JGR Atmospheres

Key Points:

- An ensemble of regional air quality models performs well in the São Paulo megacity for the main regulated pollutants (CO, NOx, O_3, SO_2, $PM_{2.5}$, and PM_{10})
- Transport of pollutants due to biomass burning events, affecting strongly the air quality of the megacity, is represented with high variability by the ensemble
- In the center of the megacity, the median of the regional model ensemble leads to the best performance for these pollutants compared to each model that composes it

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

A. Deroubaix,
Adrien.Deroubaix@mpimet.mpg.de

Author Contributions:

Formal analysis: Adrien Deroubaix
Funding acquisition: Guy Brasseur

Intercomparison of Air Quality Models in a Megacity: Toward an Operational Ensemble Forecasting System for São Paulo

Adrien Deroubaix[1,2] ⓘ, 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[1,9], Lya Lugon[1,10] ⓘ, Leila D. Martins[11], Gregori de Arruda Moreira[4], Rizzieri Pedruzzi[5] ⓘ, Nilton Rosario[12] ⓘ, and Guy Brasseur[1,13] ⓘ

[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, Natal, Brazil, [4]Instituto de Astronomia, Geofísica e Ciências Atmosféricas, University of São Paulo, 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, 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]Now at National Commission of Atomic Energy and National Council of Science and Technology, Buenos Aires, Argentina, [10]Now at CEREA, Joint Laboratory École des Ponts ParisTech/EDF R&D, Université Paris-Est, Paris, France, [11]Federal University of Technology, Londrina, Brazil, [12]Environmental Sciences Department, Federal University of São Paulo, Diadema, Brazil, [13]National Center for Atmospheric Research, Boulder, Colorado, USA

Abstract An intercomparison of four regional air quality models is performed in the tropical megacity of São Paulo with the perspective of developing a forecasting system based on a model ensemble. Modeled concentrations of the main regulated pollutants are compared with combined observations in the megacity center, after analyzing the spatial scale of representativeness of air monitoring stations. During three contrasting periods characterized by different types of pollution events, the hourly concentrations of carbon monoxide (CO), nitrogen oxides (NOx), sulfur dioxide (SO_2), and particulate matter ($PM_{2.5}$ and PM_{10}) modeled by the ensemble are in moderate agreement with observations. The median of the ensemble provides the best performance ($R \approx 0.7$ for CO, 0.7 for NOx, 0.5 for SO_2, 0.5 for $PM_{2.5}$, and 0.4 for PM_{10}) because each model has periods and pollutants for which it has the best agreement. NOx concentration is modeled with a large inter-model variability, highlighting potential for improvement of anthropogenic emissions. Pollutants transported by biomass burning events strongly affect the air quality in São Paulo and are associated with significant inter-model variability. Modeled hourly concentration of ozone (O_3) is overestimated during the day (≈ 20 ppb) and underestimated at night (≈ 10 ppb), while nitrogen dioxide (NO_2) is overestimated at night (≈ 20 ppb). The observed O_3 concentration is best reproduced by the median of the ensemble ($R \approx 0.8$), taking advantage of the variable performance of the models. Therefore, an operational air quality forecast system based on a regional model ensemble is promising for São Paulo.

Plain Language Summary Forecasting air quality in megacities is particularly difficult due to the diversity and temporal variability of emission sources. São Paulo is the largest metropolitan area in South America and has no operational air quality forecast. We perform an intercomparison of four regional air quality models with the perspective of developing an air quality forecasting system. During three contrasting periods characterized by different types of pollution events, we analyze the modeled concentrations of the main regulated pollutants (trace gases and aerosols) compared to observations from the São Paulo air quality monitoring network. The modeled concentrations of the main regulated pollutants agree well with the observations range of variation, although we show the potential in improving the treatment of anthropogenic emissions. In addition, the long-range transport of pollutants due to forest fires strongly affects the air quality in São Paulo and also reduces the performance of the models. The observed hourly ozone concentration is well reproduced by the models, and its median has the best performance, taking advantage of the capabilities of each model. Therefore, an operational air quality forecasting system for the megacity of São Paulo is promising.

Investigation: Adrien Deroubaix, Judith J. Hoelzemann, Rita Yuri Ynoue, Taciana Toledo de Almeida Albuquerque, Rafaela Cruz Alves, Maria de Fatima Andrade, Willian Lemker Andreão, Idir Bouarar, Ediclê de Souza Fernandes Duarte, Hendrik Elbern, Philipp Franke, Anne Caroline Lange, Pablo Lichtig, Lya Lugon, Leila D. Martins, Gregori de Arruda Moreira, Rizzieri Pedruzzi, Nilton Rosario, Guy Brasseur
Resources: Idir Bouarar
Visualization: Adrien Deroubaix
Writing – original draft: Adrien Deroubaix
Writing – review & editing: Adrien Deroubaix, Judith J. Hoelzemann, Rita Yuri Ynoue, Taciana Toledo de Almeida Albuquerque, Rafaela Cruz Alves, Maria de Fatima Andrade, Willian Lemker Andreão, Idir Bouarar, Ediclê de Souza Fernandes Duarte, Hendrik Elbern, Philipp Franke, Anne Caroline Lange, Pablo Lichtig, Lya Lugon, Leila D. Martins, Gregori de Arruda Moreira, Rizzieri Pedruzzi, Nilton Rosario, Guy Brasseur

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). Although global air quality forecasts are now available, such as those produced by the ECMWF (European Center for Medium-Range Weather Forecasts) (Peuch et al., 2022) and the NCAR (National Center for Atmospheric Research) (Marsh et al., 2013), the spatial resolution of these forecasts is coarse compared to the size of a megacity (Baklanov & Zhang, 2020).

São Paulo is by far the largest metropolitan area in South America, one of the largest megacities in 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 a special megacity in different respects for its geography and climate and also for vehicle emissions, because there is an important use of biofuels (Brito et al., 2018), therefore vehicle specific emission factors (Andrade et al., 2015). The proportion of secondary organic aerosols and black carbon in the $PM_{2.5}$ concentration is high due to the composition of the fuels used (Albuquerque et al., 2019). In addition, the air quality of the metropolitan area is often affected by the transport of biomass burning pollutants from remote areas (e.g., Duarte et al., 2021; Martins et al., 2018; Moreira et al., 2021; Pereira et al., 2021; Squizzato et al., 2021). Despite the emission reduction measures implemented since the 1970s, the air quality in São Paulo is neither good for $PM_{2.5}$, with about 25 exceedances per year of the WHO air quality standard (guidelines used before 2021), nor for O_3, with about 100 exceedances per year (Andrade et al., 2017), and without decreasing trends (e.g., Carvalho et al., 2015; Chiquetto et al., 2020; Schuch et al., 2019).

High-resolution modeling combining meteorology and air quality is needed to reproduce the diurnal evolution of atmospheric composition in the boundary layer of megacities (Grell & Baklanov, 2011). Previous studies have shown that where strong NOx sources are located, such as in a megacity, the O_3 concentration has a very large variation between day and night (hereafter referred to as diurnal cycle) (e.g., Monks et al., 2015; Parrish et al., 2016). In highly urbanized areas, O_3 production is mostly controlled by volatile organic compounds, that is, NOx-saturated (e.g., Monks et al., 2015; Parrish et al., 2016). In addition, the study of the (simplified) oxidant level (i.e., $Ox = NO_2 + O_3$) is interesting for monitoring O_3 production because NO_2 can act as an O_3 reservoir (Wood et al., 2009). The strong reduction of anthropogenic emissions during the COVID19 lockdowns showed that only highly urbanized areas experienced an increase in O_3 concentration associated with a decrease in NOx concentration of up to 30%, while oxidant levels remained stable (e.g., Deroubaix et al., 2021; Gaubert et al., 2021). Misrepresentation of anthropogenic emissions in air quality models can lead to a bias in modeled oxidant level in megacities (e.g., Khan et al., 2021). However, the observed increases in O_3 concentrations during the COVID19 pandemic in São Paulo and even in Rio de Janeiro (e.g., Beringui et al., 2022; Nakada & Urban, 2020), are not reproduced by the global simulations performed by Gaubert et al. (2021).

Ensembles of regional air quality models have been developed initially 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 are due to the representation of the dynamics in the planetary boundary layer (PBL), but also to inaccurate emissions and boundary conditions (Im et al., 2015; Solazzo et al., 2017). For the forecasting of air quality in megacities, the use of an ensemble of regional air quality models has two main interests: first, the inter-model range is an indicator of the uncertainty of the state-of-the-art modeling (Vautard et al., 2009), and second, its median generally yields better performance than any single model (Riccio et al., 2007). A comprehensive monitoring network is needed to evaluate the model results. Compared to European and North American air quality monitoring networks, the Brazilian one stands out because only a few states in the country have implemented air quality monitoring (Andrade et al., 2017). São Paulo has implemented the most important monitoring network in Brazil, maintained by CETESB (*Companhia Ambiental do Estado de São Paulo*), which includes 26 monitoring stations in the metropolitan area and another 63 stations throughout the state. This network is excellent for evaluating a regional air quality model ensemble in the metropolitan area, as it is well distributed spatially in the most urbanized areas of the megacity (Andrade et al., 2017).

A megacity such as São Paulo is a challenge for regional air quality models because they must be applied at a sufficiently fine resolution to represent the processes that lead to the high diurnal variability of the concentrations of the main pollutants, which also results from the interaction between the urban heat island and the

sea-breeze circulations (Andrade et al., 2015; Freitas et al., 2007). Operational air quality forecasts based on model ensembles are available for 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 with expertise in modeling, planning, monitoring and social learning,*" aims to develop such an ensemble forecasting system for South America based on these two previous experiences. As a preliminary step to the development of this system, this article evaluates the performance of state-of-the-art regional air quality models, focusing on the metropolitan area of São Paulo, for which no dedicated air quality forecasting system is currently operational.

The air quality model intercomparison is performed by focusing on the center of São Paulo, supported by the CETESB monitoring network (Section 2). Four chemistry-transport models are involved to form a regional air quality model ensemble at high resolution (Section 3). We assess the strengths and weaknesses of the regional model ensemble for the main regulated pollutants over three contrasting time periods (Section 4). The potential of the regional model ensemble is evaluated for O_3 and $PM_{2.5}$ (Section 5). Finally, the perspectives for the development of an operational air quality forecasting system based on a regional model ensemble for São Paulo are discussed (Section 6).

2. Pollutant Concentrations in the Center of São Paulo

First this section presents the São Paulo air quality monitoring network (Section 2.1), second the influence of the spatial representativeness of the stations used to represent the observed concentrations in the center of the megacity (Section 2.2), third the selection of three contrasting 15-day periods marked by different types of pollution events for the model intercomparison (Section 2.3).

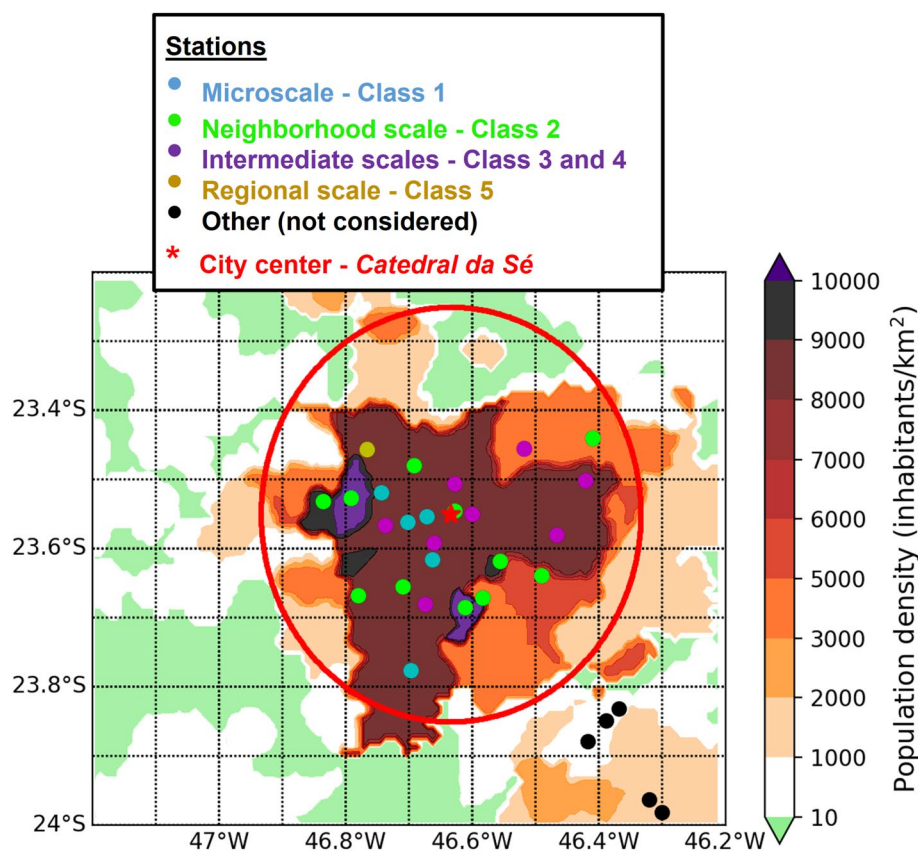### 2.1. Air Quality Monitoring Network in the Megacity

We study the concentrations of the main regulated pollutants, which are carbon monoxide (CO), nitrogen monoxide (NO), nitrogen dioxide (NO_2), nitrogen oxides (NOx), ozone (O_3), sulfur dioxide (SO_2), as well as particles with a diameter less than 2.5 and 10 μm ($PM_{2.5}$ and PM_{10}). The year 2019 was chosen because it represents typical conditions, marked by a weak "El Niño" phenomenon and occurring just before the start of the COVID-19 pandemic. Concentrations are measured by stations in the air quality monitoring network operated by CETESB in the metropolitan area of São Paulo. For stations in the center of São Paulo, all of which are within a 30 km radius of the traditional center located at Catedral da Sé (locations given in Table S1 in Supporting Information S1 and are shown in Figure 1), the CETESB (2022) classification of their spatial scale of representativeness is available, which is composed of five classes: 1—Micro-scale, 2—Neighborhood, 3—Urban, 4—Medium, 5—Regional.

To assess this classification in the context of our study, we calculate correlation coefficients between the hourly concentrations measured at each individual station and the average of all stations throughout 2019 (Table S1 in Supporting Information S1). For the most short-lived pollutants measured near sources, which is NO among the examined pollutants, the correlation coefficients of Micro-scale stations are comparable to those of other representativeness classes. Moreover, there is a high level of agreement for all variables considered (indicated by R above 0.7, and for O_3 R exceeding 0.9). Only the station classified at the Regional scale (representative beyond the megacity) presents a weak agreement with the average of all stations (R systematically lower than 0.4, with the exception of O_3). Consequently, this station is excluded from the study.

The high level of agreement between stations suggests that, given the current air quality monitoring network, an average of stations can adequately represent hourly variations of pollutant concentrations in the center of São Paulo. However, it is important to investigate the influence of Micro-scale stations, and to a lesser extent Neighborhood-class stations, due to their proximity to specific emission sources.

2.2. Spatial Representativeness of the Stations

Various methodologies have been proposed for comparing air quality models with observations (e.g., Wong et al., 2004; Yu et al., 2018). In the preceding section, we employed the straightforward approach of utilizing the average of all stations to assess station classifications. This method, however, does not consider the spatial structure of the monitoring network. In contrast, methods such as Kriging and Inverse Distance Weighting (IDW)

Figure 1. Map of the population density of the center of São Paulo with the locations of the stations of the air quality monitoring network (dots), distinguishing the 26 stations inside the most densely populated area of the megacity (inside the red circle). The color of the stations represents their class of spatial representativeness. The center of São Paulo is located at the Catedral da Sé (red star), which is used as the center of the circle (with a radius of 30 km).

interpolation account for spatial structure. Kriging is suited for regional scales marked by spatial heterogeneity in pollution sources and network structure, while IDW is more appropriate for local scales with homogeneous pollution sources and spatial network heterogeneity (Brasseur & Jacob, 2017). The São Paulo air quality monitoring network comprises 26 measurement stations located within a 30 km radius from the traditional center of the metropolitan area. This coverage encompasses the most densely populated area of the megacity (Figure 1). Consequently, we expect stronger temporal co-variation in concentrations (across all stations) from 1 hour to another compared to spatial variability among stations for a given hour. This expectation arises from the high correlation observed among the monitored pollutants between measurement stations (Table S1 in Supporting Information S1). Therefore, we use the IDW interpolation method to compare observations with the modeled ones, which means that we analyze the temporal variability of air quality in the center of the megacity with a single average.

The IDW interpolation method is based on the distance of stations to a given location, where the weight is the inverse of the distance to this location. In order to focus on the center of the São Paulo metropolitan area (Figure 1), we use the traditional city center (*Catedral da Sé*, latitude: $-23.5503°$, longitude: $-46.6339°$). The concentration at the city center ($Conc_{CC}$) is calculated as follows:

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

where the weights are:

$$w_s = 1/d(s, CC)^p \tag{2}$$
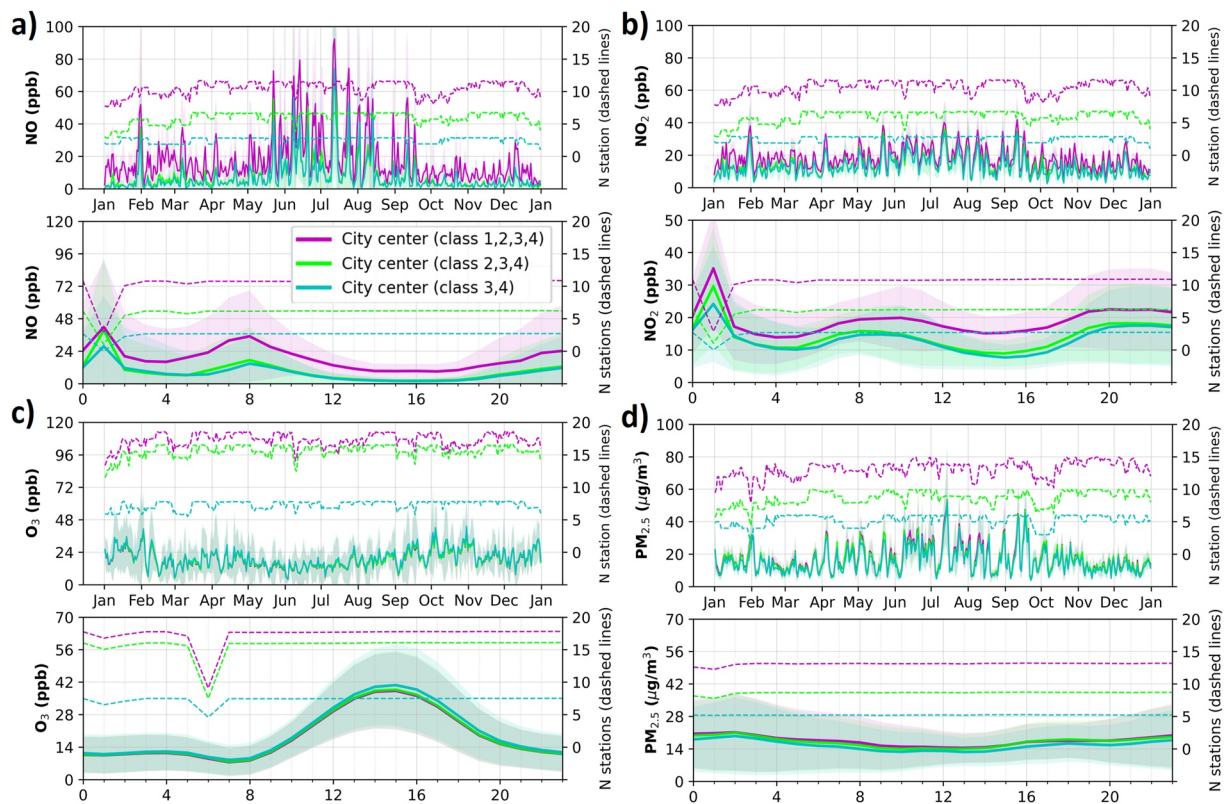
Figure 2. Time series of observed concentrations of daily averages (top) and hourly average diurnal cycles (bottom) for (a) NO, (b) NO_2, (c) O_3 and (d) $PM_{2.5}$ for the year 2019 using local time. The city averages are calculated based on an inverse distance weighting interpolation method with different stations of the São Paulo measurement network using a classification of their spatial scale of representativeness: (i) with all stations (class 1, 2, 3, 4), (ii) without Micro-scale stations (class 2, 3, 4), (iii) without Micro-scale and Neighborhood-scale stations (class 3, 4). The colored shadings represent (top) the standard deviation of daily concentrations and (bottom) the standard deviation of hourly concentrations of the stations over the year. The number of stations available for each average is given on the right axis with the associated color. Hours of the day are given in local time.

$Conc_s$ represents the concentration measured at each station, and p denotes the power factor, influencing the weight of stations (w_s) closest to the city center. Careful consideration must be given to selecting an appropriate p value, which depends on the size of the area studied, as well as avoiding stations too close to the city center (de Mesnard, 2013).

We investigate the range of station weights calculated with p equal to 1 and 2. Given the air quality monitoring network of São Paulo, p equal to 2 (or more) is not an appropriate choice. It leads to overweighting of stations near the city center, while strongly reducing the influence of stations further away. With p equal to 1, the weighting range is less than an order of magnitude, expect for 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. For this reason, the weight of this station has been changed to a weight corresponding to 3 km.

Using the IDW spatial interpolation method within the São Paulo metropolitan area, we investigate the influence of station representativeness classes. We compare city averages derived from stations of different representativeness classes: (a) including all stations (1, 2, 3, 4), (b) excluding Micro-scale stations (class 2, 3, 4), and (c) further excluding Neighborhood-class stations (class 3, 4). We consider both daily averages and hourly average diurnal cycles. The number of stations with available measurements is also examined (Figure 2; Figure S1 in Supporting Information S1).

Comparing the three city averages, we see that the averages lead to the same temporal variability (Figure 2). City averages derived from all stations show higher concentrations for NO, NO_2, and NOx, and to a lesser extent for CO and SO_2, which is due to the inclusion of stations near strong emission sources (Micro-scale stations). By using two city averages with all stations and excluding Micro-scale stations, we can establish a consistent concentration range that is representative of the megacity and suitable for comparison with different model outputs.

For each pollutant, analysis of average diurnal cycles reveals specific hours with reduced available station data, attributed to simultaneous instrument calibration across most stations (Figure 2; Figure S1 in Supporting Information S1). These hours are characterized by values deviating from the average diurnal cycles at (in local time) 05:00 for CO, 01:00 for NO, NO_2, and NOx, and 04:00 for SO_2, resulting in their removal from the study.

2.3. Selection of Three Time Periods

Three 15-day periods are retained during the year 2019 which are associated with a strong deterioration of the air quality. We aim to focus on two contrasting periods containing high O_3 events in winter and summer. In addition, another period is selected concerning pollution due to the long-range transport of burning biomass pollutants (Pereira et al., 2021). The three selected 15-day periods are:

1. From 27 January to 12 February 2019, a period of O_3 episodes, 5 days with O_3 concentration above air quality standard in São Paulo were monitored despite the precipitation occurring during this period.
2. From 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, and associated to precipitation that have formed "black rain."
3. From 6 to 20 September 2019, a period of O_3 and $PM_{2.5}$ episodes, during which the air quality standards for O_3 and $PM_{2.5}$ were exceeded for both pollutants.

In conclusion of this analysis of the measurement network of São Paulo, we have selected three periods and defined city averages to compare the observed pollutant concentrations in the center of the megacity with the outputs of the regional models and the global forecast.

3. Regional Air Quality Model Ensemble

In this section, the different chemistry-transport models employed in the study are presented (Section 3.1), with a particular focus on their configurations, especially regarding anthropogenic emissions (Section 3.2). Additionally, we present our strategy for intercomparing the regional model ensemble (Section 3.3).

3.1. Ensemble Members

In this intercomparison study, we compare a regional air quality model ensemble with global forecasts produced by ECMWF (2023) under the Copernicus Atmosphere Monitoring Service (referred to as ECMWF–CAMS) and forecasts produced by NCAR (Buchholz et al., 2019) using the Community Atmosphere Model with Chemistry (referred to as NCAR–CAMchem). Four institutions are involved in this intercomparison, one in Germany and three in Brazil. All participating institutions were tasked with providing hourly simulation outputs with high spatial resolution (less than 5 km) in configurations suitable for forecasting. Each institution uses its optimized model configuration:

1. The Max Planck Institute for Meteorology (MPI) provides simulations using the Weather Research and Forecasting model (WRF) coupled with chemistry (WRFchem) model version 4.1.2 (hereafter MPI–WRFchem). The WRFchem model integrates a mesoscale non-hydrostatic meteorological model (WRF) coupled online with an atmospheric chemistry and transport model, allowing simultaneous prediction of meteorology and atmospheric composition (Fast et al., 2006; Grell et al., 2005; Powers et al., 2017). The model is set up with meteorological nudging using the NCEP-FNL data (NCEP, 2022) above the PBL for humidity, temperature and wind speed.
2. The Universidade Federal de Minas Gerais (UFMG) provides simulations using the WRF-CMAQ model (hereafter UFMG–WRF-CMAQ). The Community Multiscale Air Quality Modeling System (CMAQ) serves as the atmospheric chemistry and transport model (Byun & Schere, 2006). The WRF model is used in conjunction with the Sparse Matrix Operator Kerner Emissions (SMOKE) model to generate meteorological and emission data (Albuquerque et al., 2019; Pedruzzi et al., 2019). The WRF model is set up with meteorological nudging using the NCEP-FNL data (NCEP, 2022) above the PBL for humidity, temperature and wind speed.
3. The Universidade Federal do Rio Grande do Norte (UFRN), in collaboration with the Rhenish Institute for Environmental Research at the University of Cologne, provides simulations using the EURopean Air Pollution and Dispersion - Inverse Model (EURAD-IM) (hereafter UFRN–EURAD-IM). EURAD-IM is an atmospheric

chemistry and transport model (Elbern et al., 2007; Hass et al., 1995; Memmesheimer et al., 2004). The model uses WRF as an offline meteorological model with meteorological nudging using NCEP-FNL data (NCEP, 2022) above the PBL for humidity, temperature, and wind speed.

4. The Universidade de São Paulo, Instituto de Astronomia, Geofísica e Ciências Atmosféricas (IAG), provides simulations using the WRFchem model version 4.0 (hereafter IAG–WRFchem). Unlike the other three institutions, the IAG uses a local anthropogenic emissions inventory limited to vehicular emissions measurements with the LAPAt model (Andrade et al., 2015), and without meteorological nudging.

To analyze the impact of meteorological inputs, MPI runs two WRFchem simulations with different meteorological initial and boundary conditions: one using the FNL (Final) operational global analysis produced by the Global Data Assimilation System of the US National Centers for Environmental Prediction (NCEP–FNL) (NCEP, 2022), and the other using the ECMWF–ERA5 reanalysis (Hersbach et al., 2020).

The model configurations used by each institution are different due to their choices of emissions inventories, of meteorological and chemical schemes and of spatial domains, which are related to their expertise in the air quality of São Paulo (Table 1). Hourly simulation outputs are bilinearly interpolated to the city center (at the first vertical level of the model). We focus exclusively on the finest resolution domain, which varies for each model. For the regional air quality model ensemble, the model outputs are provided at a frequency of 1 hr, while ECMWF–CAMS offers outputs at 3-hr intervals, and NCAR–CAMchem at 6-hr intervals.

Our strategy for the transition to an operational ensemble forecast system is to analyze the results of the individual models and the median of the regional model ensemble (referred to as the multi-model median, or MMM). The MMM is computed excluding the MPI–WRFchem-ERA5 simulation to ensure equal weighting of simulations from all four institutions. We choose the median rather than the mean to minimize the influence of outliers.

3.2. Similarities and Differences of the Model Configurations

The domains chosen by the four institutions are similar in terms of horizontal and vertical resolution (Table 1). Additionally, they use the WRF model and their meteorological configurations are similar. Anthropogenic emissions are expected to be an important source of inter-model variability due to the difference in the spatial distribution of emissions by sector (Huneeus et al., 2020). Moreover, some participating institutions use temporal or vertical profiles for sector-specific emissions data (Table 1). However, three institutions use anthropogenic emissions inventories, all from the EDGAR group, while IAG–WRFchem uses a local inventory (Andrade et al., 2015). To assess the four inventories in the São Paulo region, we compare annual average of NOx emission fluxes in terms of spatial variability and values in the center of the megacity (Figure 3).

The spatial variability of NOx emission fluxes is significant in the São Paulo region for all inventories by three orders of magnitude (Figure 3). The highest fluxes are concentrated in the metropolitan areas of São Paulo and Rio de Janeiro. Comparing the three inventories provided by the EDGAR group, the NOx emission fluxes are lowest with HTAPv2 and the highest with CAMS-GLOB-ANTv4.2. However, we see that in the most urbanized area, the spatial distributions of the highest fluxes are similar for the three inventories. The Lapat inventory differs substantially, with a greater spatial variability in NOx emission fluxes compared to the other three inventories, with emissions localized along roads. In the most urbanized area of São Paulo, the emissions are exceeding 10^{-9} kg.m^{-2}.s^{-1} (equivalent to approximately 100 kg per km^2 per day). The value in the center of São Paulo is similar for all four inventories, ranging from 1.83 to 3.16×10^{-9} kg.m^{-2}.s^{-1}, showing their consistency.

The Lapat inventory is limited to vehicular emissions measurements and is distributed along roads, which is justified by the prevalence of traffic-related emissions (Andrade et al., 2015, 2017). The absence of fire and biogenic emissions in the IAG–WRFchem simulations is compensated by the fact that the modeled domain is the smallest compared to those used by other institutions, thus leaving more influence to the boundary conditions (Table 1; Figure S2 in Supporting Information S1).

Furthermore, the long-range transport of pollutants from biomass burning significantly impacts air quality in the São Paulo region (Duarte et al., 2021; Martins et al., 2018; Pereira et al., 2021; Squizzato et al., 2021). The inclusion of fire emissions within the domain or via boundary conditions is expected to contribute significantly to inter-model variability in modeled concentrations within the megacity center during biomass burning events. The domains used by the four institutions are displayed alongside the sum of burned areas for the year 2019 obtained

Table 1
*Air Quality Model Setups Used by the Four Institutions*

Institution—model	MPI–WRFchem	UFMG–WRF-CMAQ	UFRN–EURAD-IM	IAG–WRFchem
Gas and aerosol				
Chemical scheme	MOZART4	CB06r2	RACM-MIM	CBMZ
	Emmons et al. (2010)	Inness et al. (2019)	Geiger et al. (2003)	Zaveri and Peters (1999)
Aerosol scheme	GOCART	AERO6	MADE	MOSAIC
	Chin et al. (2002)	Inness et al. (2019)	J. C. Kaiser et al. (2014)	Zaveri et al. (2008)
Chem. boundary cond.	NCAR-CAMchem	GEOS-Chem 13	ECMWF-CAMS	NCAR-CAMchem
Emissions				
Anthropogenic	CAMS-GLOB-ANTv4.2	HTAPv2	EDGARv4.3.2	LAPAt
	Granier et al. (2019)	Janssens-Maenhout et al. (2015)	Crippa et al. (2018)	Andrade et al. (2015)
Anthr. temporal profiles	Crippa et al. (2020)	Andrade et al. (2015)	Crippa et al. (2018)	Andrade et al. (2015)
Anthr. vertical profiles	Mailler et al. (2013)	None	None	None
Biogenic	MEGANv2.1	MEGANv3.1	MEGANv2.1	None
	Guenther et al. (2006)		Guenther et al. (2006)	None
Fires	FINNv1.5	FINNv1.5	GFASv1.2	None
	Wiedinmyer et al. (2011)	Wiedinmyer et al. (2011)	J. W. Kaiser et al. (2012)	None
Meteorology				
Surface scheme	Noah	Noah	Noah	Noah
	Ek et al. (2003)	Ek et al. (2003)	Ek et al. (2003)	Ek et al. (2003)
PBL scheme	Yonsei University (YSU)	Shin-Hong	YSU	YSU
	Hong et al. (2006)	Shin and Hong (2015)	Hong et al. (2006)	Hong et al. (2006)
Radiation scheme	RRTMG	RRTMG	RRTMG	RRTMG
	Mlawer et al. (1997)	Mlawer et al. (1997)	Mlawer et al. (1997)	Mlawer et al. (1997)
Micro-Physics scheme	Thompson	WSM6	WSM3	Morrison 2-mom
	Thompson and Eidhammer (2014)	Hong and Lim (2006)	Hong et al. (2004)	Morrison et al. (2009)
Convection scheme	Grell-3D	Grell-Freitas	Grell-3D	Grell-3D
	Grell and Dévényi (2002)	Grell and Freitas (2014)	Grell and Dévényi (2002)	Grell and Dévényi (2002)
Met. boundary cond.	NCEP-FNL and	NCEP-FNL	ECMWF-C-IFS	NCEP-FNL
	ECMWF-ERA5			
Domain				
Horizontal resolution	50, 10, 2 km	5 km	25 km, 5 km	3 km
Grid sizes (lon x lat)	120 × 120 (for the 3)	109 × 109	110 × 101, 326 × 241	166 × 106
Vertical levels	37 (up to 50 hPa)	41 (up to 100 hPa)	23 (up to 100 hPa)	34
Height first level	45 m	20 m	54 m	20 m

from the MCD64A1.061 product (Giglio et al., 2023). The domains differ for fire emissions, as the coarse domain of MPI–WRFchem simulations covers nearly the entire South America, while UFMG–WRF-CMAQ and IAG-WRFchem simulations focus solely on the São Paulo region (Table 1; Figure S2 in Supporting Information S1). Consequently, MPI–WRFchem simulations encompass the majority of burned areas, while UFMG–WRF-CMAQ and IAG-WRFchem simulation rely on boundary conditions to integrate the long-range transport of pollutants from biomass burning.

3.3. Intercomparison Metrics

The model inter-comparison begins by assessing the temporal correlation coefficient between modeled and observed hourly concentrations in the center of São Paulo. This metric serves as a primary indicator for assessing
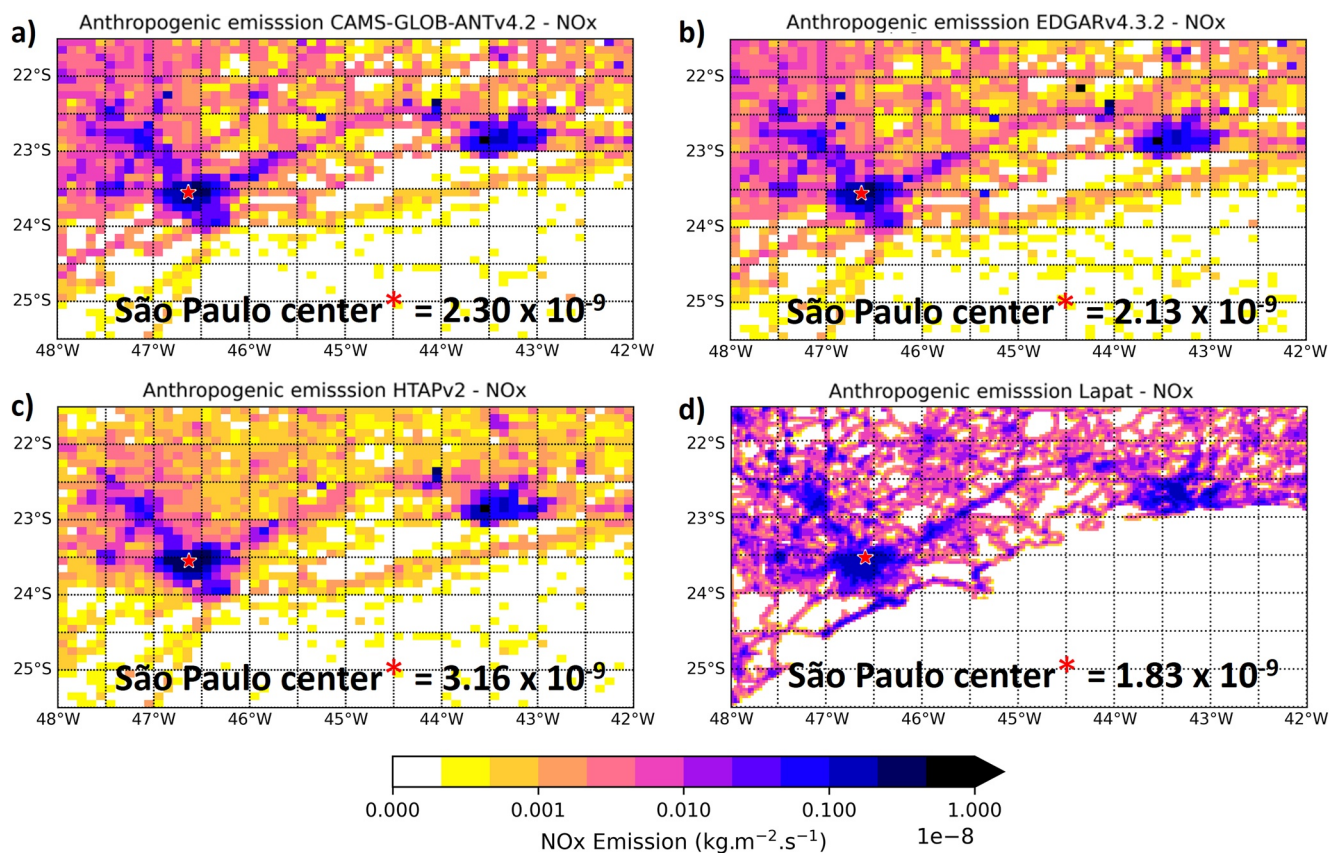
Figure 3. Maps of annual average NOx emission flux (sum of all sectors) from four anthropogenic inventory: (a) CAMS-GLOB-ANTv4.2 (representative of the year 2019), (b) EDGARv4.3.2 (representative of the year 2010) and (c) HTAPv2 (representative of the year 2010), (d) Lapat. The center of São Paulo is located at Catedral da Sé (red star).

the performance of the models. For all pollutants, a very good correlation coefficient is considered to be greater than 0.8, because the amount of predicted variability explained by observed variability (R^2) is greater than 60%. Good agreement lies between 0.8 and 0.7, moderate agreement between 0.7 and 0.5, poor agreement between 0.5 and 0.3, and a correlation coefficient below 0.3 indicates a lack of agreement. However, it is important to note that the main goal of our study is not aimed at benchmarking different air quality models in São Paulo. Rather, it consists in the following steps:

1. Analyzing the observations against the regional simulations provided by the four institutions.
2. Comparing the regional simulations with the MMM.
3. Contrasting the regional simulations and the MMM with the two global forecasts.

To initiate this analysis, we examine the correlation coefficients between modeled and observed hourly pollutant concentrations with the three 15-day periods combined (Table 2). To understand the influence of station selection, we compare the correlation coefficients using city averages with all stations and without Micro-scale stations. The highest correlation coefficients are observed for O_3 in both the regional model ensemble and the global forecasts, while $PM_{2.5}$ and PM_{10} exhibit the lowest coefficients. Inclusion or exclusion of Micro-scale stations has minimal effect on the correlations, except for NO and NO_2. Correlations for NO are stronger with the inclusion of micro-scale stations, conversely correlations for NO_2 are stronger excluding these stations. For all pollutants, the range of correlation coefficients between the different simulations is narrow. Overall, the MMM leads to the best performance in terms of temporal correlation coefficient.

To deepen our statistical analysis of the simulation results, we use scatterplots of modeled versus observed hourly concentrations for each of the three periods and for the three periods combined. We perform linear regression analyses, using the reduced principal axes regression method, and compare slopes (a), intercepts (b), correlation coefficients (R), Root Mean Square Errors (RMSE) and Mean Biases (MB). The statistical analyses are carried

Table 2
Correlation Coefficients by Pollutants Between Hourly Observations and a Regional Model Ensemble As Well As Two Global Forecasts for Three Studied 15-Day Periods Combined

Variable	Multi-model Median	MPI* (FNL) WRFchem	MPI (ERA5) WRFchem	IAG* WRFchem	UFMG* WRF-CMAQ	UFRN* EURAD-IM	NCAR CAMchem	ECMWF CAMS
			All stations					
CO	0.69	0.60	0.65	0.41	0.53	0.55	0.52	0.61
NO_2	0.63	0.56	0.58	0.53	0.58	0.53	0.43	0.55
NO	0.68	0.31	0.36	0.48	0.69	0.49	0.22	0.48
NOx	0.69	0.53	0.57	0.48	0.71	0.53	0.51	0.52
O_3	0.83	0.81	0.76	0.78	0.76	0.71	0.79	0.63
SO_2	0.58	0.44	0.45	0.36	0.55	0.28	0.46	0.37
$PM_{2.5}$	0.53	0.38	0.49	0.26	0.40	0.30	0.23	0.10
PM_{10}	0.45	0.14	0.28	0.10	0.34	0.14	0.27	0.11
			Without Micro-scale stations					
CO	0.67	0.61	0.65	0.35	0.53	0.57	0.52	0.61
NO_2	0.67	0.62	0.65	0.53	0.6	0.55	0.49	0.61
NO	0.57	0.25	0.33	0.34	0.64	0.46	0.13	0.42
NOx	0.65	0.55	0.61	0.40	0.70	0.54	0.50	0.50
O_3	0.83	0.81	0.76	0.78	0.77	0.71	0.79	0.63
SO_2	0.47	0.36	0.37	0.21	0.45	0.22	0.28	0.28
$PM_{2.5}$	0.51	0.36	0.47	0.23	0.39	0.28	0.22	0.12
PM_{10}	0.45	0.15	0.29	0.09	0.34	0.13	0.27	0.13

Note. The Multi-Model Median is calculated from the regional simulation with an asterisk (*). The observations correspond to two city averages calculated with and without the stations of Micro-scale spatial representativeness.

out using the city average for all stations, with specific figures provided for CO (Figure S5 in Supporting Information S1), NOx (Figure S10 in Supporting Information S1), O_3 (Figure S11 in Supporting Information S1), SO_2 (Figure S12 in Supporting Information S1), $PM_{2.5}$ (Figure S13 in Supporting Information S1), and PM_{10} (Figure S14 in Supporting Information S1). The two city averages, with and without the inclusion of Micro-scale stations, are considered for NO (Figures S6 and S7 in Supporting Information S1) and for NO_2 (Figures S8 and S9 in Supporting Information S1).

Our goal is to gain insight into the strengths and weaknesses of each of the four regional models studied by comparing them to each other, to the MMM, and to the global forecasts in Section 4. In order to evaluate the performance of the regional model against other studies using different models and domains, metric recommendations have been proposed for O_3 and $PM_{2.5}$ (Emery et al., 2017). These metrics are applied to the MMM of the regional model ensemble in Section 5.

4. Assessment of the Regional Air Quality Model Ensemble

The assessment of the regional model ensemble is conducted by comparing the observed and modeled temporal variability at the center of São Paulo during the three studied periods, focusing on the different meteorological conditions (Section 4.1), on the long-range transport of pollutants (Section 4.2), on the local sources of pollution (Section 4.3), and on the photochemistry (Section 4.4).

4.1. Meteorological Variability

To identify the different meteorological phases within the three selected periods, we examine relative humidity, PBL height, wind speed and direction (Figure 4; Figure S3 in Supporting Information S1). For wind and
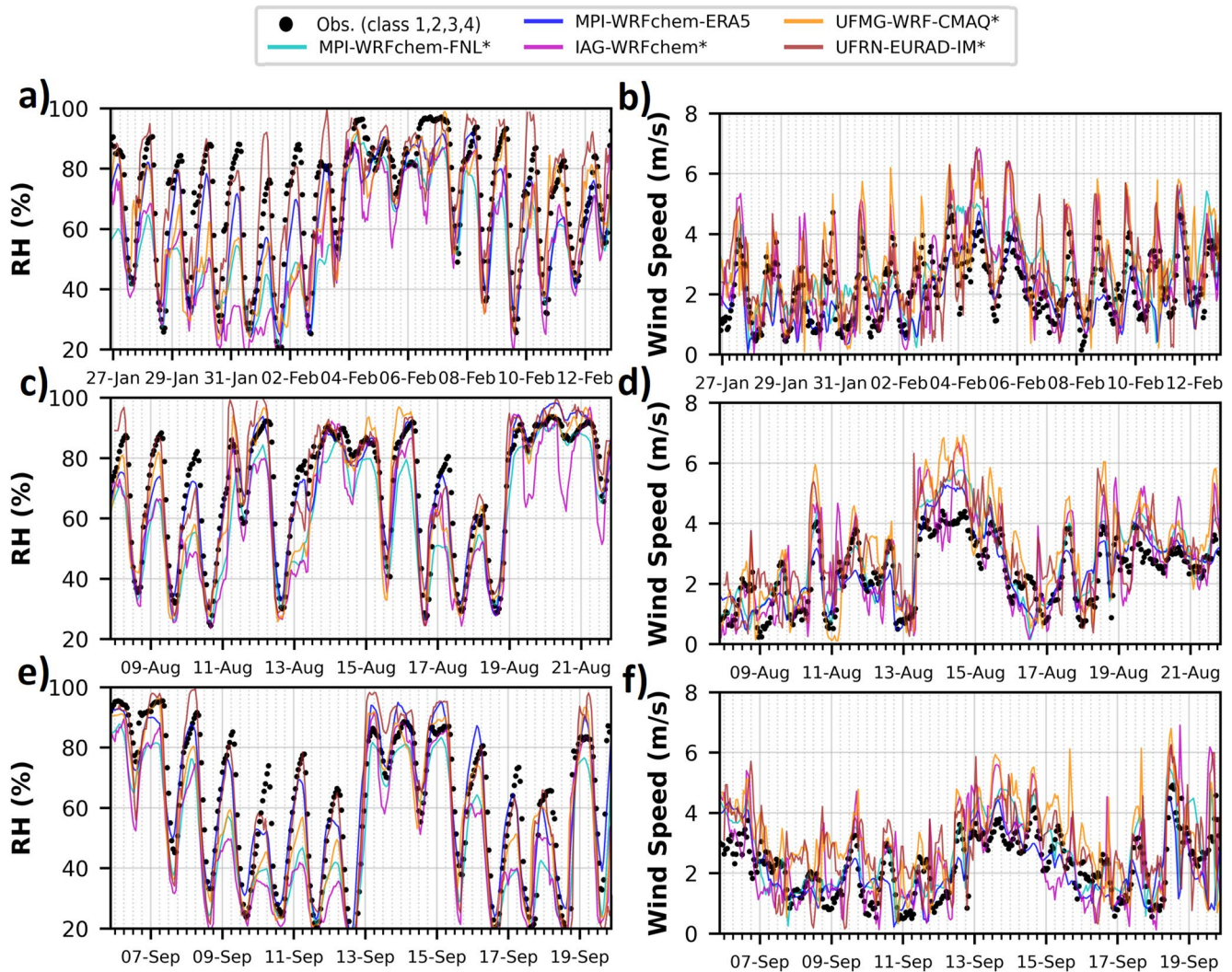
Figure 4. Time series of hourly relative humidity and wind speed observed and modeled in São Paulo for the three selected 15-day periods of the year 2019. The observations correspond to the city average calculated with all stations (class 1,2,3,4).

relative humidity, measurements from 12 to 6 stations respectively are available, and we use the city average for all stations (methodology presented in Section 2.2). To compare the 10-m wind speed diagnosed by the models with the observations made at 2 m, we multiply the observations by a factor assuming that a logarithmic profile represents the wind vertical profile, estimated with the relationship used in the YSU boundary layer scheme and described by Hu et al. (2013). The PBL height data are obtained by a LIDAR measuring the aerosol backscattered signal, located at the University of São Paulo (Moreira et al., 2019). It provides accurate data from 11:00 to 16:00, allowing the analysis of the range of the PBL height when the PBL is highly developed (Courtesy of G. de Arruda Moreira).

For each of the three periods, there are specific days shared by the four meteorological variables (relative humidity, PBL height, wind speed and direction), for which the diurnal cycles differ from other days, which are:

1. From 4 to 6 February,
2. On 12, from 14 to 16, and from 19 to 21 August, and
3. From 13 to 14 September.

These particular days are associated with high relative humidity (greater than 80%) and high wind speed (greater than 3 m/s) continuously coming from the south for several days, and with a low PBL height (lower than 1 km), which corresponds to stormy weather conditions (Figure 4; Figure S3 in Supporting Information S1). 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, there is a diurnal change from northwest at nighttime to southeast during the daytime.

For meteorological phases identified with stormy weather conditions, a greater inter-model variability is observed. Overall, for the three periods considered, the daily variability found by the four regional models agrees well with the meteorological observations of relative humidity and wind speed. There is a very good agreement for RH (R greater than 0.8), while for wind speed, the agreement is lower (R of about 0.7) and associated with RMSE below 1 m/s and MB ranging from -0.1 m/s for the MPI–WRFchem-ERA5 simulation to about 0.5 m/s for the four others. The evolution of the PBL heights modeled by the regional model ensemble is in good agreement with the observations, and the modeled daily maximum heights are in agreement with the measured ones for most of the days.

In conclusion, the regional model ensemble is in good agreement with the meteorological observations at the center of the megacity. There is a small inter-model variability, partly due to the use of the meteorological nudging which is done with the same meteorological data from NCEP-FNL, except for the MPI–WRFchem-ERA5 simulation. Therefore, differences in modeled meteorology are not responsible for persistent differences in modeled pollutant concentrations.

4.2. Long-Range Transport of Pollution

To investigate long-range transport, CO and aerosols (i.e., $PM_{2.5}$ and PM_{10}) serve as useful indicators because they are pollutants strongly emitted by biomass combustion processes (e.g., Andreae, 2019; Yokelson et al., 2008). Their long atmospheric lifetimes (typically exceeding a week) allow them to be often transported over long-range to São Paulo (e.g., Duarte et al., 2021; Martins et al., 2018; Moreira et al., 2021; Pereira et al., 2021; Squizzato et al., 2021). We compare the observed and modeled concentrations of CO and $PM_{2.5}$ (Figure 5), as well as of PM_{10} and of the $PM_{2.5}/PM_{10}$ ratio to estimate the proportion of fine aerosols compared to coarse aerosols (Figure S4 in Supporting Information S1).

The observed concentration variations range from 0.1 to 2 ppm for CO and from 10 to 80 $\mu g.m^{-3}$ for $PM_{2.5}$. Overall, the regional model ensemble reproduces well the daily variations and the amplitude of variations of CO for the three periods. Aerosols are not well reproduced, especially during the second period. Examining global forecasts, NCAR–CAMchem underestimates CO but performs well for $PM_{2.5}$, while ECMWF–CAMS overestimates both CO and $PM_{2.5}$ concentrations.

The temporal variability of PM_{10} is similar to that of $PM_{2.5}$ (Figure 5 compared to Figure S4 in Supporting Information S1). For $PM_{2.5}$ and PM_{10}, the MB of the regional model ensemble are lower than the RMSE, which reflects the difficulties to reproduce the high temporal variability of the aerosol load (for $PM_{2.5}$ Figure S13 in Supporting Information S1 and for PM_{10} Figure S14 in Supporting Information S1). The performance of the MMM is better than individual models, although achieving only moderate correlations (R of about 0.5 and 0.4 for PM2.5 and PM10, respectively). Additionally, there is a consistent overestimation of both $PM_{2.5}$ and PM_{10} concentrations by more than 5 $\mu g.m^{-3}$ for all three periods.

The observed ratio between $PM_{2.5}$ and PM_{10} typically falls between 0.4 and 0.8. Some values exceed 0.8, indicating dominance by fine particles. Conversely, some values are below 0.4, indicating dominance by coarse particles (Figure S4 in Supporting Information S1). In particular, during periods of strong southward winds (cf. Section 4.1), the ratio tends to be lower, indicating coarse particle transport (probably from Santos, where there is a major harbor area). The regional simulations exhibit distinct temporal behaviors, with a ratio remaining nearly constant at 0.8 for UFMG-WRF-CMAQ. In contrast, IAG–WRFchem and UFRN-EURAD-IM display a clear diurnal cycle in this ratio.

Synchronized increases in both pollutants are noticeable, with concentrations reaching at least 1.5 ppm for CO and 50 $\mu g.m^{-3}$ for $PM_{2.5}$. Importantly, these synchronized increases are not limited to the period selected for biomass burning events, but also occur in the other two periods. Each increase is associated with different ratios of CO to $PM_{2.5}$ and different durations, ranging from a few hours to several days. Given that São Paulo is often affected by biomass burning pollution events all year long due to agricultural practices in the surrounding rural areas (Godoy-Silva et al., 2017), deforestation, and pasture maintenance fires in distant regions (Duarte
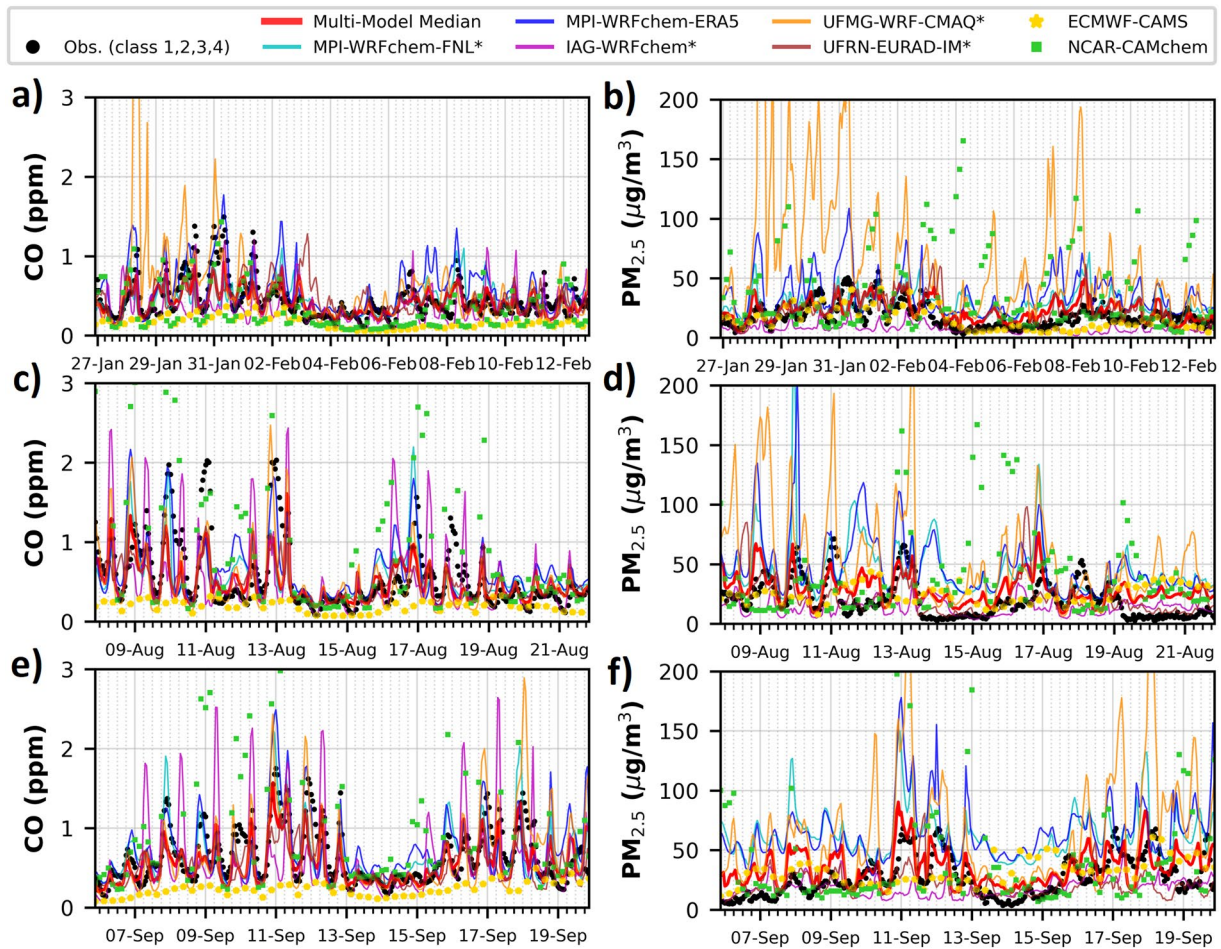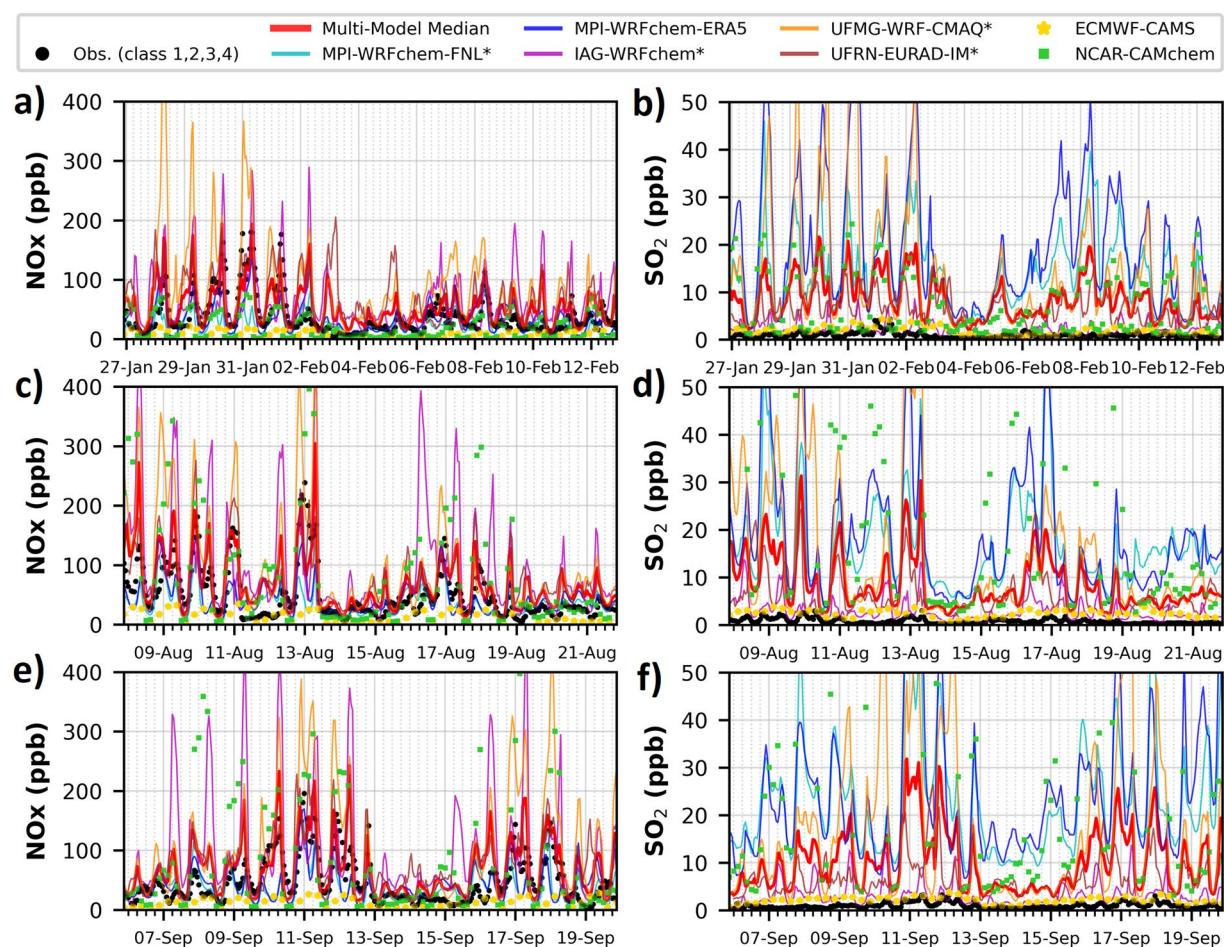
Figure 5. Time series of hourly concentrations of CO (a, c and e) and PM$_{2.5}$ (b, d and f) observed and modeled in São Paulo for the three selected 15-day periods of the year 2019. The observations correspond to the city average calculated with all stations (class 1,2,3,4). The models include simulations from two global forecasts (yellow stars and green squares) and from a regional model ensemble (colored lines) with the Multi-Model Median (red line) calculated from the simulations with an asterisk (*).

et al., 2021; Pereira et al., 2021), these synchronized increases strongly suggest biomass burning pollution events. It is important to note that these events are different from the stormy weather conditions discussed in Section 4.1. Synchronized increases in both pollutants include:

1. January 30 to February 1,
2. August 9–10, 13, and 17–18,
3. And September 10–12, 17–18.

The second period was chosen because this period contains several events of high pollutant concentrations related to biomass burning, which impacted São Paulo due to long-range transport from the Amazon (Pereira et al., 2021). For this specific period, the backward trajectories of the air masses transported toward São Paulo are analyzed using the Hysplit model on the NCEP–FNL meteorological fields (Rolph et al., 2017), and we compare them to the fire counts obtained by satellite (Figure S15 in Supporting Information S1). The backward trajectories show that the air masses came from regions where biomass burning occurs (high number of fires), namely on August 9 and 10 mainly from the Pantanal and the Cerrado, on August 13 mainly from the Cerrado and the August 17 and 18 mainly from the Pantanal, the Amazon and the Cerrado.

Increases in CO and PM$_{2.5}$ associated with biomass burning events are better represented by the regional model ensemble compared to the global forecasts (not reproduced by NCAR–CAMchem and overestimated by ECMWF–CAMS). The biomass burning pollution events are best identified by the MMM, because for each event there is at least one regional simulation in good agreement with the observations in terms of magnitude and persistence. Although the two meteorologies used for the MPI–WRFchem simulations lead to very similar simulations for CO

Figure 6. 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 observations correspond to the city average calculated with all stations (class 1,2,3,4). The models include simulations from two global forecasts (yellow stars and green squares) and from a regional model ensemble (colored lines) with the Multi-Model Median (red line) calculated from the simulations with an asterisk (*).

and $PM_{2.5}$ (Figure 5), as well as for PM_{10} and $PM_{2.5}/PM_{10}$ ratio (Figure S4 in Supporting Information S1), there are small improvements in the correlation coefficient of the MPI–WRFchem simulation using ECMWF–ERA5 compared to NCEP–FNL as meteorological input, which could be due to more accurate wind fields, improving the representation of long-range transport of pollutants.

The modeled temporal variability of aerosols is less well reproduced by the regional model ensemble than for CO, which may be caused by both the onset time of biomass burning aerosols due to long-range transport and the production of secondary aerosols, which is generally underestimated in São Paulo (Andrade et al., 2017). This section highlights the importance of long-range pollutant transport for air quality in São Paulo attributed to biomass burning sources. Additionally, the MMM offers the most comprehensive estimate for CO and aerosols.

4.3. Local Pollution

4.3.1. NOx and SO_2 Concentrations

Two key pollutants associated with anthropogenic activities are NOx and SO_2. In a megacity like São Paulo, NOx emissions primarily originate from traffic, while SO_2 is predominantly linked to industrial activities and coal-based electricity production. We analyze the observed and modeled hourly concentrations of NOx and SO_2 during the three periods in the center of São Paulo with the city average calculated with all stations (Figure 6).

Observations show that NOx concentrations vary significantly and similarly over the three periods, with low values below 10 ppb and high values often reaching up to 100 ppb. Biomass burning pollution events are associated with high NOx values, reaching at least 150 ppb (cf. Section 4.2).

The regional model ensemble is in good agreement with the observed NOx range of variation over the three periods (R ranging from 0.5 to 0.7). Biomass burning pollution events lead to an increase in the modeled NOx concentration for all models, except for NCAR–CAMchem. The magnitude of modeled NOx concentrations during these events varies significantly across models. The MMM and UFMG–WRF-CMAQ demonstrate the best agreement with observations for NOx concentrations during the three periods (Figure S10 in Supporting Information S1).

For SO_2, the picture is different from the other pollutants analyzed previously. Observations show low concentrations of SO_2 in São Paulo (less than 5 ppb). However, the regional model ensemble, as well as ECMWF–CAMS, overestimates SO_2 concentrations by nearly a factor of 10 (Figure 6).

Despite the large biases in the regional simulations (leading to a bias greater than 7 ppb for MMM), the performance of MMM is the best in terms of correlation coefficient (Figure S12 in Supporting Information S1). Interestingly, the NCAR–CAMchem forecast with coarse resolution has the lowest bias for SO_2, which may be related to its much coarser resolution of around 100 km. IAG–WRFchem has the second lowest bias, which may be related to the use of a local anthropogenic emission inventory limited to vehicular emissions measurements (Andrade et al., 2015), which only include traffic emissions (cf. Section 3.2). Comparing meteorological inputs used in MPI–WRFchem, modeled SO_2 concentrations are similar, implying that meteorological conditions may not be the primary factor driving the overestimation. Additionally, fire emissions are unlikely to be the main cause of overestimation, as consistently high SO_2 concentrations occur across all three periods.

The monitoring of SO_2 emissions by major industries became mandatory in 1982 in São Paulo, consequently many industries moved to other cities (Andrade et al., 2017). The constant decreasing trend in SO_2 concentration since 1982 is attributed first to the regulation of industries and, since the 1990s, to the reduction in the sulfur content of diesel fuel (Andrade et al., 2017; Carvalho et al., 2015). To our knowledge, there have been no recent regulatory changes (on gasoline content or industry stack emissions) that could explain the large overestimation of modeled SO_2 concentrations. Therefore, the overestimation of SO_2 suggests problems with sector-specific emission factors and spatial proxy definitions for emission sources that may not be specific to São Paulo.

### 4.3.2. Daily Variability of NO and NO_2 Concentrations

The choice of station representativeness classes has a significant influence on the city average for NOx concentrations (cf. Section 2.2). To focus on the diurnal variability of NOx, we analyze the modeled and observed concentration for NO and NO_2 with hourly average diurnal cycles for the three selected periods and using the two city averages, with or without the inclusion of the Micro-scale stations (Figure 7).

Observed NO concentrations are typically below 15 ppb during the day, rising to the range of 15–50 ppb at night, with a morning peak at 08:00 (in local time) related to traffic emissions. The morning peak at 08:00 results in the most significant differences in NO between the two city averages, up to 25 ppb. Due to the rapid conversion of NO to NO_2, the observed NO_2 concentrations vary less during the day, ranging from 10 to 35 ppb in all three periods, with differences between the two city averages being less than 10 ppb.

For NO, the models reproduce well the average diurnal cycle, with the morning peak at 08:00, and a higher concentration at night compared to the day. The variability between models is high during nighttime because the concentrations modeled by the two MPI-WRFchem simulations are underestimated, and overestimated by the other three regional simulations. This highlights the significance of vertical anthropogenic profiles used for MPI-WRF simulations during nighttime. The morning peak at 08:00 exhibits the largest inter-model variability from less than 30 ppb for the two MPI–WRFchem simulations to more than 100 ppb for the UFMG–WRF-CMAQ and the IAG–WRFchem simulations. Even taking into account the differences between the two averages for NO, the inter-model variability is much larger than the range observed in the city averages, showing the difficulties in modeling the rapid changes in PBL height and NOx emissions.

For NO_2, the regional model ensemble is in good agreement with the low concentration observed during the day (from 10:00 to 16:00), while at night NO_2 is overestimated of about 20 ppb for all regional simulations. The transition from night to day (from 06:00 to 10:00) is associated with large increase in the modeled NO_2 concentrations, up to 60 ppb, while the observed concentration increases is smaller, up to 30 ppb.

All regional and global simulations reproduce the low concentrations of NO and NO_2 from 06:00 to 10:00, when the PBL height is greater than 1 km at the center of the megacity (Moreira et al., 2019). However, there is a large inter-model variability of NO concentration during the morning peak at 08:00, and an overestimation of NO_2 concentration at night for all regional simulations, highlighting issues related to anthropogenic emissions in the
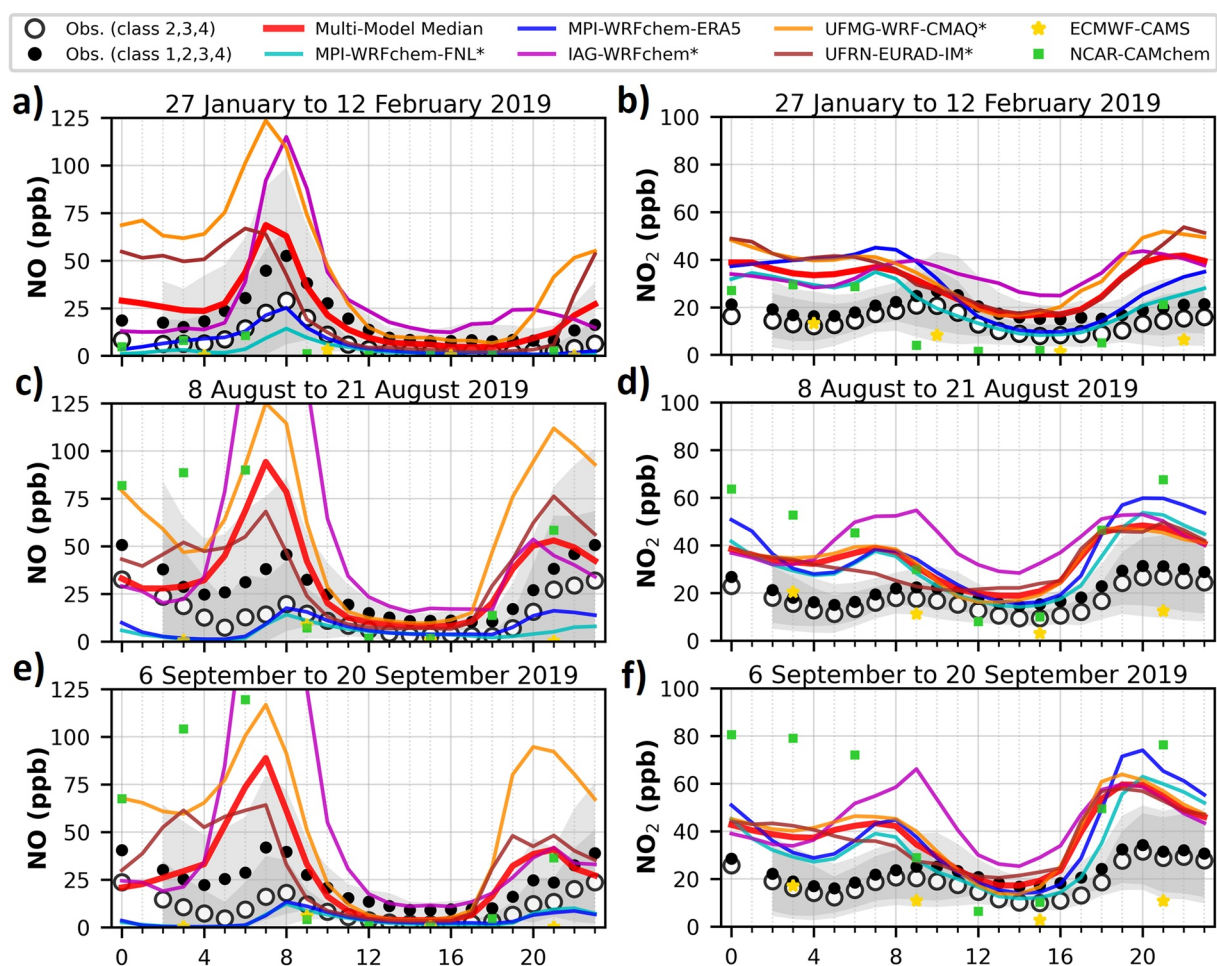
Figure 7. Hourly average diurnal cycles of concentrations of NO (a, c, and e) and NO_2 (b, d, and f) observed and modeled in São Paulo over the three selected 15-day periods of the year 2019. The observations correspond to two city averages calculated with all stations (class 1,2,3,4) and without the stations of Micro-scale representativeness (class 2,3,4). The models include simulations from two global forecasts (yellow stars and green squares) and from a regional model ensemble (colored lines) with the Multi-Model Median (red line) calculated from the simulations with an asterisk (*). The gray shadings correspond to the standard deviation of the two observed hourly city averages. Hours of the day are given in local time.

models. Despite this issues, the MMM provides the best agreement with the observed diurnal cycles of both NO and NO_2 concentrations.

4.4. Photochemistry

4.4.1. O_3 and Oxidant Level

We examine the temporal variability of O_3 concentration and oxidant level (Ox = NO_2 + O_3) as observed and modeled in São Paulo's subtropical urban environment (Figure 8). In urban areas, where NOx emissions are important, there is a competition between the loss and the production of O_3 during the day as the titration of O_3 by NO is compensated by the photolysis of NO_2 (Jacob, 1999). 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 O_3 (Wood et al., 2010). At night, Ox is not affected by the titration of O_3, making it a valuable quantity for our analysis as it should vary less between day and night than O_3.

Observed O_3 concentrations in São Paulo show a marked diurnal cycle for most days (Figure 8), with a minimum below 10 ppb at night and a maximum during the day above 50 ppb, except during certain 2-day periods associated with stormy weather conditions (cf. Section 4.1). The Ox level is approximately 20 ppb, often experiencing daytime increases corresponding to increases in O_3. The second period has a more stable Ox level compared to the other two periods (Figure 8).
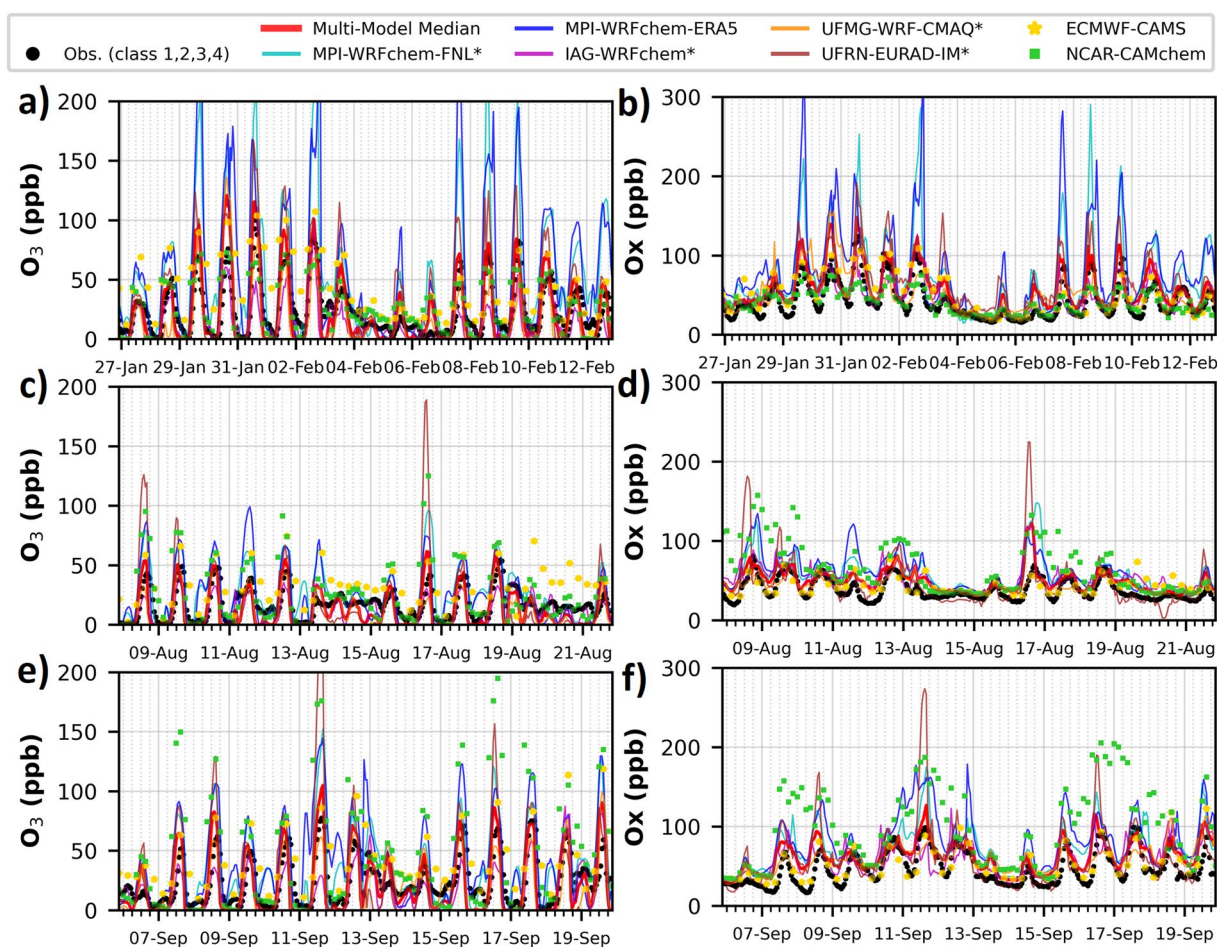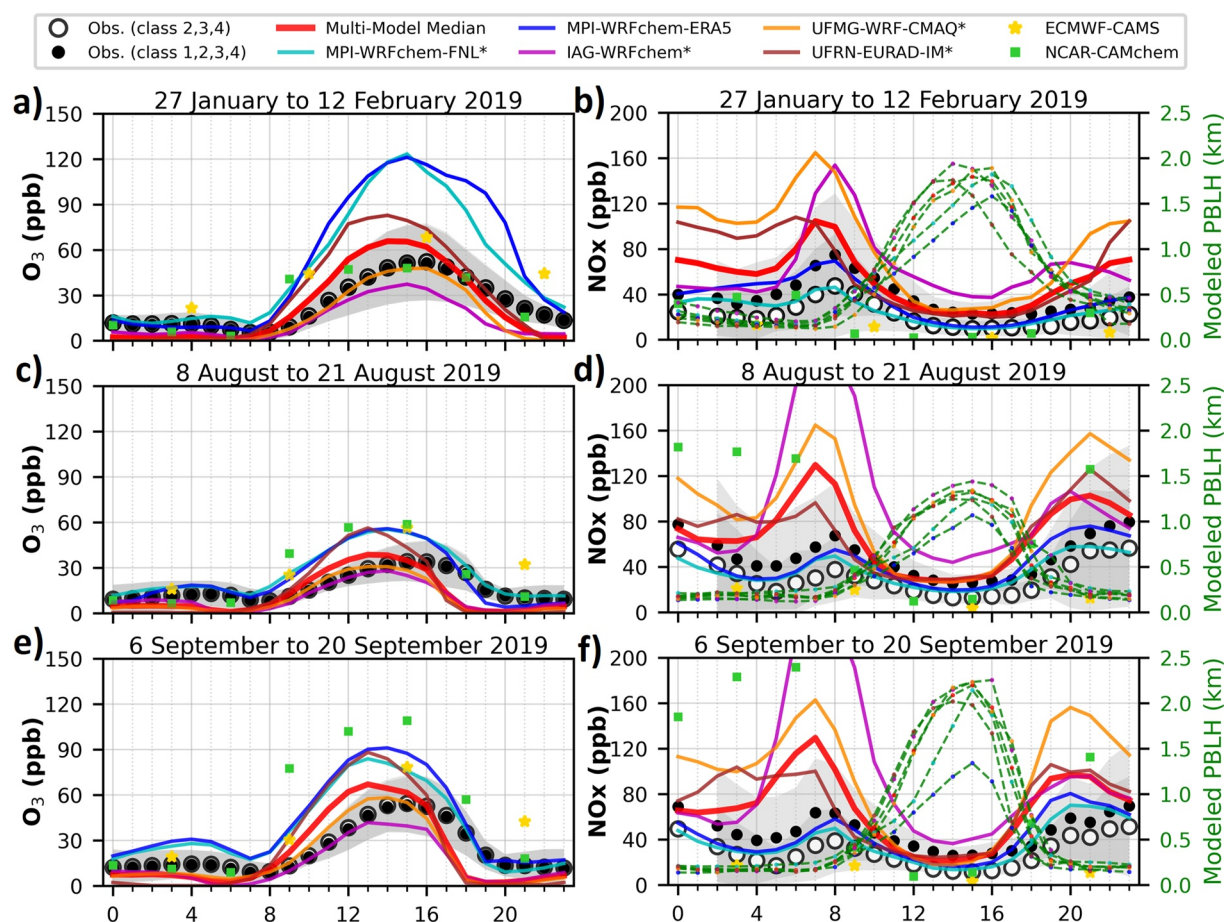
Figure 8. Time series of hourly concentrations of O_3 (a, c, and e) and oxidant, defined as $Ox = NO_2 + O_3$, (b, d, and f) observed and modeled in São Paulo for the three selected 15-day periods of the year 2019. The observations correspond to the city average calculated with all stations (class 1,2,3,4). 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 regional model ensemble closely matches the observed O_3 temporal variations (with an R of about 0.8), despite the selection of the first and third periods due to their inclusion of high O_3 events (Figure S11 in Supporting Information S1), while the Ox level is overestimated by the regional model ensemble and the global forecasts. The two global forecasts have similar performance to the regional models, although the performance is calculated with a lower number of hours for the global forecasts due to their lower output frequency.

The two meteorological inputs used with MPI–WRFchem result in small differences, such as during biomass burning pollution events (cf. Section 4.2), but both simulations have comparable O_3 and Ox concentrations versus the other simulations, showing the small influence of the meteorological inputs. It is noteworthy that IAG-WRFchem underestimates O_3, while the two MPI-WRFchem simulations overestimate it, and that UFMG-WRF-CMAQ and UFRN-EURAD-IM are in good agreement. Looking at individual days, we see that each simulation has certain periods of better agreement in terms of O_3 and Ox. Therefore, the MMM has overall the best agreement for O_3 and Ox for all three periods.

4.4.2. O_3 and NOx

O_3 and NOx concentrations show significant diurnal variability during the three studied periods, largely due to changes in anthropogenic emissions and PBL height. We analyze the hourly average diurnal cycles of O_3 and NOx concentrations along with modeled PBL heights to assess the ability of the regional model ensemble to reproduce these cycles (Figure 9). Since the city averages with and without the Micro-scale stations lead to significant differences in the observed NOx concentrations (cf. Section 2.2), the two averages are compared to the modeled concentrations.

Figure 9. Hourly average diurnal cycles of concentrations of O$_3$ (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 observations correspond to two city averages calculated with all stations (class 1,2,3,4) and without the stations of Micro-scale representativeness (class 2,3,4). The models include simulations from two global forecasts (yellow stars and green squares) and from a regional model ensemble (colored lines) with the Multi-Model Median (red line) calculated from the simulations with an asterisk (*). The modeled planetary boundary layer heights (PBLH) are the green dashed lines with colored dots corresponding to each model. The gray shadings correspond to the standard deviation of the observed hourly data. Hours of the day are given in local time.

On average, the concentration of O$_3$ observed in São Paulo has three phases: (a) below 20 ppb from midnight to 09:00, (b) increasing until 16:00, and (c) gradually decreasing until midnight As expected from the observed NO and NO$_2$ diurnal cycles (cf. Section 4.3.2), the diurnal cycle of NOx concentration shows a peak at 08:00, high concentration at night (exceeding 60 ppb) and low concentration during the day (below 40 ppb). The two city averages lead to a difference of 20 ppb throughout the day for NOx, an important consideration for interpreting modeled biases.

The regional model ensemble reproduces well the chronology of the observed phases of the O$_3$ diurnal cycle. For NOx, the morning peak is correctly modeled at 08:00, while the period of high concentration from 20:00 to 03:00 is modeled too early. During the day (from 10:00 to 16:00), UFMG–WRF-CMAQ is in good agreement for O$_3$, while O$_3$ is overestimated for MPI–WRFchem and UFRN–EURAD-IM, and O$_3$ is underestimated for IAG–WRFchem. At night (from 20:00 to 06:00), MPI–WRFchem is in good agreement for O$_3$ and for NOx, while O$_3$ is underestimated and NOx is underestimated for the other three simulations. The transitions, day to night (from 06:00 to 10:00) and night to day (from 16:00 to 20:00), are associated with very large inter-model variability for NOx, and to a lesser extent for night.

The two MPI–WRFchem simulations lie well between the observations of the two city averages for NOx, but O$_3$ is largely overestimated during the day, which is related to the modeled ratio between NOx and volatile organic compounds. Conversely, it is interesting to note that IAG–WRFchem has the highest underestimation of O$_3$ during the day, which is not only linked to its overestimation of NOx, but also to the absence of biogenic

emissions and therefore of volatile organic compounds. This result suggests that the concentration of O_3 would increase with the inclusion of biogenic emissions in the IAG-WRFchem simulation, even though the horizontal domain is the smallest.

The biases in O_3 exhibited by the two global forecasts are comparable to those observed in the four simulations of the regional model ensemble but are more pronounced for NOx. The analysis of the O_3 diurnal cycles reveals an improvement for the MMM compared to the other simulations. This improvement is attributed to the variable performance of the four simulations during the different diurnal phases. However, it is worth noting that the MMM tends to overestimate NOx concentrations.

Modeled PBL heights are similar in all regional simulations for each of the three time periods. During the night, the bias in modeled NOx concentration in each regional simulation is opposite to the bias in modeled O_3 concentration, despite the consistent modeled PBL heights. The modeled PBL, which is similar in time and height, cannot explain the large inter-model variability in O_3 concentration during the day and in NOx concentration at night, again highlighting issues related to anthropogenic emissions in the models.

4.4.3. NO_2/NOx and NO_2/Ox

The observed NO_2 to NOx and NO_2 to Ox ratios show less variability than the regional model ensemble, providing insight into the interpretation of modeled O_3 biases (Figure 10). The two city averages result in different NO_2 to NOx ratios of approximately 80% without Micro-scale stations and approximately 60% with Micro-scale stations. Conversely, both city averages result in consistent NO_2 to Ox ratios ranging from 70% at night to 20% during the day. The traffic peak at 08:00 is associated with a 20% decrease in the NO_2 to NOx ratio and a 20% increase in the NO_2 to Ox ratio.

Despite the similarity of these ratios observed for the three periods, the regional model ensemble reproduces these ratios with larger ranges, and even larger for the two global forecasts. During the day (from 10:00 to 16:00), the good agreement for O_3 of UFMG–WRF-CMAQ is associated with NO_2/NOx and NO_2/Ox ratios in better agreement than the other three simulations. At night (from 20:00 to 06:00), the good agreement for O_3 of MPI–WRFchem is associated with a NO_2/Ox ratio in better agreement than the other three simulations, while the NO_2/NOx ratio is overestimated. Compared to the four simulations of the regional model ensemble, the MMM has the best agreement for the NO_2/NOx ratio, falling well between the two city averages.

The bias in modeled NO_2/Ox ratios is closely related to the bias in modeled O3 during the day, since the more NO_2/Ox is overestimated, the more O_3 is underestimated by a model, and vice versa (Figure 9). Furthermore, the overestimation of the NO_2/NOx ratio by the MMM at night is related to the underestimation of O_3 and to the overestimation of NOx, for which we suspect the vertical anthropogenic profiles to be a key parameter.

In summary, the regional model ensemble reproduce the meteorology in very good agreement, the modeled concentrations of O_3 in good agreement, and the modeled concentrations of CO, PM, and NOx in moderate agreement. Biomass burning events have a significant impact on air quality in São Paulo, while the models have difficulties to obtain the correct magnitude of CO, NOx, PM, and SO_2 during these events. Furthermore, this section shows that all models have well reproduced periods and pollutants. Therefore, for CO, NOx, O_3, and Ox, the MMM has the best performance compared to all its members, while comparing the regional models with the global forecasts, the performance is of the same order over the three selected periods.

5. Potential of the Regional Model Ensemble

The small number of models involved in the MMM calculation, that is, the median of the four models, is an important limitation. However, Section 4 has shown that two models overestimate O_3 while the other two underestimate it, resulting in good agreement with the MMM. In addition, there is an overestimation of oxidant levels, which we will focus on in this section to understand its temporal biases (Section 5.1), before assessing the potential of the MMM for O_3 and $PM_{2.5}$ in the context of an operational ensemble forecast system for São Paulo (Section 5.2).

5.1. Oxidant Level Overestimation

The analysis continues by focusing on the temporal biases exhibited by the MMM for NO_2, O_3, and Ox, as well as their associated hourly average diurnal cycles, with the aim of distinguishing the diurnal phases and identifying the main contributors to the MMM biases (Figure 11). We compare the MMM to the city average calculated using all stations, which results in higher NO_2 and Ox observed concentrations.
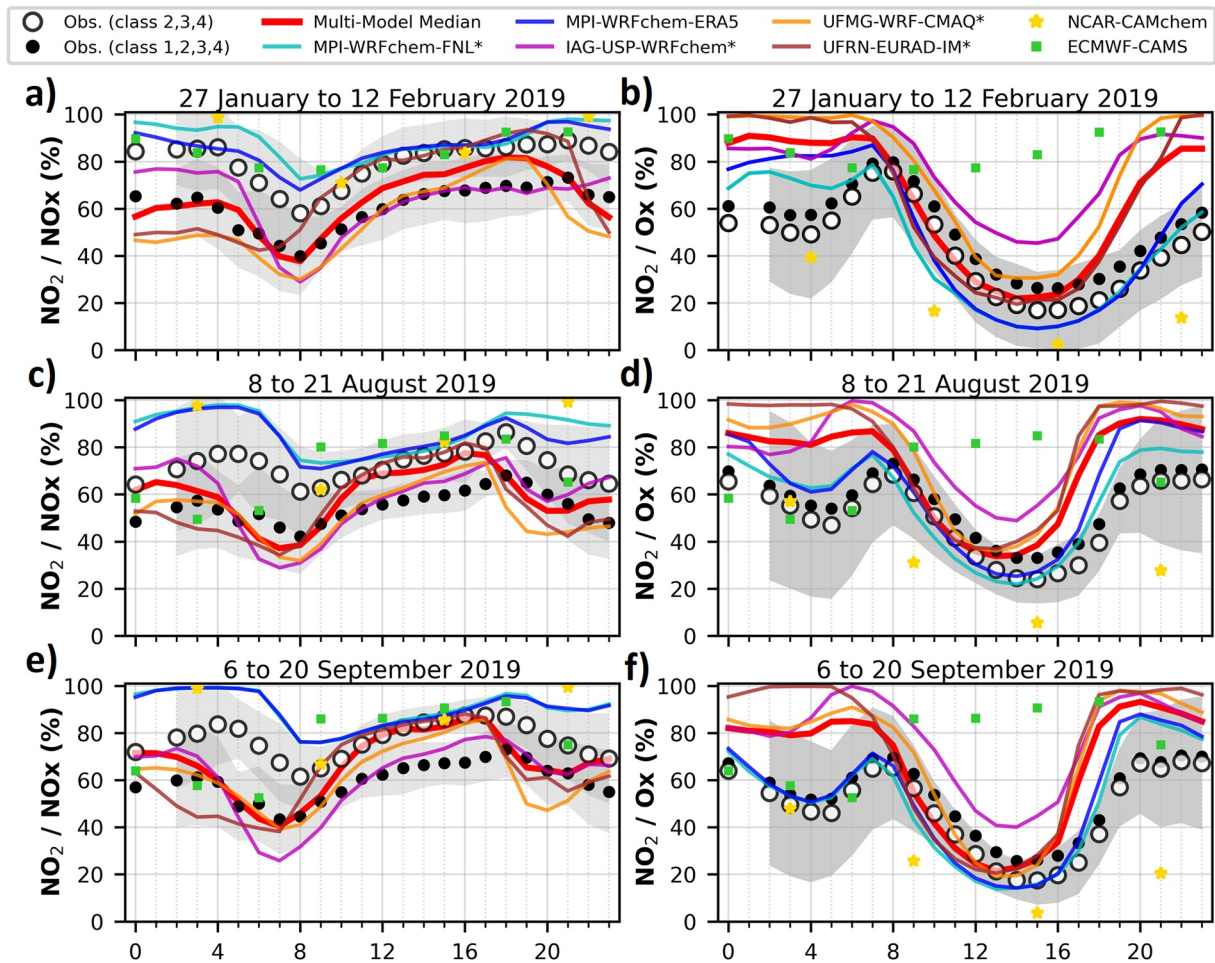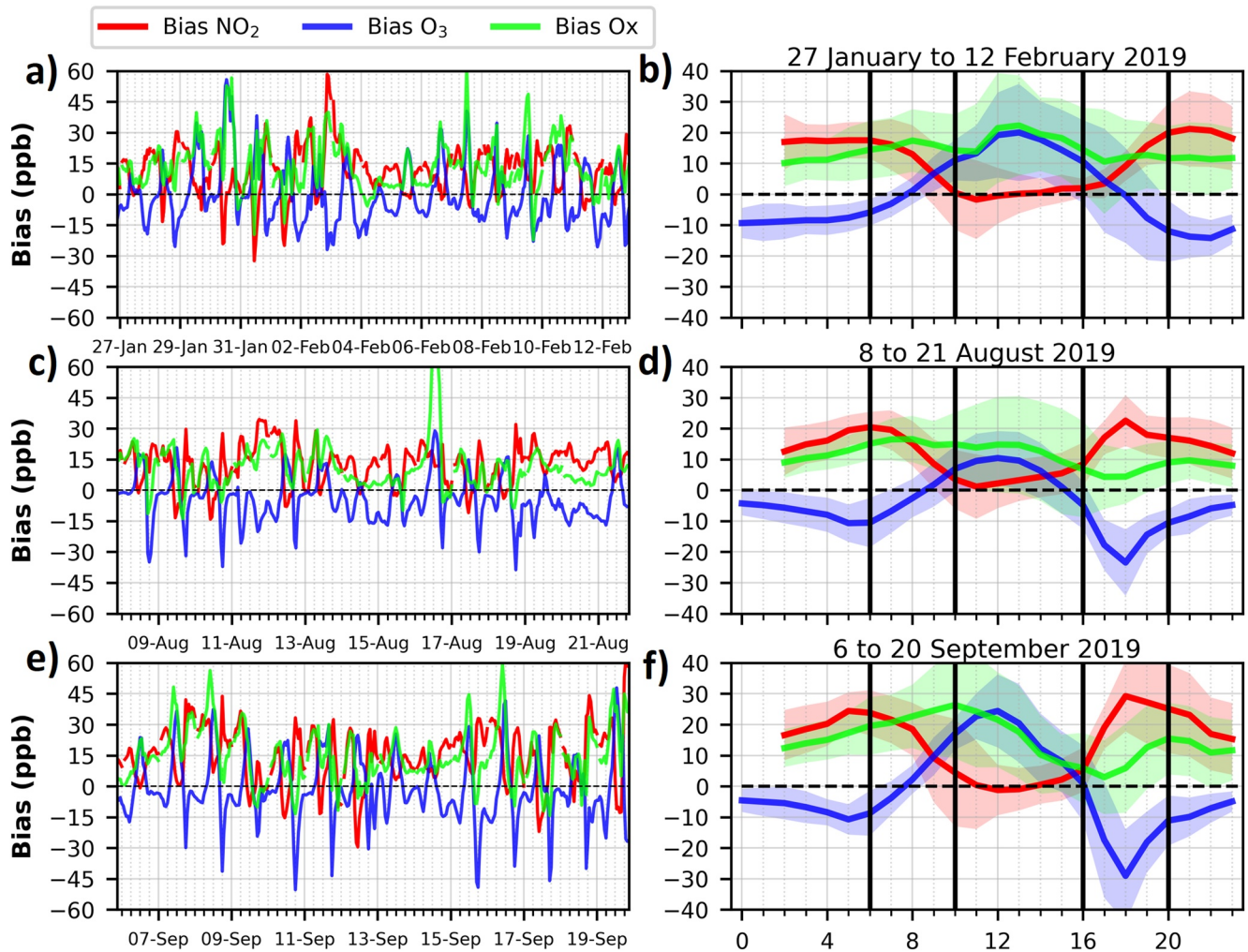
Figure 10. Hourly average diurnal cycles of ratios of NO_2 in NOx (a, c and e) and in Ox (b, d, and f) observed and modeled in São Paulo over the three selected 15-day periods of the year 2019. The observations correspond to two city averages calculated with all stations (class 1,2,3,4) and without the stations of Micro-scale representativeness (class 2,3,4). The models include simulations from two global forecasts (yellow stars and green squares) and from a regional model ensemble (colored lines) with the Multi-Model Median (red line) calculated from the simulations with an asterisk (*). The gray shadings correspond to the standard deviation of the two observed hourly city averages. Hours of the day are given in local time.

The biases in Ox shown by the MMM are consistently positive over the three periods, resulting in an overestimation of about 10 ppb. The biases in NO_2 and O_3 both exhibit distinct diurnal patterns that are mostly opposite to each other (Figures 11a, 11c, and 11e). These patterns are well reflected in the average diurnal cycles and similar for the three periods (Figures 11b, 11d, and 11f). Notably, different diurnal phases of bias emerge for NO_2 and O_3, suggesting the involvement of distinct driving factors for these biases.

1. During nighttime (from 20:00 to 06:00), an overestimation is seen for NO_2 while O_3 is consistently underestimated. The use of vertical profiles in anthropogenic emissions is a key parameter for reducing the NO_2 bias, and can therefore contribute to reducing the O_3 bias (cf. Section 4.3.2).
2. In the morning (from 06:00 to 10:00), the NO_2 and O_3 biases are high during high traffic emissions at 08:00 and decrease afterward. The peak of traffic emissions at 08:00 is associated with a large overestimation of the NO_2 concentration and an underestimation of the O_3 concentration. Despite an increasing PBL height during this period, the modeled PBL heights remain similar (cf. Section 4.4.2). Thus, the large biases again point to the treatment of anthropogenic emissions.
3. During daytime (from 10:00 to 16:00), the NO_2 bias diminishes, while O_3 continues to be significantly overestimated. Individual simulations predict daily O_3 maxima with substantial variability around the observed maxima. The similar PBL heights and low NOx concentrations suggest that the discrepancies in O_3 may be related to the ratios of NOx to volatile organic compounds between models (cf. Sections 4.4.2 and 4.4.3), primarily associated with anthropogenic and biogenic emissions.

Figure 11. Time series of the hourly bias (modeled minus observed concentration) of the Multi-Model Median for O_3, 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 observed concentration corresponds to the city average calculated with all stations (class 1,2,3,4). The modeled concentration corresponds to Multi-Model Median calculated from a regional model ensemble of four simulations. The black bars divide the diurnal cycle into four phases, with the morning and evening hours corresponding to 4-hr phases. Hours of the day are given in local time.

4. In the evening (from 16:00 to 20:00) the largest overestimation for NO_2 and O_3 underestimation of the O_3 concentration are seen. The differences in the treatment of anthropogenic emissions could contribute significantly to this bias. In addition, the biases during this period may be amplified compared to the morning due to the urban heat effect, which may maintain a higher PBL height than what the models predict (cf. Section 4.4.2).

The MMM consistently shows an underestimation of O_3 during the night and an overestimation during the day, along with a persistent overestimation of the oxidant level by about 10 ppb. This analysis of the MMM bias underscores the substantial influence of the treatment of anthropogenic emissions during each diurnal phase, and highlights their important potential for improving model performance.

5.2. Modeled O_3 and $PM_{2.5}$

5.2.1. São Paulo Against Other Areas

In this section, we use metrics that have been developed to compare the performance of air quality models used in different areas. The Normalized Mean Bias (NMB) and Normalized Mean Error (NME) are the average of the differences (or absolute differences) between paired modeled and observed concentrations, normalized to the observed average. In addition to the correlation coefficient, which indicates the degree of agreement over the

Table 3

Performance Evaluation of the Median of the Regional Air Quality Model Ensemble for MDA8-O_3 and Daily Average $PM_{2.5}$ Concentrations, Using the Air Quality Model Benchmark Metrics, Including Normalized Mean Bias (NMB) and Normalized Mean Error (NME) and Correlation Coefficient (R)

	NMB	NME	*R*
MDA8-O_3			
Goal (Criteria)	<±5% (<±15%)	<±15% (<±25%)	>0.75 (>0.5)
p1	0.62	38.92	0.88
p2	−25.44	43.06	0.78
p3	−5.33	37.19	0.85
Daily average $PM_{2.5}$			
Goal (Criteria)	<±10% (<±30%)	<±35% (<±50%)	>0.70 (>0.4)
All 3 periods	33.44	41.32	0.71

Note. The goal and criteria are given following the recommendations by Emery et al. (2017).

entire time period, the NMB and NME provide important information for estimating the quality of the results (Emery et al., 2017). These metrics mostly concern O_3 and $PM_{2.5}$, which are not based on hourly concentrations, but rather on the 8-hr maximum daily average for the concentration of O_3 (MDA8-O_3), and on the daily average for the concentration of $PM_{2.5}$. MDA8-O_3 and daily average $PM_{2.5}$ are the most controlled by WHO for air quality alerts, and thus the most used to compare the results of air quality models (Emery et al., 2017).

Recommendations are given for the appropriate use of these metrics, which concern the duration of the period, and which are different for O_3 and $PM_{2.5}$ due to the difference in averaging. In our case, the three 15-day periods are sufficient to analyze MDA8-O_3, but for the daily average $PM_{2.5}$, we have to combine them to get a sufficient number of days. Using rank-ordered distributions of NMB, NME, and R from the literature, Emery et al. (2017) proposed MDA8-O_3 and daily average $PM_{2.5}$ benchmarks for the goals and criteria of these metrics. The more restrictive (i.e., the goals) represent about one-third of the rank-ordered distributions (from the literature over the past 10 years), while the less restrictive (i.e., the criteria) represent about two-thirds. We present the MMM performance evaluation using the recommendations of Emery et al. (2017) for goals and criteria, which are therefore established in the historical context (Table 3).

For MDA8-O_3, the correlation coefficient goal is satisfied. However, NMB criteria is mostly not satisfied and NME is significantly above the criteria. Despite the higher agreement for the hourly average of O_3 concentration compared to other pollutants studied, the performance of the regional model ensemble is intermediate for MDA8-O_3, primarily due to the elevated NME (≈40% significantly above the NME criteria of <±25%). The performance evaluation of the regional model ensemble shows that improvements could be expected for the modeled MDA8-O_3, for which the previous sections suggest that anthropogenic emissions play an important role.

Conversely, despite the poor agreement with the correlation coefficients for the hourly $PM_{2.5}$ concentration (cf. Section 4.2), the performance of the regional model ensemble is acceptable for the daily average $PM_{2.5}$ since the criteria for NME and R are met. This result shows that the difficulties in modeling $PM_{2.5}$ are also present in other regions due to the multiple drivers of concentration variability.

5.2.2. Air Quality Alerts

The performance of the Multi-Model Median (MMM) is assessed with regard to O_3 and $PM_{2.5}$ alerts based on WHO air quality standards (Figure 12). These standards are based on the MDA8-O_3 and daily average $PM_{2.5}$, utilizing concentration thresholds of 50 ppb for O_3 and 25 µg.m^{-3} for $PM_{2.5}$ (Guidelines used before 2021). The MMM is compared to the city average with all stations.

Considering that an alert is triggered if the concentration threshold is exceeded during the course of a day, four cases are possible for each day when comparing model results with observations:

- 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.

The number of alerts between observations and the MMM is compared (Figure 12). Moreover, the performance of the MMM predictions for alerts can be quantified using the Probability Of Detection (POD) and the False Alarm Rate (FAR) metrics, which are calculated by comparing the number (*N*) of these four possible cases, as defined by Brasseur and Jacob (2017):

$$POD = N(CaseA)/N(CaseA + B) \qquad (3)$$

$$FAR = N(CaseD)/N(CaseA + D) \qquad (4)$$
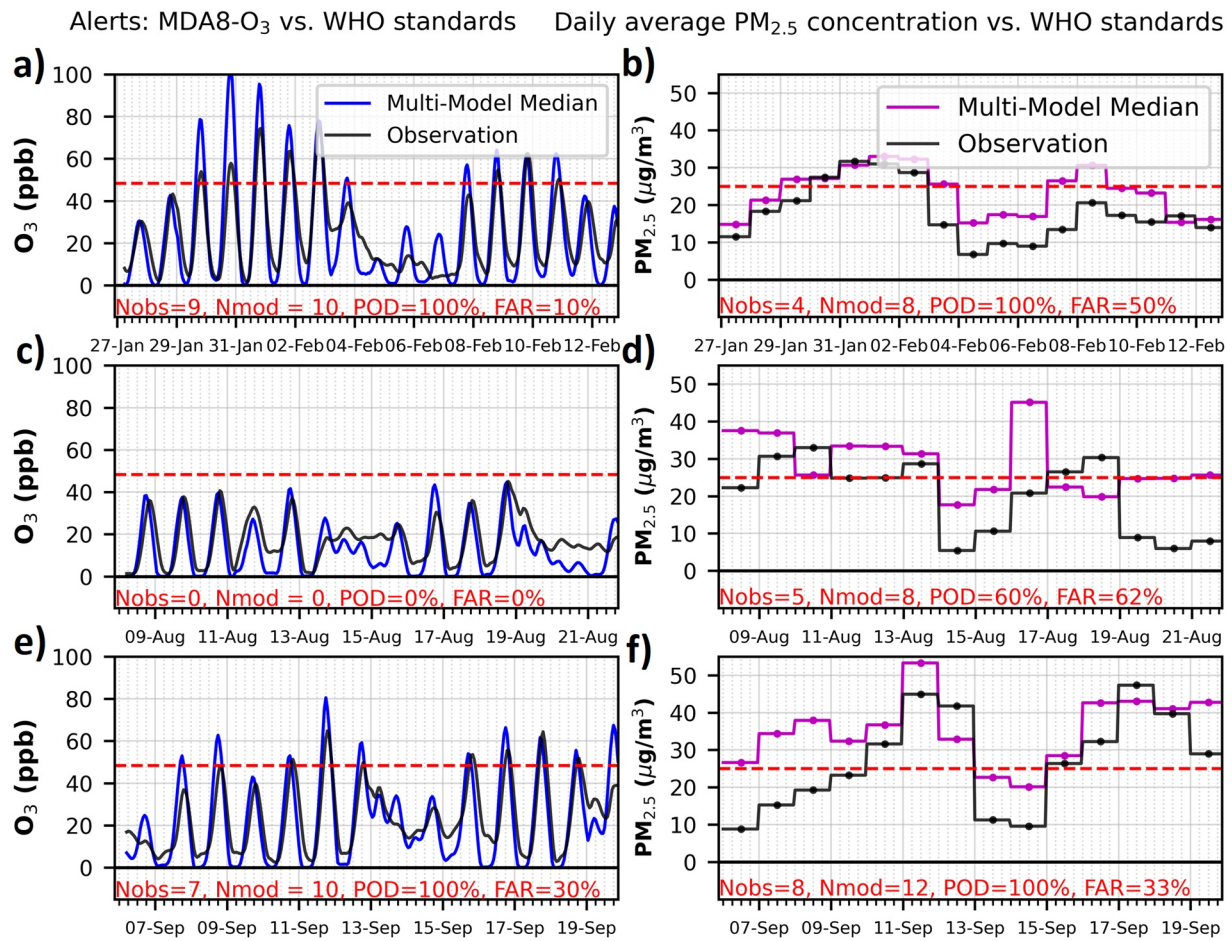
Figure 12. Comparison of the alerts observed for O_3 and $PM_{2.5}$ with those predicted by the median of the regional air quality model ensemble (Multi-Model Median). The observed concentration corresponds to the city average calculated with all stations. The modeled and observed concentrations of MDA8-O_3 (a, c, and e) and of $PM_{2.5}$ (b, d, and f) are presented for the three periods with the associated performances for alerts of the ensemble: number of observed alerts (N obs), number of modeled alerts (N mod), probability of detection (POD) and false alarm rate (FAR). The concentration thresholds defined by the WHO standards are represented by the horizontal red dashed lines.

Although MMM underestimates O_3 at night (cf. Section 5.1), the modeled MDA8-O_3 concentration leads to an overestimation of the number of O_3 alerts (10 modeled alerts for the first and 10 for the third period vs. 9 and 7 observed alerts, respectively). Consequently, the FAR is 10% for the first period and 30% for the third period, and since all alerts are predicted, the POD is 100%. The absence of alerts in the second period is also well predicted. The performance of MMM for O_3 alerts is therefore considered satisfactory.

The MMM consistently overestimates hourly $PM_{2.5}$ concentrations for all three periods (Figure S13 in Supporting Information S1), leading to an overestimation of daily average $PM_{2.5}$ concentrations. Consequently, alerts for high $PM_{2.5}$ concentrations occur too frequently during these periods. Nevertheless, the POD is high, reaching 100% for the first and third periods, indicating that all alerts are captured (Figure 12). However, alerts linked to biomass burning pollution events, occurring on August 9, 10, 13, 17, and 18 (cf. Section 4.2), are less accurately reproduced, resulting in a POD of 60% and the highest FAR of 62% across the three periods.

In terms of alerts, the MMM shows satisfying performance for O_3, and intermediate for $PM_{2.5}$ due to its overestimation by the MMM and also to the difficulty to model the biomass burning pollution events. In conclusion, the performance of the regional model ensemble is promising for the development of the air quality forecasting system.

6. 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 run by four institutions over three 15-day periods that include specific air pollution events. Focusing on the most urbanized area, we show that the median of the regional model ensemble, even with the small number of models considered, performs well for O_3 (outperforming all individual regional simulations and the global forecasts by NCAR and ECMWF), although the performance for NOx is moderate, which is related to the large inter-model variability.

For most of the pollutants considered, the performance of the median of the regional model ensemble is the best, since it benefits from the variability among the models due to the different model configurations. In addition, there is a strong overestimation of the level of oxidants (defined as $Ox = O_3 + NO_2$) in the metropolitan area of São Paulo of 10 ppb. Our results suggest that the treatment of anthropogenic emissions is an important factor in explaining the overestimation of modeled NOx and the large inter-model variability of modeled NO and NO_2 concentrations. The day to night transition is particularly biased, which may be related to the absence of the urban heat effect in the models. Our results also show the difficulty for the models to represent the long-range transport of biomass burning pollution, which strongly affects the air quality in São Paulo for a few days.

Many factors influence the performance of the regional model ensemble. For example, the model configurations for domain size and horizontal resolution were not constrained for this study. This choice is limited by the available computational time. On the one hand, the finest possible resolution is desired for the center of São Paulo. On the other hand, a large domain is needed to integrate the different pollutant sources, such as agricultural fires, which are important on a regional scale. The use of a larger domain with finer resolution, as well as more sophisticated chemical or aerosol schemes, which would cost more computational time, may not be a priority since the modeled biases are mostly related to primary emissions, for which there is an important potential for improving model performance.

In perspective, a similar intercomparison study focusing on the drivers of oxidant levels and associated aerosol composition in the PBL of São Paulo is particularly needed to complement this study by understanding the sensitivity of model outputs related to anthropogenic and biogenic emissions, urban dynamics, aerosol and gas phase chemistry, removal processes, and radiation balance.

Data Availability Statement

- 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/, last access: 4 January 2024 [Dataset];
- For ECMWF–CAMS, data are available through this website: https://ads.atmosphere.copernicus.eu/cdsapp#!/dataset/cams-global-atmospheric-composition-forecasts, last access: 4 November 2022 [Dataset];
- For NCAR–CAMchem (Buchholz et al., 2019), data are available through this website: https://www.acom.ucar.edu/cam-chem/cam-chem.shtml, last access: 1 May 2023 [Dataset].

Availability of model data: Upon reasonable request, the authors will provide model data.

References

Albuquerque, T. T. A., West, J., de Andrade, M. F., Ynoue, R. Y., Andreão, W. L., dos Santos, F. S., et al. (2019). Analysis of pm2.5 concentrations under pollutant emission control strategies in the metropolitan area of São Paulo, Brazil. *Environmental Science and Pollution Research*, *26*(32), 33216–33227. https://doi.org/10.1007/s11356-019-06447-6

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

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

Andreae, M. O. (2019). Emission of trace gases and aerosols from biomass burning—An updated assessment. *Atmospheric Chemistry and Physics*, *19*(13), 8523–8546. https://doi.org/10.5194/acp-19-8523-2019

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

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

Beringui, K., Justo, E. P. S., Falco, A. D., Santa-Helena, E., Rocha, W. F. C., Deroubaix, A., & Gioda, A. (2022). Assessment of air quality changes during COVID-19 partial lockdown in a Brazilian metropolis: From lockdown to economic opening of Rio de Janeiro, Brazil. *Air Quality, Atmosphere & Health*, *15*(7), 1205–1220. https://doi.org/10.1007/s11869-021-01127-2

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., et al. (2019). 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. https://doi.org/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). Disentangling vehicular emission impact on urban air pollution using ethanol as a tracer. *Scientific Reports*, *8*(1), 10679. https://doi.org/10.1038/s41598-018-29138-7

Buchholz, R. R., Emmons, L. K., Tilmes, S., & Team, T. C. D. (2019). CESM2.1/CAM-CHEM instantaneous output for boundary conditions [Dataset]. NCAR. https://doi.org/10.5065/NMP7-EP60

Byun, D., & Schere, K. L. (2006). 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*(2), 51–77. https://doi.org/10.1115/1.2128636

Carvalho, V. S. B., Freitas, E. D., Martins, L. D., Martins, J. A., Mazzoli, C. R., & de Fátima Andrade, M. (2015). Air quality status and trends over the metropolitan area of São Paulo, Brazil as a result of emission control policies. *Environmental Science & Policy*, *47*, 68–79. https://doi.org/10.1016/j.envsci.2014.11.001

CETESB. (2022). São Paulo air quality measurment network [Dataset]. Companhia Ambiental do Estado de SAo 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

Chin, M., Ginoux, P., Kinne, S., Torres, O., Holben, B. N., Duncan, B. N., et al. (2002). Tropospheric aerosol optical thickness from the gocart model and comparisons with satellite and sun photometer measurements. *Journal of the Atmospheric Sciences*, *59*(3), 461–483. https://doi.org/10.1175/1520-0469(2002)059⟨0461:TAOTFT⟩2.0.CO;2

Chiquetto, J. B., Ynoue, R. Y., Ibarra-Espinosa, S. A., Ribeiro, F. N. D., Cabral-Miranda, W., & Silva, M. E. S. (2020). Ozone pollution and urban mobility scenarios in the São Paulo megacity. *Ambiente & Sociedade*, *23*, 1–23. https://doi.org/10.1590/1809-4422asoc20190008r2vu2020l6ao

Crippa, M., Guizzardi, D., Muntean, M., Schaaf, E., Dentener, F., van Aardenne, J. A., et al. (2018). Gridded emissions of air pollutants for the period 1970–2012 within EDGAR v4.3.2. *Earth System Science Data*, *10*(4), 1987–2013. https://doi.org/10.5194/essd-10-1987-2018

Crippa, M., Solazzo, E., Huang, G., Guizzardi, D., Koffi, E., Muntean, M., et al. (2020). High resolution temporal profiles in the emissions database for global atmospheric research. *Scientific Data*, *7*(1), 121. https://doi.org/10.1038/s41597-020-0462-2

de Mesnard, L. (2013). Pollution models and inverse distance weighting: Some critical remarks. *Computers & Geosciences*, *52*, 459–469. https://doi.org/10.1016/j.cageo.2012.11.002

Deroubaix, A., Brasseur, G., Gaubert, B., Labuhn, I., Menut, L., Siour, G., & Tuccella, P. (2021). Response of surface ozone concentration to emission reduction and meteorology during the COVID-19 lockdown in Europe. *Meteorological Applications*, *5*(3), 28. https://doi.org/10.1002/met.1990

Duarte, E. D. S. F., Franke, P., Lange, A. C., Friese, E., da Silva Lopes, F. J., da Silva, J. J., et al. (2021). Evaluation of atmospheric aerosols in the metropolitan area of São Paulo simulated by the regional EURAD-IM model on high-resolution. *Atmospheric Pollution Research*, *12*(2), 451–469. https://doi.org/10.1016/j.apr.2020.12.006

ECMWF. (2023). Copernicus atmosphere monitoring service [Dataset]. European Center for Medium-Range Weather Forecasts. Retrieved from https://ads.atmosphere.copernicus.eu/cdsapp#!/dataset/cams-global-atmospheric-composition-forecasts

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

Elbern, H., Strunk, A., Schmidt, H., & Talagrand, O. (2007). Emission rate and chemical state estimation by 4-dimensional variational inversion. *Atmospheric Chemistry and Physics*, *7*(14), 3749–3769. https://doi.org/10.5194/acp-7-3749-2007

Emery, C., Liu, Z., Russell, A. G., Odman, M. T., Yarwood, G., & Kumar, N. (2017). Recommendations on statistics and benchmarks to assess photochemical model performance. *Journal of the Air & Waste Management Association*, *67*(5), 582–598. https://doi.org/10.1080/10962247.2016.1265027

Emmons, L. K., Walters, S., Hess, P. G., Lamarque, J.-F., Pfister, G. G., Fillmore, D., et al. (2010). Geoscientific model development description and evaluation of the model for ozone and related chemical tracers, version 4 (MOZART-4). *Geoscientific Model Development*, *3*(1), 43–67. https://doi.org/10.5194/gmd-3-43-2010

Fast, J. D., Gustafson, W. I., Easter, R. C., Zaveri, R. A., Barnard, J. C., Chapman, E. G., et al. (2006). 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*, *111*(D21), D21305. https://doi.org/10.1029/2005JD006721

Freitas, E. D., Rozoff, C. M., Cotton, W. R., & Dias, P. L. S. (2007). Interactions of an urban heat island and sea-breeze circulations during winter over the metropolitan area of São Paulo, Brazil. *Boundary-Layer Meteorology*, *122*, 43–65. https://doi.org/10.1007/s10546-006-9091-3

Galmarini, S., Bianconi, R., Addis, R., Andronopoulos, S., Astrup, P., Bartzis, J., et al. (2004). Ensemble dispersion forecasting—Part II: Application and evaluation. *Atmospheric Environment*, *38*(28), 4619–4632. https://doi.org/10.1016/j.atmosenv.2004.05.031

Gaubert, B., Bouarar, I., Doumbia, T., Liu, Y., Stavrakou, T., Deroubaix, A., et al. (2021). Global changes in secondary atmospheric pollutants during the 2020 COVID-19 pandemic. *Journal of Geophysical Research: Atmospheres*, *126*(8), e2020JD034213. https://doi.org/10.1029/2020JD034213

Geiger, H., Barnes, I., Bejan, I., Benter, T., & Spittler, M. (2003). The tropospheric degradation of isoprene: An updated module for the regional atmospheric chemistry mechanism. *Atmospheric Environment*, *37*(11), 1503–1519. https://doi.org/10.1016/S1352-2310(02)01047-6

Giglio, L., Justice, C., Boschetti, L., & Roy, D. (2023). Modis/terra+aqua burned area monthly l3 global 500m sin grid v061 [Dataset]. MODIS. https://doi.org/10.5067/MODIS/MCD64A1.061

Godoy-Silva, D., Nogueira, R. F., & Campos, M. L. A. (2017). 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. https://doi.org/10.1016/j.scitotenv.2017.07.145

Granier, C., Darras, S., van der Gon, H. D., Doubalova, J., Elguindi, N., Galle, B., et al. (2019). *The Copernicus atmosphere monitoring service global and regional emissions (April 2019 version)*. Copernicus Atmosphere Monitoring Service. https://doi.org/10.24380/d0bn-kx16

Grell, G., & Baklanov, A. (2011). Integrated modeling for forecasting weather and air quality: A call for fully coupled approaches. *Atmospheric Environment*, *45*(38), 6845–6851. https://doi.org/10.1016/j.atmosenv.2011.01.017

Grell, G. A., & Dévényi, D. (2002). A generalized approach to parameterizing convection combining ensemble and data assimilation techniques. *Geophysical Research Letters*, *29*(14), 38-1–38-4. https://doi.org/10.1029/2002GL015311

Grell, G. A., & Freitas, S. R. (2014). A scale and aerosol aware stochastic convective parameterization for weather and air quality modeling. *Atmospheric Chemistry and Physics*, *14*(10), 5233–5250. https://doi.org/10.5194/acp-14-5233-2014

Grell, G. A., Peckham, S. E., Schmitz, R., McKeen, S. A., Frost, G., Skamarock, W. C., & Eder, B. (2005). Fully coupled "online" chemistry within the WRF model. *Atmospheric Environment*, *39*(37), 6957–6975. https://doi.org/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. https://doi.org/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*(1–4), 173–200. https://doi.org/10.1007/BF01044160

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

Hong, S.-Y., Dudhia, J., & Chen, S.-H. (2004). A revised approach to ice microphysical processes for the bulk parameterization of clouds and precipitation. *Monthly Weather Review*, *132*, 103–120. https://doi.org/10.1175/1520-0493(2004)132(0103:ARATIM)2.0.CO;2

Hong, S.-Y., & Lim, J. (2006). The WRF single-moment 6-class microphysics scheme (WSM6). *Journal of the Korean Meteorological Society*, *42*, 129–151.

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. https://doi.org/10.1175/MWR3199.1

Hu, X.-M., Klein, P. M., & Xue, M. (2013). Evaluation of the updated YSU planetary boundary layer scheme within WRF for wind resource and air quality assessments. *Journal of Geophysical Research: Atmospheres*, *118*(18), 10490–10505. https://doi.org/10.1002/jgrd.50823

Huneeus, N., van der Gon, H. D., Castesana, P., Menares, C., Granier, C., Granier, L., et al. (2020). Evaluation of anthropogenic air pollutant emission inventories for South America at national and city scale. *Atmospheric Environment*, *235*, 117606. https://doi.org/10.1016/j.atmosenv.2020.117606

Im, U., Bianconi, R., Solazzo, E., Kioutsioukis, I., Badia, A., Balzarini, A., et al. (2015). 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. https://doi.org/10.1016/j.atmosenv.2014.09.042

Inness, A., Ades, M., Agusti-Panareda, A., Barré, J., Benedictow, A., Blechschmidt, A.-M., et al. (2019). The CAMS reanalysis of atmospheric composition. *Atmospheric Chemistry and Physics*, *19*(6), 3515–3556. https://doi.org/10.5194/acp-19-3515-2019

Jacob, D. J. (1999). *Introduction to atmospheric chemistry*. Princeton university press.

Janssens-Maenhout, G., Crippa, M., Guizzardi, D., Dentener, F., Muntean, M., Pouliot, G., et al. (2015). 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. https://doi.org/10.5194/acp-15-11411-2015

Kaiser, J. C., Hendricks, J., Righi, M., Riemer, N., Zaveri, R. A., Metzger, S., & Aquila, V. (2014). The MESSy aerosol submodel MADE3 (v2.0b): Description and a box model test. *Geoscientific Model Development*, *7*(3), 1137–1157. https://doi.org/10.5194/gmd-7-1137-2014

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

Khan, M. A. H., Holland, R., Foulds, A., Matthews, J. C., Panditharatne, S., Jenkin, M. E., et al. (2021). Investigating the background and local contribution of the oxidants in London and Bangkok. *Faraday Discussions*, *226*, 515–536. https://doi.org/10.1039/D0FD00086H

Mailler, S., Khvorostyanov, D., & Menut, L. (2013). Impact of the vertical emission profiles on background gas-phase pollution simulated from the EMEP emissions over Europe. *Atmospheric Chemistry and Physics*, *13*(12), 5987–5998. https://doi.org/10.5194/acp-13-5987-2013

Marécal, V., Peuch, V.-H., Andersson, C., Andersson, S., Arteta, J., Beekmann, M., et al. (2015). A regional air quality forecasting system over Europe: The MACC-II daily ensemble production. *Geoscientific Model Development*, *8*(9), 2777–2813. https://doi.org/10.5194/gmd-8-2777-2015

Marsh, D. R., Mills, M. J., Kinnison, D. E., Lamarque, J.-F., Calvo, N., & Polvani, L. M. (2013). Climate change from 1850 to 2005 simulated in CESM1(WACCM). *Journal of Climate*, *26*(19), 7372–7391. https://doi.org/10.1175/JCLI-D-12-00558.1

Martins, L. D., Hallak, R., Alves, R. C., de Almeida, D. S., Squizzato, R., Moreira, C. A., et al. (2018). Long-range transport of aerosols from biomass burning over southeastern South America and their implications on air quality. *Aerosol and Air Quality Research*, *18*(7), 1734–1745. https://doi.org/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*(1/2), 108. https://doi.org/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*(D14), 16663–16682. https://doi.org/10.1029/97JD00237

Monache, L. D., Deng, X., Zhou, Y., & Stull, R. (2006). Ozone ensemble forecasts: 1. A new ensemble design. *Journal of Geophysical Research*, *111*(D5), D05307. https://doi.org/10.1029/2005JD006310

Monks, P. S., Archibald, A. T., Colette, A., Cooper, O., Coyle, M., Derwent, R., et al. (2015). Tropospheric ozone and its precursors from the urban to the global scale from air quality to short-lived climate forcer. *Atmospheric Chemistry and Physics*, *15*, 8889–8973. https://doi.org/10.5194/acp-15-8889-2015

Moreira, G. D. A., da Silva Andrade, I., Cacheffo, A., da Silva Lopes, F. J., Yoshida, A. C., Gomes, A. A., et al. (2021). 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*(2), 425. https://doi.org/10.3390/s21020425

Moreira, G. D. A., Guerrero-Rascado, J. L., Benavent-Oltra, J. A., Ortiz-Amezcua, P., Román, R., Bedoya-Velásquez, A. E., et al. (2019). 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*(2), 1263–1280. https://doi.org/10.5194/acp-19-1263-2019

Morrison, H., Thompson, G., & Tatarskii, V. (2009). 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. https://doi.org/10.1175/2008MWR2556.1

Nakada, L. Y. K., & Urban, R. C. (2020). Covid-19 pandemic: Impacts on the air quality during the partial lockdown in São Paulo state, Brazil. *Science of the Total Environment*, *730*, 139087. https://doi.org/10.1016/j.scitotenv.2020.139087

NCEP. (2022). ds083.3 [Dataset]. US National Centers for Environmental Prediction. Retrieved from https://10.5065/D65Q4T4Z

Parrish, D. D., Xu, J., Croes, B., & Shao, M. (2016). Air quality improvement in Los Angeles—Perspectives for developing cities. *Frontiers of Environmental Science & Engineering*, *10*(5), 11. https://doi.org/10.1007/s11783-016-0859-5

Pedruzzi, R., Baek, B. H., Henderson, B. H., Aravanis, N., Pinto, J. A., Araujo, I. B., et al. (2019). 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*(16), 16125–16144. https://doi.org/10.1007/s11356-019-04953-1

Pereira, G. M., da Silva Caumo, S. E., Grandis, A., do Nascimento, E. Q. M., Correia, A. L., de Melo Jorge Barbosa, H., et al. (2021). Physical and chemical characterization of the 2019 "black rain" event in the metropolitan area of São Paulo, Brazil. *Atmospheric Environment*, *248*, 118229. https://doi.org/10.1016/j.atmosenv.2021.118229

Petersen, A. K., Brasseur, G. P., Bouarar, I., Flemming, J., Gauss, M., Jiang, F., et al. (2019). Ensemble forecasts of air quality in eastern China-part 2: Evaluation of the marcopolo-panda prediction system, version 1. *Geoscientific Model Development*, *12*(3), 1241–1266. https://doi.org/10.5194/gmd-12-1241-2019

Peuch, V. H., Engelen, R., Rixen, M., Dee, D., Flemming, J., Suttie, M., et al. (2022). The Copernicus atmosphere monitoring service from research to operations. *Bulletin of the American Meteorological Society*, *103*(12), E2650–E2668. https://doi.org/10.1175/BAMS-D-21-0314.1

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

Riccio, A., Giunta, G., & Galmarini, S. (2007). Seeking for the rational basis of the median model: The optimal combination of multi-model ensemble results. *Atmospheric Chemistry and Physics*, *7*(24), 6085–6098. https://doi.org/10.5194/acp-7-6085-2007

Rolph, G., Stein, A., & Stunder, B. (2017). Real-time environmental applications and display system: Ready. *Environmental Modelling & Software*, *95*, 210–228. https://doi.org/10.1016/j.envsoft.2017.06.025

Schuch, D., de Freitas, E. D., Espinosa, S. I., Martins, L. D., Carvalho, V. S. B., Ramin, B. F., et al. (2019). 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*(31), 31699–31716. https://doi.org/10.1007/s11356-019-06200-z

Shin, H. H., & Hong, S.-Y. (2015). Representation of the subgrid-scale turbulent transport in convective boundary layers at gray-zone resolutions. *Monthly Weather Review*, *143*(1), 250–271. https://doi.org/10.1175/MWR-D-14-00116.1

Solazzo, E., Bianconi, R., Hogrefe, C., Curci, G., Tuccella, P., Alyuz, U., et al. (2017). Evaluation and error apportionment of an ensemble of atmospheric chemistry transport modeling systems: Multivariable temporal and spatial breakdown. *Atmospheric Chemistry and Physics*, *17*(4), 3001–3054. https://doi.org/10.5194/acp-17-3001-2017

Squizzato, R., Nogueira, T., Martins, L. D., Martins, J. A., Astolfo, R., Machado, C. B., et al. (2021). Beyond megacities: Tracking air pollution from urban areas and biomass burning in Brazil. *npj Climate and Atmospheric Science*, *4*(1), 17. https://doi.org/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 Atmospheric Sciences*, *71*(10), 3636–3658. https://doi.org/10.1175/JAS-D-13-0305.1

Vautard, R., Schaap, M., Bergström, R., Bessagnet, B., Brandt, J., Builtjes, P., et al. (2009). Skill and uncertainty of a regional air quality model ensemble. *Atmospheric Environment*, *43*(31), 4822–4832. https://doi.org/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). The fire inventory from ncar (finn): A high resolution global model to estimate the emissions from open burning. *Geoscientific Model Development*, *4*(3), 625–641. https://doi.org/10.5194/gmd-4-625-2011

Wong, D. W., Yuan, L., & Perlin, S. A. (2004). Comparison of spatial interpolation methods for the estimation of air quality data. *Journal of Exposure Science and Environmental Epidemiology*, *14*(5), 404–415. https://doi.org/10.1038/sj.jea.7500338

Wood, E. C., Canagaratna, M. R., Herndon, S. C., Onasch, T. B., Kolb, C. E., Worsnop, D. R., et al. (2010). Investigation of the correlation between odd oxygen and secondary organic aerosol in Mexico City and Houston. *Atmospheric Chemistry and Physics*, *10*(18), 8947–8968. https://doi.org/10.5194/acp-10-8947-2010

Wood, E. C., Herndon, S. C., Onasch, T. B., Kroll, J. H., Canagaratna, M. R., Kolb, C. E., et al. (2009). A case study of ozone production, nitrogen oxides, and the radical budget in Mexico City. *Atmospheric Chemistry and Physics*, *9*(7), 2499–2516. https://doi.org/10.5194/acp-9-2499-2009

Yokelson, R. J., Christian, T. J., Karl, T. G., & Guenther, A. (2008). The tropical forest and fire emissions experiment: Laboratory fire measurements and synthesis of campaign data. *Atmospheric Chemistry and Physics*, *8*(13), 3509–3527. https://doi.org/10.5194/acp-8-3509-2008

Yu, H., Russell, A., Mulholland, J., Odman, T., Hu, Y., Chang, H. H., & Kumar, N. (2018). Cross-comparison and evaluation of air pollution field estimation methods. *Atmospheric Environment*, *179*, 49–60. https://doi.org/10.1016/j.atmosenv.2018.01.045

Zaveri, R. A., Easter, R. C., Fast, J. D., & Peters, L. K. (2008). Model for simulating aerosol interactions and chemistry (mosaic). *Journal of Geophysical Research*, *113*(D13), D13204. https://doi.org/10.1029/2007JD008782

Zaveri, R. A., & Peters, L. K. (1999). A new lumped structure photochemical mechanism for large-scale applications. *Journal of Geophysical Research*, *104*(D23), 30387–30415. https://doi.org/10.1029/1999JD900876