Atmospheric Environment 141 (2016) 422–429

Contents lists available at ScienceDirect

Atmospheric Environment

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

# Uncertainty of modelled urban peak $O_3$ concentrations and its sensitivity to input data perturbations based on the Monte Carlo analysis

Andrea L. Pineda Rojas [a,*], Laura E. Venegas [b], Nicolás A. Mazzeo [b]

[a] Centro de Investigaciones del Mar y la Atmósfera (CIMA/CONICET-UBA), DCAO/FCEN, UMI-IFAECI/CNRS, Ciudad Universitaria, Pabellón II, Piso 2. 1428, Buenos Aires, Argentina
[b] Department of Chemical Engineering, Avellaneda Regional Faculty, National Technological University, CONICET, Av. Ramón Franco 5050, 1874, Avellaneda, Buenos Aires, Argentina

## HIGHLIGHTS

- Modelled peak $O_3$ concentrations in the MABA during the summer are analysed.
- Larger uncertainty levels are associated with larger ozone concentrations.
- The uncertainty contributions from the model input variables vary spatially.
- That of the regional background $O_3$ concentration dominates at all analysed receptors.
- Model sensitivity responses have similarities with those obtained with complex models.

## ARTICLE INFO

*Article history:*
Received 15 January 2016
Received in revised form
23 June 2016
Accepted 7 July 2016
Available online 9 July 2016

*Keywords:*
Air quality
Model uncertainty
Sensitivity
Ozone
Monte Carlo analysis

## ABSTRACT

A simple urban air quality model [MODelo de Dispersión Atmosférica Ubana – Generic Reaction Set (DAUMOD-GRS)] was recently developed. One-hour peak $O_3$ concentrations in the Metropolitan Area of Buenos Aires (MABA) during the summer estimated with the DAUMOD-GRS model have shown values lower than 20 ppb (the regional background concentration) in the urban area and levels greater than 40 ppb in its surroundings. Due to the lack of measurements outside the MABA, these relatively high ozone modelled concentrations constitute the only estimate for the area. In this work, a methodology based on the Monte Carlo analysis is implemented to evaluate the uncertainty in these modelled concentrations associated to possible errors of the model input data. Results show that the larger 1-h peak $O_3$ levels in the MABA during the summer present larger uncertainties (up to 47 ppb). On the other hand, multiple linear regression analysis is applied at selected receptors in order to identify the variables explaining most of the obtained variance. Although their relative contributions vary spatially, the uncertainty of the regional background $O_3$ concentration dominates at all the analysed receptors (34.4–97.6%), indicating that their estimations could be improved to enhance the ability of the model to simulate peak $O_3$ concentrations in the MABA.



## 1. Introduction

Ozone ($O_3$) is known among the air pollutants to have great potential to cause adverse effects on human health and the environment (WHO, 2014). Ground-level $O_3$ concentrations are increasing in many cities of the world and their surroundings due to changing emissions of nitrogen oxides ($NO_x = NO + NO_2$) and volatile organic compounds (VOCs) from human activities and other environmental factors such as temperature (e.g., Lee et al., 2014; Paoletti et al., 2014; Wang et al., 2012). The typical horizontal distribution of ozone concentration shows a relative minimum in the urban areas (due to $O_3$ titration by NO in zones of large $NO_x$ emissions) and a maximum several kilometres downwind of cities (as a consequence of an "optimal" VOCs/$NO_x$ concentration ratio for ozone formation (Calfapietra et al., 2013)). In order to assure that its levels remain relatively low, the spatio-temporal distribution of

* Corresponding author.
*E-mail address:* pineda@cima.fcen.uba.ar (A.L. Pineda Rojas).

http://dx.doi.org/10.1016/j.atmosenv.2016.07.020
1352-2310/

ozone concentrations in urban areas should be evaluated. In places where this is not satisfied, pollution mitigation actions must be taken to reduce ozone concentrations. This type of air quality assessment is achieved through the combined use of observations providing precise information at specific monitoring sites and air quality models which allow the estimation of the concentration distribution in a given area. Since these models provide a link between emissions, meteorology and concentrations, they are also widely used by researchers to study a number of air quality issues such as process analysis (e.g., Wang et al., 2012), source apportionment (e.g., Strong et al., 2013) and the possible impact of climate change on $O_3$ levels (e.g., Athanassiadou et al., 2010), to name a few.

When modelled concentrations are used either for policy decision making or for scientific purposes, a measure of their reliability can be required. This is given by the model performance evaluation which involves different steps of varying complexity depending on the model type and specific purpose. Three common fundamental aspects of the performance evaluation of an air quality model are: the scientific evaluation, the statistical evaluation and the probabilistic evaluation (e.g., Chang and Hanna, 2005; Derwent et al., 2010). The former examines model algorithms and model assumptions in detail. The statistical evaluation refers to the comparison between modelled and observed concentrations, and plays an essential role since it provides a measure of the "total error" of the model. Finally, the probabilistic sensitivity/uncertainty evaluation aims to capture the uncertainty in model results introduced by variabilities of a specific parameter, variable, parameterisation, or a combination of them, etc. In air quality models applications, the uncertainty of the model input data is considered to be the dominant source of error (Russell and Dennis, 2000). Uncertainty and sensitivity analysis offers a tool through which the uncertainty of modelled pollutant concentrations associated to input data uncertainties can be evaluated. This is critical for policy decision makers since air quality management must be based on a range of probable results rather than on a single value whose occurrence is subject to error. On the other hand, a good understanding of the key variables associated with model output uncertainties is fundamental. This allows modellers and scientists to gain insight into model strengths and weaknesses, as well as into the variables or parameters whose estimations should be improved in order to enhance model capabilities. There are different methodologies available throughout the literature to apply uncertainty and sensitivity analysis with air quality models (see Borrego et al., 2008; Refsgaard et al., 2007), where the Monte Carlo (MC) analysis combined with multiple linear regression (MLR) analysis is one of the most widely used methods to study the uncertainty of modelled pollutant concentrations (e.g., Bergin et al., 1999; Hanna et al., 1998, 2007; Moore and Londergan, 2001; Rodriguez et al., 2007; Tang et al., 2010). Other applications of the Monte Carlo analysis include the uncertainty assessment of the impact of different pollution mitigation strategies on peak $O_3$ levels (e.g., Derwent and Murrells, 2013) and the use of different sets of observations to estimate representative average pollutant concentrations (e.g., Tan et al., 2014).

The DAUMOD-GRS (MODelo de Dispersión Atmosférica Urbana – Generic Reaction Set) model (Pineda Rojas and Venegas, 2013a) is a simple atmospheric dispersion model that allows estimation of ground-level $O_3$ concentrations resulting from area source emissions of $NO_x$ and VOCs in urban areas. It is based on the bidimensional equation of diffusion and employs a simplified photochemical scheme of the $NO_x$-VOCs-$O_3$ interactions. The model has been statistically evaluated using observations of nitrogen dioxide and ozone concentrations from twenty monitoring sites of the Metropolitan Area of Buenos Aires (MABA), Argentina, and has shown an acceptable performance (Pineda Rojas and Venegas, 2013b; Pineda Rojas, 2014). A series of features of the MABA (3830 $km^2$, ~13 million inhabitants), such as its flat terrain location or that it is surrounded by non-urban areas, support the use of simple models. The aim of the present work is to perform a probabilistic evaluation to analyse the uncertainty in modelled $O_3$ concentrations in the MABA associated to possible errors of the DAUMOD-GRS input variables. It is worth noting that previously simulated 1-h peak $O_3$ levels in the region during summer for the first time (Pineda Rojas and Venegas, 2013b) were found to be below the air quality standard for the region (120 ppb); however, values above 40 ppb [i.e., the threshold used in other parts of the world to protect vegetation (Paoletti and Manning, 2007)] were simulated for the surroundings of the MABA. Since there are no measurements outside the MABA to compare these potentially high modelled $O_3$ concentrations with, a probabilistic evaluation of these concentrations becomes critical. In this work, we implement a methodology based on the MC and MLR techniques to perform an uncertainty and sensitivity analysis of the DAUMOD-GRS model. The objectives of the present paper are 1) to evaluate the uncertainty of modelled 1-h peak $O_3$ concentrations at each receptor in the MABA region during the summer associated to uncertainties in the input variables, and 2) to determine the subset of variables explaining most of the obtained variance.

## 2. Methodology

The Monte Carlo (MC) analysis consists of performing a relatively large number of simulations (called MC runs) using different combinations of alternative values for model input variables, which are randomly obtained from their probability density functions and uncertainty ranges. As a result, a set of probable values of the modelled pollutant concentration is obtained, from which a number of statistics can be computed. The main advantages of the MC analysis are its general applicability and the relatively few assumptions that it needs. Common drawbacks are that probability density functions and uncertainty ranges of the input variables are often unknown, and that the method generates a huge amount of data that is usually difficult to analyse. The multiple linear regression (MLR) analysis offers a way of characterising the input-output transformations (i.e., the relationship between the perturbed input variables and the pollutant concentrations obtained from the MC runs) so that the uncertainty contribution of each input variable to the total uncertainty of the modelled concentration can be easily estimated. Here, the MC analysis is implemented to evaluate the uncertainty of modelled 1-h peak $O_3$ concentrations at each receptor in the MABA during the summer ($C_{max}$) associated to possible errors in the input variables; while MLR analysis is performed to estimate their relative contributions. The implementation of these techniques with DAUMOD-GRS is based on their previous applications with other air quality models (Bergin et al., 1999; Hanna et al., 1998, 2007; Moore and Londergan, 2001; Rodriguez et al., 2007), with slight modifications as described in the following sections. Section 2.1 comments on the main features of the DAUMOD-GRS model. Section 2.2 describes the implementation of the Monte Carlo analysis with the DAUMOD-GRS model and Section 2.3 the application of the multiple linear regression analysis to estimate the contribution of the uncertainty of input variables to the modelled $C_{max}$ uncertainty.

### 2.1. Model characteristics

The DAUMOD-GRS model couples the MODelo de DispersiónAtmosférica Urbana (DAUMOD) with the Generic Reaction Set (GRS). The DAUMOD model (Mazzeo and Venegas, 1991) is