18th International Conference on
Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes
9-12 October 2017, Bologna, Italy

CLUSTERING OF ATMOSPHERIC AND EMISSION CONDITIONS THAT LEAD TO MODELLED PEAK OZONE CONCENTRATIONS

Andrea L. Pineda Rojas[1] *and Nicolás A. Mazzeo*[2]

[1]Centro de Investigaciones del Mar y la Atmósfera, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, CONICET, UBA, Buenos Aires, Argentina
[2]Departamento de Ingeniería Química, Facultad Regional Avellaneda, Universidad Tecnológica Nacional, CONICET, UTN, Buenos Aires, Argentina

Abstract: An application of a simple urban air quality model (DAUMOD-GRS) shows that summer maximum O_3 hourly concentrations (C_{max}) above 40 ppb occur outside the Metropolitan Area of Buenos Aires (MABA) where the absence of observations impedes model testing. In addition, those relatively high values present the greatest model uncertainty caused by possible errors in model input variables. In order to tackle this issue, a probability assessment of C_{max} values exceeding 40 ppb is performed applying a Monte Carlo analysis. On the other hand, a non-hierarchical (k-means) clustering analyses is applied to analyse the Monte Carlo outcomes. Results show three main clusters with a marked spatial distribution resembling that of the ozone precursor species emissions, which highlights an important role of the emissions on the regimes under which modelled C_{max} values in the MABA can occur.

Key words: *air quality modelling, Buenos Aires, clustering analysis, Monte Carlo analysis, ozone.*

INTRODUCTION

Ozone (O_3) is among the air pollutants of increasing concern worldwide. Due to its photochemical nature, it usually maximises in summer and outside urban zones where an optimal ratio between their precursor species concentrations occurs (e.g., Calfapietra et al., 2013). Hence, the evaluation of summer O_3 concentrations constitutes an important aspect of any air quality assessment around urban areas.

The Metropolitan Area of Buenos Aires (MABA) concentrates nearly 30 percent (13 million inhabitants) of Argentine's population in 3830 km^2. It is located on a flat terrain and surrounded by non-urban areas and the de la Plata River. A few air sampling campaigns carried out in three urban sites have revealed ozone hourly concentration levels relatively low compared to its Air Quality Standard (120 ppb). In spite of current regulation, ozone has not been measured regularly in the MABA yet, except for one urban-industrialized site since the end of 2015. Therefore, model results constitute the only available estimates of its spatial distribution. In a previous paper, modelled summer peak O_3 hourly concentrations (C_{max}) above 40 ppb [one of the accepted thresholds to protect vegetation] were found to occur outside the MABA (Pineda Rojas and Venegas, 2013). However, those relatively high concentration values were also found to present the greatest uncertainties caused by possible errors in the model input data (Pineda Rojas et al., 2016). In this context, a probability assessment of such exceedances may provide a more robust estimate than a deterministic one (Yegnan et al., 2002).

In this work, a Monte Carlo evaluation of the probability of occurrence of peak O_3 hourly concentrations greater than 40 ppb in the MABA during a typical summer season, using the simple urban-scale atmospheric dispersion model DAUMOD-GRS, is presented. In order to overcome the limitations due to the size of the Monte Carlo outcomes, a non-hierarchical (k-means) clustering analysis is performed aiming to identify the environmental conditions under which C_{max} occurs and to gain insight on the model performance outside the MABA, where the highest values are obtained.

METHODOLOGY

The DAUMOD-GRS model (Pineda Rojas and Venegas, 2013) is an urban-scale atmospheric dispersion model which includes a simplified photochemical scheme (the Generic Reaction Set: GRS). It allows estimation of ground-level urban background concentrations of nitrogen dioxide and ozone resulting from area source emissions of nitrogen oxides (NO_x) and volatile organic compounds (VOC). A detailed description of the model and its performance evaluation in the MABA can be found in Pineda Rojas and Venegas (2013) and in Pineda Rojas (2014).

Probabilistic assessment

In this work, the DAUMOD-GRS model is applied to estimate the horizontal distributions of summer maximum peak ozone hourly concentrations (C_{max}) in the Metropolitan Area of Buenos Aires, applying a Monte Carlo analysis (Pineda Rojas et al., 2016). A large number of simulations (N) in which the input data sets (the meteorological and emission data) are perturbed randomly, is performed. Since the error probability distributions for these data are not known for the MABA, we consider functions and ranges published in the literature (e.g., Hanna et al., 1998). By performing N=100 Monte Carlo runs, the probability of having a value of C_{max} greater than 40 ppb at each receptor is estimated as the number of exceedances obtained divided by N.

Meteorological information from the station located at the domestic airport for the base case (no perturbations) belong to a typical summer (2007), while area source emission data come from the inventory developed for the MABA (Venegas et al., 2011). Nine DAUMOD-GRS input variables are perturbed: wind speed (WS) and direction (DIR), air temperature (T), sky cover (SC), total solar radiation (TSR), atmospheric stability class (KST), local NO_x emission rate (QNO_x), local VOC emission rate (QVOC) and regional background ozone concentration ($[O_3]_r$). Simulations are performed at a spatial resolution of 1 km x 1 km, in domain of 80 km x 75 km. At each run, the C_{max} value, its hour of occurrence and the associated perturbed input variables (10 variables) are stored at 4647 non-water receptors. Then, the Monte Carlo outcomes generate a dataset of 10 x 100 x 4647 = 4,647,000 values. This size clearly limits the direct observation of the data.

Clustering analysis

Clustering analysis aims for an unbiased classification of big datasets into groups containing objects with similar characteristics. The k-means algorithm is one of the most widely used methods for air quality applications (e.g., Beaver and Palazoglu, 2006; Jin et al., 2011). It consists of several steps. First, the number of clusters (k) has to be defined, and then their positions are randomly placed in a M-dimensional space, where M is determined by the number of variables describing the objects. Each object is initially assigned to a cluster based on some measure of the distance between them. Once all objects are distributed among the k clusters, the cluster centres are recalculated by averaging the positions of all members (i.e., all objects within the cluster). The two steps are repeated until they leave clusters unchanged. K-means is an heuristic algorithm that can lead to suboptimal solutions, depending on the initial conditions.

In this work, the Matlab function kmeans is used with k=4, and 100 random initializations are performed to avoid suboptimal solutions. An object is a set of conditions in which C_{max} occurs (its hour of occurrence and the perturbed input variables). Each variable is scaled subtracting its mean and dividing by its standard deviation across the whole modelling domain.

RESULTS

The probability of obtaining values of C_{max} greater than 40 ppb (not shown) is lower than 5% in most of the metropolitan area. A expected, it increases with the distance to city core. Outside the urban area, this probability is greater than 70% and reaches a maximum of 82% at 20 km southwest the MABA.

Figure 1 shows the dominant cluster at each receptor. The four clusters appear to distribute in the MABA as a function of the spatial variation of the NO_x and VOC emission rates. Clusters 1 and 4 are mostly present at places of moderate and high emission rates, respectively; while clusters 2 and 3 occur in

general at receptors with no emissions. From Figure 2, the largest difference between the mean normalised variables of clusters 1 and 4 is found in the emissions. On the other hand, clusters 2 and 3 differ mostly in the mean time of occurrence of C_{max} and in the mean sky cover.

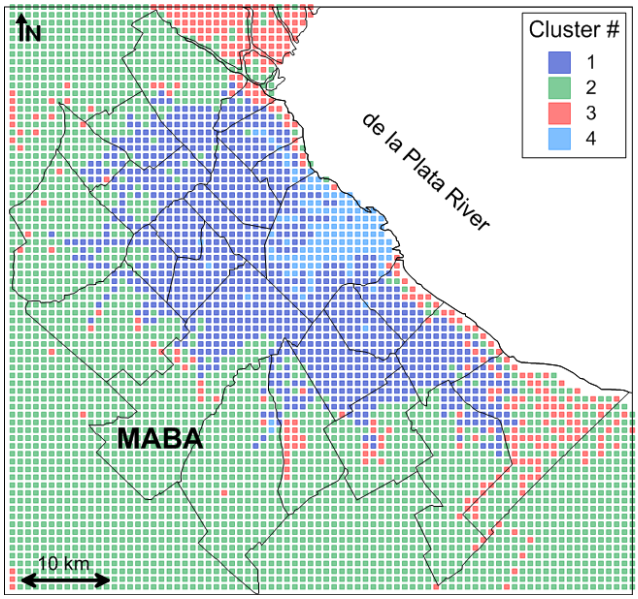

Figure 1. Dominant cluster at each receptor.

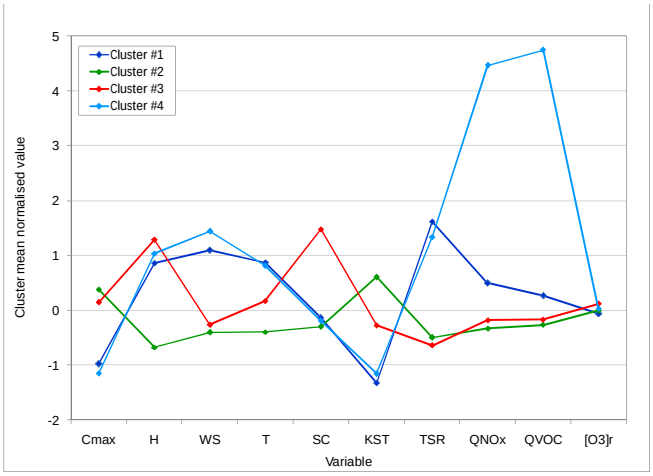

Figure 2. Normalised variables (z-score) averaged for each cluster [C_{max}: summer maximum O_3 hourly concentration, H: hour of occurrence of C_{max}, WS: wind speed, T: air temperature, SC: sky cover, KST: atmospheric stability class, TSR: total solar radiation, QNO_x: NO_x emission rate, QVOC: VOC emission rate, $[O_3]_r$: regional background O_3 concentration]

Table 1 presents the same variables as in Figure 2 with their units. The C_{max} values occur, on average, at 13-15 h for clusters 1, 3 and 4; while it does at 7 h in cluster 2. These differences can also be observed in the mean values of T and KST of the four clusters, and suggest different main drivers on ozone formation: photochemical (clusters 1, 3 and 4) vs. dispersive (cluster 2).

Table 1. Variables from Figure 2 averaged for each cluster.

Cluster #	Objects (%)	C_{max} (ppb)	H	WS (ms^{-1})	T (°C)	SC (okta)	KST	TSR (Wm^{-2})	QNO_x ($gkm^{-2}s^{-1}$)	QVOC ($gkm^{-2}s^{-1}$)	$[O_3]_r$ (ppb)
1	103036 (22)	20.2	13	5.1	27.0	1	2	854.9	1.2	0.6	20.1
2	280765 (60)	32.9	7	1.3	22.4	1	5	166.5	0.1	0.0	20.3
3	68503 (15)	30.8	15	1.7	24.5	4	4	119.8	0.3	0.1	20.6
4	12396 (3)	18.5	14	6.0	26.8	1	2	762.3	6.3	5.1	20.4

The wind roses of the clusters (Figure 3) show that the regions of Figure 1 present values of C_{max} which occur, on average, for different wind directions. At receptors of high emission rates (cluster 4), C_{max} occurs with moderate winds (3.7 ms^{-1}) from the ENE (22%) or with relatively intense winds (7.7-8.2 ms^{-1}) from the SE-SSE (27.1%) sector. At places of moderate emission rates (cluster 1), the winds that are associated to the occurrence of summer ozone peaks come mainly from the S-W sector (43.5% of the time) and have mean intensities varying between 5.1-8.0 ms^{-1}. Finally, in cells with no emissions (clusters 2 and 3), C_{max} is associated to more variable wind directions and low wind conditions (< 2 ms^{-1})

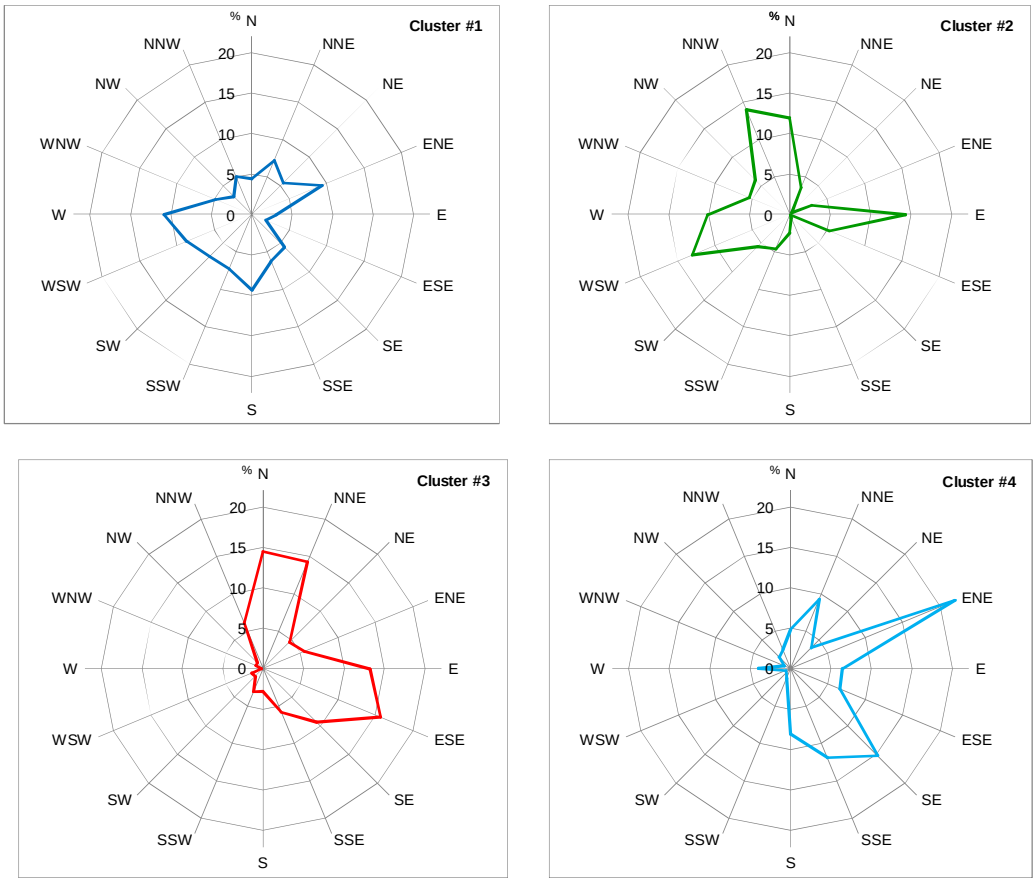

Figure 3. Wind rose of each cluster

CONCLUSIONS

A probabilistic assessment of summer maximum ozone hourly concentrations (C_{max}) greater than 40 ppb is performed for the Metropolitan Area of Buenos Aires (MABA), applying a Monte Carlo analysis with the simple urban-scale atmospheric dispersion model DAUMOD-GRS. Results show very low probabilities in the urban area and values above 70% outside the MABA. A k-means algorithm is applied to analyse the outcomes obtained from the Monte Carlo simulations. Results show three main clusters with a marked spatial distribution resembling that of the ozone precursor species emissions. Two clusters are mostly present in the most urbanised area, where C_{max} occurs around midday hours under conditions of relative high wind speeds. A third cluster is found mainly outside the MABA (no emissions) where the greatest C_{max} values occur, on average, during early morning hours with low wind intensities (< 2 ms^{-1}). The latter shows that the greatest C_{max} values obtained outside the MABA are associated to wind directions either from or to the MABA, suggesting an important role of the "memory effect" of the model.

The results obtained in this work show the potential of combining Monte Carlo simulations with clustering analysis to gain insight from modelled data that are usually not analysed beyond the estimation of uncertainty and may contain valuable information on the modelled pollutant concentration.

REFERENCES

Beaver, S. And A. Palazoglu, 2006: A cluster aggregation scheme for ozone episode selection in the San Francisco, CA Bay Area. *Atmos. Environ.*, **40**, 713–725.

Calfapietra, C., Fares, S., Manes, F., Morani, A., Sgrigna, G., Loreto, F. 2013. Role of Biogenic Volatile Organic Compounds (BVOC) emitted by urban trees on ozone concentration in cities: A review. *Environ. Pollut.* 183, 71–80.

Hanna, S.R., J.C. Chang and M.E. Fernau, 1998: Monte Carlo estimates of uncertainties in predictions by a photochemical grid model (UAM-IV) due to uncertainties in input variables. *Atmos. Environ.* **32** (21), 3619-3628.

Jin, L., R.A. Harley and N.J. Brown, 2011: Ozone pollution regimes modeled for a summer season in California's San Joaquin Valley: A cluster analysis. *Atmos. Environ.*, 45, 4707-4718.

Pineda Rojas, A.L., L.E. Venegas and N.A. Mazzeo, 2016: Uncertainty of modelled urban peak O_3 concentrations and its sensitivity to input data perturbations based on the Monte Carlo analysis. *Atmos. Environ.*, **141**, 422-429.

Pineda Rojas, A.L. 2014: Simple atmospheric dispersion model to estimate hourly ground-level nitrogen dioxide and ozone concentrations at urban scale. *Environ. Modell. Soft.*, **59**, 127-134.

Pineda Rojas, A.L. and L.E. Venegas, 2013: Spatial distribution of ground-level urban background O_3 concentrations in the Metropolitan Area of Buenos Aires, Argentina. *Environ. Pollut.*, **183**, 159-165.

Venegas, L.E., N.A. Mazzeo and A.L. Pineda Rojas, 2011: Chapter 14: Evaluation of an emission inventory and air pollution in the Metropolitan Area of Buenos Aires. In: D. Popovic (ed.) Air Quality-Models and applications, Editorial In-Tech, 261-288.

Yegnan, A., D.G. Williamson and A.J. Graettinger, 2002: Uncertainty analysis in air dispersion modeling. *Environ. Modell. Softw.*, **17**, 639–649.