Intercomparison of air quality models in a megacity: Towards an operational ensemble forecasting system for São Paulo

Adrien Deroubaix[1], Judith J. Hoelzemann[3], Rita Yuri Ynoue[4], Taciana Toledo de Almeida Albuquerque[5], Rafaela Cruz Alves[4], Maria de Fatima Andrade[4], Willian Lemker Andreão[5], Idir Bouarar[1], Ediclê de Souza Fernandes Duarte[3,6], Hendrik Elbern[7], Philipp Franke[7,8], Anne Caroline Lange[7,8], Pablo Lichtig[1], Lya Lugon[1], Leila D. Martins[9], Gregori de Arruda Moreira[4], Rizzieri Pedruzzi[5], Nilton Rosario[4], Guy Brasseur[1]

[1]Max Planck Institute for Meteorology, Hamburg, Germany
[2]Institute of Environmental Physics, University of Bremen, Bremen, Germany
[3]Graduate Program for Climate Sciences, University of Rio Grande do Norte, Brazil, Natal, Brazil
[4]Instituto de Astronomia, Geofísica e Ciências Atmosféricas, University of São Paulo, Brazil
[5]Department of Sanitary and Environmental Engineering, Federal University of Minas Gerais, Belo Horizonte, Brazil
[6]Institute of Earth Sciences, University of Évora, Portugal
[7]Rhenish Institute for Environmental Research at the University of Cologne, Cologne, Germany
[8]Institute of Energy and Climate Research - Troposphere (IEK-8), Forschungszentrum Jülich GmbH, Jülich, Germany
[9]Federal University of Technology, Londrina, Brazil

Key Points:

- An ensemble of regional air quality models performs well in Sao Paulo for the main regulated pollutants (ozone, CO, SO2, NOx, PM2.5 and PM10)
- Transport of pollutants due to biomass burning events can strongly affect the air quality of the Sao Paulo megacity
- The median of the regional model ensemble gives a better result for ozone than each model in the center of the megacity

Corresponding author: Adrien Deroubaix, `Adrien.Deroubaix@mpimet.mpg.de`

Abstract

An intercomparison of four air quality models is performed in the tropical megacity of São Paulo with the perspective of developing an air quality forecasting system based on a regional model ensemble. During three contrasting periods marked by different types of pollution events, we analyze the concentrations of the main regulated pollutants (Ozone, CO, SO_2, NOx, $PM_{2.5}$ and PM_{10}) compared to observations of a dense air quality monitoring network.

The modeled concentrations of CO, PM and NOx are in good agreement with the observations for the temporal variability and the range of variation. However, the transport of pollutants due to biomass burning pollution events can strongly affect the air quality in the metropolitan area of São Paulo with increases of CO, $PM_{2.5}$ and PM_{10}, and is associated with an important inter-model variability.

Our results show that each model has periods and pollutants for which it has the best agreement. The observed day-to-day variability of ozone concentration is well reproduced by the models, as well as the average diurnal cycle in terms of timing. Overall the performance for ozone of the median of the regional model ensemble is the best in terms of time and magnitude because it takes advantage of the capabilities of each model. Therefore, an ensemble prediction of regional models is promising for an operational air quality forecasting system for the megacity of São Paulo.

Plain Language Summary

Forecasting air quality in megacities is especially difficult because of the diversity and temporal variability of emission sources. São Paulo is the largest metropolitan area in South America, and does not have an operational air quality forecast.

We perform an intercomparison of four air quality models with the perspective of developing an air quality forecasting system. During three contrasting periods marked by different types of pollution events, we analyze the concentrations of the main regulated pollutants (Ozone, CO, SO2, NOx, PM2.5 and PM10) compared to observations from the São Paulo air quality monitoring network.

Modeled concentrations of the main regulated pollutants agree well with observations for temporal variability and range of variation (except for SO2). However, the long-range transport of pollutants due to fires can strongly affect the air quality in São Paulo, and also reduce the performance of the models.

For ozone concentration, the observed daily variability is well reproduced by the models, and the performance of the median of the models is the best in terms of time and magnitude because it takes advantage of the capabilities of each model. Therefore, an operational air quality forecasting system is promising for the megacity of São Paulo.

1 Introduction

Forecasting air quality in megacities is difficult due to the diversity and temporal variability of emission sources, as well as the specific meteorology and photochemistry of the urban boundary layer (Baklanov et al., 2016). Even though global air quality forecasts are now available, the spatial resolution of these forecasts is coarse compared to the size of a megacity (Baklanov & Zhang, 2020). For this reason, high-resolution modeling using an online approach coupling weather and air quality is needed to reproduce the diurnal evolution of air composition in megacities. (G. Grell & Baklanov, 2011).

São Paulo is by far the largest metropolitan area in South America, one of the biggest megacities of the world, located near the coast and on a plateau at about 700 m above

sea level, in a subtropical climate, characterized by a dry and a wet season. São Paulo is special in different respects, for its geography and its climate but also for vehicle emissions as there is a significant use of biofuels (Brito et al., 2018). The level of secondary particles is particularly high due to the fuel composition (Albuquerque et al., 2019). Moreover, the air quality of the metropolitan area is frequently affected by the transport of biomass burning pollutants from remote areas (Martins et al., 2018; Moreira et al., 2021; Squizzato et al., 2021). Despite emission mitigation measures in place since the 1970s, air quality is still poor in São Paulo for ozone and fine particulate levels (Andrade et al., 2017; Schuch et al., 2019).

A megacity such as São Paulo is therefore a challenge for regional air quality models: They must be applied at a resolution, which is high enough to represent the processes leading to the high concentrations and high diurnal variability of the main pollutants, and include specific vehicle emission factors (Andrade et al., 2015). In addition, comprehensive measurements are needed to evaluate the model outputs. In the case of São Paulo, an extensive measurement network in and around the megalopolis was established in the 1970s and since then has been continuously exploited and extended, constituting an excellent support for evaluating the performance of models (Andrade et al., 2017).

Ensembles of regional air quality models have been first developed for Europe (Galmarini et al., 2004) and North America (Monache et al., 2006). In these two regions, the Air Quality Model Evaluation International Initiative (AQMEII) has shown that the discrepancies between models for the main regulated pollutants (Ozone, CO, SO_2, NOx, $PM_{2.5}$ and PM_{10}) are due to the representation of the dynamics in the planetary boundary layer (PBL), but also due to inaccurate emissions and boundary conditions (Im et al., 2015; Solazzo et al., 2017). For forecasting the air quality in megacities, the use of an ensemble of regional air quality models has two main interests: firstly, the inter-model range is an indicator of the uncertainty of the state-of-the-art modeling (Vautard et al., 2009), and secondly its median generally yields better performances than each single model (Riccio et al., 2007).

Operational air quality forecasts based on model ensembles are available in Europe (Marécal et al., 2015) and East Asia (Brasseur et al., 2019; Petersen et al., 2019). The Klimapolis project, whose goal is to establish a "Joint Laboratory on Urban Climate, Water and Air Pollution: Modeling, Planning, Monitoring, Social Learning", aims to develop such an ensemble forecasting system for South America based on these two previous experiences. As a preliminary step to develop this system, this article evaluates the performance of state-of-the-art regional air quality models focusing on the metropolitan area of São Paulo.

Four chemistry-transport models are involved in this intercomparison of high-resolution (*i.e.* less than 5 km) modeling results which are described in section 2. The evaluation is supported by the São Paulo measurement network, for which we propose a methodology to compare the model outputs with a representative value for the whole megacity, discussed in section 3. We assess the strengths and weaknesses of the models for the main regulated pollutants over three contrasting time periods in section 4. In sequence, we then focus on the diurnal variability of photochemistry-related variables in section 5. Finally, we analyze the performance of the ensemble forecast regarding the prediction of ozone and $PM_{2.5}$ alerts in section 6. Conclusions and perspectives are given in section 7.

2 The air quality models

In this section, we briefly present the different chemistry-transport-models (Sect. 2.1) and we describe the main setup differences that may be important to interpret the results presented in the next sections (Sect. 2.2).

2.1 Strategy towards an operational ensemble forecasts

In this intercomparison study, a regional air quality model ensemble is compared to the global forecasts generated by the European Centre for Medium-Range Weather Forecasts through the Copernicus Atmosphere Monitoring Service (hereafter ECMWF–CAMS) and by the US National Center for Atmospheric Research using Community Atmosphere Model with Chemistry (hereafter NCAR–CAMchem).

All regional models provide hourly simulation outputs in a configuration fast enough that it can be used for forecasting, and also with high spatial resolution (less than 5 km). Four institutes are involved in this intercomparison, three of them are located in Brazil and one in Germany, using their optimal setups for their model:

1. The Max Planck Institute for Meteorology (MPI) provides simulations made with the WRFchem model.
   The Weather Research and Forecasting model (WRF) coupled with chemistry (WRFchem) is a mesoscale non-hydrostatic meteorological model online coupled with chemistry that simultaneously predicts meteorology and atmospheric composition (G. A. Grell et al., 2005; Fast et al., 2006; Powers et al., 2017). The model is based on WRF version 4.1.2, with the Model for Ozone and Related chemical Tracers, MOZART version 4, as chemical scheme (Emmons et al., 2010). The anthropogenic emissions are taken from the CAMS-GLOB-ANT version 4.2 inventory (Granier et al., 2019). The monthly emissions are distributed for each hour according to vertical profiles based on (Bieser et al., 2011; Mailler et al., 2013), and to daily and weekly profiles (Crippa et al., 2020). The biogenic emissions are calculated using the Model of Emissions of Gases and Aerosols from Nature, MEGAN version 2.1 (Guenther et al., 2006) and fire emissions using the Fire INventory from NCAR, FINN version 1.5 (Wiedinmyer et al., 2011). Dust and sea salt are parametrized online, depending on the wind intensity, using the Global Ozone Chemistry Aerosol Radiation and Transport (GOCART) model (Ginoux et al., 2001).
   For the meteorological configuration, the planetary boundary layer physics are calculated by the YSU (Yonsei University) scheme (Hong et al., 2006), the surface layer scheme is the Carlson-Boland viscous sub-layer with the surface physics calculated by the 'Noah' land surface model (Ek et al., 2003). The RRTMG radiation scheme (Mlawer et al., 1997), the Thompson and Eidhammer (2014) aerosol aware cloud microphysics scheme and the Grell-Devenyi 3D cumulus scheme (G. A. Grell & Dévényi, 2002) are selected.
   Two WRFchem simulations are carried out at the MPI using two meteorological initial and boundary conditions, one with the FNL (Final) operational global analysis produced by the Global Data Assimilation System of the US National Centers for Environmental Prediction (NCEP-FNL; ds083.3 dataset, DOI: `https://10.5065/D65Q4T4Z`), and the other one with the ECMWF-ERA5 reanalysis (Hersbach et al., 2020).
2. The Universidade Federal de Minas Gerais (UFMG) provides simulations made with the WRF-CMAQ model.
   The Community Multiscale Air Quality Modeling System (CMAQ) is a three-dimensional Eulerian atmospheric chemistry and transport, which is used by the United States Environmental Protection Agency (Byun & Schere, 2006). The anthropogenic emissions are taken from the Emissions Database for Global Atmospheric Research to study Hemispheric Transport of Air Pollution, EDGAR-HTAP inventory version 2.2 (Janssens-Maenhout et al., 2015). The WRF model and Sparse Matrix Operator Kerner Emissions (SMOKE) model were selected to generate meteorology and emissions (Albuquerque et al., 2019). Pedruzzi et al. (2019) applied the CMAQ model at a local scale over the urban and industrialized area of Vitória-ES (Brazil), and the setup used for this intercomparison is similar.

3. The Universidade Federal do Rio Grande do Norte (UFRN) together with the Rhenish Institute for Environmental Research at the University of Cologne provide simulations made with EURAD-IM model.
   The EURopean Air pollution and Dispersion - Inverse Model (EURAD-IM) is chemistry-transport model (Hass et al., 1995; Memmesheimer et al., 2004; Elbern et al., 2007), which uses WRF as offline meteorological model. The anthropogenic emissions are taken from the Emissions Database for Global Atmospheric Research, EDGAR inventory version 4.3.2 (Crippa et al., 2018). The vertical distribution of emissions and the emission strength per hour is calculated within the EURAD-IM model based on prescribed source category dependent vertical profiles and daily, weekly, and yearly time profiles. Fire emissions are from the Global Fire Assimilation System, GFAS Version 1.2 (Kaiser et al., 2012).
4. The Universidade de São Paulo, Instituto de Astronomia, Geofísica e Ciências Atmosféricas (USP-IAG) provides simulations made with the WRFchem model.
   The WRFchem model is used on version 4.0, with the Carbon-Bond Mechanism version Z (CBMZ) gas-phase chemistry mechanism (Zaveri & Peters, 1999) and the Model for Simulating Aerosol Interactions and Chemistry (MOSAIC) aerosol module (Zaveri et al., 2008). Vehicular emissions were estimated with LAPAt model (Andrade et al., 2015). The Morrison 2-moment microphysics scheme (Morrison et al., 2009) is selected.

The model configurations used by each institution are different due to their choices of emissions inventories, meteorological and chemical configuration, and spatial resolution. We consider the variability of different forecasts to be representative of the uncertainties in air quality forecasts using state-of-the-art chemistry and meteorology models.

In order to analyze the influence of the meteorological inputs, two WRFchem simulations are performed at the MPI with NCEP-GFS and with ECMWF-ERA5). We analyze the results of the individual models as well as the median of the regional model ensemble which we call Multi-Model Median, hereinafter MMM, which is calculated without the ECMWF-ERA5 simulation made at the MPI in order to have the same weight for the model simulations carried out by each of the four institutions. The median is chosen rather than the mean to reduce the influence of outliers.

2.2 Similarities and differences of the modeling setup

The main differences of model configuration chosen by the four institutes consists in the model domain, the emission datasets, the chemistry and aerosol schemes, and the meteorological parametrizations (Tab. 1).

The domains chosen by the four institutions are similar in terms of horizontal and vertical resolution. Moreover, meteorological inputs and physical parametrizations are similar for all models. Three of the institutions use similar anthropogenic emission dataset of the EDGAR database.

However, one would expect anthropogenic emissions to be a large source of model variability due to the difference in the geographical distribution of emissions by sector (Huneeus et al., 2020), and how participating groups simulate temporal or vertical profiles for the sector-specific emission input data. Moreover, long-range transport of biomass burning aerosols is important for the São Paulo region (Martins et al., 2018; Squizzato et al., 2021). Therefore, biomass burning emission integration in the domain or by boundary conditions may also be sensitive for air quality forecast inside the megacity.

3 A distance-weighted average for São Paulo

This section firstly presents the air quality measurement network of São Paulo (Sect. 3.1), secondly analyzes the inter-station variability of the pollutant concentrations in 2019 (Sect. 3.2), and thirdly describes the three 15-day periods that we selected for the model intercomparison (Sect. 3.3). The year 2019 is selected as sufficiently representative of typical conditions, because it was a weak 'El Niño' year and not affected by, but shortly before the COVID-19 pandemic.

We study the use of a distance-weighted average to represent the air quality in the São Paulo megacity, which can be questionable in particular for the most short-lived pollutants measured near sources, which is NO among our studied pollutants. Of course, it is not possible to define the true value that represents a megacity because the concentrations vary spatially. However, we focus on hourly concentrations and, from one hour to another, we can expect a stronger temporal co-variation of the concentrations (for all the stations) than of its spatial variability of all the stations (for a given hour). Nevertheless, it is essential to avoid stations located too close to the sources, as they are not representative for a large area.

3.1 Measurements of the CETESB air quality network

The São Paulo measurement network, maintained by CETESB (Companhia Ambiental do Estado de São Paulo, `https://cetesb.sp.gov.br/ar/qualar/`), is composed of 26 stations within the metropolitan area and another 63 within the state of São Paulo mostly in or near other cities (Fig. 1). This network is excellent as it is well distributed spatially and well maintained for several decades (Andrade et al., 2017). The number of stations is large, for comparison there are 58 stations in the Île-de-France region (which includes the Paris megacity).

Although we mainly focus on (1) the metropolitan area of São Paulo, two other surrounding localities are studied (2) Santos, and (3) Campinas (Fig. 1). We define a city center for these three locations by choosing their traditional center, such as (1) São Paulo center at *Catedral da Sé* (latitude: -23.5503^o, longitude: -46.6339^o), (2) Santos center at *Paróquia Sagrada Família* (latitude: -23.9427^o, longitude: -46.3783^o), and (3) Campinas center at *Catedral Metropolitana de Campinas* (latitude: -22.9060^o, longitude: -47.0605^o).

Stations located within a radius of 15 km to the São Paulo city center are selected (and within a radius of 10 km for the two other locations). For São Paulo, we have a classification of stations composed of 5 classes, which depend on their spatial scale of representativeness: 1 - Microscale, 2 - Neighborhood, 3 - Urban, 4 - Medium, 5 - Regional, (based on CETESB report and characteristics of each station place) (CETESB, 2022). In order to remove the stations not representative for the megacity, we compare the average of all the stations with the concentrations measured at each station using the correlation coefficients over the entire year 2019 (Tab. A1).

The only station associated with the regional scale (higher representativeness scale than the megacity) is weakly correlated with the average of all the stations (R ¡ 0.4 except for ozone). This station is removed to calculate an accurate average concentration of the megacity. Conversely, the stations associated with the microscale class could lead to a false representation of the whole megacity because they are close to specific emission sources. This applies to six stations, which can largely contribute to the average of the available stations, and which are removed from the analysis as well (Tab. A1).

It should also be noted that the level of agreement between the stations is high for all the variables considered, as evidenced by the correlation coefficients greater than 0.7, with the highest for ozone (greater than 0.9). This result shows that, given the current

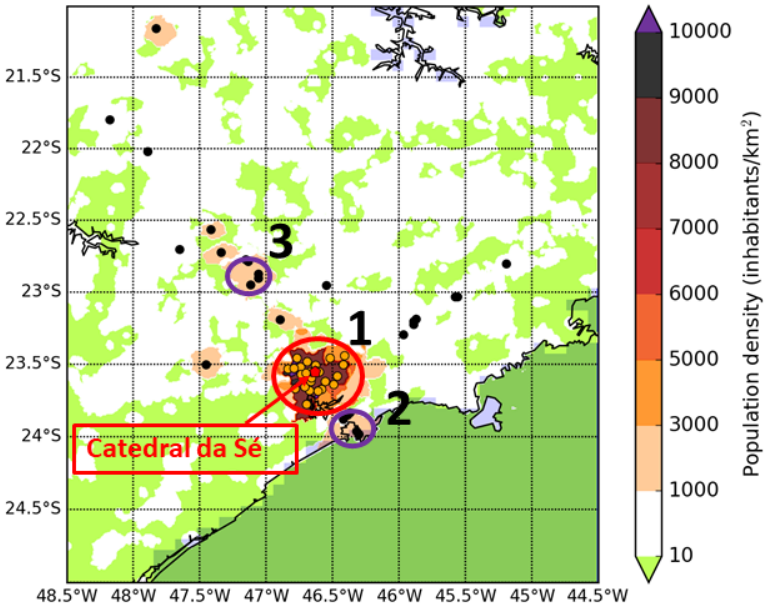

Figure 1. *Population density map showing the locations of São Paulo state measurement network stations (dots) with distinguished metropolitan area stations (orange dots). The numbers indicate the three cities studied: (1) São Paulo, (2) Santos, and (3) Campinas. The radius of the circles (in purple and red) represent the stations included to calculate the distance-weighted average of pollutant concentrations for the three cities. The city center of São Paulo is located at Catedral da Sé (red dot).*

measurement network, it is possible to consider the average of the stations to represent the hourly variation of the concentrations for the metropolitan area of São Paulo.

3.2 Spatial representativeness of the stations

Using stations from classes 2, 3 and 4, we compare two methods to calculate the average of each pollutant concentration for the megacity, (i) a simple method which consists in averaging the selected stations, and (ii) a distance-weighted average using the distance from station to the city center, where the weight is based on the inverse of the distance to a specific location (here the city center, CC). The concentration at the city center ($Conc_{CC}$) is calculated as follows:

$$Conc_{CC}(t) = \left(\sum_{s=1}^{s=N} w_s \times Conc_s(t)\right) / \sum_{s=1}^{s=N} w_s \quad (1)$$

where the weights are:

$$w_s = 1/d(s, CC)^p, \quad (2)$$

$Conc_s$ is the concentration measured at each station, and p is the power factor, which changes the importance of the stations located the closest to the CC.

The range of station weights calculated with p equal to 2 or 3 is five orders of magnitude (Tab. 2). Therefore, given the São Paulo network, p equal to 2 or 3 is not an appropriate choice giving to much weight to the stations close to the city center while the influence of more distant stations is highly reduced. With p equal to 1, the weight range is less than two orders of magnitude, which is already significant (Tab. 2). Indeed, the closest station to the city center (*Parque Dom Pedro II*) is 840 m away, much closer than all the other stations, which are at least more than 3 km away. This causes this station to contribute more than 30 % of the city center average calculated with a distance-weighted average using the classes 2, 3 and 4.

We compare the averages obtained with two methods for the NO concentration (the shortest lifetime of the pollutant studied) during the year 2019 with and without class 2. In addition, we plot the average of all the stations (as a reference to compare) in order to estimate the influence of the selection of the stations based on their spatial scale of representativeness. From the raw hourly data, we present the daily average and the averaged hourly diurnal cycle (Fig. A1).

NO concentrations are higher from May to September (during the colder and dryer months) than during the rest of the year, often above 20 ppb (Fig. A1). Moreover, the highest concentrations occur at night, with two peaks at 01:00 and 08:00, suggesting the combined effect of traffic emissions and a strong diurnal evolution of the PBL height. Note that the peak at 01:00 is surprising because neither the emissions nor the height of the PBL are likely to change so drastically during a single hour (averaged over a year). In fact, this is due to the configuration of the automatic NOx analyzers, most of which are calibrated at 01:00 (personal communication with CETESB by Maria De Fatima Andrade).

By comparing the average of all the stations ('Stations mean' in Fig. A1) with the average of the selected stations ('Selected mean' in Fig. A1), we note a greater difference for classes 3 and 4 (panels a and c) than for classes 2, 3 and 4 (panels b and d). This shows that class 2 stations largely influence the average.

By comparing the distance-weighted average ('City center' in Fig. A1) and the average of the selected stations, we see that the diurnal cycles are different for classes 2, 3 and 4, while it is the same for classes 3 and 4. This result shows that the distance-weighted average for classes 2, 3 and 4 (with our CC defined at *Catedral da Sé*) is influenced by the *Parque Dom Pedro II* station. Therefore, class 2 stations are excluded from the distance-weighted average calculations used in the following.

From this analysis, we see also that the distance-weighted average and the average of the selected stations lead to similar NO concentrations using the stations class 3 and 4. To conclude, using the stations class 3 and 4, it is possible to define a consistent value of concentration representing the megacity that can be used to evaluate the different models.

3.3 Selection of three time periods

We select three 15-day periods that are:

1. 27 January to 12 February 2019, a period of ozone episodes, five days with ozone concentration above air quality standard in São Paulo were monitored despite the precipitation occurring during this period.
2. 8 to 21 August 2019, a period of aerosol episodes from long-range transport, during which biomass burning aerosols from the Amazon basin and central areas of Brazil transported to São Paulo, have created 'black rain'.
3. 6 to 20 September 2019, a period of ozone and $PM_{2.5}$ episodes, during which the air quality standards for ozone and $PM_{2.5}$ were exceeded for both pollutants.

These three periods are presented for ozone and $PM_{2.5}$ with the daily averages and the averaged hourly diurnal cycles (Fig. 2). We notice for ozone and $PM_{2.5}$ that the averages calculated with the two methods lead to closer results than for NO, which is expected due to their longer lifetime. The correlation coefficient of the two methods is equal to 0.84 for NO, whereas it is 0.92 for $PM_{2.5}$ and 0.99 for ozone. Consequently, the averages calculated with the two methods should lead to the same interpretation for $PM_{2.5}$ and for ozone (and to a lesser extent for NO).

In conclusion of this analysis of the measurement network of São Paulo, we have selected three periods and defined a method for calculating the concentrations of pollutants representative of the city. Distance-weighted average to the city center is convenient for the model intercomparison because it allows model outputs to be interpolated only to a single location (instead of all station locations). In the following, observed concentrations are calculated using distance-weighted average (applied to class 3 and 4 stations for São Paulo city center).

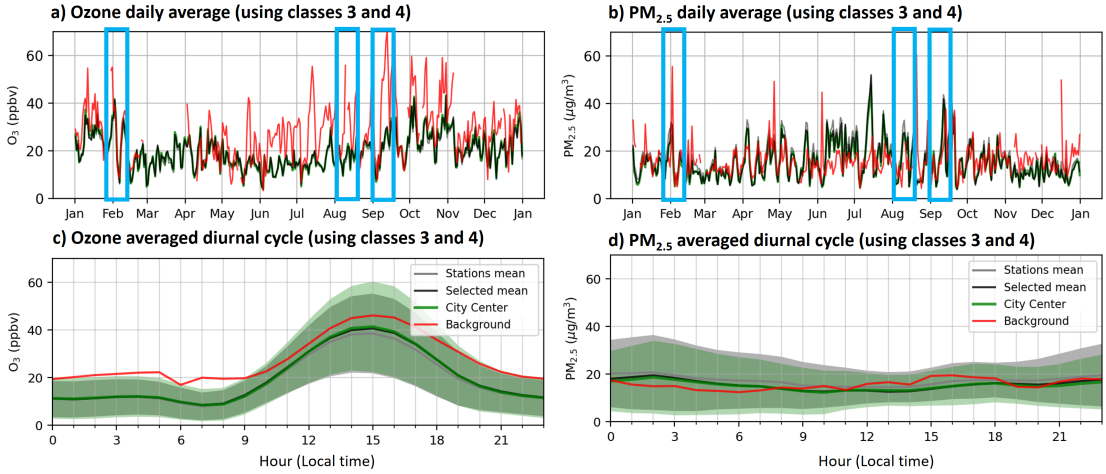

Figure 2. *Time series of the daily average (top) and the average hourly diurnal cycle (bottom) of ozone and $PM_{2.5}$ concentrations for the year 2019 from the CETESB measurement network. The three selected periods are marked by blue rectangles. The concentrations are calculated from the average of all the stations ('Stations mean' the gray line), from the average of the stations selected from a classification of their spatial scale of representativeness ('Selected mean' with classes 3 and 4, black line), from an average of the selected stations weighted by the distance between the station and the center of São Paulo ('City center', green line), and for the concentration at the background station ('background', red line). The color shadings (bottom) represent the standard deviation of hourly concentrations over the year.*

4 Performance of the regional model ensemble

We start the intercomparison by studying the general performances of the air quality models at the center of São Paulo (Sect. 4.1), and we focus on the temporal variation of selected variables relevant for meteorology (Sect. 4.2), the long-range transport (Sect. 4.3), and anthropogenic emissions (Sect. 4.4). We aim to understand the strengths and weaknesses of each of the four regional models studied in comparison with the three others, and also with the global forecasts.

4.1 General performance

The general performance of the models is assessed for the main regulated pollutants (Ozone, CO, SO_2, NOx, $PM_{2.5}$ and PM_{10}) using the correlation coefficients of the hourly observations and the different model outputs over the first, second and third studied periods (Tab. 3, 4 and 5, respectively) as well as the root mean square error (RMSE) (Tab. A2) and the mean bias (Tab. A3). In addition, we define the 'oxidant' concentration as: $Ox = NO_2 + O_3$.

Overall, all models perform well with a majority of correlation coefficients greater than 0.5 (although a low correlation coefficient may be due to some outliers, a value greater than 0.5 means that the model reproduced part of the observed variability), and both the RMSE and the mean biases are small for most variables (because they are of the same order of magnitude as the observation mean). It is also interesting to note that all models have episodically periods and pollutants with very good evaluation scores. For NO_2, ozone and Ox, we notice that the MMM has in some cases a higher correlation than all the members that compose it. Comparing the regional models with the global forecasts, we note that the scores are of the same order. However the MMM has the best scores over the three periods for these three pollutants.

Looking at the individual variables, the correlation coefficients of CO are intermediate (R close to 0.5) with a low RMSE and biases (compared to the observation mean). Aerosols are not well reproduced, especially during the second period. There is an improvement in the correlation coefficients with the ECMWF-ERA5 reanalysis compared to the NCEP-FNL forecast, which could be due to more accurate wind fields, improving the representation of the pollutant transport.

For $PM_{2.5}$ and PM_{10}, the correlation coefficients are less than 0.5, the biases are low and the RMSE are high, which may reflect the high temporal variability of the aerosol load (Tab. A2 and A3). This indicates that the modeled variability range is in good agreement while the modeled temporal variability is not well reproduced, which may be caused by the advent time of aerosols due to long-range transport. Moreover, the production of secondary aerosols is generally underestimated in São Paulo, and this could lead to a time-offset (Andrade et al., 2017). However, we notice that the correlation coefficients for $PM_{2.5}$ are slightly higher than for PM_{10}.

For SO_2, the correlation coefficients are low and the bias is several times higher than the average concentration observed over each period, which may be due to the magnitude of anthropogenic emissions. For the nitrogenous species (NO, NO_2 and NOx), the correlation coefficients are low and the RMSE is high (compared to the observation mean) but the biases are low, which may be due to inaccurate hourly profiles applied to the anthropogenic emissions.

Ozone is in good agreement with observations even though the first and third periods were chosen because they include high ozone events. For all three periods, the MMM ozone concentration has the best evaluation scores, and the UFMG–WRF-CMAQ scores are the best of the regional model ensemble. It should be noted that the scores of the global forecasts are similar to those of the regional models, but the correlations are calculated with a smaller number of hours for the global forecasts due to their lower output frequency (3 hours for ECMWF–CAMS and 6 hours for NCAR–CAMchem). For Ox, the correlation coefficients are close to that of ozone with increased biases. All models overestimate Ox over the three selected periods, which may be due to their lower diurnal variability. We also note that the ozone biases are mostly of the opposite sign to NO (Tab. A3).

The remarks made in this section will be analyzed in the following by looking at the temporal variability of the different variables.

4.2 Meteorological variability

To investigate the differences of the regional models, we start by analyzing the temporal variability of relative humidity, PBL height, wind speed and direction during the three periods (Fig. A2 and A3) in order to identify the different meteorological conditions occurring during this study. The PBL height data is obtained by a LIDAR measuring the aerosol backscattered signal, which is located at the university of São Paulo (Moreira et al., 2019). It provides accurate data from 11:00 to 16:00 using quality criteria (Courtesy of G. de Arruda Moreira), allowing the analysis of the range of the PBL height. To compare the 10-m wind speed diagnosed by the models with the observations made at 2 meters, we multiply the observations by a factor of 4/3 (assuming that a logarithmic profile represents well the wind).

There are specific days shared by the four meteorological variables (RH, PBL height, wind speed and direction) for each period, for which the values for this day differ from other days: (i) 5 February, (ii) 12, 15 and 20 August, (iii) 14 September. These particular days are associated with high relative humidity (¿ 80 %) and high wind speed (¿ 3 m/s) continuously coming from the south for two days, and with a low height of PBL (¡ 1 km), which corresponds to stormy weather conditions (Fig. A3). Excluding these specific days, we notice a clear diurnal cycle of relative humidity, wind speed and PBL height with a minimum at night and a maximum during the day. For the direction of the wind, we notice there is often a change from north west to south east.

During these three periods, we see that the temporal variability found by the models corresponds well to the observations. The models overestimate wind speed, especially during the daytime. During the days with the stormy weather conditions, a greater inter-model variability can be observed.

In conclusion, it seems that the models agree well with the meteorological observations. Therefore the differences in the modeled meteorology may not be responsible for persistent differences in the simulated concentrations by the models. These differences are rather to be found on the side of emissions or long-range transport.

4.3 Long-range transport of pollution

In order to focus on long-range transport, we analyze CO and $PM_{2.5}$ concentrations, which are two pollutants notably emitted by combustion processes and transported due to their long lifetime (greater than a week) in São Paulo (Fig. 3) and in Campinas (Fig. A5). In addition, we analyze PM_{10} and the ratio of $PM_{2.5}$ against PM_{10} (Fig. A4).

The amplitude of variation for CO ranges from 0.1 to 2.4 ppm and for $PM_{2.5}$ from 10 to 80 $\mu g.m^{-3}$. There are large increases synchronized for both pollutants (reaching at least 1.5 ppm for CO and 50 $\mu g.m^{-3}$ for $PM_{2.5}$) for the three time periods. These increases are associated with different ratios of CO to $PM_{2.5}$, and different persistence over time from some hours to one day. Considering that São Paulo is frequently affected by biomass burning events throughout the year, either due to agricultural practices in the surrounding rural areas, or by deforestation and pasture-maintainance fires from remote regions (Godoy-Silva et al., 2017), this suggests biomass burning events. We note these events on (i) 30, 31 January and 1 February, on 10, 11, 13 and 17 August, and (iii) on 11, 12, 17 and 18 September (which are different from the meteorological events; *cf.* Sect. 4.2).

By excluding these biomass burning events, the models reproduce well the amplitude of variation for CO. $PM_{2.5}$ is overestimated by the simulations of UFMG–WRF-CMAQ and MPI–WRFchem, whereas it is in good agreement for IAG-USP–WRFchem and UFRN–EURAD-IM. Biomass burning pollution events are identified by MMM because, for each event, there is at least one simulation in good agreement with the obser-

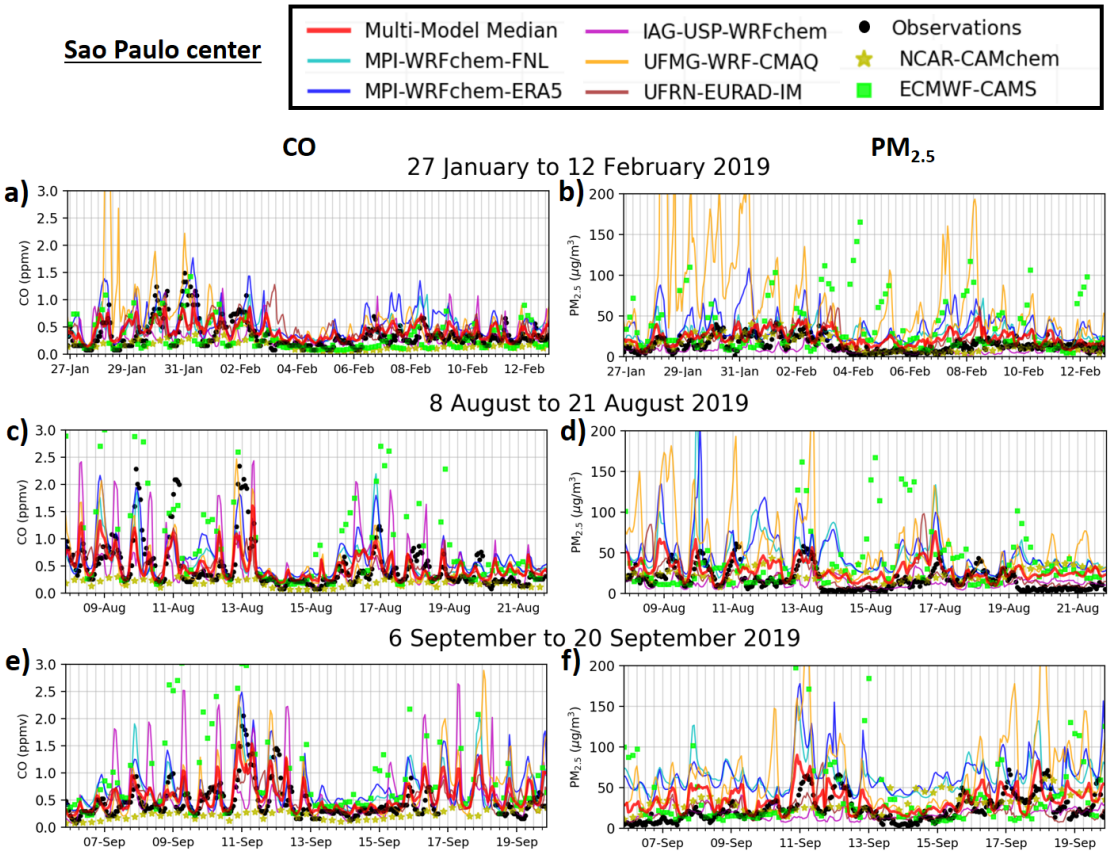

Figure 3. *Time series of hourly concentrations of CO (a, c and e) and $PM_{2.5}$ (b, d and f) observed and modeled in São Paulo for the three selected 15-day periods of the year 2019. The models include data from two global forecasts (yellow stars and green squares) and a regional model ensemble of five simulations (colored lines) with the Multi-Model Median (red line).*

vations. However, the overall CO concentration during biomass burning event is generally underestimated by the MMM. The two meteorological datasets used with WRFchem (MPI-WRFchem-ERA5 and MPI-WRFchem-FNL) lead to close results for CO, $PM_{2.5}$ and PM_{10}, although there is an improvement with ERA5 during some biomass burning events, which may explain the slightly greater correlation coefficients (*cf.* Sect. 4.1).

For global models, NCAR-CAMchem underestimates CO, while the variation range of $PM_{2.5}$ is in agreement with observations. Increases in CO and $PM_{2.5}$ associated with biomass burning events are not reproduced by NCAR–CAMchem. ECMWF–CAMS reproduces well the average concentration of CO and $PM_{2.5}$, however there are very high concentrations, in particular during biomass burning events, for which the bias is the highest, and which may be related to the GFAS biomass burning emissions.

The observed temporal variability of PM_{10} is similar to that of $PM_{2.5}$, which is also the case for the four regional simulations (Fig. A4). As for $PM_{2.5}$, PM_{10} is overestimated by all models except UFRN–EURAD-IM. The observed ratio of $PM_{2.5}$ against PM_{10} is ranging mostly between 0.4 and 0.8. There are a few values above 0.8, *i.e.* dominated by fine particles, and below 0.4, *i.e.* dominated by coarse particles. Biomass burning pollution events are not clearly associated with a low value of this ratio, but during the periods of strong wind coming from the South (*cf.* Sect. 4.2), the value of the ratio is low

which indicates a transport of large particles (to the south is a large harbor area in Santos). In general, the regional models have very different temporal behaviors with UFMG–WRF-CMAQ nearly constant at 0.8, and UFRN–EURAD-IM with a clear diurnal cycle. The regional models reproduce the variation range of $PM_{2.5}$ against PM_{10} ratio.

In Campinas (Fig. A5), the level of CO and $PM_{2.5}$ is slightly lower than in São Paulo, and the same events are also observed for the two pollutants, which reinforces the interpretation of these events as being related to the long-range transport of pollution caused by biomass burning. The models underestimate the CO concentrations by about 0.2 ppm, while the modeled $PM_{2.5}$ level is well reproduced. However, for both pollutants, most of the biomass burning events are not reproduced neither by the regional models nor by the global forecasts at Campinas.

This section shows the importance of pollutant transport for air quality in São Paulo, especially from biomass burning sources. Each model reproduces certain events well in terms of magnitude and persistence. Therefore, the median of the regional model ensemble (*i.e.* MMM) produces overall the best estimate for CO and PM.

4.4 Anthropogenic pollution

Two characteristic pollutants of anthropogenic activities and their emissions are NOx and SO_2. In a megacity, NOx is mainly emitted by traffic, while SO_2 is mainly related to industries and electricity production from coal. We analyze here their temporal variability during the three periods in São Paulo (Fig. 4) and in Santos (Fig. A6).

The NOx observations show significant variability over the three periods. The diurnal variability shows an amplitude of about 30 ppb with daily minimums below 10 ppb. Biomass burning pollution events (*cf.* Sect. 4.3) are associated with high NOx values, reaching at least 150 ppb, and with a maximum reaching 300 ppb on 13 August.

For NOx, the models are in good agreement over the range of variation over the three periods. Pollution events related to biomass burning lead to an increase in the modeled NOx concentration for all models except NCAR–CAMchem. The magnitude of NOx concentration during biomass burning events is reproduced with large inter-model variability. Therefore, the MMM has the best agreement with the observations.

For SO_2, the picture is different from that of the other compounds presented previously. The observations range from 0 to 5 ppb in São Paulo, while there is almost a factor of 10 overestimation by the regional models and ECMWF–CAMS. Interestingly, the NCAR–CAMchem forecast run with coarse resolution has the best agreement. Additionally, comparing the meteorology used with MPI–WRFchem, the modeled SO_2 concentrations are very similar.

SO_2 is also produced by fire emissions. Note that during biomass burning pollution events, the observed concentration of SO_2 increases (up to 5 ppb). However, there is a constant bias over time for regional models using high resolution in the center of São Paulo as well as for ECMWF–CAMS. Only NCAR–CAMchem is in good agreement, which may be related to its much coarser resolution of about 100 km. So this points towards the anthropogenic inventory and the proxy used to downscale the emissions as main cause for the overestimation.

We further investigate concentrations in the industrialized area of Santos, where emissions from ships and industry are high compared to emissions from the traffic and residential sectors. The modeled SO_2 concentrations are in good agreement with the observations in Santos, while the modeled NOx concentrations are underestimated by the regional model ensemble. This points towards the industry sector which seems to be to important in the metropolitan area of São Paulo. We also note very high concentrations of NOx and SO_2 modeled by ECMWF–CAMS during biomass burning events in both

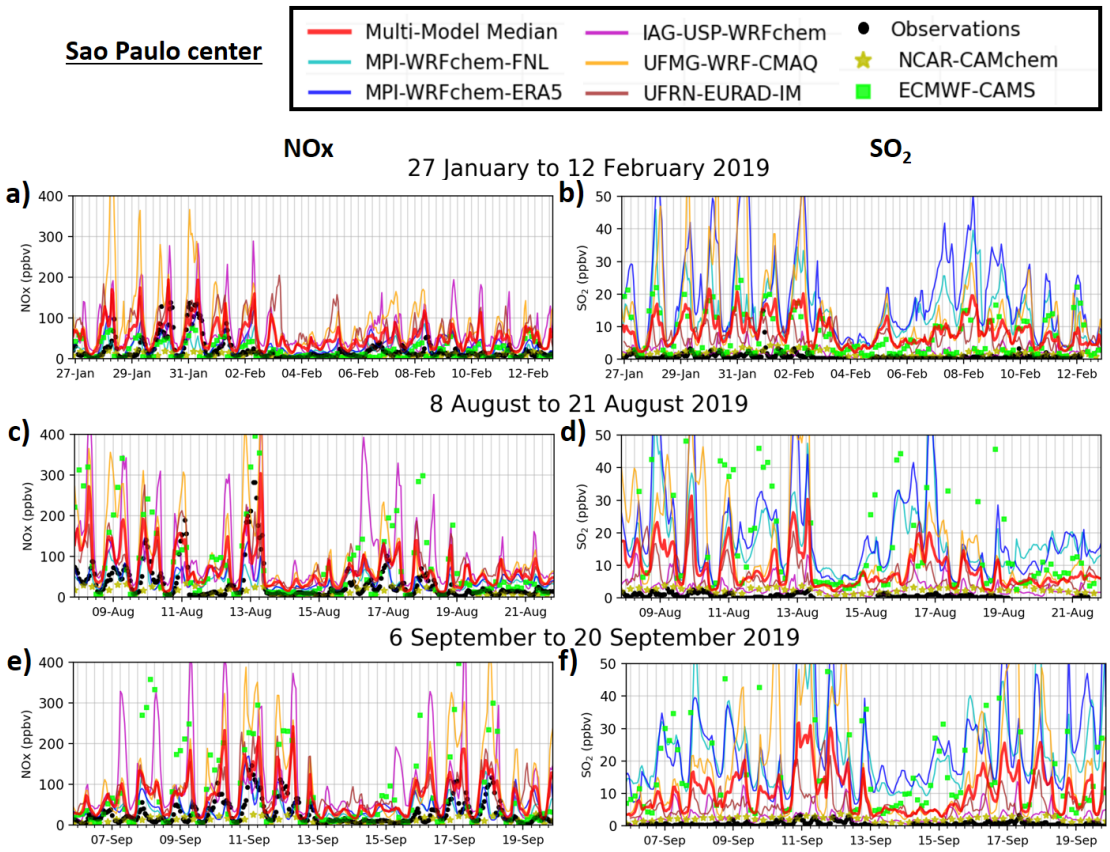

Figure 4. *Time series of hourly concentrations of NOx (a, c and e) and SO_2 (b, d and f) observed and modeled in São Paulo for the three selected 15-day periods of the year 2019. The models include data from two global forecasts (yellow stars and green squares) and a regional model ensemble of five simulations (colored lines) with the Multi-Model Median (red line).*

São Paulo and Santos, again suggesting an overestimation of the GFAS emissions for this type of event.

To our knowledge, there have been no major regulatory changes that could explain the large overestimation of modeled SO_2 concentrations (on gasoline content or industry stack emissions). Therefore, we suspect anthropogenic emissions (rather than fire emissions), and more specifically the industrial sector (rather than traffic), to be responsible for the large model bias, which may be related to emission factors and to the spatial proxy defining source locations.

In summary, the models reproduce the meteorology well and the modeled concentrations of CO, PM and NOx are in good agreement when there is no biomass burning pollution event. This section shows the importance of these events for the air quality in the São Paulo region as well as the difficulty for the models to obtain the correct magnitude of CO, NOx, PM and SO_2 during these events.

5 Assessment of the modeled photochemistry

This section is dedicated to the evaluation of the photochemistry that the models reproduce in the tropical and urban environment of São Paulo. We expect the São

Paulo center to be saturated with NOx and ozone production to be controlled by the level of volatile organic compounds (Schuch et al., 2019; Rudke et al., 2021; Squizzato et al., 2021).

The level of oxidant (*i.e.* Ox = NO_2 + O_3) is an interesting quantity for our analysis because it should vary less between day and night (Wood et al., 2010). In urban areas, where NOx emission are important, there is a competition between the loss and the production of ozone during the day (the titration of ozone by NO is compensated by the photolysis of NO_2). As a result, there is a partitioning between NO_2 and O_3 due to the daytime photo-stationary state, thus an increase of Ox during the day corresponds more likely to the formation of ozone. At night, Ox is not affected by the titration of ozone.

We analyze the ozone and Ox concentrations in São Paulo, Santos and Campinas during the three studied periods (Sect. 5.1), and we focus on the averaged diurnal variability in São Paulo (Sect. 5.2).

5.1 Ozone and oxidant levels

We investigate the temporal variability of ozone and Ox concentrations in São Paulo (Fig. 5), Santos (Fig. A8) and Campinas (Fig. A7).

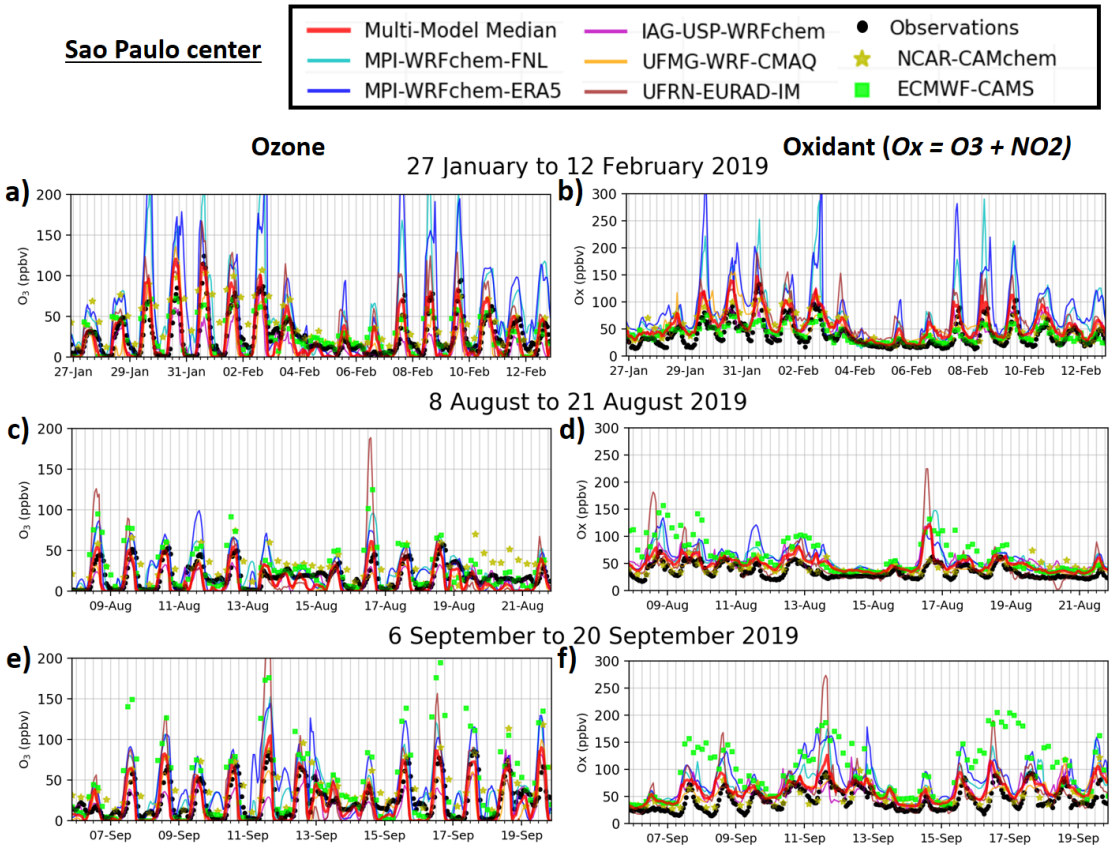

Figure 5. *Time series of hourly concentrations of ozone (a, c and e) and oxidant (b, d and f) observed and modeled in São Paulo for the three selected 15-day periods of the year 2019. The models include data from two global forecasts (yellow stars and green squares) and a regional model ensemble of five simulations (colored lines) with the Multi-Model Median (red line).*

Ozone observations in São Paulo show a clear diurnal cycle for most days, with a daily minimum below 10 ppb at night and a daily maximum above 50 ppb, except during certain 2-day periods associated with storms (*cf.* Section 4.2). For Ox, there is a background level of around 20 ppb, and there are often increases during the day that match the ozone increases. The second period has a more consistent oxidant level compared to the other two periods, which were chosen because they contain high ozone events. Looking at Santos and Campinas, ozone concentrations also show a clear diurnal cycle with a smaller amplitude, and the oxidant level is more constant than in São Paulo, with the same background level of around 20 ppb for the three periods. It is noted that in Campinas, the ozone concentration is often high at night, above 20 ppb, which is not observed in the other two places.

For the three locations, the models of the regional ensemble are in good agreement with the temporal variation of the observed ozone concentrations. It can be seen that the level of oxidant is overestimated by the regional model ensemble and the two global forecasts. NCAR-CAMchem is the most in agreement regarding the range of concentrations. Each model of the regional ensemble has days for which the modeled value is higher than the maximum observed ozone concentration, suggesting that the modeled ozone production reaches an intensity that is not observed. For ECMWF–CAMS, the three periods are not found with the same quality because during the third, the ozone is largely overestimated (much more than for all the other models) in São Paulo, in Santos and to a lesser extent in Campinas.

For all models, the oxidant level is overestimated in the metropolitan area of São Paulo (Fig. 5) compared to Santos (Fig. A8) and Campinas (Fig. A7). We note that the two WRFchem simulations run at MPI overestimate ozone and Ox, and that this overestimation is greater with the ERA5 reanalysis. Moreover, we note that IAG-USP–WRFchem underestimates ozone, and that UFRN–EURAD-IM and UFMG–WRF-CMAQ have good agreement. Focusing on individual days, we also note that each individual simulation has certain periods for which ozone is in better agreement. Therefore, the MMM has overall the best agreement for ozone over all three time periods.

The two meteorological inputs used at MPI with WRFchem lead to significant magnitude differences for certain days, for example during biomass burning pollution events (*cf.* Sect. 4.2), which could be due to differences in the air masses transported to the megacity. We further investigate the relationship between ozone and wind direction to identify sectors of wind direction associated with high or low ozone concentrations, and compare those with modeled results (Fig. 6 and Fig. A9).

The wind direction observed is mainly from West to North sectors (more than 80 % of the hourly occurrence) and sometimes from East to South sectors (less than 15 %) for the three periods. Low (below 16 ppb) and high (above 50 ppb) ozone concentrations are associated with west-north sectors, while high (above 50 ppb) concentrations are associated with east-south sectors.

The MMM reproduces well the occurrence of the wind direction as well as the observed distribution of ozone concentrations (Fig. 6). The main wind direction is well reproduced except for the third period where there is a shift (coming from N-NE instead of N-NW). However, the individual simulations have significant biases regarding the occurrence of wind direction and the distribution of ozone concentrations (Fig. A9). This analysis is limited by the difficulty of defining a wind direction when the wind speed is low, especially in a megacity. Nevertheless, we still notice that the MMM is in better agreement with the observation than each of its members.

To synthesize the results of the different simulations, we plot the modeled and observed ozone and Ox concentrations in a scatter plot with the regression line of each regional model using the reduced major axis method (Fig. 7). For each model of the re-

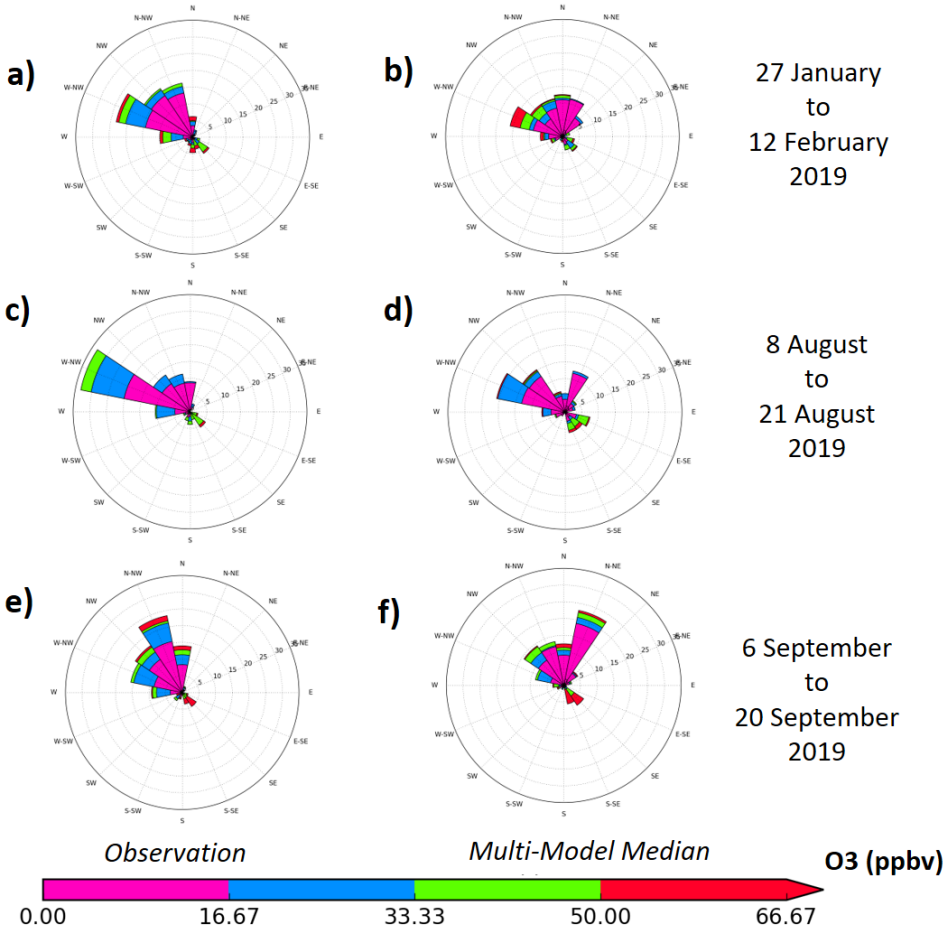

Figure 6. *Pollution roses obtained from the hourly occurrence of the observed and modeled wind direction (Multi-Model Median) by direction sector (in %) using 16 sectors, for the three selected 15-day periods of the year 2019. Each pollution rose shows the predominant direction of the pollution transport. For each wind direction sector, the distribution of ozone concentrations is given separated into four concentration ranges (color code).*

gional ensemble, the regression lines are similar (in terms of agreement of slope with respect to the line Mod=Obs) for the three periods for ozone and for Ox. For ozone, the best agreement is obtained for the MMM, then UFMG–WRF-CMAQ, whereas the UFRN–EURAD-IM and MPI–WRFchem simulations overestimate it and that of the IAG-USP–WRFchem underestimates it. For Ox, we again observe the overestimation of the models because the vast majority of the points are located above the line Mod=Obs, and therefore the regression lines are shifted. For the two pollutants, the slopes are correct for IAG-USP–WRFchem and UFMG–WRF-CMAQ whereas for UFRN–EURAD-IM and MPI-WRFchem they are overestimated, which seems to indicate that ozone production is too high.

5.2 Average diurnal cycles

The concentrations of NOx and ozone show marked diurnal variability over the three periods studied, which is notably due to the evolution during the day of anthropogenic

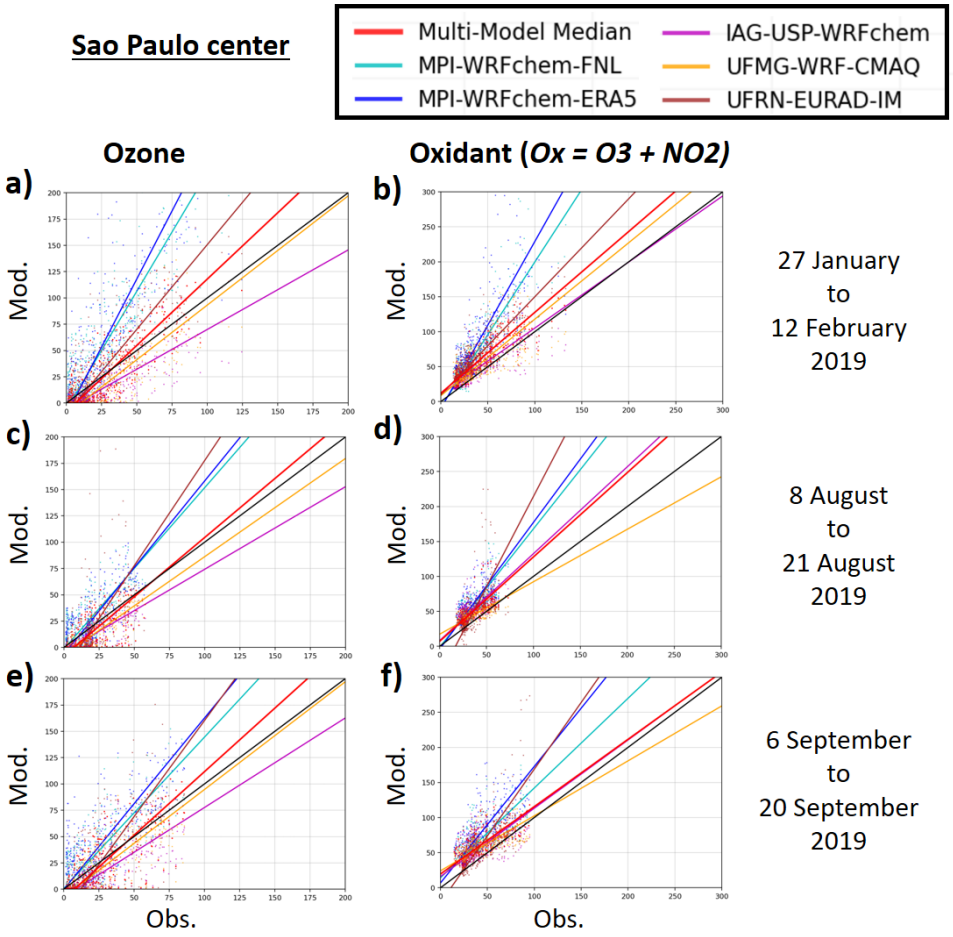

Figure 7. *Ozone (a, c, and e) and oxidant (b, d, and f) scatter plots of observed versus modeled hourly concentrations for the three selected 15-day periods of the year 2019. The regression lines are calculated using the reduced major axis method for each model. The models include data from a regional model ensemble from five simulations (colored lines) with the Multi-Model Median (red line).*

emissions and of the height of PBL. We continue by analyzing the average diurnal cycles of ozone, NOx concentrations with the modeled PBL heights (Fig. 8) as well as NO and NO_2 (Fig. A10).

On average, the ozone concentration in São Paulo has three phases: (i) it is below 20 ppb from midnight to 9h, (ii) it increases until 16h, up to 50 ppb, 35 ppb and 50 ppb for the first, second and respectively the third periods, (ii) it decreases slowly until midnight for the first period, while the decreases are faster (until 19h) for the second and third periods.

The diurnal cycle of NOx is opposite to that of ozone for the three periods because high concentrations are observed at night (reaching 50 ppb) and low concentrations during the day (below 25 ppb). The concentration of NOx, as well as NO and NO_2, presents a peak at 8h-9h, which seems to correspond to the morning peak of traffic emissions. There is another period of high concentration in the evening which lasts longer and differs between periods (comparing Fig. 8 and Fig. A10). NOx concentrations are higher from

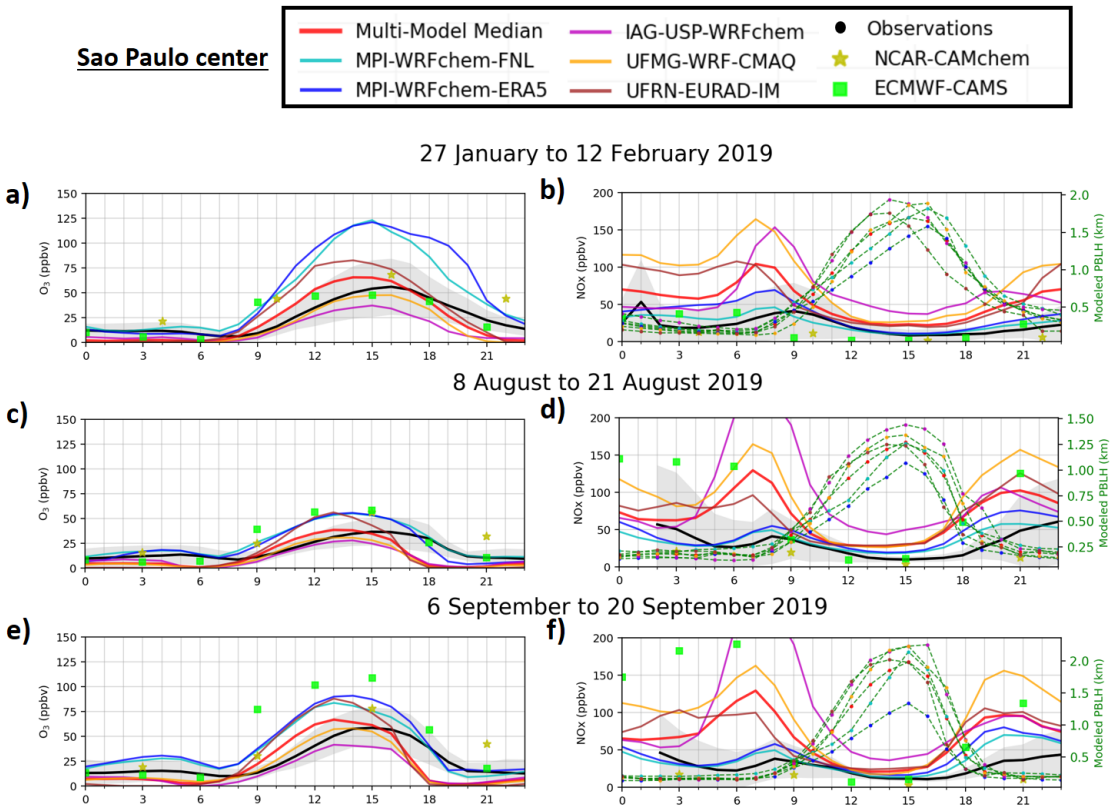

Figure 8. *Average diurnal cycles of hourly concentrations of ozone (a,c and e) and NOx (b,d and f) observed and modeled in São Paulo over the three selected 15-day periods of the year 2019. The models include data from two global forecasts (yellow stars and green squares) and a regional model ensemble of five simulations (colored lines) with the Multi-Model Median (red line). The modeled planetary boundary layer heights (PBLH) are the green dashed lines with colored dots corresponding to the models. The black line is observation average and the gray shadings correspond to the standard deviation.*

19h to 3h during the second and third periods compared to the first, which is driven by a difference in NO. It should also be noted that the morning peak is observed around 8h for NO and around 10h for NO_2, while in the evening, a long period of high concentrations of NO and NO_2 from 19h to 3h.

The models reproduce well the chronology of the observed phases of the mean diurnal cycle of ozone. For NOx, the traffic peak is well modeled around 8h, while the period of high NOx in the evening (observed between 19h to 3h) is modeled too early. During daytime, low NOx correspond well to the PBL height greater than 1 km. Looking at the magnitudes of the diurnal cycles, we see that:

- For MPI–WRFchem, ozone is overestimated (day and night), and NOx is in good agreement;
- For IAG-USP–WRFchem, ozone is underestimated (day and night), and NOx is overestimated at night;
- For UFMG–WRF-CMAQ, ozone is in good agreement during the day and underestimated at night, and NOx is overestimated at night;

- For UFRN–EURAD-IM, ozone is overestimated during the day and underestimated at night, and NOx is overestimated at night;
- For NCAR–CAMchem, ozone is overestimated (day and night), and NOx is underestimated;
- For the ECMWF–CAMS, ozone is overestimated during the day and underestimated at night, and NOx is overestimated at night.

In addition, the modeled PBL heights are similar for the regional model ensemble over the three time periods. The PBL height modeled with ERA5 reanalysis (MPI–WRFchem-ERA5) is the lowest. The modeled PBL height is highest during the day-to-night transition for the first period, which could explain the lower modeled NOx concentrations. However, the modeled PBL, being similar in time and height, cannot explain the large inter-model variability observed for ozone and NOx, which is particularly true from 6h to 9h.

Regarding NO and NO_2 (Fig. A10), the differences between the models are more important for NO than for NO_2, and they seem related to modeled ozone biases because we see that:

- For MPI–WRFchem, NO is underestimated (at night) and NO_2 is overestimated (at night);
- For IAG-USP–WRFchem, NO and NO_2 are overestimated (day and night);
- For UFMG–WRF-CMAQ, NO and NO_2 are overestimated (night);
- For UFRN–EURAD-IM, NO and NO_2 are overestimated (night);
- For NCAR–CAMchem, NO and NO_2 are underestimated (day and night);
- For ECMWF–CAMS, NO and NO_2 are overestimated (night).

At night, for all models, the biases in modeled NO concentrations are opposite to the biases in modeled ozone concentrations, despite the consistency between the modeled PBL height. Consequently, the proportion of NO to NO_2 appears to be related to the modeled ozone biases. We thus analyze the diurnal cycles of the proportion of NO_2 in NOx and in Ox predicted by the regional model ensemble compared to observation (Fig. 9).

- For MPI–WRFchem, the proportion of NO_2 in NOx is overestimated, and in Ox is in good agreement;
- For IAG-USP–WRFchem, the proportion of NO_2 in NOx is underestimated, and in Ox is overestimated;
- For UFMG–WRF-CMAQ, the proportion of NO_2 in NOx is is underestimated, and in Ox is overestimated (at night);
- For UFRN–EURAD-IM, the proportions of NO_2 in NOx and of NO_2 in Ox are underestimated (at night).

The MMM has the best agreement for ozone because two models overestimate it and the other two underestimate it. The level of oxidant is especially overestimated in the metropolitan area of São Paulo (Fig. 5) compared to the two surrounding localities studied (Fig. A8 and Fig. A7), and this for all models. Understanding this overestimation may be essential to improve the modeled ozone variability in the PBL of São Paulo.

However, from this analysis it is not possible to identify the main drivers of the variability of ozone and the level of oxidant, which are related to anthropogenic and biogenic emissions, urban dynamics in the PBL, to the chemistry, to the deposition, to the radiation or to the configuration of the models. Thus, each institution should conduct sensitivity studies to improve its simulation using the results of this intercomparison to assess their performances.

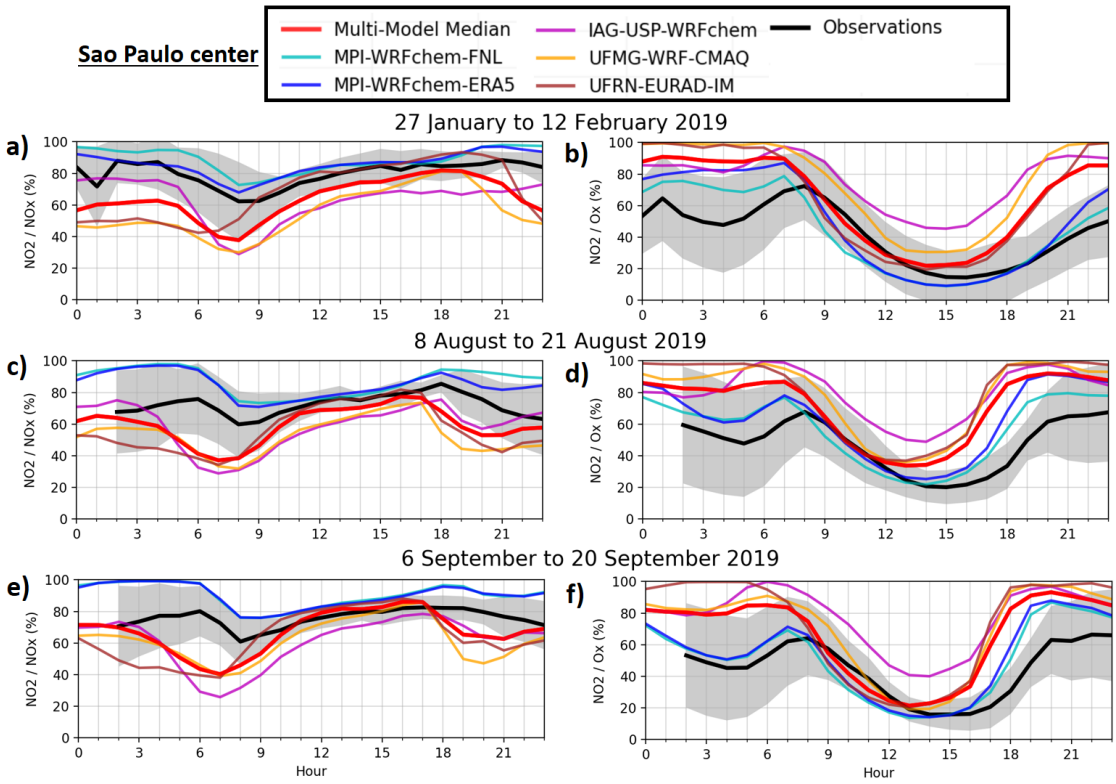

Figure 9. *Average diurnal cycles of hourly proportion of NO_2 in NOx (a,c and e) and in Ox (b,d and f) observed and modeled in São Paulo over the three selected 15-day periods of the year 2019. The models include data from a regional model ensemble of five simulations (colored lines) with the Multi-Model Median (red line). The black line is observation average and the gray shadings correspond to the standard deviation.*

In conclusion, there is a large inter-model variability in the magnitude of modeled daily maximum of ozone (approximately ± 20 ppb around the observed value). The ozone bias of the models seems to be related to the relative proportions of NO and NO_2 as well as to the amount of NOx. Overall, the Multi-Model Median has the best agreement.

6 Potential of the regional model ensemble

Of course, the small number of models involved in the calculation of the MMM, *i.e.* the median of the four models, is an important limitation. However the previous section showed that two models overestimate ozone, and the other two underestimate, leading to good scores for the MMM. This section proposes to focus on the MMM to finely analyze the temporal biases of Ox and NOx (Sect. 6.1), and to evaluate the potential of the MMM in the perspective of an early warning system for ozone and aerosol alerts (Sect. 6.2).

6.1 Ox and NOx temporal biases

We analyze the temporal biases, *i.e.* the modeled minus observed concentration, for Ox (Fig. 10) and NOx (Fig. A11) as well as the average diurnal cycles in order to distinguish the phases which occur during the day.

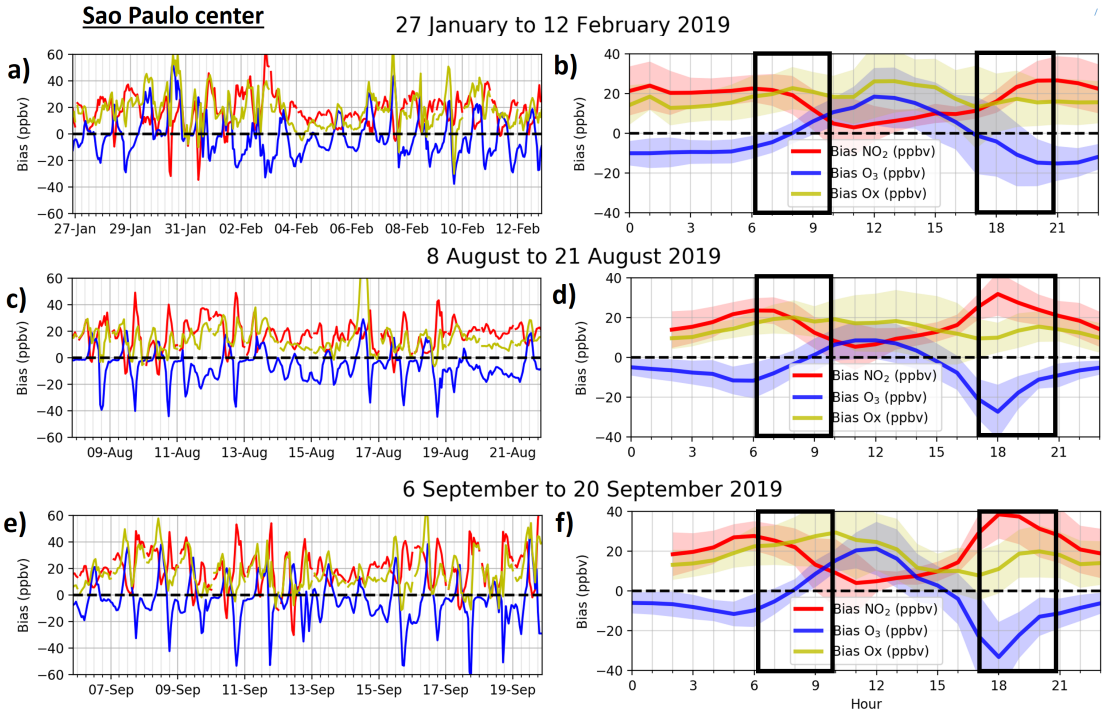

Figure 10. *Time series of hourly bias (difference in modeled and observed concentration) of the Multi-Model Median for ozone, NO_2 and Ox (a, c and e) and their associated average diurnal cycles (b , d and f) in São Paulo for the three selected 15-day periods of the year 2019. The Multi-Model Median is calculated from a regional model ensemble of four simulations. The black boxes mark the morning and evening hours.*

We note that the concentration of Ox is overestimated during the three periods and that there is an opposition of the bias in NO_2 and ozone, which seems to take place on most days, and which is well represented in the average diurnal cycles. It follows that it seems possible to define different diurnal phases of the bias in NO_2 and ozone, such as:

1. At night (21h to 6h), the NO_2 bias is positive (overestimation) and that of ozone is negative (underestimation);
2. In the morning (from 6h to 10h), the NO_2 and ozone biases are large at 6h and then decrease;
3. During the day (from 10h to 17h), the ozone bias becomes positive while the NO_2 bias is weak;
4. In the evening (from 17h to 21h), the biases are strongest, NO_2 is overestimated and ozone is underestimated.

The evening period exhibits biases similar to the morning but stronger, which could be related to the urban heat effect which would in fact keep the height of the PBL higher than in the models. Looking at the NOx biases (Fig. A11), we see that the NO bias is much stronger than the NO2 bias, especially in the morning and evening. The same diurnal phases are noted for NOx as for Ox, suggesting that different factors or processes are responsible for these biases during each phase:

1. At night (21h to 6h): this phase is linked to nocturnal chemistry, when the height of the PBL is low (a few hundred meters). During this phase, the MMM has a strong NOx bias. The results of the individual simulations showed a high inter-model variability for NOx concentrations as well as for the proportion of NO_2 in NOx and in Ox (Sect. 5). This suggests that the treatment of anthropogenic emissions (in terms of sector or NO/NO_2 ratio at the emission) and nocturnal chemistry play an important role;
2. In the morning (from 6h to 10h): this phase is linked to the peak of morning traffic and the transition from night to day, with an increasing PBL height. During this phase, the bias of ozone becomes positive while the bias of NO_2 decreases. The results of the individual simulations were similar for the height of the PBL but there is a strong inter-model variability for NO and NO_2. This suggests that there are significant differences in the magnitude (and hourly profile) of anthropogenic emissions associated with the traffic sector between models;
3. During the day (from 10h to 17h): this phase is related to the active period of photochemistry, with a high PBL up to about 2 km. During this period, the bias of ozone is positive and that of NO_2 is weak. Individual simulations predict daily ozone maxima with high variability, while PBL heights and low NOx concentrations are similar. This suggests that ozone production is different, hence the ratios of NOx to volatile organic compounds between models, which are related to anthropogenic and biogenic emissions;
4. In the evening (from 17h to 21h): this phase is linked to the evening traffic peak and the transition from day to night, with a decreasing PBL height. As for the morning phase, there is an underestimation of ozone and an overestimation of NO_2, but it is the phase with the largest biases. In addition, there is high inter-model variability of NO and NO_2, indicating large differences in emissions from the traffic sector.

In conclusion, our regional model ensemble shows an underestimation of ozone at night and an overestimation during the day. This section indicates that anthropogenic emissions are linked to the biases of each diurnal phase, particularly in the morning and afternoon, and their treatment seems to be one of the keys to improving the models.

6.2 Air quality alerts

This section analyzes the performance of the median of the regional model ensemble in terms of ozone and $PM_{2.5}$ alerts. The WHO air quality standards are based on the maximum daily average for 8 hours (MDA8) for the concentration of ozone, and on the daily average for the concentration of $PM_{2.5}$. We use the WHO standards, *i.e.* threshold of concentration, of 50 ppb for ozone and of 25 $\mu g.m^{-3}$ for $PM_{2.5}$ (guidelines used before 2021). If the WHO threshold is exceeded during a day, then there is an alert. There are therefore four cases for each day:

- Case A: an alert is observed and modeled;
- Case B: an alert is observed and not modeled;
- Case C: an alert is neither observed nor modeled;
- Case D: an alert is not observed but modeled.

Moreover, in order to quantify the performance of MMM predictions, the probability of detection (POD) and the false alarm rate (FAR) are calculated following Brasseur and Jacob (2017) such that:

$$POD = N(CaseA)/N(CaseA + B) \tag{3}$$

$$FAR = N(CaseD)/N(CaseA + D) \tag{4}$$

We compare the number of alerts and non-alerts between observations and the MMM (Fig. 11).

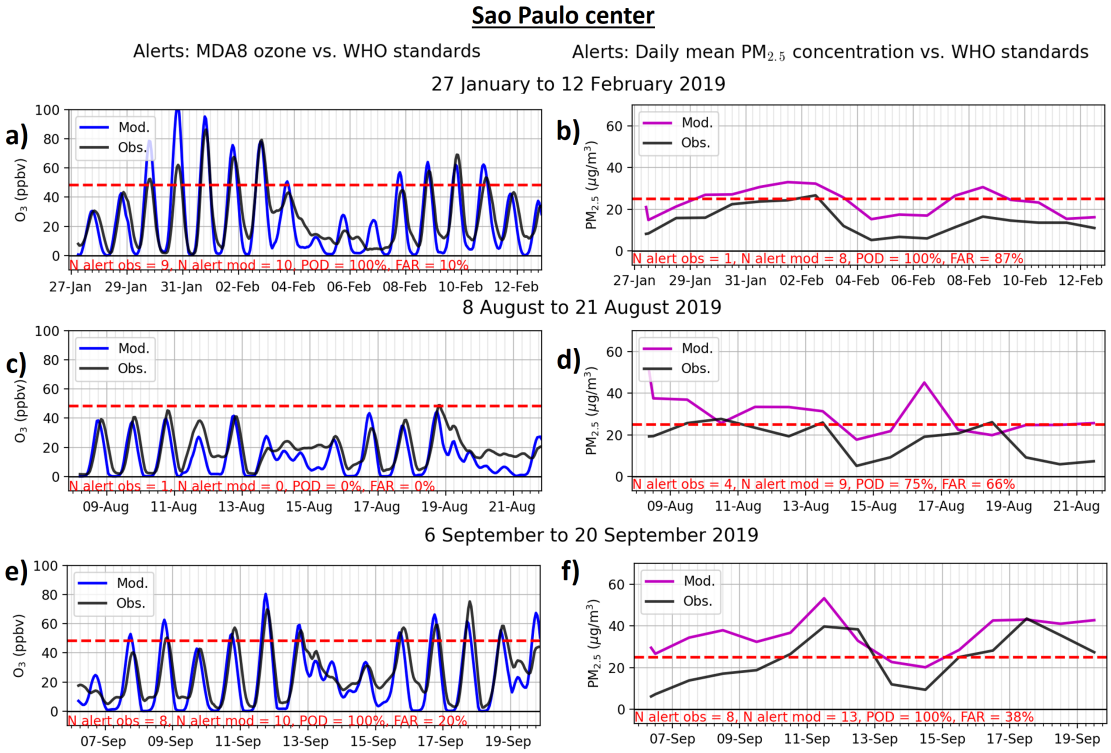

Figure 11. *Modeled and observed MDA8 ozone concentrations (a, c and e) and $PM_{2.5}$ concentrations (b, d and f) for the three periods. The thresholds defined by the WHO standards are represented by the horizontal red dotted lines.*

The median of the regional model ensemble shows good performance for ozone and poor performance for $PM_{2.5}$ due to its constant overestimation. The number of alerts is well predicted for ozone, even for the second period which is predicted without any alert while one was observed (close to the threshold). The first and third periods have low FAR and maximum POD for ozone concentration. For $PM_{2.5}$, the overestimation is of the order of 10 µg.m^{-3} for the three periods, which implies that there is too often an alert for the three periods. Alerts associated with days of biomass burning pollution events are less well reproduced (*cf.* Sect. 4.3).

In conclusion, the performance of the regional model ensemble is promising for the development of the air quality warning forecast system, in terms of alerting the population as the quality is good for ozone and for $PM_{2.5}$ on condition of improving the forecast of pollution due to biomass burning.

7 Conclusions

This study addresses the development of an air quality forecasting system based on a regional model ensemble for the megacity of São Paulo. We compare the results of regional air quality models carried out by four institutes, over three 15-day periods that include particular air pollution events. We focus on the heavily urbanized area, where we expect anthropogenic emissions to be dominant. We show that the median of the re-

gional model ensemble, even with the low number of models we considered, performs well for ozone (better than compared to the global forecasts made at NCAR and ECMWF), although the performance for NOx is poor due to the large inter-model variability.

Our results suggest that the treatment of anthropogenic emissions is an important factor in explaining the variability of modeled NO and NO_2 concentrations. There is a strong overestimation of the level of oxidant (defined as $Ox = O_3 + NO_2$) in the metropolitan area of São Paulo compared to the surrounding localities. The transition from day to night is particularly biased, which could be linked to the absence of urban heat effect. The overestimation of NO_2 concentration made by all models in the evening should be reduced with increased PBL height taking into account this effect. A study focusing on the drivers of the level of oxidant in the PBL of megacities is particularly needed to understand the sensitivity related to anthropogenic and biogenic emissions, urban dynamics, chemistry, deposition, or radiation.

Nevertheless, many other factors influence the performance of the regional model ensemble. For example, the model configurations for the size domain and the horizontal resolution were not constrained for this study. This choice is limited by available computing time. On the one hand, the finest possible resolution is desired for the center of São Paulo. On the other hand, a vast area integrating the different sources of pollutants such as agricultural fires which are important on a regional scale is needed. For most of the pollutants considered, the score of the median of the regional model ensemble is the best because it seems to benefit of the different model configurations.

The use of more sophisticated chemical schemes or aerosol schemes, which would cost more computation time, may not be the priority because the modeled biases are mostly associated with primary emissions. Indeed, our results demonstrated the importance of biomass burning pollution events occurring at the regional scale for the air quality of São Paulo, as well as the difficulty for the model to represent these events. The use of satellite information and its integration, in particular through data assimilation techniques, should improve the forecasting of these events in São Paulo. In perspective, a similar study on the composition of aerosols, and related to the meteorological systems, to the removal processes and to the radiative balance would be interesting in addition to this study.

Open Research Section

- For the observational data, we thank CETESB (Companhia Ambiental do Estado de São Paulo) for sharing the data, which are available through this website: `https://cetesb.sp.gov.br/ar/qualar/`;
- For ECMWF–CAMS, data are available through this website: `https://ads.atmosphere.copernicus.eu/cdsapp#!/dataset/cams-global-atmospheric-composition-forecasts`, last access: November 4, 2022;
- For NCAR–CAMchem, data are available through this website: `https://www.acom.ucar.edu/cam-chem/cam-chem.shtml`, last access: November 4, 2022.

Availability of model data: Upon acceptance of the manuscript, the model data will be made accessible.

Acknowledgments

This article is a direct contribution to the research themes of the Klimapolis Laboratory (klimapolis.net), which is funded by the German Federal Ministry of Education and Research (BMBF). A.D. acknowledge the European Union's Horizon 2020 research and innovation programme for supporting this work under the Marie Skłodowska-Curie grant agreement No 895803 (MACSECH — H2020-MSCA-IF-2019).

Authors contribution:

AD designed the study, performed the analysis and wrote the first draft. AD, IB, PL, LL and GB produced the MPI–WRFchem simulations. JJH, ESFD, HE, ACL and PF produced the UFRN–EURAD-IM simulations. TTAA, WLA and RP produced the UFMG–WRF-CMAQ simulations. RYY, MFA and RAC produced the IAG-USP–WRFchem simulations. GAM provided the PBL height data. LDM provided the classification of the CETESB stations. All authors contributed to the final version of the manuscript.

The computation of the simulations presented in this work were completed by different supercomputers:

- For MPI–WRFchem, the authors gratefully acknowledge the computing time granted by DKRZ (German Climate Computing Centre);
- For UFRN–EURAD-IM, the authors gratefully acknowledge the computing time granted by the JARA Vergabegremium and provided on the JARA Partition part of the supercomputer JURECA at Forschungszentrum Jülich.

References

Albuquerque, T. T. d. A., West, J., de F. Andrade, M., Ynoue, R. Y., Andreão, W. L., dos Santos, F. S., ... Moreira, D. M. (2019, 11). Analysis of pm2.5 concentrations under pollutant emission control strategies in the metropolitan area of são paulo, brazil. *Environmental Science and Pollution Research*, *26*, 33216-33227. doi: 10.1007/s11356-019-06447-6

Andrade, M. d. F., Kumar, P., de Freitas, E. D., Ynoue, R. Y., Martins, J., Martins, L. D., ... Zhang, Y. (2017, 6). Air quality in the megacity of são paulo: Evolution over the last 30 years and future perspectives. *Atmospheric Environment*, *159*, 66-82. doi: 10.1016/j.atmosenv.2017.03.051

Andrade, M. d. F., Ynoue, R. Y., Freitas, E. D., Todesco, E., Vela, A. V., Ibarra, S., ... Carvalho, V. S. B. (2015, 2). Air quality forecasting system for southeastern brazil. *Frontiers in Environmental Science*, *3*. doi: 10.3389/fenvs.2015.00009

Baklanov, A., Molina, L. T., & Gauss, M. (2016). Megacities, air quality and climate. *Atmospheric Environment*, *126*, 235-249. doi: 10.1016/j.atmosenv.2015.11.059

Baklanov, A., & Zhang, Y. (2020, 1). Advances in air quality modeling and forecasting. *Global Transitions*, *2*, 261-270. doi: 10.1016/j.glt.2020.11.001

Bieser, J., Aulinger, A., Matthias, V., Quante, M., & van der Gon, H. D. (2011, 10). Vertical emission profiles for europe based on plume rise calculations. *Environmental Pollution*, *159*, 2935-2946. doi: 10.1016/j.envpol.2011.04.030

Brasseur, G. P., & Jacob, D. J. (2017). *Modeling of atmospheric chemistry*. Cambridge University Press.

Brasseur, G. P., Xie, Y., Petersen, A. K., Bouarar, I., Flemming, J., Gauss, M., ... Zhou, G. (2019, 1). Ensemble forecasts of air quality in eastern China-Part 1: Model description and implementation of the MarcoPolo-Panda prediction system, version 1. *Geoscientific Model Development*, *12*, 33–67. doi: 10.5194/gmd-12-33-2019

Brito, J., Carbone, S., dos Santos, D. A. M., Dominutti, P., de Oliveira Alves, N., Rizzo, L. V., & Artaxo, P. (2018, 12). Disentangling vehicular emission impact on urban air pollution using ethanol as a tracer. *Scientific Reports*, *8*, 10679. doi: 10.1038/s41598-018-29138-7

Byun, D., & Schere, K. L. (2006, 3). Review of the governing equations, computational algorithms, and other components of the models-3 community multiscale air quality (cmaq) modeling system. *Applied Mechanics Reviews*, *59*, 51-77. doi: 10.1115/1.2128636

CETESB. (2022). São paulo air quality measurment network. *Companhia Ambiental do Estado de São Paulo*. Retrieved from `https://cetesb.sp.gov.br/ar/wp-content/uploads/sites/28/2020/07/Relat%C3%B3rio-de-Qualidade-do-Ar-2019.pdf,[Accessed:27/09/2022]`

Crippa, M., Guizzardi, D., Muntean, M., Schaaf, E., Dentener, F., van Aardenne, J. A., ... Janssens-Maenhout, G. (2018, 10). Gridded emissions of air pollutants for the period 1970–2012 within edgar v4.3.2. *Earth System Science Data*, *10*, 1987-2013. doi: 10.5194/essd-10-1987-2018

Crippa, M., Solazzo, E., Huang, G., Guizzardi, D., Koffi, E., Muntean, M., ... Janssens-Maenhout, G. (2020, 12). High resolution temporal profiles in the emissions database for global atmospheric research. *Scientific Data*, *7*, 121. doi: 10.1038/s41597-020-0462-2

Ek, M. B., Mitchell, K. E., Lin, Y., Rogers, E., Grunmann, P., Koren, V., ... Tarpley, J. D. (2003). Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. *Journal of Geophysical Research: Atmospheres*, *108*(D22), n/a–n/a. doi: 10.1029/2002JD003296

Elbern, H., Strunk, A., Schmidt, H., & Talagrand, O. (2007, 7). Emission rate and chemical state estimation by 4-dimensional variational inversion. *Atmospheric Chemistry and Physics*, *7*, 3749-3769. doi: 10.5194/acp-7-3749-2007

Emmons, L. K., Walters, S., Hess, P. G., Lamarque, J.-F., Pfister, G. G., Fillmore, D., ... Kloster, S. (2010). Geoscientific model development description and evaluation of the model for ozone and related chemical tracers, version 4 (mozart-4). *Geosci. Model Dev*, *3*, 43-67.

Fast, J. D., Gustafson, W. I., Easter, R. C., Zaveri, R. A., Barnard, J. C., Chapman, E. G., ... Peckham, S. E. (2006, 11). Evolution of ozone, particulates, and aerosol direct radiative forcing in the vicinity of houston using a fully coupled meteorology-chemistry-aerosol model. *Journal of Geophysical Research Atmospheres*, *111*. doi: 10.1029/2005JD006721

Galmarini, S., Bianconi, R., Addis, R., Andronopoulos, S., Astrup, P., Bartzis, J., ... der Auwera, L. V. (2004, 9). Ensemble dispersion forecasting—part ii: application and evaluation. *Atmospheric Environment*, *38*, 4619-4632. doi: 10.1016/j.atmosenv.2004.05.031

Ginoux, P., Chin, M., Tegen, I., Prospero, J. M., Holben, B., Dubovik, O., & Lin, S.-J. (2001, 9). Sources and distributions of dust aerosols simulated with the gocart model. *Journal of Geophysical Research: Atmospheres*, *106*, 20255-20273. doi: 10.1029/2000JD000053

Godoy-Silva, D., Nogueira, R. F., & Campos, M. L. A. (2017, 12). A 13-year study of dissolved organic carbon in rainwater of an agro-industrial region of são paulo state (brazil) heavily impacted by biomass burning. *Science of the Total Environment*, *609*, 476-483. doi: 10.1016/j.scitotenv.2017.07.145

Granier, C., Darras, S., van der Gon, H. D., Doubalova, J., Elguindi, N., Galle, B., ... Sindelarova, K. (2019). The copernicus atmosphere monitoring service global and regional emissions (april 2019 version). doi: 10.24380/d0bn-kx16

Grell, G., & Baklanov, A. (2011, 12). Integrated modeling for forecasting weather and air quality: A call for fully coupled approaches. *Atmospheric Environment*, *45*, 6845-6851. doi: 10.1016/j.atmosenv.2011.01.017

Grell, G. A., & Dévényi, D. (2002, 7). A generalized approach to parameterizing convection combining ensemble and data assimilation techniques. *Geophysical Research Letters*, *29*, 38-1-38-4. doi: 10.1029/2002GL015311

Grell, G. A., Peckham, S. E., Schmitz, R., McKeen, S. A., Frost, G., Skamarock, W. C., & Eder, B. (2005, 12). Fully coupled "online" chemistry within the wrf model. *Atmospheric Environment*, *39*, 6957-6975. doi: 10.1016/j.atmosenv.2005.04.027

Guenther, a., Karl, T., Harley, P., Wiedinmyer, C., Palmer, P. I., & Geron, C. (2006). Estimates of global terrestrial isoprene emissions using MEGAN (Model of Emissions of Gases and Aerosols from Nature). *Atmospheric Chemistry and Physics Discussions*, *6*, 107–173. doi: 10.5194/acpd-6-107-2006

Hass, H., Jakobs, H. J., & Memmesheimer, M. (1995). Analysis of a regional model (eurad) near surface gas concentration predictions using observations from networks. *Meteorology and Atmospheric Physics*, *57*, 173-200. doi: 10.1007/BF01044160

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., ... Thépaut, J.-N. (2020). The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999-2049. doi: 10.1002/qj.3803

Hong, S.-Y., Noh, Y., & Dudhia, J. (2006). A new vertical diffusion package with an explicit treatment of entrainment processes. *Monthly Weather Review*, *134*(9), 2318–2341. doi: 10.1175/MWR3199.1

Huneeus, N., van der Gon, H. D., Castesana, P., Menares, C., Granier, C., Granier, L., ... Ynoue, R. Y. (2020, 8). Evaluation of anthropogenic air pollutant emission inventories for south america at national and city scale. *Atmospheric Environment*, *235*, 117606. doi: 10.1016/j.atmosenv.2020.117606

Im, U., Bianconi, R., Solazzo, E., Kioutsioukis, I., Badia, A., Balzarini, A., ... Galmarini, S. (2015, 8). Evaluation of operational on-line-coupled regional air quality models over europe and north america in the context of aqmeii phase 2. part i: Ozone. *Atmospheric Environment*, *115*, 404-420. doi: 10.1016/j.atmosenv.2014.09.042

Janssens-Maenhout, G., Crippa, M., Guizzardi, D., Dentener, F., Muntean, M., Pouliot, G., ... Li, M. (2015, oct). HTAP_v2.2: a mosaic of regional and global emission grid maps for 2008 and 2010 to study hemispheric transport of air pollution. *Atmospheric Chemistry and Physics*, *15*(19), 11411–11432. doi: 10.5194/acp-15-11411-2015

Kaiser, J. W., Heil, A., Andreae, M. O., Benedetti, A., Chubarova, N., Jones, L., ... van der Werf, G. R. (2012, 1). Biomass burning emissions estimated with a global fire assimilation system based on observed fire radiative power. *Biogeosciences*, *9*, 527-554. doi: 10.5194/bg-9-527-2012

Mailler, S., Khvorostyanov, D., & Menut, L. (2013, 6). Impact of the vertical emission profiles on background gas-phase pollution simulated from the emep emissions over europe. *Atmospheric Chemistry and Physics*, *13*, 5987-5998. doi: 10.5194/acp-13-5987-2013

Marécal, V., Peuch, V.-H., Andersson, C., Andersson, S., Arteta, J., Beekmann, M., ... Ung, A. (2015, 9). A regional air quality forecasting system over europe: the macc-ii daily ensemble production. *Geoscientific Model Development*, *8*, 2777-2813. doi: 10.5194/gmd-8-2777-2015

Martins, L. D., Hallak, R., Alves, R. C., de Almeida, D. S., Squizzato, R., Moreira, C. A., ... Martins, J. A. (2018, 6). Long-range transport of aerosols from biomass burning over southeastern south america and their implications on air quality. *Aerosol and Air Quality Research*, *18*, 1734-1745. doi: 10.4209/aaqr.2017.11.0545

Memmesheimer, M., Friese, E., Ebel, A., Jakobs, H., Feldmann, H., Kessler, C., & Piekorz, G. (2004). Long-term simulations of particulate matter in europe on different scales using sequential nesting of a regional model. *International Journal of Environment and Pollution*, *22*, 108. doi: 10.1504/IJEP.2004.005530

Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J., & Clough, S. a. (1997). Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *Journal of Geophysical Research*, *102*, 16663. doi: 10.1029/97JD00237

Monache, L. D., Deng, X., Zhou, Y., & Stull, R. (2006, 3). Ozone ensemble forecasts: 1. a new ensemble design. *Journal of Geophysical Research*, *111*, D05307. doi: 10.1029/2005JD006310

Moreira, G. D. A., da Silva Andrade, I., Cacheffo, A., da Silva Lopes, F. J., Yoshida, A. C., Gomes, A. A., ... Landulfo, E. (2021, 1). Influence of a biomass-burning event in pm2.5 concentration and air quality: A case study in the metropolitan area of são paulo. *Sensors*, *21*, 425. doi: 10.3390/s21020425

Moreira, G. D. A., Guerrero-Rascado, J. L., Benavent-Oltra, J. A., Ortiz-Amezcua, P., Román, R., Bedoya-Velásquez, A. E., ... Alados-Arboledas, L. (2019, 1). Analyzing the turbulent planetary boundary layer by remote sensing systems: The doppler wind lidar, aerosol elastic lidar and microwave radiometer. *Atmospheric Chemistry and Physics*, *19*, 1263-1280. doi: 10.5194/acp-19-1263-2019

Morrison, H., Thompson, G., & Tatarskii, V. (2009, 3). Impact of cloud microphysics on the development of trailing stratiform precipitation in a simulated squall line: Comparison of one- and two-moment schemes. *Monthly Weather Review*, *137*, 991-1007. doi: 10.1175/2008MWR2556.1

Pedruzzi, R., Baek, B. H., Henderson, B. H., Aravanis, N., Pinto, J. A., Araujo, I. B., ... de Almeida Albuquerque, T. T. (2019, 6). Performance evaluation of a photochemical model using different boundary conditions over the urban and industrialized metropolitan area of vitória, brazil. *Environmental Science and Pollution Research*, *26*, 16125-16144. doi: 10.1007/s11356-019-04953-1

Petersen, A. K., Brasseur, G. P., Bouarar, I., Flemming, J., Gauss, M., Jiang, F., ... Zhou, G. (2019, 4). Ensemble forecasts of air quality in eastern china-part 2: Evaluation of the marcopolo-panda prediction system, version 1. *Geoscientific Model Development*, *12*. doi: 10.5194/gmd-12-1241-2019

Powers, J. G., Klemp, J. B., Skamarock, W. C., Davis, C. A., Dudhia, J., Gill, D. O., ... Duda, M. G. (2017). The weather research and forecasting model: Overview, system efforts, and future directions. *Bulletin of the American Meteorological Society*, *98*(8), 1717-1737. doi: 10.1175/BAMS-D-15-00308.1

Riccio, A., Giunta, G., & Galmarini, S. (2007, 12). Seeking for the rational basis of the median model: the optimal combination of multi-model ensemble results. *Atmospheric Chemistry and Physics*, *7*, 6085-6098. doi: 10.5194/acp-7-6085-2007

Rudke, A., Martins, J., de Almeida, D., Martins, L., Beal, A., Hallak, R., ... de A. Albuquerque, T. (2021, 7). How mobility restrictions policy and atmospheric conditions impacted air quality in the state of são paulo during the covid-19 outbreak. *Environmental Research*, *198*, 111255. doi: 10.1016/j.envres.2021.111255

Schuch, D., de Freitas, E. D., Espinosa, S. I., Martins, L. D., Carvalho, V. S. B., Ramin, B. F., ... de Fatima Andrade, M. (2019, 11). 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*, *26*, 31699-31716. doi: 10.1007/s11356-019-06200-z

Solazzo, E., Bianconi, R., Hogrefe, C., Curci, G., Tuccella, P., Alyuz, U., ... Galmarini, S. (2017, 2). Evaluation and error apportionment of an ensemble of atmospheric chemistry transport modeling systems: multivariable temporal and spatial breakdown. *Atmospheric Chemistry and Physics*, *17*, 3001-3054. doi: 10.5194/acp-17-3001-2017

Squizzato, R., Nogueira, T., Martins, L. D., Martins, J. A., Astolfo, R., Machado, C. B., ... de Freitas, E. D. (2021, 12). Beyond megacities: tracking air pollution from urban areas and biomass burning in brazil. *npj Climate and Atmospheric Science*, *4*, 17. doi: 10.1038/s41612-021-00173-y

Thompson, G., & Eidhammer, T. (2014). A study of aerosol impacts on clouds and precipitation development in a large winter cyclone. *Journal of the Atmo-*

spheric Sciences, *71*(10), 3636-3658. doi: 10.1175/JAS-D-13-0305.1

Vautard, R., Schaap, M., Bergström, R., Bessagnet, B., Brandt, J., Builtjes, P., ... Wind, P. (2009, 10). Skill and uncertainty of a regional air quality model ensemble. *Atmospheric Environment*, *43*, 4822-4832. doi: 10.1016/j.atmosenv.2008.09.083

Wiedinmyer, C., Akagi, S. K., Yokelson, R. J., Emmons, L. K., Al-Saadi, J. A., Orlando, J. J., & Soja, A. J. (2011, 7). The fire inventory from ncar (finn): a high resolution global model to estimate the emissions from open burning. *Geoscientific Model Development*, *4*, 625-641. doi: 10.5194/gmd-4-625-2011

Wood, E. C., Canagaratna, M. R., Herndon, S. C., Onasch, T. B., Kolb, C. E., Worsnop, D. R., ... Williams, E. J. (2010, 9). Investigation of the correlation between odd oxygen and secondary organic aerosol in mexico city and houston. *Atmospheric Chemistry and Physics*, *10*, 8947-8968. doi: 10.5194/acp-10-8947-2010

Zaveri, R. A., Easter, R. C., Fast, J. D., & Peters, L. K. (2008, 7). Model for simulating aerosol interactions and chemistry (mosaic). *Journal of Geophysical Research*, *113*, D13204. doi: 10.1029/2007JD008782

Zaveri, R. A., & Peters, L. K. (1999, 12). A new lumped structure photochemical mechanism for large-scale applications. *Journal of Geophysical Research: Atmospheres*, *104*, 30387-30415. doi: 10.1029/1999JD900876

Appendix A Supplemental Material

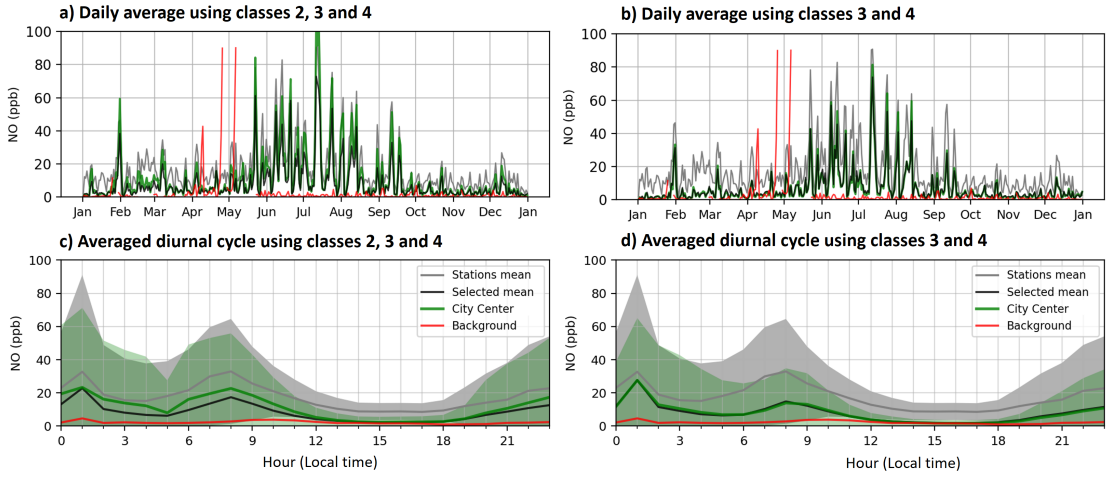

Figure A1. *Time series of the average daily diurnal cycle (top) and of the average daily hourly cycle (bottom) of the NO concentration for the year 2019. The stations are selected according to a classification of their spatial scale of representativeness, 1 being the microscale and 5 being the background. Concentrations are calculated from the average of all the stations ('Stations mean', gray line), from the average of the selected stations from the classification ('selected stations', black line) for classes 2, 3 and 4 (left) and for classes 3 and 4 (right), from an interpolation of the selected stations weighted by the distance between the station and the center of São Paulo ('City center DWI', green line), and for the concentration at the background station ('background', red line). The color shadings (bottom) represent the standard deviation of hourly concentrations over the year.*

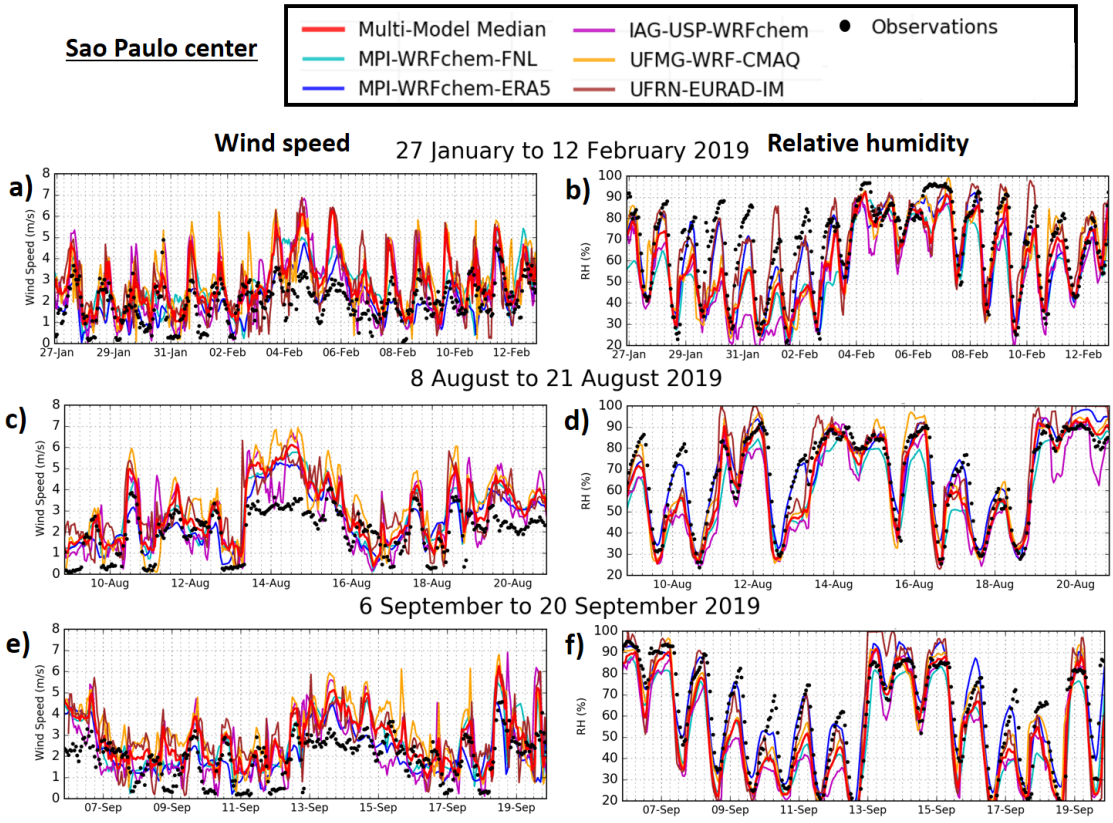

Figure A2. *Time series of hourly relative humidity (RH) and wind speed (WS) observed and modeled in São Paulo for the three selected 15-day periods of the year 2019. The models include data from a regional model ensemble of five simulations (colored lines) with the Multi-Model Median (red line).*

Table 1. *Air quality model setups used by the four institutions.*

Institution - Model	MPI–WRFchem	UFMG–WRF-CMAQ	UFRN–EURAD-IM	IAG-USP–WRFchem
		Domain		
Horizontal resolution	2 km	5 km	5 km	3 km
grid size	120x120	109x109	351×251	166x106
Vertical levels	37 (up to 50 hPa)	41 (up to 100 hPa)	23 (up to 100 hPa)	34
Output frequency	1h	1h	1h	1h
		Emission		
Anthropogenic emissions	CAMS-GLOB-ANT 4.2	EDGAR-HTAP 2.2	EDGAR 4.3.2	LAPAt
Biogenic	MEGAN 2.1	MEGAN 3.1	MEGAN 2.1	MEGAN 2.1
Fires	FINN 1.5	FINN 1.5	GFAS	none
Mineral dust	GOCART	none	GOCART	none
Sea salt	GOCART	Inline	GOCART	none
		Gas and aerosol		
Chemical mechanism	MOZART 4	CB06r2 AERO6	RACM-MIM	CBMZ
Aerosol scheme	GOCART	AERO6	MADE	MOSAIC
Boundary conditions	NCAR-CAMchem	GEOS-Chem 13	ECMWF-CAMS	NCAR-CAMchem
		Meteorology		
Meteo Boundary conditions	NCEP-FNL + ECMWF-ERA5	NCEP-FNL	ECMWF-C-IFS	NCEP-FNL
Surface	Noah	Noah	Noah	Noah
PBL	YSU	Shin-Hong	YSU	YSU
Radiation	RRTMG	RRTMG	RRTMG	RRTMG
Micro-Physics	Thompson	WSM6	Single-Moment 3-class	Morrison 2-mom
Convection	Grell-3D	Grell-Freitas	Grell-3D	Grell-3D
Nudging	only PBL	only PBL	only PBL	only PBL

manuscript submitted to *JGR: Atmospheres*

Table 2. *Names of the air quality monitoring stations corresponding to the metropolitan area of São Paulo with their spatial representativeness classes and coordinates. The distances from each station (s) are given with regard to the city center (CC) located at Catedral da Sé (latitude: -23.5503°, longitude: -46.6339°). The weights of the stations defined by the inverse of the distance to the city center are given for the power factors 1, 2 and 3.*

Station	Num. class	Classification	Latitude (°)	Longitude (°)	d(s,CC) (km)	1/d(s,CC)	$1/d(s,CC)^2$	$1/d(s,CC)^3$
Cerqueira César	1	Microscale	-23.5531	-46.6723	3.85	0.2597	0.0675	0.0175
Congonhas	1	Microscale	-23.6159	-46.6630	7.17	0.1395	0.0195	0.0027
Grajaú - Parelheiros	1	Microscale	-23.7763	-46.6975	23.47	0.0426	0.0018	0.0001
Marginal Tietê	1	Microscale	-23.5187	-46.7434	11.39	0.0878	0.0077	0.0007
Osasco	1	Microscale	-23.5263	-46.7916	15.95	0.0627	0.0039	0.0002
Pinheiros	1	Microscale	-23.5611	-46.7016	6.85	0.1460	0.0213	0.0031
Capão Redondo	2	Neighborhood	-23.6684	-46.7800	18.78	0.0532	0.0028	0.0002
Carapicuíba	2	Neighborhood	-23.5314	-46.8358	20.27	0.0493	0.0024	0.0001
Diadema	2	Neighborhood	-23.6855	-46.6113	13.7	0.0730	0.0053	0.0004
Guarulhos - Pimentas	2	Neighborhood	-23.4401	-46.4099	24.97	0.0400	0.0016	0.0001
Mogi das Cruzes	2	Neighborhood	-23.5267	-46.7921	15.99	0.0625	0.0039	0.0002
Nossa Senhora do Ó	2	Neighborhood	-23.4796	-46.6916	9.13	0.1095	0.0120	0.0013
Parque Dom Pedro II	2	Neighborhood	-23.5449	-46.6277	0.84	1.1905	1.4172	1.6872
Santo Amaro	2	Neighborhood	-23.6546	-46.7096	12.87	0.0777	0.0060	0.0005
Santo André - Capuava	2	Neighborhood	-23.6394	-46.4912	16.82	0.0595	0.0035	0.0002
São Bernardo do Campo	2	Neighborhood	-23.6709	-46.5843	13.04	0.0767	0.0059	0.0005
São Bernardo - Paulicéia	2	Neighborhood	-23.6185	-46.5563	10.33	0.0968	0.0094	0.0009
Cid. Universitária USP	3	Urban	-23.5662	-46.7375	10.47	0.0955	0.0091	0.0009
Guarulhos - Paço	3	Urban	-23.4555	-46.5185	14.94	0.0669	0.0045	0.0003
Ibirapuera	3	Urban	-23.5914	-46.6602	4.87	0.2053	0.0422	0.0087
Interlagos	3	Urban	-23.6805	-46.6750	13.64	0.0733	0.0054	0.0004
Itaim Paulista	3	Urban	-23.5016	-46.4208	21.87	0.0457	0.0021	0.0001
Itaquera	3	Urban	-23.58	-46.4666	16.99	0.0589	0.0035	0.0002
Móoca	4	Medium	-23.5497	-46.6004	3.35	0.2985	0.0891	0.0266
Santana	4	Medium	-23.5056	-46.6285	4.51	0.2217	0.0492	0.0109
Pico do Jaraguá	5	Regional	-23.4561	-46.7663	16.25	0.0615	0.0038	0.0002

manuscript submitted to *JGR: Atmospheres*

Table 3. *Correlation coefficients by variables for the first period (27 January to 12 February 2019) between hourly observations and different model outputs. The Multi-Model Median is calculated from the model outputs with an asterisk (*).*

Variable	Multi-Model Median	MPI* WRFchem-FNL	MPI WRFchem-ERA5	IAG-USP* WRFchem	UFMG* WRF-CMAQ	UFRN* EURAD-IM	NCAR CAMchem	ECMWF CAMS
	First period: 27 January to 12 February 2019							
CO	0.62	0.48	0.65	0.39	0.39	0.33	0.59	0.44
NO_2	0.54	0.45	0.56	0.45	0.61	0.35	0.43	0.38
NO	0.52	0.13	0.22	0.48	0.50	0.31	0.33	0.15
NOx	0.55	0.39	0.55	0.49	0.56	0.34	0.48	0.29
O_3	0.85	0.84	0.77	0.78	0.78	0.80	0.70	0.69
Ox	0.87	0.83	0.77	0.75	0.66	0.82	0.80	0.74
SO_2	0.30	0.17	0.19	0.20	0.27	0.15	0.33	0.20
$PM_{2.5}$	0.51	0.37	0.51	0.55	0.29	0.44	0.58	0.02
PM_{10}	0.37	0.05	0.46	0.29	0.32	0.27	0.69	-0.04

Table 4. *Correlation coefficients by variables for the second period (8 to 21 August 2019) between hourly observations and different model outputs. The Multi-Model Median is calculated from the model outputs with an asterisk (*).*

Variable	Multi-Model Median	MPI* WRFchem-FNL	MPI WRFchem-ERA5	IAG-USP* WRFchem	UFMG* WRF-CMAQ	UFRN* EURAD-IM	NCAR CAMchem	ECMWF CAMS
	Second period: 8 to 21 August 2019							
CO	0.53	0.49	0.59	0.20	0.53	0.64	0.52	0.64
NO_2	0.69	0.67	0.70	0.55	0.61	0.63	0.58	0.74
NO	0.46	0.12	0.27	0.23	0.55	0.46	0.07	0.59
NOx	0.58	0.41	0.52	0.31	0.65	0.61	0.48	0.66
O_3	0.70	0.74	0.72	0.72	0.72	0.46	0.73	0.59
Ox	0.70	0.72	0.75	0.49	0.70	0.61	0.59	0.66
SO_2	0.58	0.32	0.41	0.34	0.71	0.19	0.28	0.53
$PM_{2.5}$	0.45	0.23	0.35	-0.06	0.58	0.31	-0.24	0.08
PM_{10}	0.31	-0.05	0.04	-0.31	0.44	0.12	-0.30	0.06

manuscript submitted to *JGR: Atmospheres*

Table 5. *Correlation coefficients by variables for the third period (6 to 20 September 2019) between hourly observations and different model outputs. The Multi-Model Median is calculated from the model outputs with an asterisk (*).*

Variable	Multi-Model Median	MPI* WRFchem-FNL	MPI WRFchem-ERA5	IAG-USP* WRFchem	UFMG* WRF-CMAQ	UFRN* EURAD-IM	NCAR CAMchem	ECMWF CAMS
	Third period: 6 to 20 September 2019							
CO	0.61	0.65	0.68	0.20	0.57	0.51	0.31	0.46
NO_2	0.61	0.56	0.52	0.48	0.58	0.53	0.58	0.63
NO	0.48	0.17	0.12	0.22	0.65	0.44	0.13	0.30
NOx	0.59	0.51	0.43	0.31	0.72	0.49	0.61	0.46
O_3	0.78	0.78	0.74	0.75	0.72	0.65	0.85	0.69
Ox	0.76	0.67	0.73	0.53	0.67	0.69	0.85	0.53
SO_2	0.49	0.40	0.39	0.39	0.48	0.21	0.56	0.35
$PM_{2.5}$	0.45	0.33	0.37	0.18	0.62	0.20	0.13	0.14
PM_{10}	0.45	0.12	0.34	-0.26	0.57	0.07	0.08	0.26

manuscript submitted to *JGR: Atmospheres*

Table A1. *Station names, their classification number and name, and the correlation coefficient obtained between the measured pollutant concentration and the average of all the sites.*

Station	Num. class	Classification	CO	NOx	NO_2	NO	O3	SO_2	$PM_{2.5}$	PM_{10}
Cerqueira César	1	Microscale	0.82	0.87	0.92	0.83	nan	0.72	nan	0.86
Congonhas	1	Microscale	0.79	0.72	0.72	0.72	nan	0.65	0.87	0.81
Grajaú - Parelheiros	1	Microscale	0.70	nan	nan	nan	0.88	nan	0.75	0.81
Marginal Tietê	1	Microscale	0.88	0.84	0.84	0.82	nan	0.65	0.80	0.87
Osasco	1	Microscale	0.80	0.82	0.83	0.80	nan	0.64	0.83	0.74
Pinheiros	1	Microscale	0.93	0.88	0.90	0.87	0.95	nan	0.86	0.86
Capão Redondo	2	Neighborhood	nan	nan	nan	nan	0.94	nan	nan	0.85
Carapicuíba	2	Neighborhood	0.88	nan	nan	nan	0.94	nan	nan	0.83
Diadema	2	Neighborhood	nan	nan	nan	nan	0.95	nan	nan	0.69
Guarulhos - Pimentas	2	Neighborhood	0.78	0.79	0.81	0.74	0.93	0.61	0.73	0.77
Mogi das Cruzes	2	Neighborhood	nan	nan	nan	nan	nan	nan	nan	nan
Nossa Senhora do Ó	2	Neighborhood	nan	nan	nan	nan	0.94	nan	nan	0.77
Parque Dom Pedro II	2	Neighborhood	0.89	0.91	0.94	0.89	0.97	nan	0.86	0.89
Santo Amaro	2	Neighborhood	0.82	nan	nan	nan	0.95	nan	nan	0.87
Santo André - Capuava	2	Neighborhood	nan	0.82	0.89	0.74	0.92	0.64	0.76	0.77
São Bernardo do Campo	2	Neighborhood	0.87	0.82	0.88	0.71	0.93	nan	0.74	nan
São Bernardo - Paulicéia	2	Neighborhood	nan	nan	nan	nan	nan	nan	nan	0.81
Cid. Universitária USP	3	Urban	nan	nan	nan	nan	0.96	nan	0.83	nan
Guarulhos - Paço	3	Urban	nan	nan	nan	nan	0.93	nan	0.82	0.85
Ibirapuera	3	Urban	0.87	0.85	0.90	0.78	0.98	nan	0.82	nan
Interlagos	3	Urban	nan	0.88	0.89	0.82	0.95	0.62	nan	0.83
Itaim Paulista	3	Urban	nan	0.84	0.84	0.78	0.93	nan	0.81	0.85
Itaquera	3	Urban	nan	nan	nan	nan	0.94	nan	nan	nan
Móoca	4	Medium	0.87	nan	nan	nan	0.96	nan	0.89	nan
Santana	4	Medium	nan	nan	nan	nan	0.96	nan	0.78	nan
Pico do Jaraguá	5	Regional	nan	0.07	0.31	0.04	0.74	nan	0.32	nan

manuscript submitted to *JGR: Atmospheres*

Table A2. *Root mean square error by variable between hourly observations and different model outputs for the three studied periods. The Multi-Model median is calculated from the model outputs with an asterisk (*). The observation mean is given for each variable and period.*

Variable	Obs. mean	Multi-Model Median	MPI* WRFc-FNL	MPI WRFc-ERA5	IAG-USP* WRFc	UFMG* WRF-CMAQ	UFRN* EURADim	NCAR CAMchem	ECMWF CAMS
		First period: 27 January to 12 February 2019							
CO (ppmv)	0.21	0.25	0.33	0.26	0.61	0.28	0.29	0.28	0.35
NO_2 (ppbv)	20.43	15.61	21.85	24.80	27.27	26.08	14.12	14.92	13.67
NO (ppbv)	25.96	16.58	17.25	41.45	70.98	37.25	12.30	16.11	6.43
NOx (ppbv)	43.08	25.05	28.36	59.29	89.59	57.72	26.09	27.39	21.00
O_3 (ppbv)	15.47	43.00	52.78	18.73	18.49	23.70	27.43	17.58	25.15
Ox (ppbv)	22.36	46.06	59.47	17.92	23.30	31.48	19.64	15.40	39.14
SO_2 (ppbv)	9.02	15.95	23.19	2.08	18.48	6.64	2.14	9.45	0.85
$PM_{2.5}$ ($\mu g.m^{-3}$)	12.48	17.45	26.09	10.16	100.69	12.74	8.38	41.39	14.50
PM_{10} ($\mu g.m^{-3}$)	18.27	26.76	30.77	19.68	137.95	22.32	15.48	39.40	24.40
		Second period: 8 to 21 August 2019							
CO (ppmv)	0.37	0.40	0.40	0.56	0.41	0.34	0.55	1.15	0.49
NO_2 (ppbv)	20.05	20.73	25.23	29.37	20.54	20.79	12.43	35.61	16.59
NO (ppbv)	42.45	34.35	32.68	85.64	75.69	41.62	41.43	62.32	14.62
NOx (ppbv)	55.55	39.79	39.02	106.31	88.04	54.28	50.42	89.44	32.32
O_3 (ppbv)	13.00	16.83	17.21	12.79	13.29	25.48	19.06	22.57	18.38
Ox (ppbv)	17.94	27.30	28.91	21.86	12.13	28.65	13.77	40.15	35.31
SO_2 (ppbv)	11.17	21.20	25.71	4.00	22.97	8.17	2.00	59.09	0.77
$PM_{2.5}$ ($\mu g.m^{-3}$)	17.59	37.44	39.60	15.76	49.38	16.89	18.85	44.06	17.54
PM_{10} ($\mu g.m^{-3}$)	30.81	66.29	65.94	37.68	55.91	37.18	42.28	46.74	35.43
		Third period: 6 to 20 September 2019							
CO (ppmv)	0.27	0.32	0.40	0.51	0.36	0.28	0.37	1.48	0.46
NO_2 (ppbv)	24.50	23.87	29.55	31.99	27.43	25.46	14.52	53.81	18.12
NO (ppbv)	35.60	16.62	16.67	85.62	75.78	37.12	16.95	76.48	8.76
NOx (ppbv)	54.26	28.12	32.81	108.53	94.64	57.28	29.42	122.09	27.99
O_3 (ppbv)	17.03	22.99	29.03	18.36	17.76	30.35	22.01	44.20	25.71
Ox (ppbv)	22.41	32.84	42.92	24.21	20.94	33.95	14.61	69.83	44.36
SO_2 (ppbv)	11.64	26.57	28.11	3.66	22.33	7.53	1.66	61.96	0.67
$PM_{2.5}$ ($\mu g.m^{-3}$)	19.59	43.39	46.32	17.17	55.66	17.47	20.49	55.93	24.26
PM_{10} ($\mu g.m^{-3}$)	29.47	54.14	51.86	39.59	62.53	38.22	41.46	53.17	45.49

manuscript submitted to *JGR: Atmospheres*

Table A3. *Mean bias by variable between hourly observations and different model outputs for the three studied periods. The Multi-Model Median is calculated from the model outputs with an asterisk (*). The observation mean is given for each variable and period.*

Variable	Obs. mean	Multi-Model Median	MPI* WRFc-FNL	MPI WRFc-ERA5	IAG-USP* WRFc	UFMG* WRF-CMAQ	UFRN* EURADim	NCAR CAMchem	ECMWF CAMS
		First period: 27 January to 12 February 2019							
CO (ppmv)	0.35	0.08	0.10	0.23	0.02	0.22	0.11	-0.19	-0.02
NO_2 (ppbv)	13.67	16.62	8.25	14.84	20.40	21.89	20.04	-7.19	1.36
NO (ppbv)	6.43	15.40	-3.21	-0.06	23.73	41.00	20.02	-4.56	-2.55
NOx (ppbv)	21.00	32.26	4.14	13.87	43.23	61.99	39.16	-12.69	-2.11
O_3 (ppbv)	25.15	-1.70	25.25	28.54	-11.81	-9.76	4.85	19.30	1.79
Ox (ppbv)	39.14	18.10	34.14	44.28	8.52	12.22	25.39	12.11	3.15
SO_2 (ppbv)	0.85	8.03	13.71	19.35	1.32	13.45	5.88	0.65	6.96
$PM_{2.5}$ ($\mu g.m^{-3}$)	14.50	8.89	13.80	21.42	-7.10	54.65	6.28	1.50	24.09
PM_{10} ($\mu g.m^{-3}$)	24.40	9.84	16.92	24.89	-13.03	67.34	9.83	7.06	14.74
		Second period: 8 to 21 August 2019							
CO (ppmv)	0.49	0.00	0.07	0.15	0.04	0.01	-0.01	-0.34	0.68
NO_2 (ppbv)	16.59	17.11	14.61	18.83	25.23	16.94	15.89	-5.74	23.54
NO (ppbv)	14.62	19.80	-9.59	-6.95	40.82	43.25	21.49	-15.42	27.71
NOx (ppbv)	32.32	36.23	3.90	10.76	64.93	59.06	36.27	-22.43	50.06
O_3 (ppbv)	18.38	-5.67	8.29	5.41	-8.20	-8.35	-4.01	14.75	8.15
Ox (ppbv)	35.31	14.44	23.10	24.56	16.79	8.48	12.14	9.01	31.68
SO_2 (ppbv)	0.77	9.53	17.94	21.45	2.99	16.74	6.57	1.87	44.93
$PM_{2.5}$ ($\mu g.m^{-3}$)	17.54	11.28	27.76	31.18	-4.11	35.33	2.97	6.71	22.86
PM_{10} ($\mu g.m^{-3}$)	35.43	6.21	39.55	44.18	-12.59	31.02	-3.20	12.53	4.93
		Third period: 6 to 20 September 2019							
CO (ppmv)	0.46	0.08	0.17	0.27	0.11	0.12	0.05	-0.25	0.88
NO_2 (ppbv)	18.12	19.78	14.70	20.00	26.46	21.66	19.34	-8.48	32.53
NO (ppbv)	8.76	19.83	-4.71	-4.42	44.31	43.69	22.13	-8.03	33.32
NOx (ppbv)	27.99	38.83	8.87	14.45	69.64	64.23	40.35	-17.70	64.75
O_3 (ppbv)	25.71	-3.79	11.77	15.88	-11.32	-7.07	-1.21	17.37	23.63
Ox (ppbv)	44.36	17.83	26.62	36.15	14.83	14.54	18.61	8.89	56.16
SO_2 (ppbv)	0.67	9.88	23.48	24.21	2.73	15.96	6.37	1.52	43.12
$PM_{2.5}$ ($\mu g.m^{-3}$)	24.26	11.00	38.10	40.23	-5.57	37.12	-3.32	8.28	21.35
PM_{10} ($\mu g.m^{-3}$)	45.49	1.70	37.43	38.36	-15.82	32.10	-12.66	19.31	0.55

manuscript submitted to *JGR: Atmospheres*

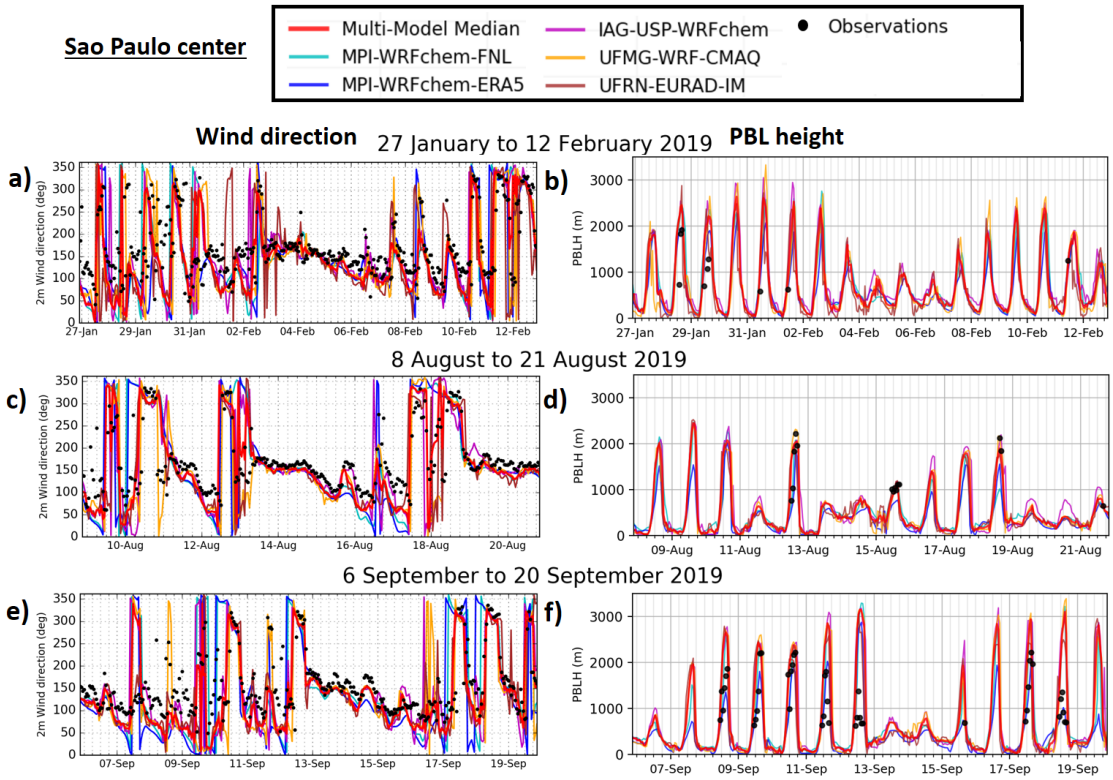

Figure A3. *Time series of hourly wind direction (degree) and PBL height (m) observed and modeled in São Paulo for the three selected 15-day periods of the year 2019. The models include data from a regional model ensemble of five simulations (colored lines) with the Multi-Model Median (red line).*

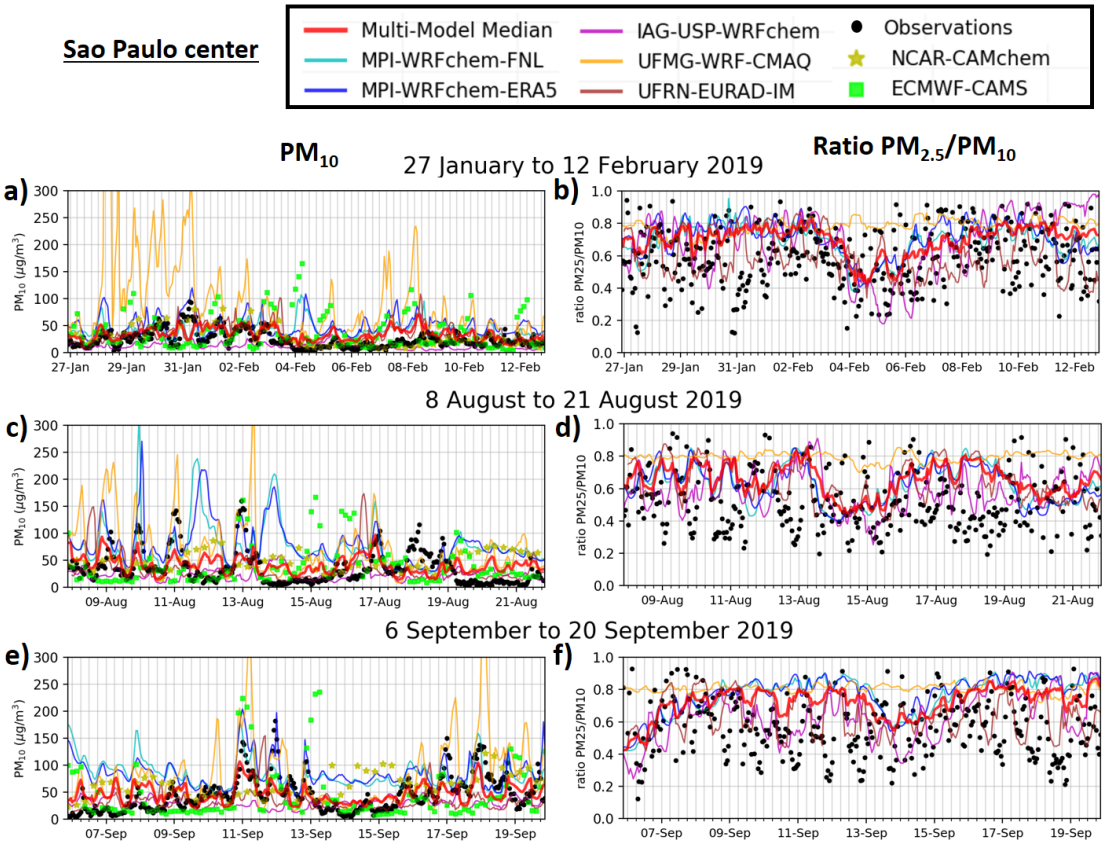

Figure A4. *Time series of hourly concentrations of PM_{10} (a, c and e) and $PM_{2.5}/PM_{10}$ (b, d and f) observed and modeled in Campinas for the three selected 15-day periods of the year 2019. The models include data from the two global forecasts (yellow stars and green squares) and a regional model ensemble of five simulations (colored lines) with the Multi-Model Median (red line). $PM_{2.5}/PM_{10}$ ratios are not presented for the global forecasts.*

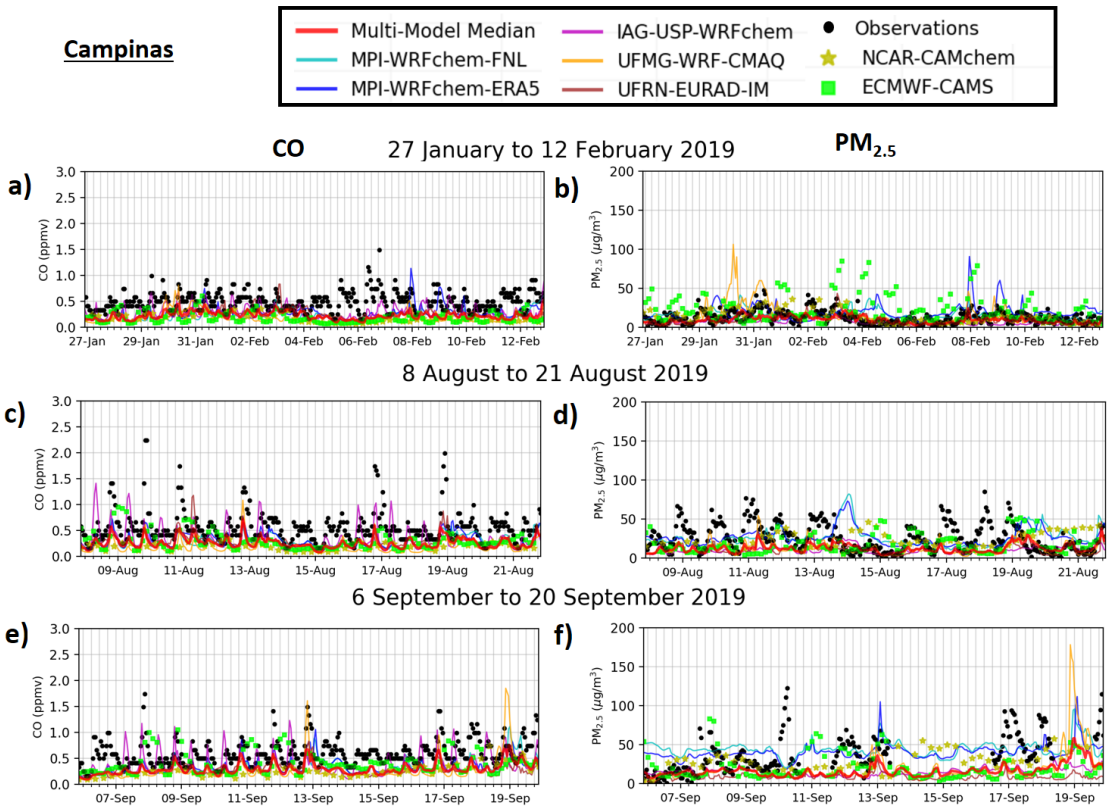

Figure A5. *Time series of hourly concentrations of CO (a, c and e) and $PM_{2.5}$ (b, d and f) observed and modeled in Campinas for the three selected 15-day periods of the year 2019. The models include data from two global forecasts (yellow stars and green squares) and a regional model ensemble of five simulations (colored lines) with the Multi-Model Median (red line).*

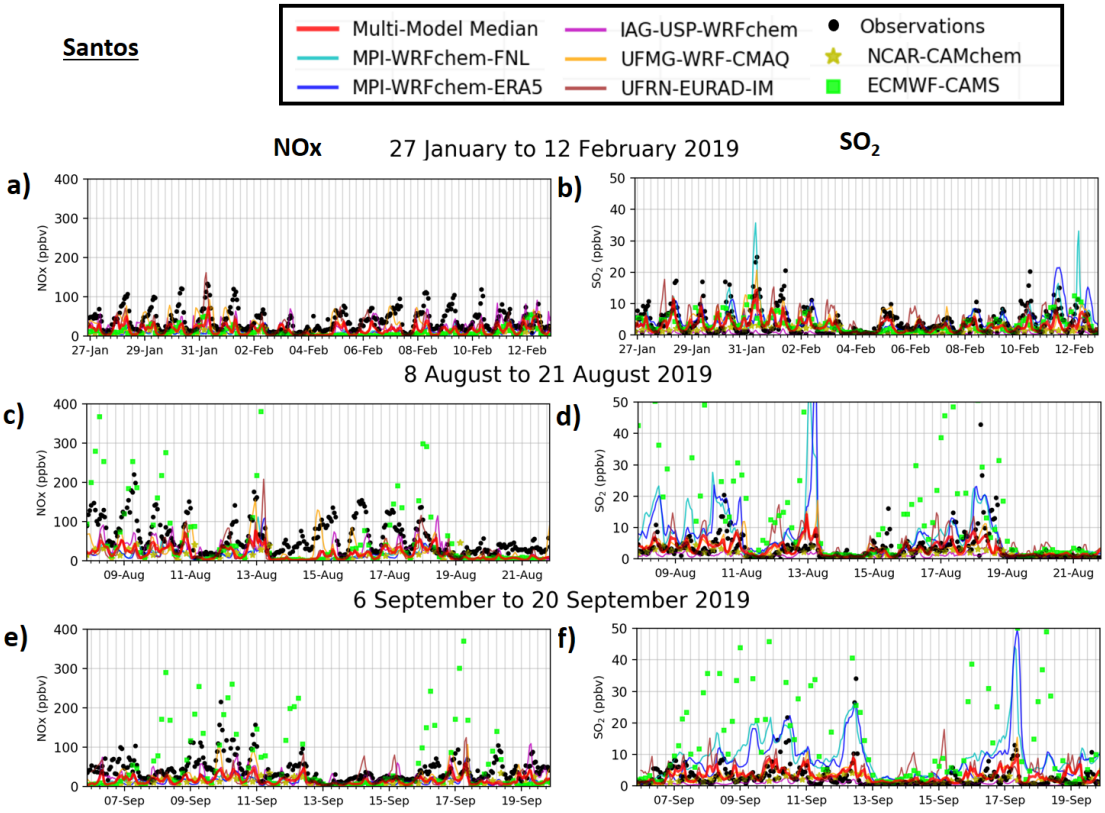

Figure A6. *Time series of hourly concentrations of NOx (a, c and e) and SO_2 (b, d and f) observed and modeled in Santos for the three selected 15-day periods of the year 2019. The models include data from two global forecasts (yellow stars and green squares) and a regional model ensemble of five simulations (colored lines) with the Multi-Model Median (red line).*

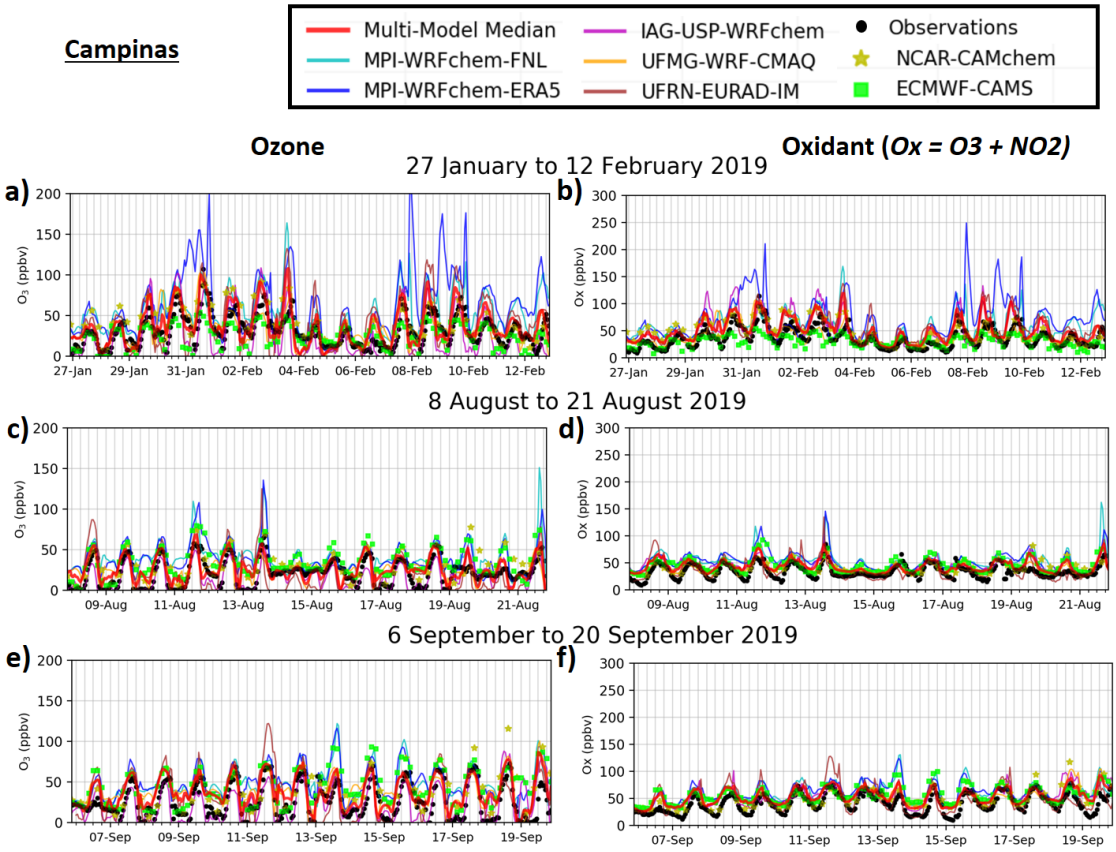

Figure A7. *Time series of hourly concentrations of ozone (a, c and e) and oxidant (b, d and f) observed and modeled in Campinas for the three selected 15-day periods of the year 2019. The models include data from two global forecasts (yellow stars and green squares) and a regional model ensemble of five simulations (colored lines) with the Multi-Model Median (red line).*

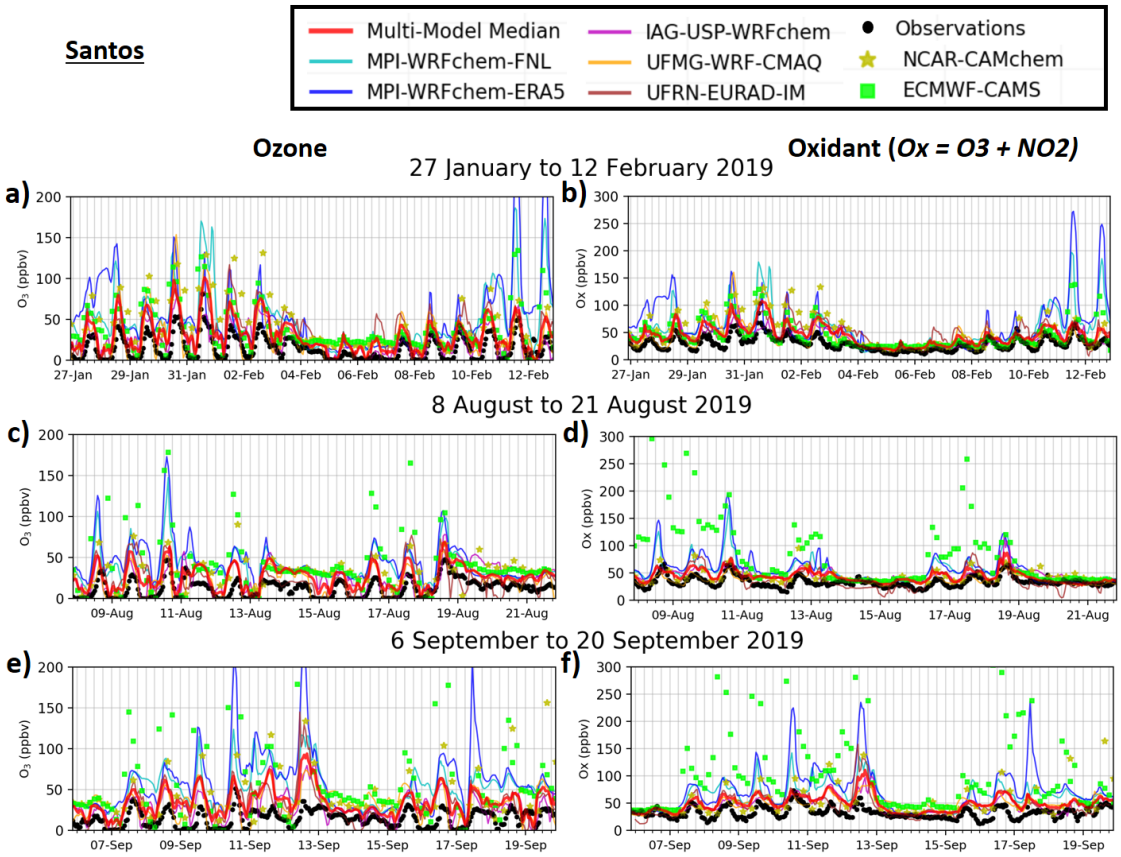

Figure A8. *Time series of hourly concentrations of ozone (a, c and e) and oxidant (b, d and f) observed and modeled in Santos for the three selected 15-day periods of the year 2019. The models include data from two global forecasts (yellow stars and green squares) and a regional model ensemble of five simulations (colored lines) with the Multi-Model Median (red line).*

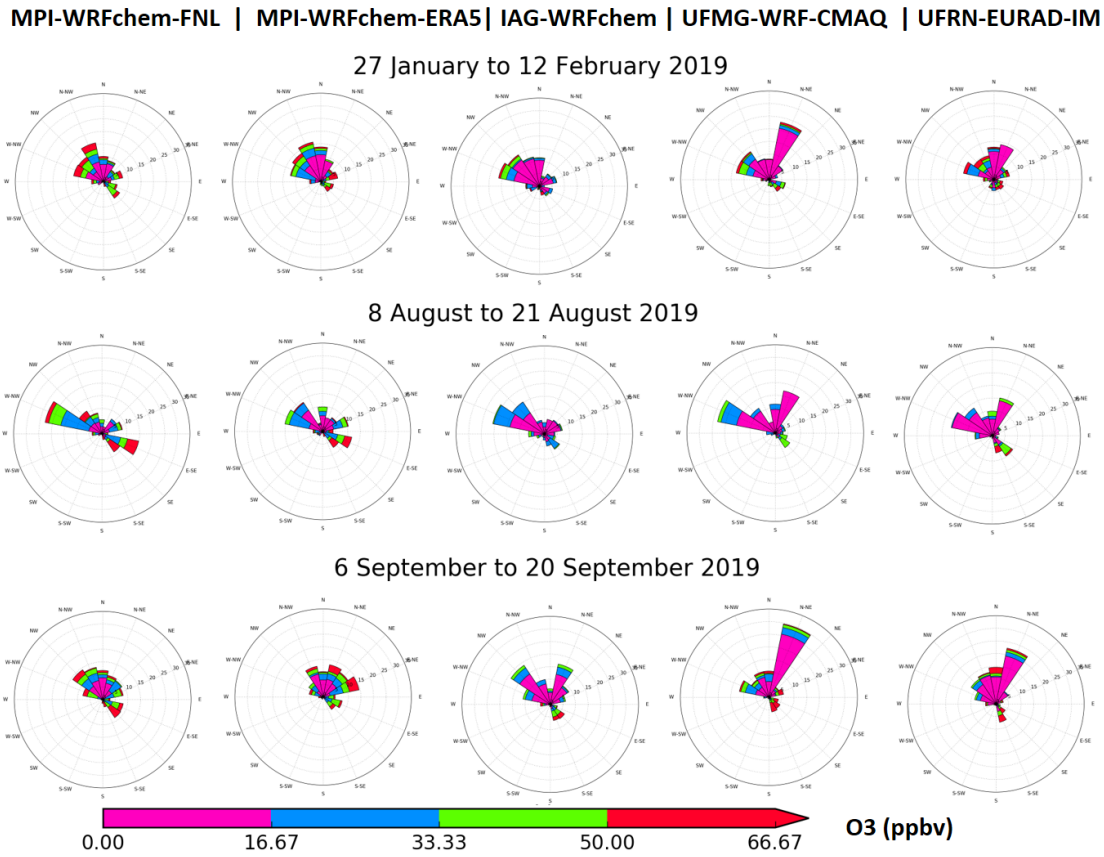

Figure A9. *Pollution roses obtained from the hourly occurrence of the observed and modeled wind direction by direction sector (in %) using 16 sectors, for the three selected 15-day periods of the year 2019. Each pollution rose shows the predominant direction of the pollution transport. For each wind direction sector, the distribution of O3 concentrations is given separated into four concentration ranges (color code).*

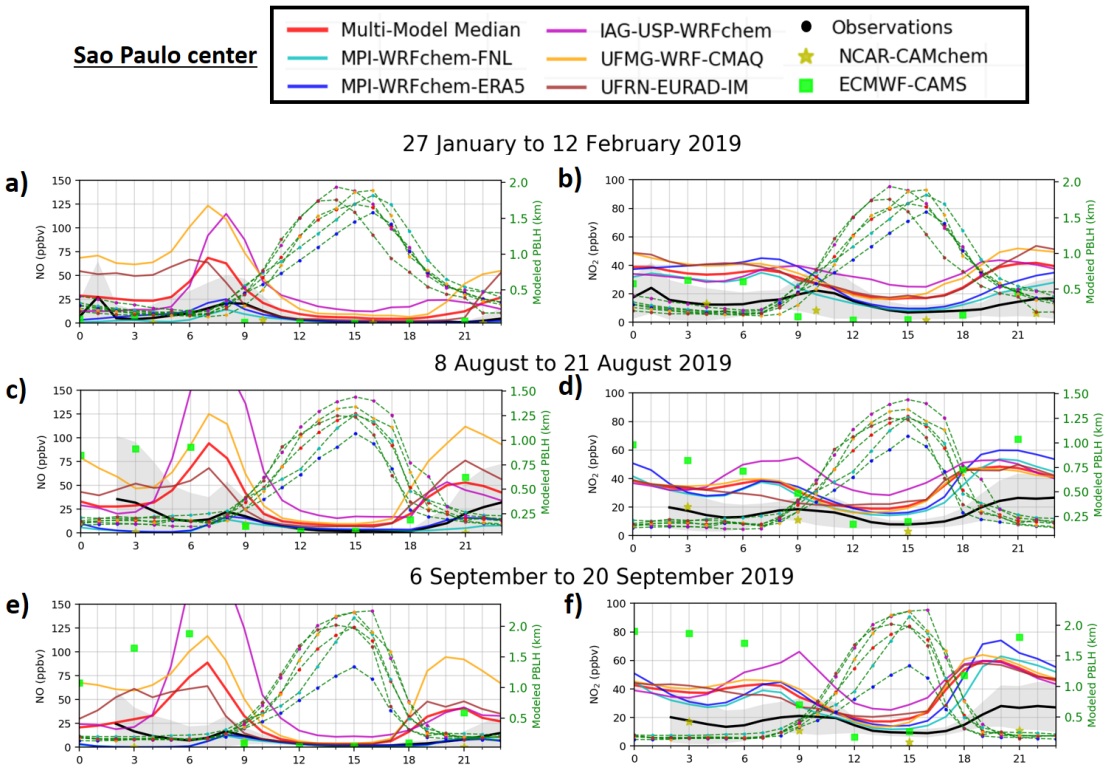

Figure A10. *Average diurnal cycles of hourly concentrations of NO (a,c and e) and NO_2 (b,d and f) observed and modeled in São Paulo over the three selected 15-day periods of the year 2019. The models include data from two global forecasts (yellow stars and green squares) and a regional model ensemble of five simulations (colored lines) with the Multi-Model Median (red line). The modeled planetary boundary layer heights (PBLH) are the green dashed lines with colored dots corresponding to the models. The gray shadings correspond to the standard deviation of the observed hourly data.*

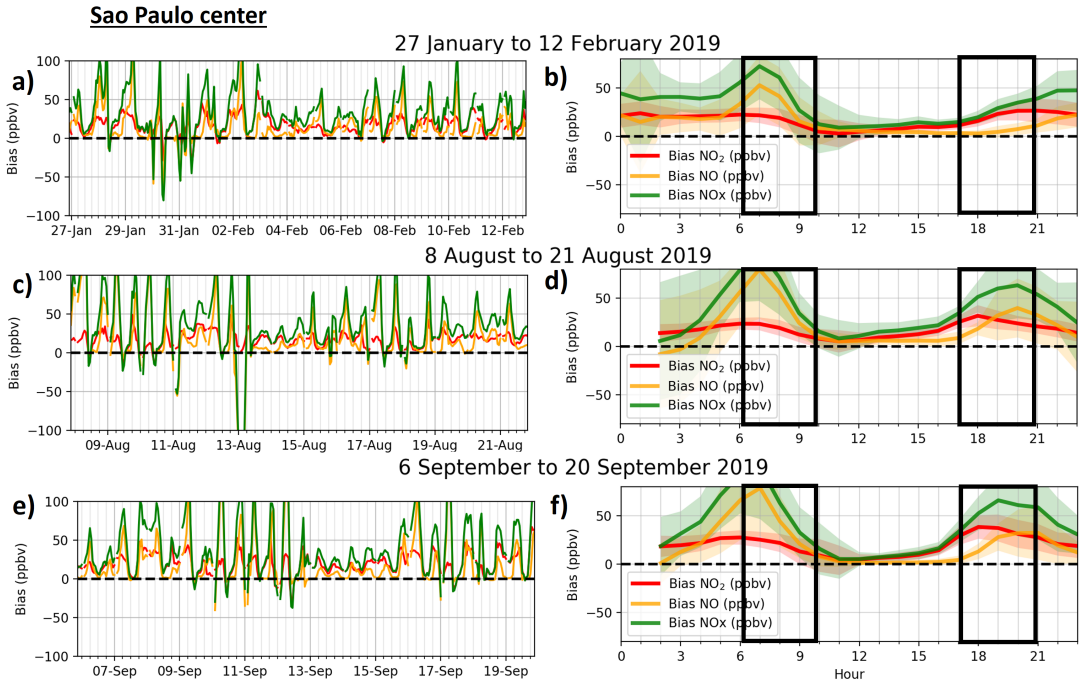

Figure A11. *Time series of hourly bias (difference in modeled and observed concentration) of the Multi-Model Median for NO, NO_2 and NOx (a, c and e) and their associated average diurnal cycles (b, d and f) in São Paulo for the three selected 15-day periods of the year 2019. The Multi-Model Median is calculated from a regional model ensemble of four simulations. The black boxes mark the morning and evening hours.*