20th International Conference on
Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes
14-18 June 2021, Tartu, Estonia

EXPLORING ERROR TYPES AND PERFORMANCE OF AN AIR QUALITY MODEL THROUGH CLUSTERING ANALYSIS

Andrea L. Pineda Rojas[1] *and Emilio Kropff*[2]

[1]Centro de Investigaciones del Mar y la Atmósfera (CIMA), UMI-IFAECI/CNRS, FCEyN, Universidad de Buenos Aires – UBA/CONICET, Buenos Aires, Argentina
[2]Fundación Instituto Leloir – IIBBA/CONICET, Buenos Aires, Argentina

Abstract: In this work, a simple approach to identify input data conditions that are associated with different model performance levels is presented. Using four years nitrogen dioxide (NO_2) hourly concentrations measured at three air quality sites in the city of Buenos Aires and DAUMOD-GRS model results, a clustering analysis is applied over three performance metrics (FB, NMSE and R) to group days according to their levels of model performance. Four clusters are found to better describe such differences at the three sites. At the urban background site, wind speed and air temperature present the largest statistical differences between model performance clusters. In turn, at the residential industrial site, clusters show clear significant differences in most meteorological variables which suggest a potential role from the emissions coming from the power plants that are located on the coast. Overall, a better understanding of the DAUMOD-GRS model performance and how it changes with different conditions is obtained.

Key words: *Buenos Aires, clustering analysis, DAUMOD-GRS, model performance evaluation.*

INTRODUCTION

Performance evaluation is a key aspect in the development of air quality models. In a typical evaluation, performance metrics are applied over the complete data set and a single performance measure is obtained. However, when only a few monitoring sites are available, this analysis may not allow for the identification of individual model features. A common alternative is to obtain metrics for different data subsets (e.g., by ranges of meteorological variables). However, this approach has two drawbacks: i) it is very sensitive to the election of subsets and ii) it may not capture different performance degrees resulting from a combination of conditions rather than the occurrence of one parameter or variable only.

The performance of the urban scale atmospheric dispersion model DAUMOD-GRS (Pineda Rojas and Venegas, 2013) has recently been evaluated to estimate nitrogen dioxide (NO_2) concentration using the first available long-term (four-year) air quality record at three sites in the city of Buenos Aires (Pineda Rojas and Borge, 2019). The results show a relatively good ability of the model to estimate hourly NO_2 concentrations at the three sites. In this work, we apply a simple clustering analysis to commonly used metrics using these series, to characterise typical model errors by grouping days according to the performance level. The objective is to assess whether such levels are associated with particular meteorological conditions in order to better understand DAUMOD-GRS model behaviour under different input data conditions.

METHODOLOGY

The DAUMOD-GRS model is applied over the Metropolitan Area of Buenos Aires (3830 km^2) considering four years (2009-2012) of surface hourly meteorological data from the domestic airport and emissions of nitrogen oxides and volatile organic compounds from the high resolution (1 km x 1 km) emissions inventory developed by Venegas et al. (2011). Given that the MABA is surrounded by non-urban areas, clean air concentration values are considered as regional background levels.

Model results are compared with hourly concentration values of NO_2 recorded by the local environmental protection agency (APRA) at three air quality monitoring sites [Parque Centenario (CEN, urban background), Córdoba (COR, urban traffic) and La Boca (LB, residential industrial)]. Three model performance metrics are used for the statistical comparison (Chang and Hanna, 2005): fractional bias (FB), normalised mean square error (NMSE) and correlation coefficient (R). These metrics are computed daily at each monitoring site, considering days having complete data (24 hourly values of NO_2 concentration) within the four-year period. A k-means algorithm (e.g., Pineda Rojas et al., 2019) is applied at each site individually to classify days according to their model performance metrics. The silhouette criterion (Rouseeuw, 1987) is used to determine a suitable number of clusters for the three sites. To simplify the analysis, clusters are ordered from "best" to "worst" performing metrics considering increasing values of the sum:

$$S_j = \overline{|FB|} + \overline{NMSE} + \left(1 - \overline{|R|}\right)$$

where the vertical bars denote absolute value and the over bar indicates the average over all members of cluster j. Once days are labelled (i.e., grouped according to their model performance cluster), the values of input variables (WS: wind speed, WD: wind direction, T: air temperature, SC: sky cover, TSR: solar radiation, KST: PGT atmospheric stability class) are analysed in order to identify whether different model performance levels are associated with different patterns of meteorological input data conditions.

RESULTS

Applying a k-means algorithm, we obtain that k = 4 is an adequate number of clusters for the three sites. **Figure 1** presents box plots of the metrics for each cluster and monitoring site (CEN, COR and LB). Since the analysis is performed individually at each site, their classifications are not comparable between each other. A relatively good separation is obtained at the three sites. At CEN, ”best” performing days with respect to the three metrics are grouped in clusters 1 and 2; while those of ”worse” model performance are divided between cluster 3 (lowest R) and cluster 4 (largest NMSE and absolute value of FB). At COR and LB, ”best” performing days are included in cluster 1; days of lowest correlation are grouped in cluster 2; and the largest differences between model results and observations are concentrated in a relative small group of days in cluster 4 (see **Table 1**).

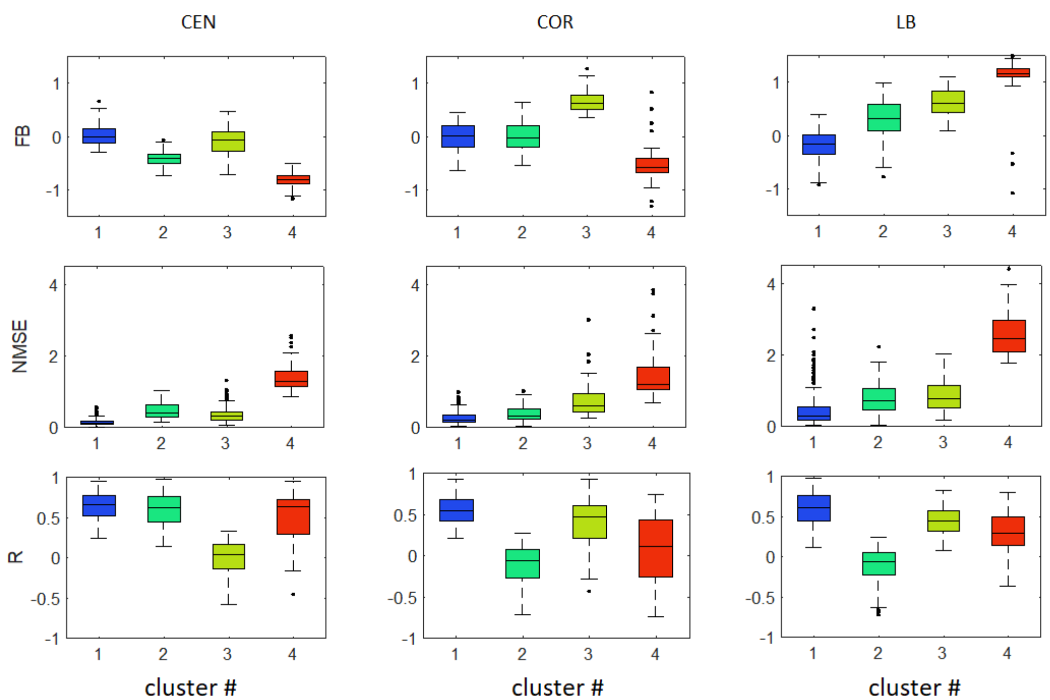
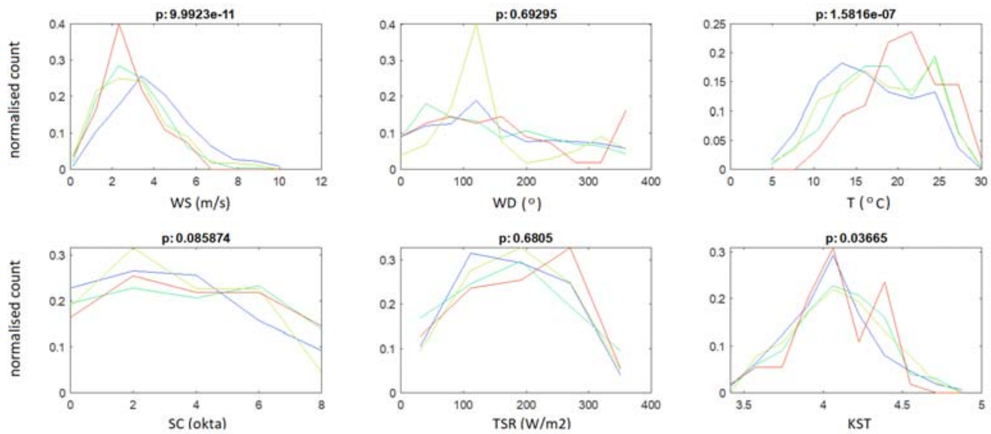


Figure 1. Box plots of three metric (FB, NMSE, R) values by cluster at each air quality monitoring site (CEN: Parque Centenario, COR: Córdoba, LB: La Boca)

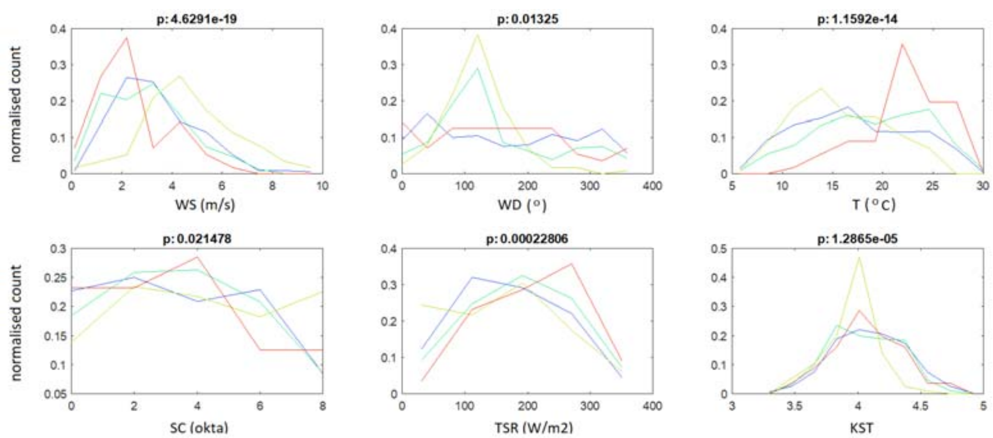
Table 1. Number of days within each cluster

Site	Cluster number				Total
	1	2	3	4	
CEN	325	231	177	55	788
COR	340	255	115	56	766
LB	364	213	265	80	922

Once the algorithm has grouped the days sharing (one or more) metric similarities, we want to understand whether metric differences between clusters (days) are associated with different input data conditions. A first inspection of this can be done by looking at the cluster distributions of daily mean meteorological variables (**Figure 2**). While some overlap is evident, differences between the ”best” and ”worst” performance distributions are observed. As indicated with the p-value obtained by a Kruskal-Wallis test, at CEN and COR, worst performing days (red curves) occur with relatively lower wind speed (WS) and higher air temperature (T). Differences between other meteorological variables are not signifficant at these sites. At LB, in turn, signifficant statistical differences between clusters are found for all variables, except for sky cover. Days with worse model performance to estimate NO_2 concentrations present relatively larger WS, winds from the 1st and 2nd quadrants, higher T and TSR and lower KST values. This suggests a role of non-local sources on model performace at this site.

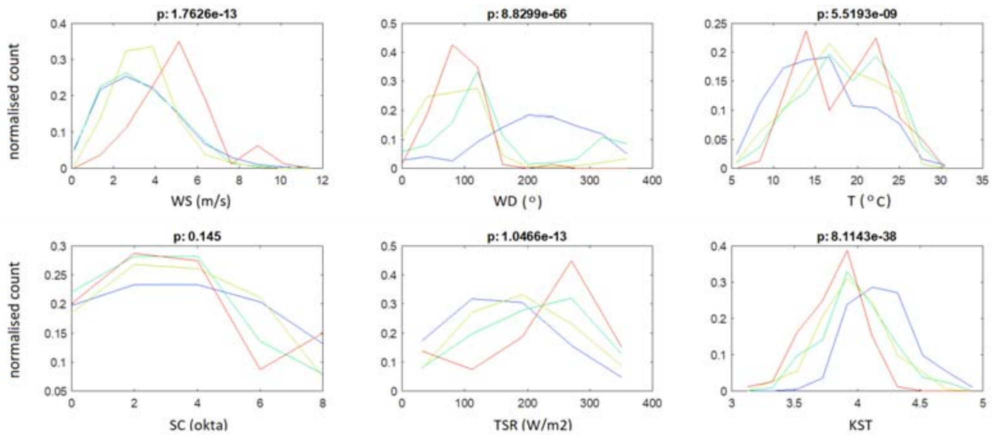


(a) CEN



(b) COR

Figure 2. Cluster distributions of daily mean meteorological variables (WS: wind speed, WD: wind direction, T: air temperature, SC: sky cover, TSR: solar radiation, KST: atmospheric stability class) at each monitoring site: (a) CEN, (b) COR and (c) LB. The largest statistical difference between the cluster distributions is indicated with the p-value (Kruskal-Wallis test).



(c) LB
Figure 2. Cont.

Distributions of clusters over different meteorological variable planes shows considerable superposition (not shown). In turn, when ploting the average hourly ratio (Cm-Co)/(Cm+Co) on the polar plane (X, WD) for variables X = WS, T, SC, TSR and H (hour), some interesting features are found. At CEN and COR, the bivariate polar plot of this ratio varies with the cluster. For example, at CEN, some overestimation is obtained for most meteorological conditions in clusters 2 and 4, while smaller differences between modelled and observed concentrations are found for some wind directions in cluster 3 (not shown). At LB, a relatively large underestimation with winds from the N→SE is obtained for all clusters. This sistematic underprediction and the fact that this occur mostly with relatively larger wind speeds suggest that emissions from distant sources are probably underestimated. This is consistent with previous results (Pineda Rojas et al., 2020) that point to the power plants that are located in the coast of the city.

Finally, the method is used to assess the impact of a previously proposed model change (Pineda Rojas and Borge, 2019) in terms of model performance metrics. As an example, **Figure 3** shows the distributions of days over different metric planes at the CEN site, under conditions of a standard run and when the ”memory effect” of the model is removed. By using the same classification in the two simulations, it is possible to observe whether and how the points (representing days) "move" in the metric planes.

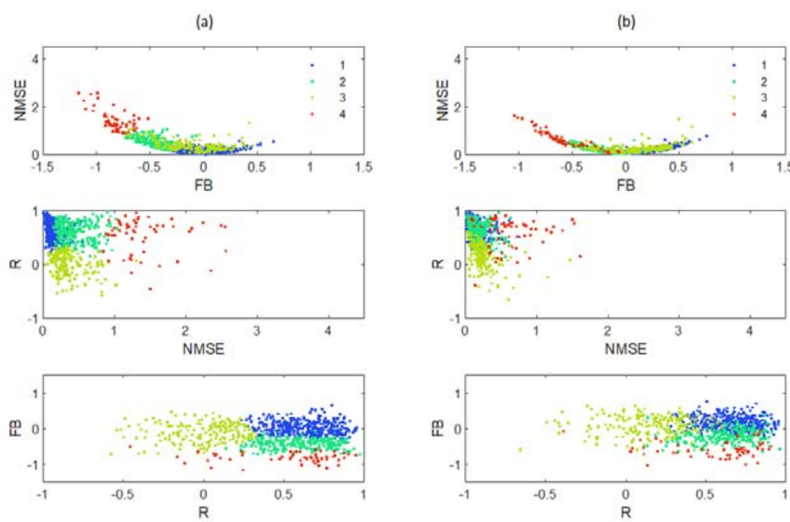


Figure 3. Distributions of days over different metric planes by cluster at CEN, obtained with (a) the standard simulation and (b) removing the memory effect in the standard run.

CONCLUSIONS

A simple approach to analyse the DAUMOD-GRS model performance using long term series of NO_2 hourly concentrations measured at the three air quality sites of the city of Buenos Aires is presented. A k-means algorithm is applied over three relevant performance metrics computed daily in order to classify days based on model performance metric similarities. A good separation is obtained at the three sites. At the urban background site, where the best overall model performance is obtained, results show that the largest statistical differences between ”best” and ”worst” performing days are found between the distributions of wind speed and air temperature daily values. The largest overestimation is concentrated in a relatively small group of days. At the residential industrial site, the clustering analysis shows statistically signifficant inter-cluster differences for most analysed meteorological variables. In this case, distinct meteorological conditions are more clearly associated to ”worst” performing days. Bivariate polar plots show a sistematic underprediction when the wind comes from the coast. This suggests that the contribution of the three power plants located on the coast should be assessed in future modeling studies. Overall, the method allows to better understand how the model performance metrics vary under different input data conditions.

REFERENCES

Chang, J.C. and Hanna, S.R. 2005: Technical descriptions and user’s guide for the BOOT statistical model evaluation software package, version 2.0, p. 64. Available on. http://www.harmo.org/Kit/Download/BOOT_UG.pdf

Pineda Rojas, A.L. and Borge, R. 2019. Statistical evaluation of the urban atmospheric dispersion model DAUMOD-GRS to estimate NO_2 concentrations using new available data from Buenos Aires. Proceedings of the '19th International Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes (Harmo19)', 3-6 June 2019, Bruges, Belgium, H19-061, 5pp.

Pineda Rojas, A.L., Borge, R., Mazzeo, N.A., Saurral, R.I., Matarazzo, B.N., Cordero, J.M. and Kropff, E. 2020: High PM_{10} concentrations in the city of Buenos Aires and their relationship with meteorological conditions. *Atmos Environ.,* **214**, 117773.

Pineda Rojas, A.L., Leloup, J.A. and Kropff, E. 2019: Spatial patterns of conditions leading to peak O_3 concentrations revealed by clustering analysis of modeled data. *Air Qual Atmos Health.,* **12(6)**, 743–754.

Pineda Rojas, A.L. and Venegas, L.E. 2013: Upgrade of the DAUMOD atmospheric dispersion model to estimate urban background NO_2 concentrations. *Atmos Res*., **120-121**:147–154

Rouseeuw, P.J. 1987: Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math.,* **20(1)**: 53–65.

Venegas, L.E., Mazzeo, N.A., Pineda Rojas, A.L. 2011: Chapter 14: evaluation of an emission inventory and air pollution in the metropolitan area of Buenos Aires. In: Popovic (ed) Air Quality-models and Applications, Editorial In-tech, pp 261–288.