lines in the other panels represent the planetary boundary layer (PBL).The black background of each panel represents the topography of the surface.

Chapter 6

Conclusions and Future work

The objective of this thesis was to represent pollutant transport events that affect air quality both in the MASP and in neighboring areas. Section 6.1, consolidates the contributions made by the researcher and Section 6.2 points up the directions in which future research studies can be focused.

6.1 General conclusions

In this study, ground observations, chemical transport model results, and satellite data were combined to further study the impact that anthropogenic activities over large urban areas produce on rural areas. This study is novel in several aspects. First, we have analyzed a six-year study period to validate five MODIS AOD products with AERONET estimates over southeastern Brazil, focusing on the Metropolitan Area of São Paulo. Considering the greater availability of MODIS data, we have showed the best MODIS product to represent AOD values over southeast Brazil by using a matching methodology. This methodology uses the closest pixel (from satellite) to the location of an AERONET station ($<$ 1km, 3km and 10 km) and the AERONET data in an approximate satellite overpass time ($\pm$ 15, $\pm$ 30 and $\pm$ 60 minutes). The results showed that in the annual analysis, MODIS mean AOD data within 10 km distance from each AERONET station and in combination with 60 minutes time interval (before and after the Terra and Aqua satellites overpassing times) agreed better with AERONET observations. However, different spatiotemporal windows were found seasonally. The presence of rain and clouds over our study area resulted in the lack of AOD data during the summer and early fall seasons, implying that a greater number of MODIS and AERONET observations was observed between March and November.

Under this condition, the validation approach was applied to the average MODIS retrievals within 10x10 km^2 centered at each site, in combination with average of AERONET measurements within ± 60 minutes around the time of the satellite overpass.

Once the most appropriate space-time windows have been determined to validate AOD data between MODIS and AERONET, temporal series of AOD values (MODIS - AERONET) were analyzed over the four selected stations. The time series showed that the Dark Target product better represented the AOD values compared to the other products. The MAIAC product had deficiencies to represent the high AOD values between September and October, and the Deep Blue product overestimated the AERONET AOD values throughout the study period. In relation to correlation coefficient, Dark Target product with 3 and 10 km spatial resolution showed the best representation of AOD values over all the AERONET stations (e.g. São Paulo station - Terra: R = 0.80 and 0.78; Aqua: R = 0.90 and 0.91, respectively). Instead, Deep Blue product presented lower correlation coefficient over all the AERONET stations (e.g. São Paulo station - Terra: R = 0.45; Aqua: R = 0.78). Despite MAIAC having higher resolution compared to the other products, it is not as efficient as the Dark Target product to represent the AOD values over the AERONET stations in southeastern Brazil. This is because MAIAC uses an inappropriate aerosol model for our region which presents land use characteristics different from those found in United States cities. MAIAC aerosol model considers non-absorbing aerosols as dominant, while, in our study region, absorbing aerosols prevail. Bearing in mind that MODIS Aerosol products reach a good accuracy when more than 66 % of the retrievals fall within EE envelope, the Dark Target product was the only one fulfilling this threshold. For this reason WRF-Chem AOD simulations were compared to the Dark Target product with 10 km x 10 km spatial resolution.

Due to the fact that meteorological conditions are relevant to air pollution dispersion, first the validation between the simulated meteorological variables against the observed data was carried out. Four Weather Research and Forecasting with Chemistry (WRF-Chem) model simulations (June 2017, July 2018, June 2019 and August 2019) were performed and compared against observational data from air quality stations distributed throughout the State of São Paulo. Due to the large number of air quality station, both observed and

simulated variables were grouped into four categories of land use: regional urban, urban park, urban and industrial. The IAG and Botucatu stations were classified within the urban park and regional urban areas, respectively.

Overall, the model satisfactorily represented the strong changes in temperature and relative humidity at 2 meters observed in the four study periods, although it failed in representing the maximum and minimum peaks. The model underestimated the temperature over regional urban, urban, industrial areas, while it overestimated the temperature over urban park areas for the majority of the periods with the exception of August 2019. The stations: Ibirapuera, Capão Redondo, Cidade Universitária-USP-Ipen and Itaquera, are classified as urban park, areas that were poorly represented by the model. The model overestimated the temperature with an average bias of less than 1 $^{\circ}C$ for all periods, result that is in line with previous WRF-Chem studies over southeast Brazil. For relative humidity, the model overestimated the observed values over the industrial (in July 2018 and June 2019) and urban (in June and August 2019) areas, while it underestimated the observed values over regional urban and urban park areas for most periods. Simulated temperature and relative humidity correlated well with their corresponding observed values for the AQS (R = 0.95 and 0.87, respectively), and for IAG-USP (R = 0.84 and 0.68, respectively) and Botucatu (R = 0.88) stations, fulfilling the expected performance benchmarks proposed by Monk et al. (2019).

Simulated wind speed and wind direction at 10 m were well represented by the model in terms of spatial and temporal variation and trends. Both observed and simulated wind speed were more intense during the daytime and relatively calm during nighttime. Model-observation discrepancies are strongly associated with the misrepresentation of land use and stability, as well as with differences in geographical locations between model grid-points and air quality stations; however, our modeling results fulfilled the expected performance benchmarks.

Regarding air pollutants, the model was able to represent the temporal variations and trends of most of the chemical species considered in this study, including CO, O_3, $PM_{2.5}$ and NO_x. However, discrepancies were also noticeable, with NO_x concentrations, for

instance, underestimation over regional urban and urban areas, and NO_x overestimation over industrial areas. The CO concentrations were significantly underestimated most of the time. In general, underestimation of NO_x and CO is related to the emission sources, while an overestimation in the concentration of these pollutants could be related to either inaccurate representations of emissions or meteorology, or a combination of both. As in the case of CO, the simulated $PM_{2.5}$ concentrations were also underestimated most of the time. In addition to inaccurate representations of emissions and meteorology, model bias in O_3 and $PM_{2.5}$ simulations depends on the photo-chemistry and VOCs speciation assumptions in the chemical mechanisms.

Furthermore, the model was also validated against total column data. The comparison between re-gridded OMI-retrieved and WRF-Chem tropospheric NO_2 VCD and MOPITT-retrieved CO column during the four study periods showed that the model agreed well with satellite values, even though the model underestimated or overestimated the retrieved values. Both the model and OMI captured higher NO_2 VCD values in regions of industrial activities and/or dense population and lower NO_2 VCD values over less urbanized areas. Regarding CO column data, both the model and the satellite showed a homogeneous spatial distribution during all study periods, resulting in high correlation coefficient (mean: R = 0.71). For aerosol products, WRF-Chem and re-gridded Terra- and Aqua-retrieved AOD integrated in the entire atmospheric column detected higher AOD values over the ocean and lower values over the continent. However, on days with high MODIS AOD values over the continent, the model was not able to represent them. The inclusion of sea salt emissions in the model simulations contributed to a better representation of aerosol loadings over the ocean, while over the continent, the aerosol module was not able to represent the formation of organic aerosol, resulting in poorer comparison. The spatial correlation coefficients were from moderate to high, with values between 0.44 and 0.75. The WRF-Chem model showed deficiency in representing aerosols at the four specific AERONET stations, presenting lower correlation coefficients compared to those found for satellite estimates. Discrepancies between model and observations at site-specific locations at both surface and total-column can be also related to a misrepresentation of local conditions, not only in terms of emissions but also in terms of land-use and atmospheric stability. Overall and in spite of model-observation discrepancies, we can conclude that the WRF-Chem model was

able to detect high aerosol concentration events at regional scale. Model underestimations are in general related to an underestimation in the emissions calculation, highlighting the need for more accurate aerosol representation over a metropolitan area with complex and distinct anthropogenic contributions.

Finally, O_3 was the air pollution better simulated in terms of temporal variability, with the characteristic large and small amplitudes observed over the MASP (as urban area) and Botucatu (as rural area) being well represented by the model. The model also represented the high O_3 concentrations observed at the Botucatu station satisfactorily, indicating the importance of using comprehensive modeling tools to evaluate the impact of the regional transport of emissions from large urban centers on air pollutant levels over downwind areas. The high concentrations observed over Botucatu were strongly influenced by the winds. O_3 formation was enhanced in downwind areas due to the contributions of O_3 precursors from anthropogenic, biogenic, and biomass burning emissions along the air mass trajectories.

This work represents a first effort combining different ground-based observations, remote sensing and modeling tools to improve the understanding of how pollutant emissions impact the air quality over the metropolitan area of São Paulo and inland areas.

6.2 Future Works

Based on the results presented here, some suggestions for future works are proposed:

1. The poor representation of the land use type considered in the WRF-Chem simulations generated large uncertainties in the model results. Modifications of the land use type can possibly improve the urban land surface parameters.

2. In this work, the industrial inventory was not considered, either due to the lack of information (at a local scale) or due to poor representation (at a regional scale) on our study domain. For a regional and/or local study, it is important to consider "local"inventories to reduce uncertainties in the results, since these provide more adequate information. However, in this work the emission inventories used consider information collected from more than 6 years ago, therefore, updating this information to build more realistic inventories should be taken into account for future simulations.

3. In this work, several tests were carried out in order to find a chemical mechanism and aerosol module that fits the good representation of atmospheric pollutants (mainly for $PM_{2.5}$) over our study domain. Preliminary results showed greater uncertainty compared to the one used in this work. Due to the little scientific information on how to build the chemical-aerosol mechanisms. In addition to social distancing due to the pandemic, it was difficult to find a mechanism that well represents all the pollutants considered in this work. Each mechanism (both gas and aerosol) is usually focused on to study certain chemical species and aerosol particles. Therefore, in this work, which considers several pollutants, it is expected that the inclusion of, at least two chemical-aerosol mechanisms, would improve the modeling simulatons. On the other hand, secondary aerosol formation is also important to consider, especially if the study focuses on MASP.

Bibliography

Ahrens C. D., , 2011 Essentials of meteorology: An invitation to the atmosphere. 4 edn. Thomson Brooks/Cole p. 473

Albuquerque T. T. A., Andrade M. F., Ynoue R. Y., Moreira D. M., Andreao W. L., Dos Santos F. S., Nascimento E. G. S., WRF-SMOKE-CMAQ modeling system for air quality evaluation in São Paulo megacity with a 2008 experimental campaign data, Environmental Science and Pollution Research, 2018, vol. 25, p. 36555–36569

Ali M., Assiri M., et al., Analysis of AOD from MODIS-merged DT-DB products over the Arabian Peninsula, Earth Systems and Environment, 2019, vol. 3, p. 625

Alonso M. F., Longo K. M., Freitas S. R., Mello da Fonseca R., Marécal V., Pirre M., Klenner L. G., An urban emissions inventory for South America and its application in numerical modeling of atmospheric chemical composition at local and regional scales, Atmospheric Environment, 2010, vol. 44, p. 5072

Alvim D. S., Gatti L. V., Corrêa S. M., Chiquetto J. B., Santos G. M., de Souza Rossatti C., Pretto A., H. S. M., Yamazaki Amelia O. J. P., Main ozone-forming VOCs in the city of Sao Paulo: observations, modelling and impacts., Air Quality Atmosphere and Health, 2017, vol. 10, p. 421

Alvim D. S., Gatti L. V., Corrêa S. M., Chiquetto J. B., Santos G. M., de Souza Rossatti C., Pretto A., Rozante J. R., Figueroa S. N., Pendharkar J., Nobre P., Determining VOCs Reactivity for Ozone Forming Potential in the Megacity of São Paulo, Aerosol and Air Quality Research, 2018, vol. 18, p. 2460

Anderson L. O., Latorre M. L., Shimabukuro Y. E., Arai E., Carvalho Júnior O. D., Sensor MODIS: Uma Abordagem Geral, São José dos Campos: INPE, 2003, p. 58

Andrade M. F., Fornaro A., Miranda R. M., Kerr A., Oyama B., Andre P. A., Saldiva P., Vehicle emissions and $PM_{2.5}$ mass concentrations in six Brazilian cities, Air Quality, Atmosphere & Health, 2012, vol. 5, p. 79

Andrade M. F., Kumar P., de Freitas E. D., Ynoue R. Y., Martins J., Martins L. D., Nogueira T., Perez-Martinez P., de Miranda R. M., Albuquerque T., Gonçalves F. L. T., Oyama B., Zhang Y., Air quality in the megacity of São Paulo: Evolution over the last 30 years and future perspectives, Atmospheric Environment, 2017, vol. 159, p. 66

Andrade M. F., Ynoue R. Y., Freitas E. D., Todesco E., Vara-Vela A., Ibarra S., Martins L. D., Martins J. A., Carvalho V. S. B., Air quality forecasting system for Southeastern Brazil, Frontiers in Environmental Science, 2015, vol. 3, p. 1

Andrews D. G., , 2010 Atmospheric radiation 2 edn. Cambridge University Press p. 52–93

Andreão W. L., Alonso M. F., Kumar P. e., Top-down vehicle emission inventory for spatial distribution and dispersion modeling of particulate matter, Environmental Science and Pollution Research, 2020, vol. 27, p. 35952–35970

Angelevska B., Atanasova V., Andreevski I., Urban air quality guidance based on measures categorization in road transport, Civil Engineering Journal, 2021, vol. 7, p. 253

Angstrom A., On the atmospheric transmission of Sun radiation and on dust in the air, George Ann, 1929, vol. 12, p. 130–159

Arakawa A., Lamb V. R., , 1977 in Methods in Computational Physics: Advances in Research and Applications, Vol. 17, General Circulation Models of the Atmosphere. Elsevier pp 173–265

Arbex M. A., Santos U. d. P., Martins L. C., Saldiva P. H. N., Pereira L. A. A., Braga A. L. F., Air pollution and the respiratory system, Jornal brasileiro de pneumologia, 2012, vol. 38, p. 643

Arneth A., Schurgers G., Lathiere J., Duhl T., Beerling D. J., Hewitt C. N., Martin M., Guenther A., Global terrestrial isoprene emission models: sensitivity to variability in climate and vegetation, Atmospheric Chemistry and Physics, 2011, vol. 11, p. 8037

Aura Ozone Monitoring Instrument (OMI), 2021

Banta R. M., Seniff C. J., Nielsen-Gammon J., Darby L. S., Ryerson T. B., Alvarez R. J., Sandberg S. P., Williams E. J., Trainer M., A bad air day in Houston, Bulletin of the American Meteorological Society, 2005, vol. 86, p. 657

Boersma K. F., Eskes H. J., Dirksen R. J., Van Der A. R. J., Veefkind J. P., Stammes P., Huijnen V., Kleipool Q. L., Sneep M., Claas J., et al., An improved tropospheric NO_2 column retrieval algorithm for the Ozone Monitoring Instrument, Atmospheric Measurement Techniques, 2011, vol. 4, p. 1905

Brito J., Carbone S., A. Monteiro dos Santos D., Dominutti P., de Oliveira Alves N., V. Rizzo L., Artaxo P., Disentangling vehicular emission impact on urban air pollution using ethanol as a tracer, Scientific Reports, 2018, vol. 8, p. 1

Brito J., Wurm F., Yáñez-Serrano A. M., de Assunção J. V., Godoy J. M., Artaxo P., Vehicular Emission Ratios of VOCs in a Megacity Impacted by Extensive Ethanol Use: Results of Ambient Measurements in São Paulo, Brazil, Environmental Science Technology, 2015, vol. 49, p. 11381

Carvalho L., Jones C., Silva A., Liebmann B., Silva Dias P., The South American Monsoon System and the 1970s climate transition, International Journal of Climatology, 2011, vol. 31, p. 1248

Carvalho L. M. V., Jones C., Silva Dias M. A. F., Intraseasonal large-scale circulations and mesoscale convective activity in tropical South America during the TRMM-LBA campaign, Journal of Geophysical Research: Atmospheres, 2002, vol. 107, p. LBA 9

CETESB Companhia de Tecnologia de Saneamento Ambiental: Relatorio Anual de Qualidade do Ar no Estado de Sao Paulo 2015,Sao Paulo, Brazil technical report edn, 2015

CETESB Companhia de Tecnologia de Saneamento Ambiental: Relatorio Anual de Qualidade do Ar no Estado de São Paulo 2017,Sao Paulo, Brazil technical report edn, 2017

CETESB Companhia de Tecnologia de Saneamento Ambiental: Relatorio Anual de Qualidade do Ar no Estado de São Paulo 2017,Sao Paulo, Brazil technical report edn, 2019

CETESB Companhia de Tecnologia de Saneamento Ambiental: Relatorio Anual de Qualidade do Ar no Estado de São Paulo 2021,Sao Paulo, Brazil technical report edn, 2021

Chaithanya D. J., Vikas S., S. T. Akhil R., Madhavan B. L., M. Venkat R., Local emission and long-range transport impacts on the CO, CO_2, and CH_4 concentrations at a tropical rural site, Atmospheric Environment, 2021, vol. 254, p. 118397

Chance K., Martin R. V., Spectroscopy and Radiative Transfer of Planetary Atmospheres. Oxford University Press, 2017

Cheng W. Y. Y., Steenburgh W. J., Evaluation of surface sensible weather forecasts by the WRF and the Eta models over the western United States, Weather and Forecasting, 2005, vol. 20, p. 812

Chin M., Ginoux P., Kinne S., Torres O., Holben B. N., Dun-can B. N., Martin R. V., Logan J. A., Higurashi A., Naka-jima T., Tropospheric aerosol optical thickness from the GOCART model and comparisons with satellite and sun photometer measurements, Journal of the Atmospheric Sciences, 2002, vol. 59, p. 461–483

Chin M., Rood R. B., Lin S.-J., Müller J.-F., Thompson A. M., Atmospheric sulfur cycle simulated in the global model GOCART: Model description and global properties, Journal of Geophysical Research: Atmospheres, 2000, vol. 105, p. 24671

Chiquetto J., F. N. D R., Alvim D., Ynoue R., Silva J., Silva M., Transport of Pollutants by the Sea Breeze in São Paulo under the South Atlantic High, 2018, pp 148–161

Chudnovsky A., Lyapustin A., Wang Y.and Tang C., Schwartz J., Koutrakis P., High Resolution Aerosol Data from MODIS Satellite for Urban Air Quality Studies, Central European Journal of Geosciences, 2014, vol. 6, p. 17

Cohen A. E., Cavallo S. M., Coniglio M. C., Brooks H. E., A Review of Planetary Boundary Layer Parameterization Schemes and Their Sensitivity in Simulating Southeastern U.S. Cold Season Severe Weather Environments, Weather and Forecasting, 2015, vol. 30, p. 591

Cohen A. J., Brauer M., Burnett R., Anderson H. R., Frostad J., Estep K., Balakrishnan K., Brunekreef B., Dandona L., Dandona R., et al., Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015, The Lancet, 2017, vol. 389, p. 1907

Cuspilici A., Monforte P., Ragusa M. A., Study of Saharan dust influence on PM10 measures in Sicily from 2013 to 2015, Ecological Indicators, 2017, vol. 76, p. 297

Damascena A. S., Yamasoe M. A., Martins V. S., Santana J. R., Benavente N. R., Sanchez M. P., Tanaka N. I., Nascimento S. P. H., Exploring the relationship between high-resolution aerosol optical depth values and ground-level fine particulate matter concentrations in the metropolitan area of São Paulo, Atmospheric Environment, 2020, vol. 244, p. 1

Darby L. S., Cluster analysis of surface winds in Houston, Texas, and the impact of wind patterns on ozone, Journal of Applied Meteorology and Climatology, 2005, vol. 44, p. 1788–1806

Deeter M. N., Emmons L. K., Francis G. L., Edwards D. P., Gille J. C., Warner J. X., Khattatov B., Ziskin D., Lamarque J.-F., Ho S.-P., et al., Operational carbon monoxide retrieval algorithm and selected results for the MOPITT instrument, Journal of Geophysical Research: Atmospheres, 2003, vol. 108

Dekker I. N., Houweling S., Aben I., Röckmann T., Krol M., Martínez-Alonso S., Deeter M. N., Worden H. M., Quantification of CO emissions from the city of Madrid using MOPITT satellite retrievals and WRF simulations, Atmospheric chemistry and physics, 2017, vol. 17, p. 14675

Dominutti P., Nogueira T., Fornaro A., Borbon A., One decade of VOCs measurements in São Paulo megacity: Composition, variability, and emission evaluation in a biofuel usage context, Science of The Total Environment, 2020, vol. 738, p. 139790

Dubovik O., Holben B. N., King M. D., Smirnov A., Eck T. F., Kinne S., Slutsker I., A flexible inversion algorithm for retrieval of aerosol optical properties from sun and sky radiance measurements, Journal of Geophysical Research, 2000, vol. 20, p. 673

Eck T. F., Holben B. N., Reid J. S., Dubovik O., Smirnov A., O'Neill N. T., Slutsker I., Kinne S., Wavelength dependence of the optical depth of biomass burning, urban and desert dust aerosols, Journal of Geophysical Research, 1999, vol. 104, p. 31333–31349

Edwards D. P., Halvorson C. M., Gille J. C., Radiative transfer modeling for the EOS Terra satellite Measurement of Pollution in the Troposphere (MOPITT) instrument, Journal of Geophysical Research: Atmospheres, 1999, vol. 104, p. 16755

Emery C., Liu Z., Russell A. G., Odman M. T., Yarwood G., Kumar N., Recommendations on statistics and benchmarks to assess photochemical model performance, Journal of the Air and Waste Management Association, 2017, vol. 67, p. 582

Emmons L. K., Walters S., Hess P. G., Lamarque J.-F., Pfister G. G., Fillmore D., Granier C., Guenther A., Kinnison D., Laepple T., Orlando J., Tie X., Tyndall G., Wiedinmyer C., Baughcum S. L., Kloster S., Description and evaluation of the Model for Ozone and Related chemical Tracers, version 4 (MOZART-4), Geoscientific Model Development, 2010, vol. 3, p. 43–67

Fares S., Vargas R., Detto M., Goldstein A. H., Karlik J., Paoletti E., Vitale M., Tropospheric ozone reduces carbon assimilation in trees: estimates from analysis of continuous flux measurements, Global Change Biology, 2013, vol. 19, p. 2427

Feng Y., Penner J. E., Global modeling of nitrate and ammonium: Interaction of aerosols and tropospheric chemistry, Journal of Geophysical Research: Atmospheres, 2007, vol. 112

Finlayson-Pitts B., Pitts J., Chemistry of Upper and Lower Atmosphere, 2000

Freitas S. R., Longo K. M., Chatfield R., Latham D., Silva D. M. A. F., Andreae M. O., Prins E., Santos J. C., Gielow R., Carvalho J. J. A., Including the sub-grid scale plume rise of vegetation fires in low resolution atmospheric transport models, Atmospheric Chemistry and Physics, 2007, vol. 7, p. 3385

Fu D., Kulawik S. S., Miyazaki K., Bowman K. W., Worden J. R., Eldering A., Livesey N. J., Teixeira J., Irion F. W., Herman R. L., et al., Retrievals of tropospheric ozone profiles from the synergism of AIRS and OMI: methodology and validation, Atmospheric Measurement Techniques, 2018, vol. 11, p. 5587

Gavidia-Calderón M., Vara-Vela A., Crespo N. M., Andrade M. F., Impact of time-dependent chemical boundary conditions on tropospheric ozone simulation with WRF-Chem: An experiment over the Metropolitan Area of São Paulo, Atmospheric Environment, 2018, vol. 195, p. 112

Gavidia-Calderón M. E., Ibarra-Espinosa S., Kim Y., Zhang Y., Andrade M. d. F., Simulation of O 3 and NO x in São Paulo street urban canyons with VEIN (v0. 2.2) and MUNICH (v1. 0), Geoscientific Model Development, 2021, vol. 14, p. 3251

Ginoux P., Chin M., Tegen I., Prospero J. M., Holben B., Dubovik O., Lin S. J., Sources and distributions of dust aerosols simulated 570 with the GOCART model, Journal of Geophysical Research Atmospheres, 2001

Godish T., Air Quality. vol. 4, 2004, 1

Gouveia N., Fletcher T., Time series analysis of air pollution and mortality: effects by cause, age and socioeconomic status, Journal of epidemiology and community health, 2000, vol. 54, p. 750

Grell G. A., Devenyi D., A Generalized Approach to Parameterizing Convection Combining Ensemble and Data Assimilation Techniques, Geophysical Research Letters, 2002, vol. 29, p. 1693

Grell G. A., Freitas S. R., Stuefer M., J. F., Inclusion of biomass burning in WRF-Chem: impact of wildfires on weather forecasts, Atmospheric Chemistry and Physics, 2011, vol. 11, p. 5289

Grell G. A., Peckham S. E., Schmitz R., McKeen S. A., Frost G., Skamarock W. C., Eder B., Fully coupled "on-line" chemistry within the WRF model, Atmospheric Environment, 2005, vol. 39, p. 6957–6976

Guenther A., Karl T., Harley P., Wiedinmyer C., Palmer P. I., Geron C., Estimates of global terrestrial isoprene emissions using MEGAN (Model of Emissions of Gases and Aerosols from Nature), Atmospheric Chemistry and Physics, 2006, vol. 6, p. 3181

Health E., Healthy housing reference manual rev. 2006. edn. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for

Environmental Health, U.S. Department of Housing and Urban Development Atlanta, 2006, 1

Hertel O., Salmond J., Bloss W., Air quality in urban environments. vol. 28, Royal Society of Chemistry, 2009

Hoelzemann J. J., Longo K. M., Fonseca R. M., do Rosário N. M. E., Elbern H., Freitas S. R., Pires C., Regional representativity of AERONET observation sites during the biomass burning season in South America determined by correlation studies with MODIS Aerosol Optical Depth, Journal of Geophysical Research: Atmospheres, 2009, vol. 114, p. D13301,1

Holben B., Tanre D., Smirnov A., Eck T., Slutsker I., Abuhassan N., Newcomb W. W., Schafer J., Chatenet B., Lavenue F., Kaufman Y. J., Castle J. V., Setzer A., Markham B., Clark D., Frouin R., Halthore R., Karnieli A., O'Neill N. T., Zibordi G., An emerging ground-based aerosol climatology: Aerosol Optical Depth from AERONET, Journal of Geophysical Research, 2001, vol. 106, p. 12,067

Holben B. N., Eck T. F., Slutsker I., Tanré D., Buis J. P., Setzer A., Vermote E., Reagan J. A., Kaufman Y. J., Nakajima T., Lavenu F., Jankowiak I., Smirnov A., AERONET - A Federated Instrument Network and Data Archive for Aerosol Characterization, Remote Sensing of Environment, 1998, vol. 66, p. 1

Holman C., , 1999 in Holgate S. T., Samet J. M., Koren H. S., Maynard R. L., eds, , Air Pollution and Health. Academic Press London pp 115–148

Hong Song-You Yign N., Dudhia J., A new vertical diffusion package with an explicit treatment of entrainment processes, Monthly Weather Review - American Meteorological Society, 2006, vol. 134, p. 2318

Hoshyaripour G., Brasseur G., Andrade M. d. F., Gavidia-Calderón M. E., Bouarar I., Ynoue R. Y., Prediction of ground-level ozone concentration in São Paulo, Brazil: Deterministic versus statistic models, Atmospheric environment, 2016, vol. 145, p. 365

Hsu A., Reuben A., Shindell D., de Sherbinin A., Levy M., Toward the next generation of air quality monitoring indicators, Atmospheric Environment, 2013, vol. 80, p. 561

Hsu N., Tsay S., King M., Herman J., Aerosol properties over bright-reflecting source regions, IEEE Transactions on Geoscience and Remote Sensing, 2004, vol. 42, p. 557–569

Hsu N. C., Lee J., Sayer A. M., Carletta N., Chen S.-H., Tucker C. J., Holben B. N., Tsay S.-C., Retrieving near-global aerosol loading over land and ocean from AVHRR, Journal of Geophysical Research: Atmospheres, 2017, vol. 122, p. 9968–9989

Huffman G. J., Bolvin D. T., Braithwaite D., Hsu K., Joyce R., Kidd C., Nelkin E. J., Xie P., NASA Global Pre-cipitation Measurement Integrated Multi-Satellite Retrievalsfor GPM (IMERG), Algorithm Theoretical Basis Document, version4.5, 2015, p. 30

Ibrahim S., Halounova L., Statistical study of MODIS algorithms in estimating Aerosol Optical Depth over The Czech Republic, Civil Engineering Journal, 2019

Ichoku C., Chu D., Mattoo S., Kaufman Y., Remer L., Tanre D., Slutsker I., Holben B., A spatio-temporal approach for global validation and analysis of MODIS aerosol products, Geophysical Research Letters, 2002, vol. 29, p. MOD1

IEMA, 2022 Qualidade do ar no municipio de Sao Paulo: Nota tecnica Maio 2022

Jang Y., Woo J., Kim Y., Kim J., Park M., Kim B., Impact of Land Cover and Leaf Area Index on BVOC Emissions over the Korean Peninsula. In AGU Fall Meeting Abstracts , vol. 2019, 2019, p. A33I

Jiménez P. A., Dudhia J., On the Ability of the WRF Model to Reproduce the Surface Wind Direction over Complex Terrain, Journal of Applied Meteorology and Climatology, 2013, vol. 52, p. 1610

Joiner J., Bhartia P. K., The determination of cloud pressures fromrotational-Raman scattering in satellite backscatter ultraviolet measurements, Journal of Geophysical Research, 1995, vol. 1002, p. 23019

Joyce R. J., Janowiak J. E., Arkin P. A., Xie P., CMORPH:A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution, Journal of Hydrometeorology - American Meteorological Society, 2004, vol. 5, p. 487–503

Junyan Y., Beixiang S., Yi S., Simon M., Yi Z., Geyang X., Air pollution dispersal in high density urban areas: Research on the triadic relation of wind, air pollution, and urban form, Sustainable Cities and Society, 2020, vol. 54, p. 101941

Kanakidou M., Seinfeld J. H., Pandis S. N., Barnes I., Dentener F. J., Facchini M. C., Van Dingenen R., Ervens B., Nenes A., Nielsen C. J., et al., Organic aerosol and global climate modelling: a review, Atmospheric Chemistry and Physics, 2005, vol. 5, p. 1053

Kaufman Y. J., Tanré D., Remer L. A., Vermote E., Chu A., N.Holben B., Operational remote sensing of tropospheric aerosol over land from EOS Moderate Resolution Imaging Spectroradiometer, Journal of Geophysical Research, 1997, vol. 102, p. 17,051

Kopacz M., Jacob D. J., Fisher J., Logan J., Zhang L., Megretskaia I., Yantosca R., Singh K., Henze D. K., Burrows J., Buchwitz M., Khlystova I., McMillan W. W., Gille J. C., Edwards D. P., Eldering A., Thouret V., Nedelec P., Global estimates of CO sources with high resolution by adjoint inversion of multiple satellite datasets (MOPITT, AIRS, SCIAMACHY, TES), Atmospheric Chemistry and Physics, 2010, vol. 10, p. 855

Koppmann R., Volatile organic compounds in the atmosphere. John Wiley & Sons, 2008

Kumar N., Chu A., Foster A., An empirical relationship between $PM_{2.5}$ and aerosol optical depth in Delhi Metropolitan, Atmospheric Environment, 2007, vol. 41, p. 4492

Kumar R., Naja M., Pfister G., Barth M., Wiedinmyer C., Brasseur G., Simulations over South Asia using the Weather Research and Forecasting model with Chemistry (WRF-Chem): Chemistry evaluation and initial results, Geoscientific Model Development, 2012, vol. 5, p. 619

Lamarque J. F., Emmons L. K., Hess P. G., Kinnison D. E., Tilmes S., Vitt F., Heald C. L., Holland E. A., Lauritzen P. H., Neu J., Orlando J. J., Rasch P. J., Tyndall G. K., CAM-chem: description and evaluation of interactive atmospheric chemistry in the Community Earth System Model, Geoscientific Model Development, 2012, vol. 5, p. 369

Laprise R., The Euler Equations of Motion with Hydrostatic Pressure as an Independent Variable, Monthly Weather Review, 1992, vol. 120, p. 197

Lee C. B., Kim J.-C., Belorid M., Zhao P., Performance evaluation of four different land surface models in WRF, Asian Journal of Atmospheric Environment, 2016, vol. 10, p. 42

Lenoble J., Atmospheric radiative transfer. A. Deepak Pub., 1993

Levelt P. F., van den Oord G. H. J., Dobber M. R., Malkki A., Visser H., de Vries J., Stammes P., Lundell J. O. V., Saari H., The ozone monitoring instrument, IEEE Transactions on Geoscience and Remote Sensing, 2006, vol. 44, p. 1093

Levy R., Remer L., Dubovik O., Global aerosol optical properties and application to Moderate Resolution Imaging Spectroradiometer aerosol retrieval over land, Journal of Geophysical Research-Atmosphere, 2007a, vol. 112, p. D13210

Levy R., Remer L., Mattoo S., Vermote E., Kaufman Y. J., Second-generation operational algorithm: Retrieval of aerosol properties over land from inversion of Moderate Resolution Imaging Spectroradiometer spectral reflectance, Journal of Geophysical Research-Atmosphere, 2007b, vol. 112, p. D13211

Levy R. C., Mattoo S., Munchak L. A., Remer L. A., Sayer A., Hsu N. C., The Collection 6 MODIS aerosol products over land and ocean, Atmospheric Measurement Techniques Discussions, 2013, vol. 6, p. 159–259

Li Y., Shi G., Sun Z., Evaluation and improvement of MODIS aerosol optical depth products over China, Atmospheric Environment, 2020, vol. 223, p. 117251

Liou K.-N., An introduction to atmospheric radiation. vol. 84, Elsevier, 2002

Liu C., Chen R., Sera F., Vicedo-Cabrera A. M., Guo Y., Tong S., Coelho M. S., Saldiva P. H., Lavigne E., Matus P., Valdes Ortega N., Osorio Garcia S., Pascal M., Stafoggia M., Scortichini M., Hashizume M., Honda Y., Hurtado-Díaz M., Cruz J., Nunes B., Teixeira J. a. P., Kim H., Tobias A., Íñiguez C., Forsberg B., Åström C., Ragettli M. S., Guo Y.-L., Chen B.-Y., Bell M. L., Wright C. Y., Scovronick N., Garland R. M., Milojevic A., Kyselý J., Urban A., Orru H., Indermitte E., Jaakkola J. J., Ryti N. R., Katsouyanni K., Analitis A., Zanobetti A., Schwartz J., Chen J., Wu T., Cohen A., Gasparrini A., Kan H., Ambient Particulate Air Pollution and Daily Mortality in 652 Cities, New England Journal of Medicine, 2019, vol. 381, p. 705

Lorenzini G., Saitanis C., , 2003 in , Abiotic stresses in plants. Springer pp 205–229

Lucht W., Schaaf C. B., Strahler A. H., An algorithm for the retrieval of albedo from space using semiempirical BRDF models, IEEE Transactions on Geoscience and Remote Sensing, 2000, vol. 38, p. 977–998

Lyapustin A., Wang Y., Korkin S., Huang D., MODIS Collection 6 MAIAC algorithm, Atmospheric Measurement Techniques, 2018, vol. 11, p. 5741–5765

Lyapustin A., Wang Y., Laszlo I., Korkin S., Improved cloud and snow screening in MAIAC aerosol retrievals using spectral and spatial analysis, Atmospheric Measurement Techniques, 2012c, vol. 5, p. 843–850

Mar K. A., Ojha N., Pozzer A., Butler T. M., Ozone air quality simulations with WRF-Chem (v3.5.1) over Europe: model evaluation and chemical mechanism comparison, Geoscientific Model Development, 2016, vol. 9, p. 3699–3728

Martin R. V., Jacob D. J., Chance K., Kurosu T. P., Palmer P. I., Evans M. J., Global inventory of nitrogen oxide emissions constrained by space-based observations of NO_2 columns, Journal of Geophysical Research: Atmospheres, 2003, vol. 108

Martínez-Alonso S., Deeter M. N., Baier B. C., McKain K., Worden H., Borsdorff T., Sweeney C., Aben I., Evaluation of MOPITT and TROPOMI carbon monoxide retrievals using AirCore in situ vertical profiles, Atmospheric Measurement Techniques, 2022, vol. 15, p. 4751

Martins L. D., Andrade M. F., Freitas Edmilson D., Pretto A., Gatti L. V., Albuquerque [U+FFFD] L., Tomaz E., Guardani M. L., Martins M. H., Junior Olimpio M. A., Emission factors for gas-powered vehicles traveling through road tunnels in São Paulo, Brazil, Environmental science & technology, 2006, vol. 40, p. 6722

Martins L. D., Hallak R., Alves R. C., de Almeida D. S., Squizzato R., Moreira C. A., Beal A., da Silva I., Rudke A., Martins J. A., Long-range Transport of Aerosols from Biomass Burning over Southeastern South America and their Implications on Air Quality, Aerosol and Air Quality Research, 2018, vol. 18, p. 1734

Martins M. C., Fatigati F. L., Véspoli T. C., Martins L. C., Pereira L. A., Martins M. A., Saldiva P. H., Braga A. L., Influence of socioeconomic conditions on air pollution adverse health effects in elderly people: an analysis of six regions in São Paulo, Brazil, Journal of epidemiology and community health, 2004, vol. 58, p. 41–46

Martins V. S., Lyapustin A., deCarvalho L. A. S., Barbosa C. C. F., Novo E. M. L. M., Validation of high-resolution MAIAC aerosol product over South America, Journal of Geophysical Research-Atmosphere, 2017, vol. 122, p. 7537–7559

Mass C., Ovens D., WRF model physics: Problems, solutions and a new paradigm for progress. In Preprints, WRF Users' Workshop, Boulder, CO, NCAR , 2010

Mass C., Ovens D., Fixing WRF's high speed wind bias: A new subgrid scale drag parameterization and the role of detailed verification. In 24th conference on weather and forecasting and 20th conference on numerical weather prediction, preprints, 91st American meteorological society annual meeting , vol. 23727, 2011

Mhawish A., Sorek-Hamer M., Banerjee T., Lyapustin A., Broday D. M., Chatfield R., Intercomparison and evaluation of MODIS C6 MAIAC, DB and DT AOD over South Asia, Remote Sensing of Environment, 2018, vol. 40

Miranda R. M., de Fatima Andrade M., Dutra Ribeiro F. N., Mendonça Francisco K. J., Pérez-Martínez P. J., Source apportionment of fine particulate matter by positive matrix factorization in the metropolitan area of São Paulo, Brazil, Journal of Cleaner Production, 2018, vol. 202, p. 253–263

Mlawer E. J., Taubman S. J., Brown P. D., Iacono M. J., Clough S. A., Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated–k model for the longwave, Journal of Geophysical Research, 1997, vol. 102, p. 16663

Monk K., Guérette E.-A., Paton-Walsh C., Silver J. D., Emmerson K. M., Utembe S. R., Zhang Y., Griffiths A. D., Chang L. T. C., Duc H. N., Trieu T., Scorgie Y., E. C. M., Evaluation of Regional Air Quality Models over Sydney and Australia: Part 1—Meteorological Model Comparison, Atmosphere, 2019, vol. 10

Morrison H., G. T., V. T., Impact of cloud microphysics on the development of trailing

stratiform precipitation in asimulated squall line: Comparison of one- and two-moment schemes, Monthly Weather Review, 2009, vol. 3, p. 991–1007

Myhre G., Myhre C., Samset B. H., Storelvmo T., Aerosols and their relation to global climate and climate sensitivity, Nature Education Knowledge, 2013, vol. 4, p. 7

Nandi I., Srivastava S., Yarragunta Y., Kumar R., Mitra D., Distribution of surface carbon monoxide over the Indian subcontinent: Investigation of source contributions using WRF-Chem, Atmospheric Environment, 2020, vol. 243, p. 117838

Nascimento J. P., Barbosa H. M. J., Banducci A. L., Rizzo L. V., Vara-Vela A. L., Meller B. B., Gomes H., Cezar A., Franco M. A., Ponczek M., et al., Major Regional-Scale Production of O_3 and Secondary Organic Aerosol in Remote Amazon Regions from the Dynamics and Photochemistry of Urban and Forest Emissions, Environmental Science & Technology, 2022, vol. 56, p. 9924

Nogueira T., Kamigauti L. Y., Pereira G. M., Gavidia-Calderon M. G., Ibarra-Espinosa S., Oliveira G. L., Miranda R. M., Vasconcelos P. C., Freitas E. D., Andrade M. F., BEvolution of Vehicle Emission Factors in a Megacity Affected by Extensive Biofuel Use: Results of Tunnel Measurements in São Paulo, Brazil, npj Climate and Atmospheric Science, 2021, vol. 55(10), p. 6677

Oke T. R., Mills G., Christen A., Voogt J. A., Urban climates. Cambridge University Press, 2017

Oliveira E. M. G., , 2014 Parametrizações convectivas no modelo WRF e sua relação com a precipitação durante ciclogêneses no sudeste da América do Sul

Orlando J. P., Alvim D. S., Yamazaki A., Corrêa S. M., Gatti L. V., Ozone precursors for the São Paulo Metropolitan Area, Science of The Total Environment, 2010, vol. 408, p. 1612

Oyama B. S., Andrade M. F., Herckes P., Dusek U., Rockmann T., Holzinger R., Chemical characterization of organic particulate matter from on-road traffic in São Paulo, Brazil, Atmospheric Chemistry and Physics, 2016, vol. 16, p. 14397

Pascal M., Corso M., Chanel O., Declercq C., Badaloni C., Cesaroni G., Henschel S., Meister K., Haluza D., Martin-Olmedo P., Medina S., Assessing the public health impacts of urban air pollution in 25 European cities: Results of the Aphekom project, Science of The Total Environment, 2013, vol. 449, p. 390

Pey J., Pérez N., Querol X., Alastuey A., Cusack M., Reche C., Intense winter atmospheric pollution episodes affecting the Western Mediterranean, Science of the total environment, 2010, vol. 408, p. 1951

Ponche J.-L., Vinuesa J.-F., Emission scenarios for air quality management and applications at local and regional scales including the effects of the future European emission regulation (2015) for the upper Rhine valley, Atmospheric Chemistry and Physics, 2005, vol. 5, p. 999

Popescu F., Ionel I., , 2010 in Kumar A., ed., , Air Quality. IntechOpen Rijeka Chapt. 1

Pérez-Martínez P. J., F. A. M., de Miranda R. M., Traffic-related air quality trends in São Paulo, Brazil, Journal of Geophysical Research: Atmospheres, 2015, vol. 120, p. 6290

Randall D. A., Wood R. A., Bony S., Colman R., Fichefet T., Fyfe J., Kattsov V., Pitman A., Shukla J., Srinivasan J., et al.,, 2007 in , Climate change 2007: The physical science basis. Contribution of Working Group I to the Fourth Assessment Report of the IPCC (FAR). Cambridge University Press pp 589–662

Remer L. A., Kaufman Y. J., Tanre D., Mattoo S., Chu D. A., Martins J. V., Li R. R., Ichoku C., Levy R. C., Kleidman R. G., Eck T. F., Vermote E., Holben B. N., The MODIS aerosol algorithm, products, and validation, Journal of the Atmospheric Sciences, 2005, vol. 62, p. 947

Remer L. A., Mattoo S., Levy R. C., Munchak L. A., MODIS 3 km aerosol product: algorithm and global perspective, Atmospheric Measurement Techniques, 2013, vol. 6, p. 1829

Richmond-Bryant J., Chris Owen R., Graham S., Snyder M., McDow S., Oakes M., Kimbrough S., Estimation of on-road NO_2 concentrations, NO_2/NO_X ratios, and related roadway gradients from near-road monitoring data, Air Quality, Atmosphere & Health, 2017, vol. 10, p. 611

Rojas Benavente N., , 2014 Estudio de la dinámica del espesor óptico de los aerosoles producido por las fuentes naturales y antropogénicas a partir de las imágenes del sensor modis a bordo del satélite Terra y Aqua sobre américa del Sur (2000-20012)

Roux G., Liu Y., Monache L. D., Sheu R.-S., Warner T. T., Verification of high resolution WRF-RTFDDA surface forecasts over mountains and plains. In 10th WRF users' workshop , 2009, p. 20

Roy P. S., Behera M. D., Srivastav S. K., Satellite remote sensing: sensors, applications and techniques, Proceedings of the National Academy of Sciences, India Section A: Physical Sciences, 2017, vol. 87, p. 465

Rozante J. R., Rozante V., Souza Alvim D., Ocimar Manzi A., Barboza Chiquetto J., Siqueira D'Amelio M. T., Moreira D. S., Variations of Carbon Monoxide Concentrations in the Megacity of São Paulo from 2000 to 2015 in Different Time Scales, Atmosphere, 2017, vol. 8

Saldiva P. H. N., Clarke R. W., Coull B. A., Stearns R. C., Lawrence J., Murthy G. G. K., Diaz E., Koutrakis P., Suh H., Tsuda A., et al., Lung inflammation induced by concentrated ambient air particles is related to particle composition, American journal of respiratory and critical care medicine, 2002, vol. 165, p. 1610

Santiago A., , 2009 Simulação da Camada Limite Planetária sobre a Região Metropolitana da Grande Vitória com uso do modelo WRF

Santos T., Dominutti P., Pedrosa G., Coelho M., Nogueira T., Borbon A., Souza S., Fornaro A., Isoprene in urban Atlantic forests: Variability, origin, and implications on the air quality of a subtropical megacity, Science of The Total Environment, 2022, vol. 824, p. 153728

Sarkar M., Venkataraman C., Guttikunda S., Sadavarte P., Indian emissions of technology-linked NMVOCs with chemical speciation: an evaluation of the SAPRC99 mechanism with WRF-CAMx simulations, Atmospheric Environment, 2016, vol. 134, p. 70–83

Schneider C., Kessler C., Moussiopoulos N., Influence of emission input data on ozone level predictions for the Upper Rhine valley, Atmospheric Environment, 1997, vol. 31, p. 3187

Schuch D., de Freitas E., Espinosa S. I., et al. A two decades study on ozone variability and trend over the main urban areas of the São Paulo state, Brazil, Environmental Science and Pollution Research, 2019, vol. 26, p. 31699–31716

Schuch D. A., Ibarra S., Freitas E. D., Andrade M. F., emissV: A Preprocessor for WRF-Chem Model, Journal of Atmospheric Science Research, 2018, vol. 1, p. 1

Schwela D., Haq G., Huizenga C., Han W.-J., Fabian H., Ajero M., Urban air pollution in Asian cities: status, challenges and management. Routledge, 2012

Seinfeld J. H., Pandis S. N., Atmospheric Chemistry and Physics: From Air Pollution to Climate Change. Wiley, 2016

Sekiya T., Miyazaki K., Eskes H., Sudo K., Takigawa M., Kanaya Y., A comparison of the impact of TROPOMI and OMI tropospheric NO_2 on global chemical data assimilation, Atmospheric Measurement Techniques, 2022, vol. 15, p. 1703

Silva R. S., Andrade M. d. F., Validação de poluentes fotoquímicos e inclusão do inventário de emissões no modelo de qualidade do ar WRF/CHEM, para a região metropolitana de São Paulo, Revista Brasileira de Meteorologia, 2013, vol. 28, p. 105

Simpson D., Malik S., Photochemical oxidant modelling and source-receptor relationships for ozone, EMEP MSC-W Report, 1996, vol. 1, p. 96

Skamarock W. C., Klemp J. B., Dudhia J., Gill D. O., Barker D. M., Duda M. G., Huang X., Wang W., Powers J. G., A Description of the Advanced Research WRF Version 3, University Corporation for Atmospheric Research, 2008, vol. 8

Song H.-J., Sohn B.-J., An Evaluation of WRF Microphysics Schemes for Simulating the Warm-Type Heavy Rain over the Korean Peninsula, Asia-Pacific Journal of Atmospheric Sciences, 2018, vol. 54, p. 1

Sportisse B., Fundamentals in air pollution: from processes to mod-elling. Atmospheric Boundary Layer, Springer, 2010, 93–132

Squizzato R., Nogueira T., Martins L. D., Martins J. A., Astolfo R., Machado C. B., Andrade M. F., Freitas E. D., Beyond megacities: tracking air pollution from urban

areas and biomass burning in Brazil, npj Climate and Atmospheric Science, 2021, vol. 4, p. 17

Stensrud D. J., Parameterization schemes: keys to understanding numerical weather prediction models. Cambridge University Press, 2009

Stocker T. F., Qin D., Plattner G. K., Tignor M., Allen S. K., Boschung J., Nauels A., Xia Y., Bex V., Midgley P. M., , 2013 IPCC, 2013: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change

Stull R. B., Ahrens C. D., Meteorology for Scientists and Engineers. Earth Science Series, Brooks/Cole, 2000

Sánchez-Ccoyllo O. R., Silva Dias P. L., de Fátima Andrade M., Freitas S. R., Determination of O_3-, CO- and PM_{10}-transport in the metropolitan area of São Paulo, Brazil through synoptic-scale analysis of back trajectories, Meteorology and Atmospheric Physics, 2006, vol. 92, p. 83

Taghavi S., Barac T., Greenberg B., Borremans B., Vangronsveld J., van der Lelie D., Horizontal gene transfer to endogenous endophytic bacteria from poplar improves phytoremediation of toluene, Applied and environmental microbiology, 2005, vol. 71, p. 8500

Tang W., Worden H., Deeter M., Edwards D., Emmons L., Martínez-Alonso S., Gaubert B., Buchholz R., Diskin G., Dickerson R., Ren X., He H., Kondo Y., Assessing Measurements of Pollution in the Troposphere (MOPITT) carbon monoxide retrievals over urban versus non-urban regions, Atmospheric Measurement Techniques, 2020, vol. 13, p. 1337

Tao D., Jia G., Yuan Y., Zhao H., A Digital Sensor Simulator of the Pushbroom Offner Hyperspectral Imaging Spectrometer, Sensors (Basel, Switzerland), 2014, vol. 14, p. 23822

Tewari M., Chen F., Wang W., Dudhia J., LeMone M. A., Mitchell K., Ek M., Gayno G., Wegiel J., Cuenca R. H., Implementation and verification of the unified NOAH land sur-

face model in the WRF model, 20th conference on weather analysis and forecasting/16th conference on numerical weather prediction, 2004, pp 11–15

Thielmann A., Prévôt A. S. H., Staehelin J., Sensitivity of ozone production derived from field measurements in the Italian Po basin, Journal of Geophysical Research, 2002, vol. 107, p. 8194

Turpin B. J., Saxena P., Andrews E., Measuring and simulating particulate organics in the atmosphere: problems and prospects, Atmospheric Environment, 2000, vol. 34, p. 2983

Valin L., Russell A., Hudman R., Cohen R., Effects of model resolution on the interpretation of satellite NO_2 observations, Atmospheric Chemistry and Physics, 2011, vol. 11, p. 11647

Van Donkelaar A., Martin R. V., Park R. J., Estimating ground-level PM2.5 using aerosol optical depth determined from satellite remote sensing, Journal of Geophysical Research: Atmospheres, 2006, vol. 111

Vara-Vela A., Andrade M. F., Kumar P., Ynoue R. Y., Muñoz 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, 2016, vol. 16, p. 777

Vara-Vela A., Andrade M. F., Zhang Y., Kumar P., Ynoue R. Y., Souto-Oliveira C. E., da Silva L. F., Landulfo E., Modeling of atmospheric aerosol properties in the Sao Paulo Metropolitan Area: Impact of biomass burning, Journal of Geophysical Research, 2018, vol. 123, p. 9935–9956

Vara-Vela A. L., Herdies L. D., Alvim D. S., Vendrasco E. P., Figueroa S. N., Pendharkar J., Fernandez J. P. R., A New Predictive Framework for Amazon Forest Fire Smoke Dispersion over South America, Bulletin of the American Meteorological Society, 2021, vol. 102, p. 1700

Vera C., Higgins W., Amador J., Ambrizzi T., Garreaud R., Gochis D., Gutzler D., Lettenmaier D., Marengo J., Mechoso C. R., et al., Toward a unified view of the American monsoon systems, Journal of climate, 2006, vol. 19, p. 4977

Villányi V., Turk B., Batic F., Csintalan Z., Ozone pollution and its bioindication, Air pollution, 2010, vol. 153

Wang J., Christopher S. A., Intercomparison between satellite-derived aerosol optical thickness and PM2. 5 mass: Implications for air quality studies, Geophysical research letters, 2003, vol. 30

Wang L. K.-P., Pereira N. C., Hung Y.-T., Air Pollution Control Engineering, 2004

Wang Y., Zhuang G., Sun Y., An Z., The variation of characteristics and formation mechanisms of aerosols in dust, haze, and clear days in Beijing, Atmospheric Environment, 2006, vol. 40, p. 6579

WHO The World Health Report 2002: Reducing Risks, Promoting Healthy Life, 2002

WHO WHO global air quality guidelines: particulate matter ($PM_{2.5}$ and PM_{10}), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide, 2021, xxi, 273 p.

Wiedinmyer C., Akagi S. K., Yokelson R. J., Emmons L. K., Al-Saadi J. A., Orlando J. J., Soja A. J., The Fire Inventory from Ncar (FINN): A High Resolution Global Model to Estimate the Emissions from Open Burning, Geoscientific Model Development, 2011, vol. 4, p. 625

Xu L., Penner J. E., Global simulations of nitrate and ammonium aerosols and their radiative effects, Atmospheric Chemistry and Physics, 2012, vol. 12, p. 9479

Yadav I. C., Linthoingambi Devi N., Li J., Syed J. H., Zhang G., Watanabe H., Biomass burning in Indo-China peninsula and its impacts on regional air quality and global climate change-a review, Environmental Pollution, 2017, vol. 227, p. 414

Zaveri R. A., Berkowitz C. M., Kleinman L. I., Springston S. R., Doskey P. V., Lonneman W. A., Spicer C. W., Ozone production efficiency and NO_x depletion in an urban plume: Interpretation of field observations and implications for evaluating O_3 - NO_x - VOC sensitivity, Journal of Geophysical Research, 2003, vol. 108, p. 4436

Zaveri R. A., Peters L. K., A new lumped structure photochemical mechanism for large-scale applications, Journal of Geophysical Research, 1999, vol. 104, p. 30387–30415

Zhang H., Zhao S., Wang Z., Zhang X., Song L., The updated effective radiative forcing of major anthropogenic aerosols and their effects on global climate at present and in the future, International Journal of Climatology, 2016, vol. 36, p. 4029

Zhao W., , 2013 Study of the Radiation and Energy Balances Reproduced by WRF Using 10-year (2001-2010) Observations at Cabauw

Appendix

Appendix A

A brief description of the atmospheric WRF model components

A.1 General mathematical formulation

The WRF-ARW has fully compressible, Eulerian and non-hydrostatic equations with a run-time hydrostatic option which are conservative for scalar variables (Skamarock et al., 2008). The governing equations are based on the principles of conservation of momentum, mass and energy, from which it is possible to describe the flow of fluids and the relationships between changes in pressure, energy and density experienced, in this case, by an atmospheric parcel when subjected to different forces (Oke et al., 2017). Euler's equations are a simplification of the Navier-Stokes equations, based on the consideration of air as a non-viscous fluid. The details of the governing equations as well as the numerical methods used for their resolution can be consulted in Skamarock et al. (2008). The model uses terrain-following, hydrostatic-pressure vertical coordinate, called η coordinates, with constant pressure surface as top of the model boundary (Skamarock et al., 2008). A coordinate is assigned to each vertical level, ranging from one (on the surface) to zero (up to the top of the upper boundary of the model) (Santiago, 2009). Mathematically, the vertical coordinates are calculated from the normalization of the hydrostatic component of pressure, as expressed in Equation (A.1) (Laprise, 1992):

$$\eta = \frac{P_z - P_{ht}}{P_{hs} - P_{ht}}, \tag{A.1}$$

where P_z is the hydrostatic component of pressure at altitude z; P_{ht} and P_{hs} are the pressure values at the highest (top) and lowest (surface) level of the model, respectively. In Figure

A.1, a representation of the η coordinates is illustrated, in which it is possible that the vertical spacing is non-uniform, allowing a greater resolution close to the surface. Such a configuration is necessary for a better discretization of the variation of meteorological parameters caused by the interaction of the soil with the planetary boundary layer, as well as the interface region between troposphere and stratosphere, mainly due to rapid vertical variations in the temperature field.

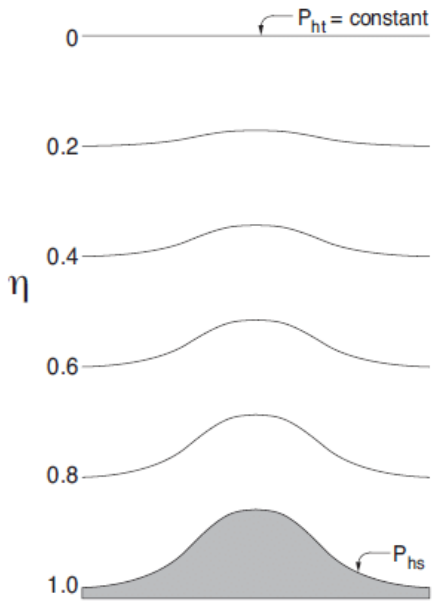

Figure A.1: Vertical coordinate system η of the WRF model.(Adapted from (Skamarock et al., 2008))

The simulation domain of the WRF model is the physical space represented in space by a three-dimensional calculation grid. Spatial discretization is done using an Arakawa C-type staggering grid (Arakawa and Lamb, 1977). The output of the modeled variables, both in the horizontal and vertical direction, for each grid cell are scaled as indicated in Figure A.2. In this type of grid, the thermodynamic and humidity variables are calculated for the grid center or center of mass (indicated by θ), while dynamic variables are allocated on the sides. The indices (i, j, k) refer to the location axis of each variable, whether in the direction of the x axis, y axis or η (vertical) levels. The variables u, v and w indicate the horizontal and vertical components of velocity, respectively. The numerical values that accompany the indices indicate relative distance, in this way the normal velocities are half the grid length away from the thermodynamic variables. The spacing Δx and Δy are constant, and their correction for different types of projections is made by a map factor

(Skamarock et al., 2008).

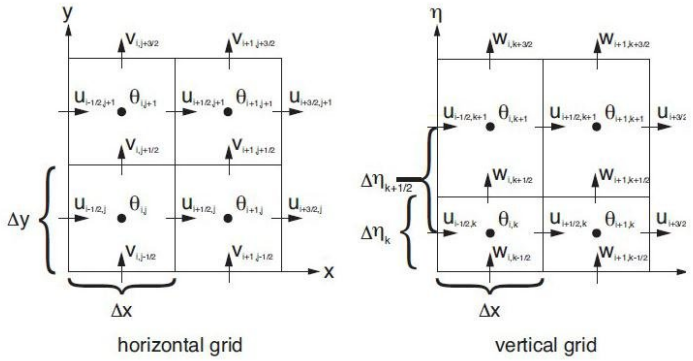

Figure A.2: Horizontal (left) and vertical (right) grid of Arakawa type C.(Adapted from (Skamarock et al., 2008))

A.2 Geographic allocation, initial and boundary conditions

The definition of the study area in the WRF meteorological model is carried out through the allocation of domains or grids in a region of the globe based on a central coordinate. The model also allows nesting configuration, allowing smaller domain with a finer grid within a larger domain with a broader grid. The reasons for using nesting are often related to the need to represent phenomena at a local scale without running the simulation for very large and high-resolution domains, which demand greater computer processing capacity, time and memory for storage, as well as the need to use initial and contour data in a better resolution. After defining the area of interest, some additional information is still needed, such as static geographic data, initial and boundary conditions. Such data can be acquired from a database supported by the model or entered by the user himself, in case he wants more updated or better resolution specific information. In addition to information about the surface, it is also necessary to determine the initial conditions. As the name implies, these provide initial values for the model variables for all grid points in all directions, depicting the state of the atmosphere at time t=0. Boundary conditions, on the other hand, aim to integrate the influence of advected flows from regions adjacent to the sides of the defined grids, and can be obtained from two sources, from data modeled by a broader domain (global or regional) or originating from more external domains (parent domain), if this is a nested grid, as can be seen in Figure A.3.

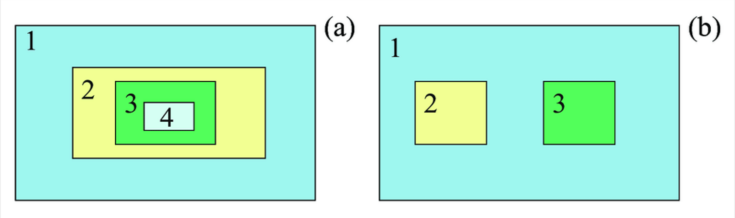

Figure A.3: Various nest configurations for multiple grids. (a) Telescoping nests and (b) nests at the same level with respect to a parent grid. (Adapted from (Skamarock et al., 2008))

The interaction between the outer and inner grids can be unidirectional (one-way) or bi-directional (two-way). In both options, the boundary conditions of the inner grid (that is, the side boundaries) are interpolated from the preview of the outer grid. For unidirectional mode, this is the only information exchange between them. In the second case, the interaction between the most external and internal domains is allowed both at the lateral limits and at the points of the external grid that are inside the internal grid, in a kind of feedback (Skamarock et al., 2008).

A.3 Physical parameterizations

Many of the atmospheric processes occur on a smaller scale than the grid spacing of meteorological models, and it is not possible to resolve them explicitly. Due to the limited computing power currently available, the use of grids with a scale of the same order in which such processes occur is prohibitive. The physical parameterization, therefore, can be understood as a representation of sub-grid scale phenomena, being resolved, therefore, implicitly. Although guided to some extent by theory, parameterizations are largely empirical in nature (Stensrud, 2009). In this sense, several physical schemes were developed to represent atmospheric phenomena, with different approximations and degrees of complexity. The uncertainties associated with such representations and their subsequent influence on the resolved scale, however, are a primary source of forecast errors (Stensrud, 2009). The physical processes to be parameterized in the WRF model are subdivided into six: cloud microphysics, cumulus convection, planetary boundary layer, surface layer, terrestrial surface and radiation (shortwave and longwave).

- Cloud Microphysics

  Cloud microphysics schemes address different interactions between types of precipi-

tation, being essential in improving rainfall prediction in higher resolution domains (a few kilometers) (Song and Sohn, 2018). Scheme options can include up to six hydrometeor classes: water vapor, cloud water, cloud ice, rain, snow and hail/graupel (Skamarock et al., 2008). The interactions between them include several processes such as nucleation, coalescence, accretion, freezing, precipitation, among others; and determine their interconversions (emergence or consumption). The differences between the microphysics schemes, therefore, lie in the types of hydrometeors considered and the interactions between them.

- Cumulus Convection

  The cumulus parameterization schemes are responsible for the sub-grid scale effects of convective clouds and/or shallow clouds, being applicable only in lower resolution grids (above 10 km) (Skamarock et al., 2008). They are intended to represent vertical flows due to unresolved updraft and downdraft currents, including the processes of entrainment and detrainment of saturated air in the cloud, resulting in the precipitation related to convective processes (Skamarock et al., 2008). The transition from such effects to larger-scale variables, which will be explicitly resolved by the model, takes place from the change in the vertical distribution of heat associated with the condensation of steam. This process occurs from the transport of moisture from the low to the upper levels by the updrafts and in the radiative balance of the atmosphere, modified by the presence of convective clouds (Oliveira, 2014).

- Radiation

  Radiation schemes calculate atmospheric heating due to radiative flux divergence and long-wave and short-wave downward surface radiation. Shortwave radiation comes from the Sun, however, the processes taken into account here include all the phenomena that occur in the Earth-atmosphere system, such as absorption, reflection, and scattering (Skamarock et al., 2008). Longwave radiation includes infrared or thermal radiation absorbed and emitted by gases and surfaces. In the case of radiative flux, the upward flux of short-wave radiation is due to surface albedo and long-wave radiation is determined by the surface emissivity, which, in turn, depends on the type of land use and the soil surface temperature (Zhao, 2013). Within the atmosphere the radiation responds to model-predicted cloud and water vapor distributions, as

well as specified carbon dioxide, ozone, and (optionally) trace gas concentrations (Skamarock et al., 2008).

- Surface Layer

  The surface layer models are responsible for calculating the friction velocities and exchange coefficients that allow the determination of heat and moisture fluxes by the land-surface models and surface stress by the planetary boundary layer scheme (Skamarock et al., 2008). The surface layer is the lowest portion of the atmosphere where mechanical turbulence is generated over convective turbulence. In this layer, friction due to drag forces, heat exchange by conduction, and evaporation cause substantial variations in wind speed, temperature, and humidity with height, although turbulent flows are approximately constant (Stull and Ahrens, 2000; Oke et al., 2017).

- Land Surface

  The land surface models make use of information from other parameterization categories. The atmospheric information come from the surface layer scheme, the radiative forcings from the radiation schemes and the precipitation forcings from the microphysics and cumulus convection schemes. Along with state variables and soil characteristics, heat and moisture fluxes over land and ocean are calculated, providing variables for atmospheric processes at the bottom of the boundary layer (Skamarock et al., 2008; Lee et al., 2016). Land surface models consider different degrees of complexity when calculating heat and mass fluxes, for which the number of soil layers can be changed as well as the effect of different phenomena produced by vegetation, urban layer, snow cover, among others (Skamarock et al., 2008).

- Planetary Boundary Layer

  The planetary boundary layer (PBL) schemes is to represent the vertical flows caused by the transport of eddies, based on information provided by the surface layer and land surface schemes (Skamarock et al., 2008). These eddies occur in spatio-temporal scales that cannot be explicitly represented in the grid and in the time step usually used in mesoscale models, being, for this reason, parameterized through PBL schemes (Cohen et al., 2015). For this reason, it becomes possible to determine the vertical distribution of temperature, humidity (including clouds) and horizontal momentum

throughout the atmospheric column.

The parameterization schemes are important because they considerably influence the predictions and despite being modular physical schemes, they interact with each other indirectly, through changes in the variables estimated by the model (see Figure A.4).

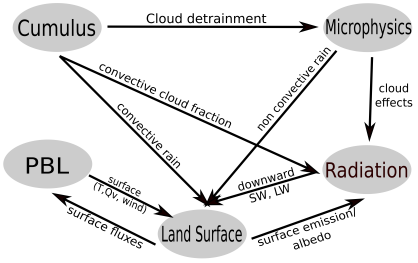

Figure A.4: Interactions between physical parameterizations in WRF model.

Appendix B

Pollutant transport from the MASP to other regions

B.1 Episode - 17 June 2019

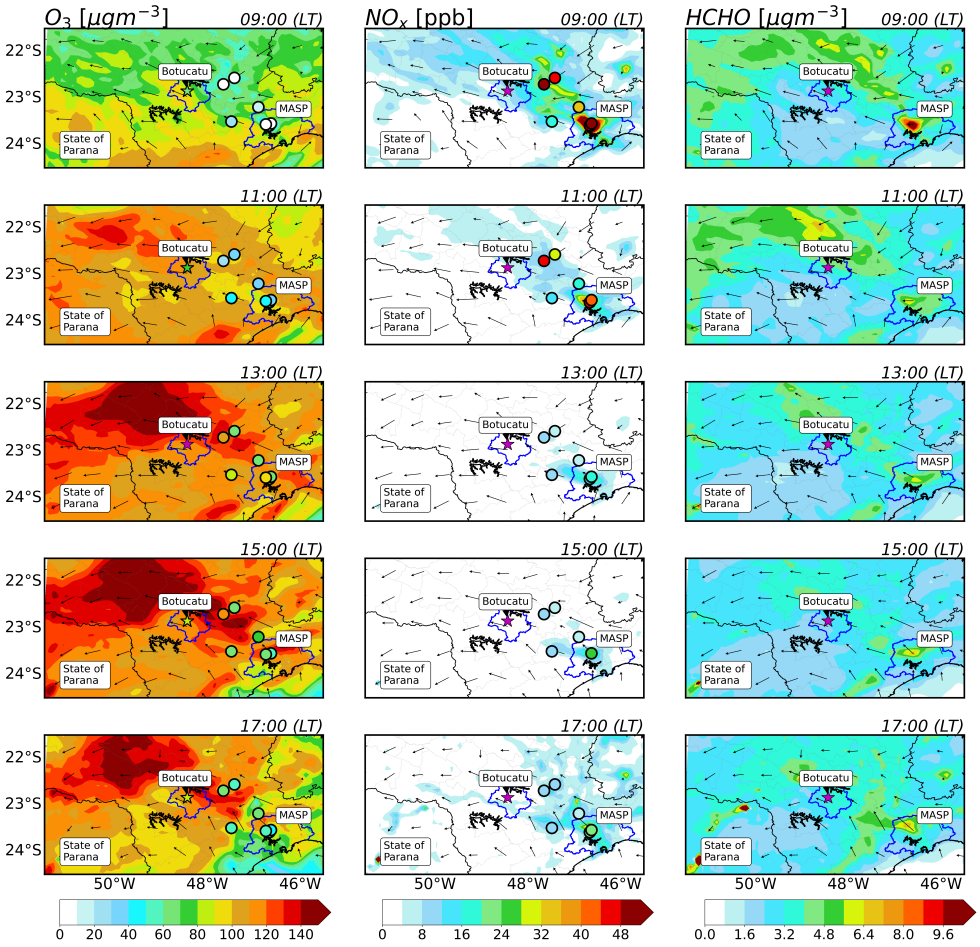

Figure B.1: Overlay of observed and predicted O_3 (in μg m^{-3}) (left panels), NO_x (in ppb) (middle panels) and HCHO (in μg m^{-3}) (right panels) spatial distributions during daytime for June 17, 2019. The observed values are indicated by the colored star and circles. The blue lines represent Botucatu city and the MASP.

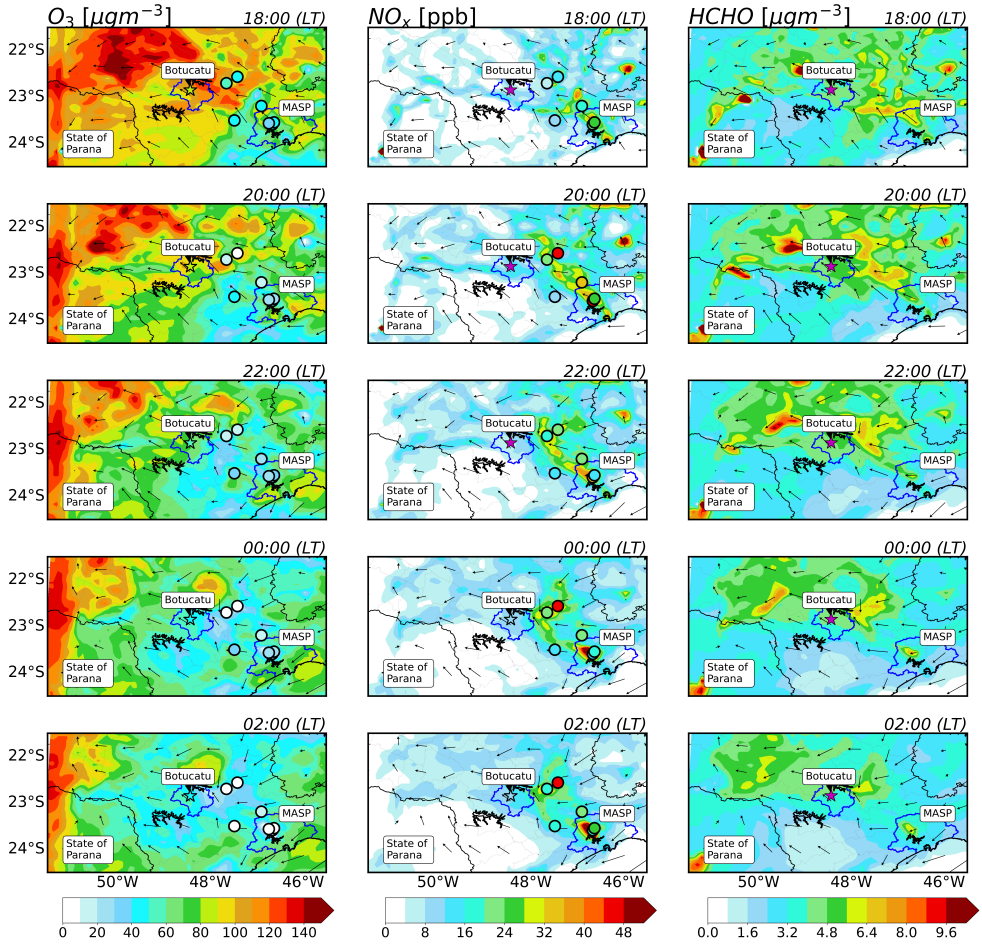

Figure B.2: Overlay of observed and predicted O_3 (in $\mu g\ m^{-3}$) (left panels), NO_x (in ppb) (middle panels) and HCHO (in $\mu g\ m^{-3}$) (right panels) spatial distributions during nighttime for June 17, 2019. The observed values are indicated by the colored star and circles. The blue lines represent Botucatu city and the MASP.

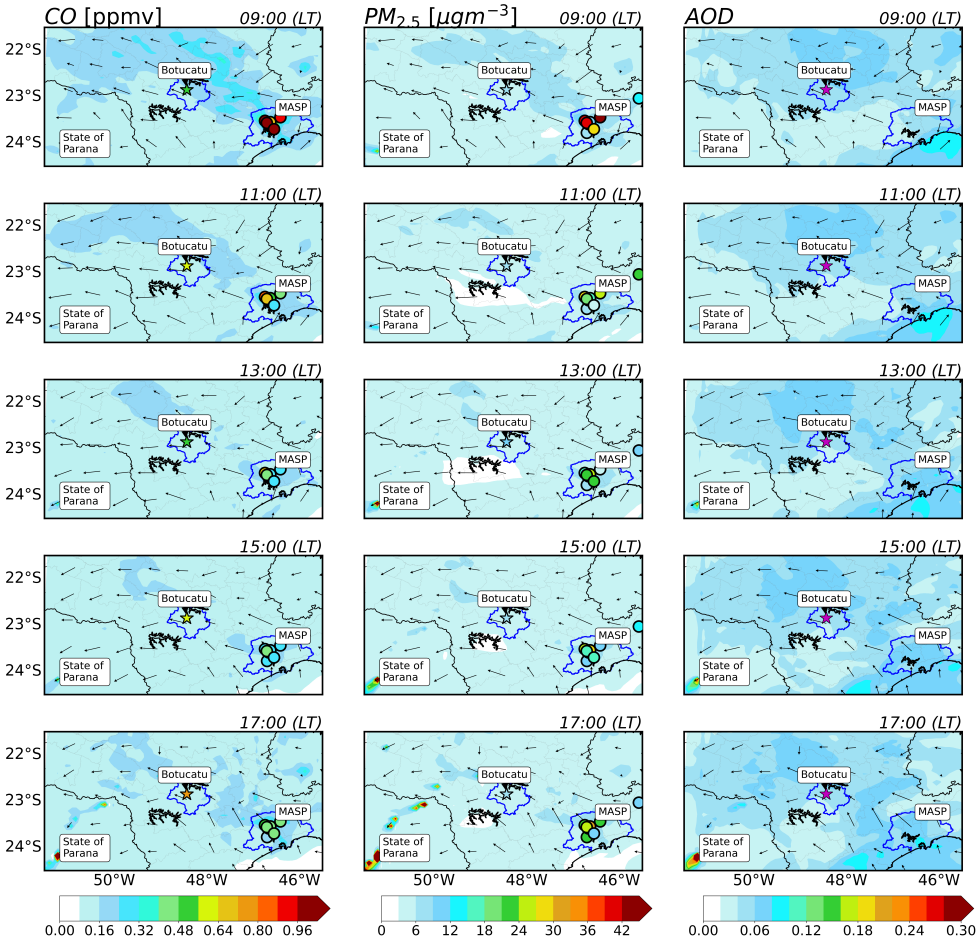

Figure B.3: Overlay of observed and predicted CO (in ppmv) (left panels), $PM_{2.5}$ (in μg m^{-3}) (middle panels) and AOD (right panels) spatial distributions during daytime for June 17, 2019. The observed values are indicated by the colored star and circles. The blue lines represent Botucatu city and the MASP.

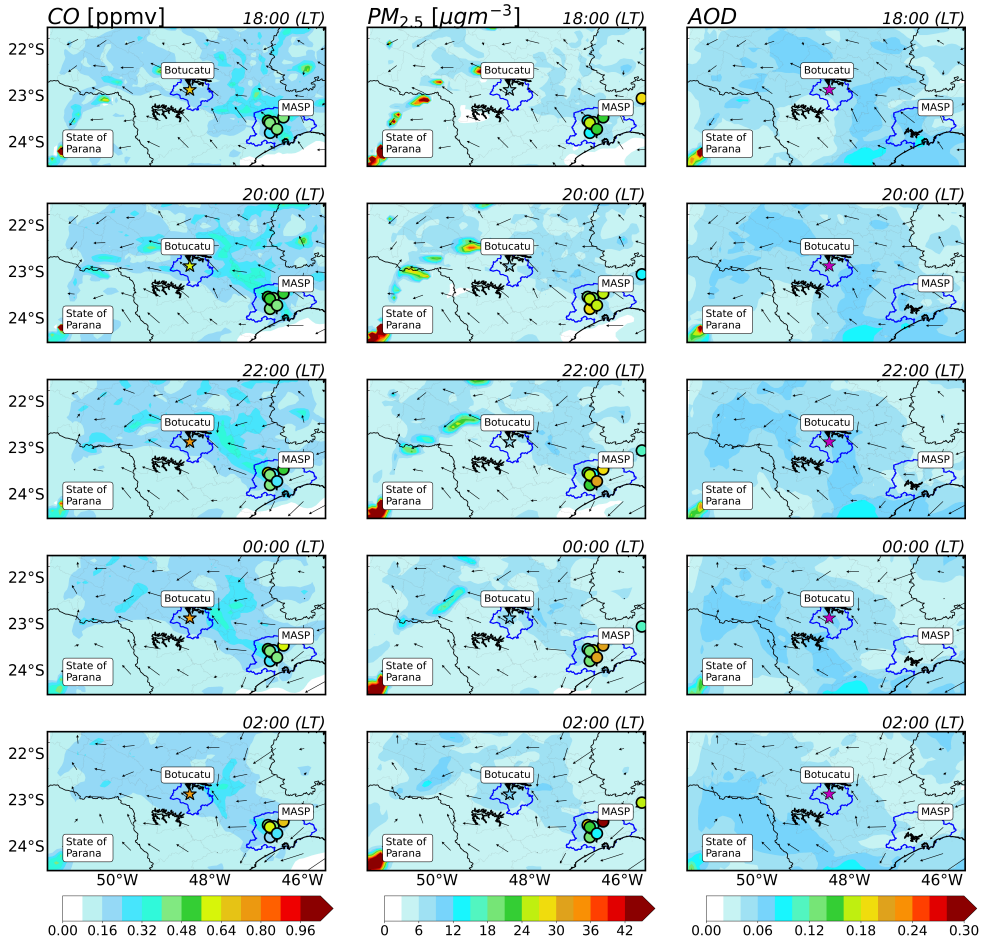

Figure B.4: Overlay of observed and predicted CO (in ppmv) (left panels), $PM_{2.5}$ (in $\mu g\ m^{-3}$) (middle panels) spatial distributions during nighttime for June 17, 2019. The observed values are indicated by the colored star and circles. The blue lines represent Botucatu city and the MASP.

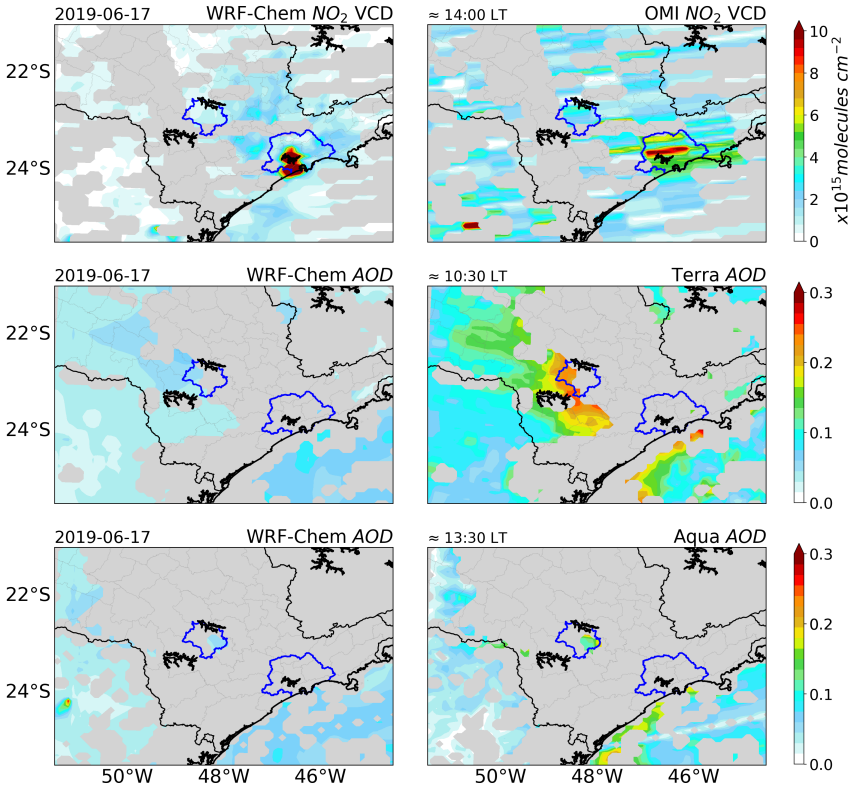

Figure B.5: Spatial distribution of model-simulated and OMI-retrieved tropospheric NO_2 VCD, Terra- and Aqua-retrieved AOD at 550 nm integrated into the atmosphere for June 17, 2019.

B.2 Episode - 11 August 2019

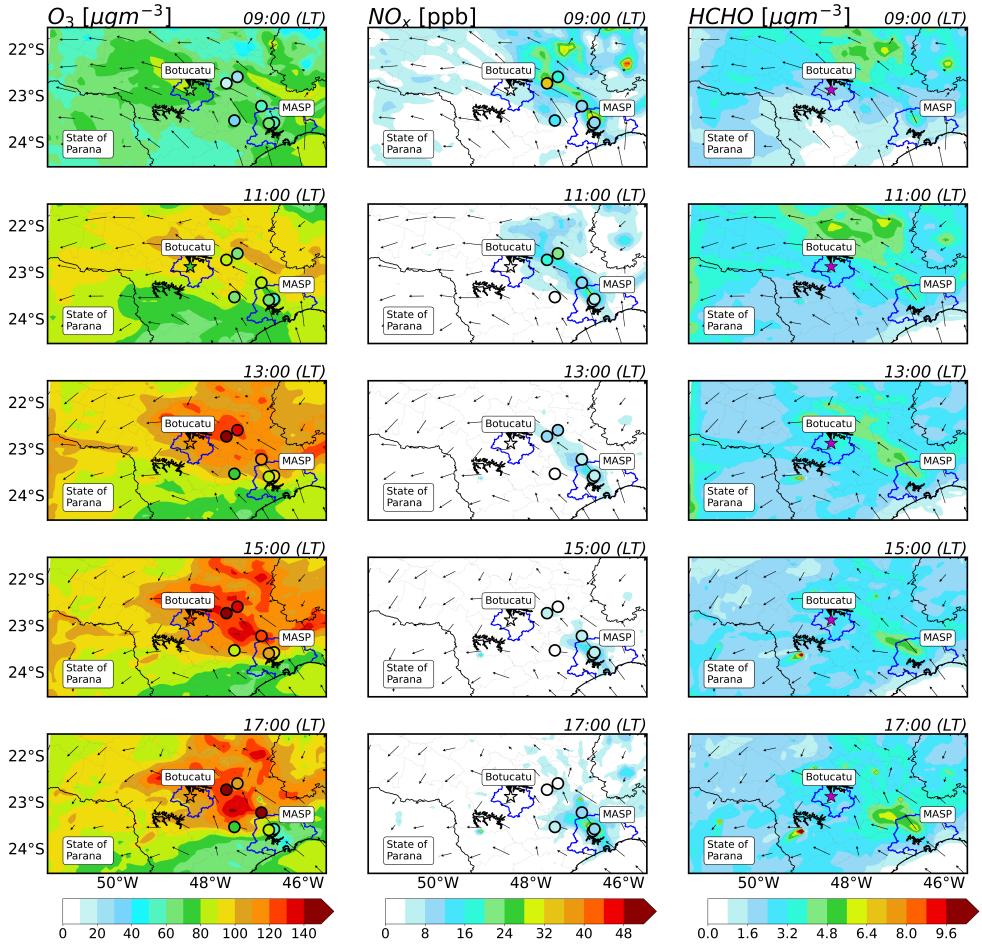

Figure B.6: Overlay of observed and predicted O_3 (in μg m^{-3}) (left panels), NO_x (in ppb) (middle panels) and HCHO (in μg m^{-3}) (right panels) spatial distributions during daytime for August 11, 2019. The observed values are indicated by the colored star and circles. The blue lines represent Botucatu city and the MASP.

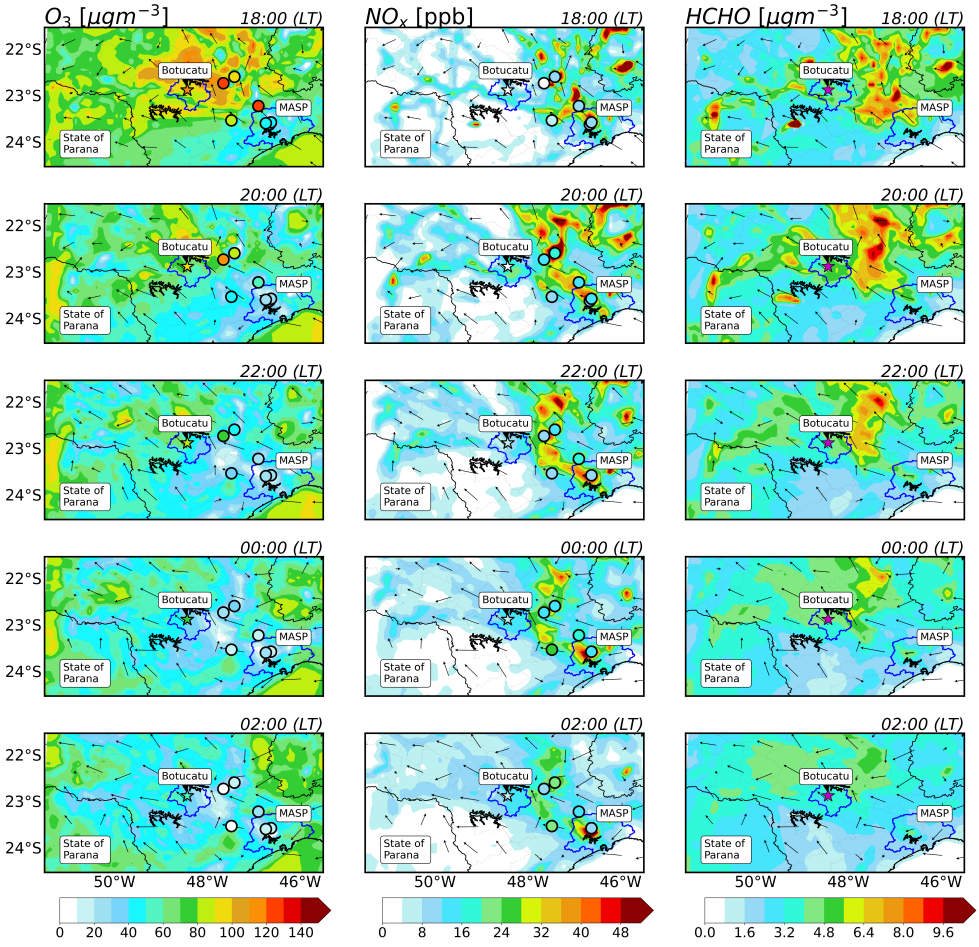

Figure B.7: Overlay of observed and predicted O_3 (in μg m^{-3}) (left panels), NO_x (in ppb) (middle panels) and HCHO (in μg m^{-3}) (right panels) spatial distributions during nighttime for August 11, 2019. The observed values are indicated by the colored star and circles. The blue lines represent Botucatu city and the MASP.

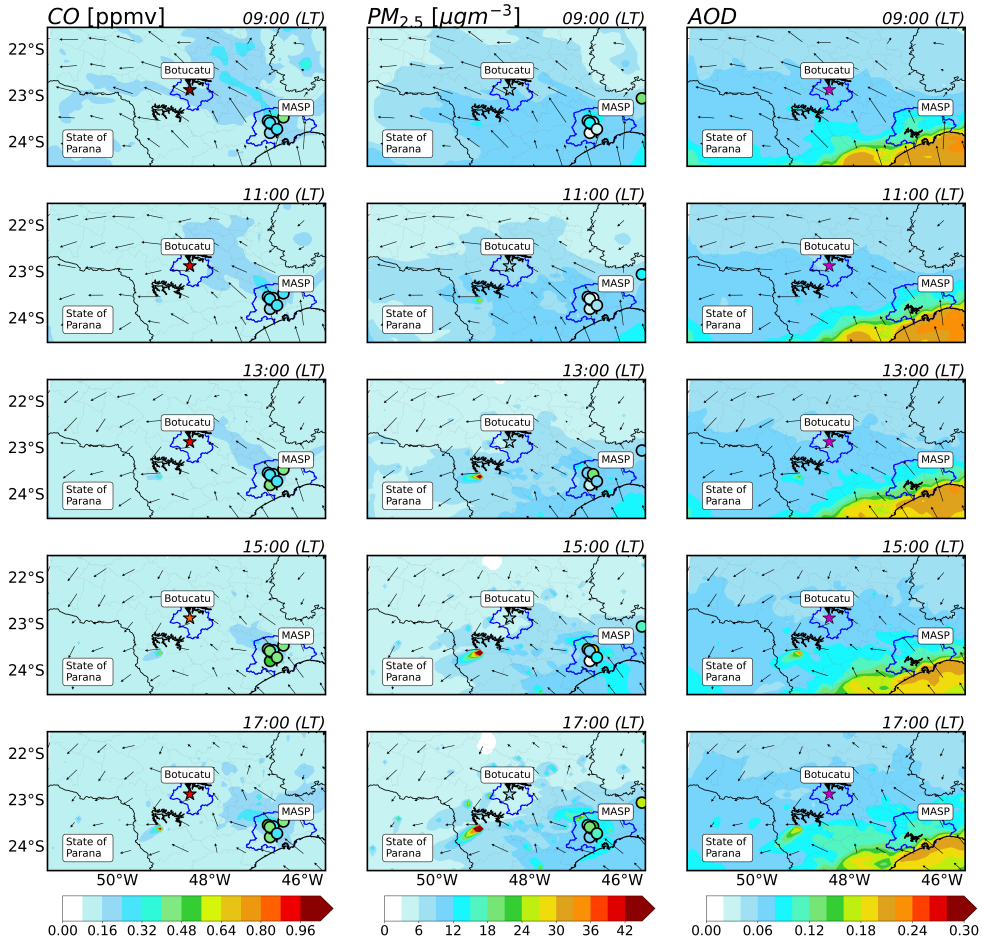

Figure B.8: Overlay of observed and predicted CO (in ppmv) (left panels), $PM_{2.5}$ (in μg m^{-3}) (middle panels) and AOD (right panels) spatial distributions during daytime for August 11, 2019. The observed values are indicated by the colored star and circles. The blue lines represent Botucatu city and the MASP.

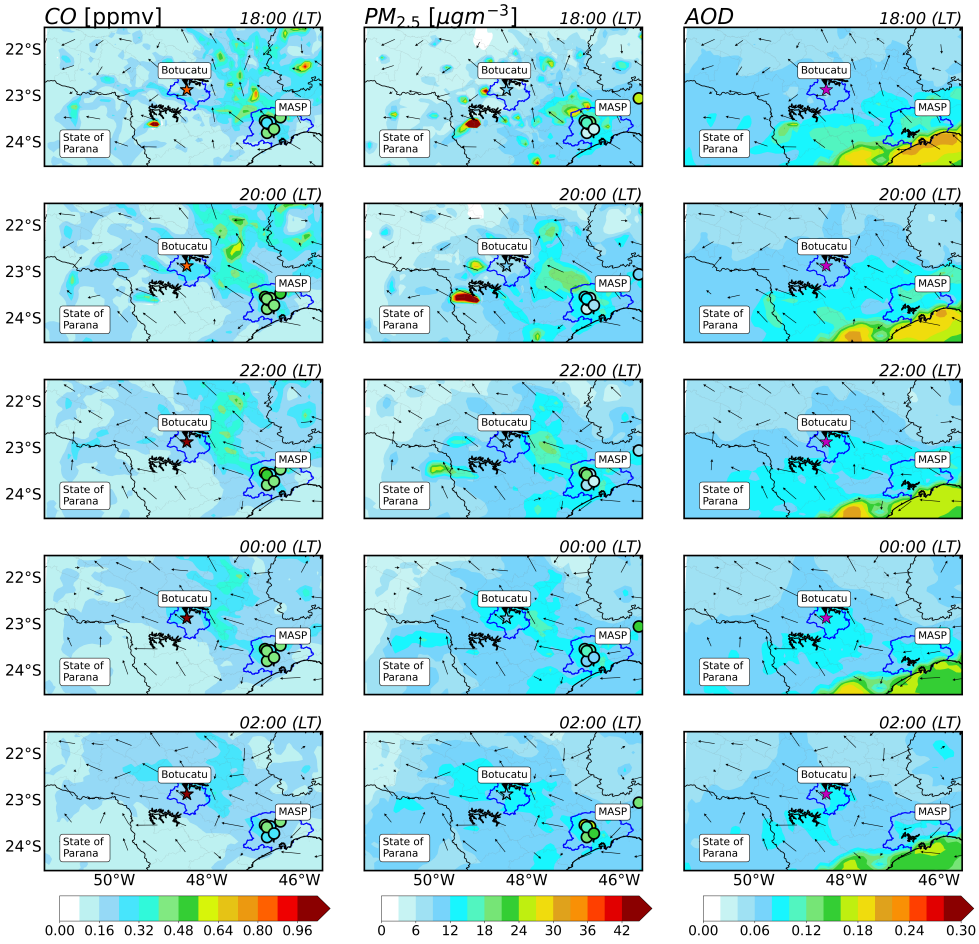

Figure B.9: Overlay of observed and predicted CO (in ppmv) (left panels), $PM_{2.5}$ (in $\mu g\ m^{-3}$) (middle panels) and AOD (right panels) spatial distributions during nighttime for August 11, 2019. The observed values are indicated by the colored star and circles. The blue lines represent Botucatu city and the MASP.

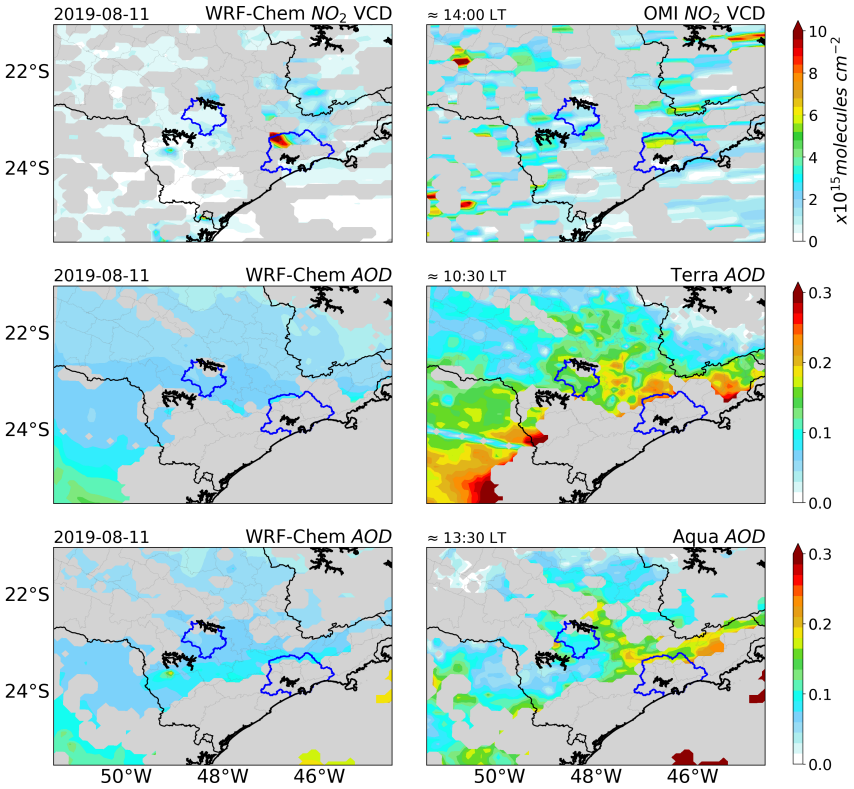

Figure B.10: Spatial distribution of model-simulated and OMI-retrieved tropospheric NO_2 VCD, Terra- and Aqua-retrieved AOD at 550 nm integrated into the atmosphere for August 11, 2019.