Atmospheric Environment 145 (2016) 365–375

Contents lists available at ScienceDirect

Atmospheric Environment

journal homepage: www.elsevier.com/locate/atmosenv

Prediction of ground-level ozone concentration in São Paulo, Brazil: Deterministic versus statistic models

G. Hoshyaripour [a, b, *], G. Brasseur [b, c], M.F. Andrade [d], M. Gavidia-Calderón [d], I. Bouarar [b], R.Y. Ynoue [d]

[a] Institute of Geophysics, Center for Earth System Research and Sustainability (CEN), Universität Hamburg, Germany
[b] Max Planck Institute for Meteorology, Hamburg, Germany
[c] National Center for Atmospheric Research, Boulder, CO, USA
[d] Institute of Astronomy, Geophysics and Atmospheric Sciences, University of São Paulo, Brazil

HIGHLIGHTS

- Two deterministic and statistic models for ozone prediction are compared.
- Deterministic model predicts the ozone episodes more satisfactorily.
- Statistic model is slightly advantageous only due to its lower computational costs.
- Mechanistic insights are achievable only in the deterministic model.

ARTICLE INFO

Article history:
Received 3 May 2016
Received in revised form
22 September 2016
Accepted 23 September 2016
Available online 24 September 2016

Keywords:
Air quality modeling
WRF-Chem
Neural networks
Southeastern Brazil
Tropospheric ozone
São Paulo

ABSTRACT

Two state-of-the-art models (deterministic: Weather Research and Forecast model with Chemistry (WRF-Chem) and statistic: Artificial Neural Networks: (ANN)) are implemented to predict the ground-level ozone concentration in São Paulo (SP), Brazil. Two domains are set up for WRF-Chem simulations: a coarse domain (with 50 km horizontal resolution) including whole South America (D1) and a nested domain (with horizontal resolution of 10 km) including South Eastern Brazil (D2). To evaluate the spatial distribution of the chemical species, model results are compared to the Measurements of Pollution in The Troposphere (MOPITT) data, showing that the model satisfactorily predicts the CO concentrations in both D1 and D2. The model also reproduces the measurements made at three air quality monitoring stations in SP with the correlation coefficients of 0.74, 0.70, and 0.77 for O_3 and 0.51, 0.48, and 0.57 for NO_x. The input selection for ANN model is carried out using Forward Selection (FS) method. FS-ANN is then trained and validated using the data from two air quality monitoring stations, showing correlation coefficients of 0.84 and 0.75 for daily mean and 0.64 and 0.67 for daily peak ozone during the test stage. Then, both WRF-Chem and FS-ANN are deployed to forecast the daily mean and peak concentrations of ozone in two stations during 5–20 August 2012. Results show that WRF-Chem preforms better in predicting mean and peak ozone concentrations as well as in conducting mechanistic and sensitivity analysis. FS-ANN is only advantageous in predicting mean daily ozone concentrations considering its significantly lower computational costs and ease of development and implementation, compared to that of WRF-Chem.

1. Introduction

Recent industrialization and urbanization in Brazil have notably exacerbated the air pollution with adverse impacts on human health and socio-economic systems (UNEP, 2003; Marcilio and Gouveia, 2007). These negative effects are more pronounced in the Brazil's greatest metropolitan area, São Paulo (SP) that accommodates approximately 20 million people (IBGE, 2010) being continuously exposed to the health risks of air pollution (Gouveia and Fletcher, 1998; Braga et al., 2001; Ribeiro and Cardoso, 2003;

* Corresponding author. Institute of Geophysics, Universität Hamburg, Bundesstr. 55, 20146 Hamburg, Germany.
E-mail address: gholamali.hoshyaripour@uni-hamburg.de (G. Hoshyaripour).

http://dx.doi.org/10.1016/j.atmosenv.2016.09.061
1352-2310/

Martins et al., 2004). The enduring emission reduction policies mainly for the vehicular emissions in SP have improved the air quality with respect to all pollutants except ground-level ozone (and also fine Particulate Matter ($PM_{2.5}$)), which is often well above the national air quality standards (Andrade et al., 2012; Carvalho et al., 2015). Consequently, there is a strong demand for developing ever-better assessment mechanisms and tools to forecast the surface ozone concentrations at different temporal and spatial scales in SP. These tools are expected to assist the policy-making process to minimize the health hazards and socio-economic damages of the air pollution.

The scientific attempts to develop air pollution prediction systems have generally used two distinct approaches: deterministic (or mechanistic) and statistic (or data-driven). While the former deploys mathematical representation of various physical and chemical mechanisms to simulate the atmospheric environment, the latter utilizes the statistical analysis of the observational data to evaluate the behavior of the atmospheric systems (Seinfeld and Pandis, 2006). For SP, both approaches have been used in independent researches that are summarized below.

Bischoff-Gauß et al. (1998) pioneered the deployment of deterministic models to study air pollution transport in SP, which was followed by later studies on SP using three-dimensional photochemical grid models (Andrade et al., 2004; Vivanco and Andrade, 2006). As ground-level ozone became the major air quality concern in SP (CETESB, 2005), researchers focused on predicating the O_3 concentrations using photochemistry models (Freitas et al., 2005; Sánchez-Ccoyllo et al., 2006; Martins and Andrade, 2008). These studies often observed significant prediction errors, which were attributed mainly to the huge uncertainties in precursor concentrations (NO_x and VOCs in emission inventories) or intermediate species involved in the ozone formation process over time (Vivanco and Andrade, 2006; Borges et al., 2012). Recently, Andrade et al. (2015) deployed an enhanced traffic emission inventory (CETESB, 2012) within WRF-Chem (Weather Research and Forecasting System with Chemistry (Grell et al., 2005)) and BRAMS-SPM (Brazilian developments on the Regional Atmospheric Modeling System with Simplified Photochemical Module (Freitas et al., 2005) to predict O_3 together with NO_x and PM in southeastern Brazil focusing on SP. Results show that the improved modeling system captures the general trends of the surface ozone satisfactorily (with overall correlation coefficient (R) of 0.84) but not the peak values. This uncertainty may stem from three major sources: 1) the comparison made between the model results in grid cells and the local measurements at the air quality stations; 2) discounting the emissions from industrial, domestic and energy sectors in SP and 3) chemical and meteorological mechanisms used in the model. These especially account for the complex and nonlinear chemistry of ozone, which is known to be very sensitive to the small changes in not only the precursor concentrations but also the meteorological conditions (Wilson et al., 2012). It is observed that deterministic models tend to severely under-predict peak O_3 values that are of major concern for air quality forecast and control policy applications (Im et al., 2015). These difficulties in deterministic ozone prediction in addition to its extensive computational burdens have motivated the researchers to investigate the statistical air pollution prediction methods, which are arguably easier to develop, set-up and operate with much lower computational costs (Luna et al., 2014).

Unlike to the deterministic air quality models, statistic tools have not been that popular in studying the air pollution in SP. Guardani et al. (1999) conducted the first statistical study to correlate the ozone levels with air quality data in SP using an artificial neural network (ANN) model. Later studies confirmed the capability of statistic models, especially ANN, in predicting the ground-level ozone satisfactorily (Guardani et al., 2003; Guardani and Nascimento, 2004). For instance, Guardani et al. (2003) and later, Borges et al. (2012) utilized ANN to predict ozone at the air quality monitoring stations in SP and reported a strong correlation between the modeled and observed ozone concentrations with $R > 0.90$. However, the major challenge these researches faced were the optimization of the ANN system in terms of the network structure (e.g., number of neurons, number of layers) and selection of the input variables (e.g., type and frequency of the meteorological and air quality data) in order to obtain the highest prediction skills (Borges et al., 2012). The significant impacts of these parameters on general performance of ANN-based models are well studied resulting in design and implementation of helpful preprocessing tools and measures to tackle these challenges. Noori et al. (2010), for example, reported that coupling ANN with linear and nonlinear input selection modules significantly enhance its performance in predicting carbon monoxide (CO) concentrations in the urban atmosphere.

Comparative studies implementing both deterministic and statistic air quality prediction models in SP are lacking. Here, for the first time, the state-of-the-art deterministic and statistic air quality modeling systems are deployed simultaneously to predict the ground-level ozone in SP. The main objective of the research is to compare the performance of deterministic and statistic modeling systems concerning ozone prediction and to evaluate their potential to support policy-making process. In the following sections, first the materials and methods are presented. Then the results are discussed and conclusions are given.

2. Methodology

2.1. Study area

The Metropolitan Area of São Paulo (MASP) is the largest urban area in South America, having its growth associated to the economic importance of the southeast area of Brazil. MASP, composed by 39 municipalities, is among the 10 most populous regions in the world. The region is situated in a plateau with 700 m height and 50 km distance from the coast. Rainy summer and dry winter characterize the climate of the MASP. The average temperature is 19.3 °C and the annual precipitation and sunshine are 1454,8 mm, 1948 h, respectively (Inmet-National Institute of Meteorology-www.inmet.gov.br).

MASP has a fleet of more than seven million motor vehicles (DENATRAN, 2015). About 55% of light-duty vehicles (2,4 million cars) burn a mixture of 78% gasoline and 22% ethanol (called gasohol), 4% (170 thousand cars) use hydrous ethanol (95% ethanol and 5% water), 38% (3,1 million cars) are flex-fuel vehicles that are capable of burning both gasohol and hydrous ethanol and 3% (881 thousand) use diesel (diesel + 5% bio-diesel). About 100 thousand motorcycles significantly contaminate the atmosphere in MASP as only recently the emission control program was implemented on them. According to the State Environmental Agency (CETESB, 2014) the vehicular fleet is responsible for the emission of about 97% of CO, 87% of VOCs and 80% of NO_x emissions in 2013.

MASP is also affected by biomass burning from the agricultural areas of the state, mainly during the harvesting of the sugar cane to produce alcohol and sugar. During the dry season it is also common to have the burning of waste in the suburbs. The industrial emission inventory is not generally well described for the whole region. Most of the industries are located at the Baixada Santista and Vale do Paraíba Metropolitan Areas.