CALM ANALYSIS USING A ROBUST METHOD

G. Ratto [a, b], F. Videla [a, b, c, *], R. Maronna [d] and J. Reyna Almandos [a, c, e]
[a] CIOp (Centro de Investigaciones Opticas) - La Plata- Argentina
[b] Facultad de Ingeniería, Universidad Nacional de La Plata- La Plata- Argentina
[c] CIC BA (Comisión de Investigaciones Científicas de la Pcia. de Buenos Aires)- Argentina
[d] Facultad de Ciencias Exactas, Universidad Nacional de La Plata- Argentina
[e] Facultad Regional La Plata, Universidad Tecnológica Nacional (UTN)- Argentina
fabianv@ciop.unlp.edu.ar

Abstract

La Plata city and surroundings (Argentina) -around 800 000 inhabitants- possesses high industrial activity and intense vehicular traffic but there is no governmental air pollution network and basic meteorological parameters have been scarcely studied. This paper focuses on wind calm structure and wind patterns after the end of calms in order to assess the potential accumulation and fate of released airborne pollutants. A robust correlation method using an M-estimator is employed to compare seasonal observations at three monitoring sites located in key parts of the city covering 1995- 2006. Results show that calms are on average 14.7% in summer, 19.1% in autumn, 12.8% in winter and 11.6% in spring. The three sites under study were highly correlated indicating a generalized calm pattern for the city and surroundings. The robust coefficient allows inferring the lineal character of dependence among observations at the three sites. Calm structure revealed that approximately 50.6 % of the calms lasted 1 hour while calms lasting 5 hours or less constitute around 90% of calm occurrences. Calms lasting two hours or more are more probably to occur during wee and evening hours indicating when the accumulation of released air pollutants is likely to happen. Wind roses named "outgoing of calm wind roses" (time consuming to estimate) showed the wind direction patterns more probable to occur after a calm. Overall correlations between seasonal averaged wind roses and those corresponding to the outgoing of calms showed strong linear correlation for all the seasons. This allows concluding that winds after calms may be well represented by their corresponding seasonal averaged patterns (easy to compute).

Introduction

High industrial activity and intense vehicular traffic together with the lack of governmental air quality information (WHO, 1988) and the fragmentary character of the meteorological data available in La Plata city and surroundings (around 800 000 inhabitants) justify the characterization of basic meteorological parameters that play an important role on air pollutants. Ratto et al. (2010) were focused on wind direction frequencies and velocities. The present report is devoted to analyze calms observed at three sites covering different time periods from 1995 to 2006. Consecutives hours of calms can constitute a propitious meteorological condition for the later transport of large amounts of pollutants towards a given place. This phenomenon is called "accumulation effect" (Alvarez Morales and Alvarez Escudero, 2000) and can be characterized by analyzing calm locations and durations as well as wind directions and velocities immediately after calms. Frequency wind roses called "outgoing of calm wind roses" (OOCWR) are proposed as an approach to study wind patterns when calms end; this tool together with the knowledge of wind velocities after calms allows characterizing the potential fate of air pollutants accumulated during calm events. Outliers occur very frequently in large real data sets (Rousseeuw and Leroy, 1987). Although not traditionally applied to meteorological data, robust estimates are of great value and increasingly recommended (EPA, 2009) in order to diminish the distortion on the results that may cause atypical values (Maronna et al., 2006).

Characteristics of the region under study

The city of La Plata (35 S Lat 58´W Long) located in the northeastern part of Buenos Aires province (Argentina) is placed in a typical plain of the "Pampa", being on average 15 m above sea level. Its center (see Point B in Figure 1) is about 11 km far from La Plata River bank (marked as Point C). A Petrochemical Pole (see the rectangle in Figure 1) is located between the river and the city, where the country's main oil refinery and six petrochemical plants among other industries are placed. The climate corresponds to a humid temperate zone, with an annual average temperature of 16.2 ºC and an annual average rainfall of 1079.5 mm.

Instruments and data characteristics

Points A (covering 1997-2003) and J (covering 1997-2006) employed Davis® weather stations. Point K (covering 1995-2005) belongs to the National Meteorological Service network (La Plata Airport). Throughout this paper, hourly averages imply hourly blocks (e.g. 00:00-00:59 h. is equivalent to "hour 0" Local Time).

Methods

To obtain correlations that are not affected by outliers multivariate M-estimates were used (Maronna, 1976). From a sample of *n* observations of *p*-dimensional data given by the column vectors $\mathbf{x}_i = (x_{i1},.....,x_{ip})$ *(i=1,….,n)*, our goal was to define a *p*-vector μ and a *p* x *p* matrix Σ that are robust versions of the sample mean vector and sample covariance matrix, respectively. To this end, given **x**, μ and Σ, define the Mahalanobis distance $d^2(\mathbf{x},\mu,\Sigma) = (\mathbf{x} - \mu)^{\mathbf{t}}\Sigma^{-1}(\mathbf{x} - \mu)$ where in general $\mathbf{x}^{\mathbf{t}}$ denotes the transpose of **x**. Let *W*(d) be a nonnegative function defined for *d* ≥ 0. Then the M-estimates were defined implicitly as a weighted mean and a weighted covariance matrix:

$$\mu = \frac{1}{\sum_{i=1}^{n} w_i} \sum_{i=1}^{n} w_i \, \mathbf{x}_i \quad \text{and} \quad \Sigma = \sum_{i=1}^{n} w_i (\mathbf{x}_i - \mu)(\mathbf{x}_i - \mu)^t \qquad (1)$$

where the weights are $w_i = W(d(\mathbf{x}_i, \mu, \Sigma))$ (2)

Note that when $W(d) \equiv 1$, then μ and Σ are the classical mean vector and covariance matrix. To make the estimate robust, we employed a weight function *W* that decreases to zero at infinity, namely $W(d) = \frac{p+1}{1+d^2}$. The M-estimate corresponding to this *W* is the Maximum Likelihood estimate for the multivariate Cauchy distribution. Note that the Mahalanobis distances are a measure of the "outlyingness" of an observation, and therefore the estimate automatically gives less weight to atypical observations. The given implicit definition suggests an iterative procedure for computing the estimate: start with some initial μ and Σ, compute the Mahalanobis distance and the weights in Eq. 2, update μ and Σ through Eq. 1, and so on until convergence. Robust correlations are derived from the robust covariances in the same way as with the classical estimates: the robust correlation between variables *j* and *k* is $\rho_{jk} = \frac{\sigma_{jk}}{\sqrt{\sigma_{jj}\sigma_{kk}}} = \rho_{M-estimator}$ (3) where σ_{jk} denotes the elements of Σ. The confidence intervals (CI) for each ρ_{jk} were calculated with the well- known Fisher *z* transformation.

Results and Discussion

Total calms through the day

Figure 2 shows the typical curves of calms during the day for summer (the rest of the seasons are not shown due to space constraints). All the seasons showed a similar shape with two maximums, one around sunrise (build up of the diurnal boundary layer) and the other in the evening (build up of the nocturnal boundary layer); during the night higher values of calms are found due to nocturnal stability. A broad valley is observed between the maximums according to the development of the mixing height when wind become important during the daytime. The general average for calms (i.e. the average computed throughout the hours) was 14.7% in summer, 19.1% in autumn, 12.8% in winter and 11.6% in spring. The three sites were in general highly correlated as shows Table 1.

Calms location according to their duration

Figure 3 shows the distribution of accumulated calm durations computed in one-hour steps up to 20 hours long covering the whole range of calms found. The number of calms for a given duration respect to all calms during the day for a given season is expressed as an accumulated percentage in this figure. These curves show similar values for the seasons; calms lasting 1 hour covered on average 50.6%, lasting 2 hours covered 20.1%, lasting 3 hours covered 9.5%, lasting 4 hours covered 6.2% and lasting 5 hours covered 3.7% of the total of calms; the rest of each single longer durations had less than 2.2 % of the total of calms. 90.1% of the calm events lasted less than 5 hours, therefore the group of calms up to 5 hours were chosen to deepen the analysis.

Calm distribution within the day for different durations (lasting from 1 hour to 5 hours) were built on seasonal basis but no seasonal structure was found. As a consequence an average curve involving all the seasons for each of the durations was computed. Figure 4 depicts the percentage of calms observed (Y axis) for a given starting hour (X axis) and duration (parameter) respect to the total of calms distributed along all the durations (calm structure). For example, in hour 9 72.2 % of calms lasted for 1 hour, 15.2 % lasted for 2 hours, 6.2 % lasted for 3 hours, 2.4 % lasted for 4 hours and 1.5 % lasted for 5 hours (these five hours added 97.5 %, the rest were distributed in longer durations). Figure 4a shows that throughout the day there is a range of hours (between hour 7 and 12) that contains major peaks. For longer durations (from Figure 4b to Figure 4e) there are two particular regions that reveal the pattern: the first in wee hours and the second in the evening. Therefore, regarding calms lasting 2 hours or more, wee and evening hours constitute parts of the day when accumulation of released air pollutants is likely to occur.

Calms and winds after calms

We computed (seasonally but not hourly) the first directions that appeared immediately after a calm and their corresponding wind velocities. The frequency count for the directions (later averaged seasonally) provided a wind direction frequency rose that we called “outgoing of calm wind rose” (OOCWR). An example of this kind of wind rose is represented in Figure 5 (red curve).

Seasonal OOCWR for each monitoring site were computed. As wind direction frequency roses segregated by the module of wind velocities were unknown (not computed) we naturally recurred to the average wind direction roses involving all wind velocity modules to make a comparison with OOCWR. Then we assessed the similarity between such pair of wind roses employing the robust correlation coefficient as shown in Table 2. On the whole, OOCWR and averaged wind roses show similar patterns. Low correlation coefficients indicates the presence of non- linearities, while high coefficients reveal the linear character of dependence between the wind roses involved.

As an example, Figure 5 shows overall seasonal averages embracing the three monitoring sites (blue line) together with the corresponding averaged OOCWR (red line) for summer. The robust correlation coefficient regarding this figure is 0.9640 while for autumn is 0.8729, for winter is 0.7977 and for spring is 0.9458 (figures not shown).

Table 2 and Figure 5 allows inferring that when calms end, in the local scale, the probability of winds blowing from a particular direction follow, in general, the averaged wind direction patterns. Averaged wind velocities corresponding to Figure 5 and to the rest of the seasons are shown in Table 3. Wind velocities after calms are on average around half of those corresponding to the seasonal average. Figure 5 shows the high presence of wind direction occurrences between N and E (clockwise) that must be taken into account because they transport industrial air pollutants towards exposed population (see Figure 1). Table 3 and Figure 5 allow suspecting that wind direction frequency roses are somewhat independent from the modules of the wind velocities.

Conclusions

General average for calms respect to the total of occurrences was 14.7% in summer, 19.1% in autumn, 12.8% in winter and 11.6% in spring. The three sites under study were highly correlated indicating a generalized calm pattern for the city and surroundings. The robust coefficient employed allows inferring the lineal character of dependence among observations at the three sites. Calm structure revealed that approximately 50.6 % of the calms lasted 1 hour while calms lasting 5 hours or less constitute around 90% of calm occurrences. Wee and evening hours indicate when the accumulation of released air pollutants is likely to happen.

The OOCWR (time consuming to estimate) show the wind direction patterns more probable to occur after a calm. As overall robust correlations between seasonal averaged wind roses and those corresponding to the outgoing of calms showed high values for all the seasons, winds after calms may be well represented by their corresponding seasonal averaged patterns (easy to compute). As wind direction frequencies between N and E (clockwise) –that transport industrial air pollutants towards exposed population - have high occurrences and wind velocities after the end of calm are low (around 5 km h^{-1}) the accumulation effect appears as an important issue. Future studies should add air pollutant assessment as well as some aspects of the boundary layer meteorology.

Table 1

Correlated sites	Summer	Autumn	Winter	Spring	Average
A,J	**0,9846** 0,9641 [a] 0,9934 [b]	**0,9567** 0,9010 0,9814	**0,9293** 0,8413 0,9693	**0,9742** 0,9404 0,9890	**0,9612**
A,K	**0,9891** 0,9745 0,9954	**0,9766** 0,9458 0,9900	**0,9112** 0,8029 0,9613	**0,9925** 0,9824 0,9968	**0,9674**
J,K	**0,9520** 0,8906 0,9793	**0,7415** 0,4824 0,8813	**0,8020** 0,5892 0,9107	**0,9752** 0,9426 0,9894	**0,8677**
Average	**0.9752**	**0.8916**	**0.8808**	**0.9806**	**0.9321**

Table 1: Robust correlation coefficients for calms (bolded) with their corresponding confidence interval (CI) for each of the seasons and monitoring sites. Lower case letters indicate: [a] lower limit of the CI and [b] upper limit of the CI. The averages are simply displayed to help synthesizing the results. The second column corresponds to the curves

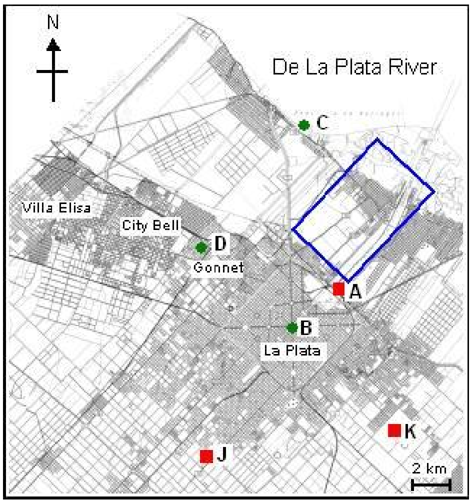

Figure 1: Map covering parts of La Plata City and surroundings. Measurement points are indicated with a red square and the other reference points with a green circle.

Table 2					
Point	Summer	Autumn	Winter	Spring	Average
A	**0,8026** 0,5903 [a] 0,9110 [b]	**0,7375** 0,4756 0,8793	**0,4751** 0,0887 0,7372	**0,7589** 0,5123 0,8899	**0,6935**
J	**0,9770** 0,9467 0,9902	**0,9830** 0,9605 0,9927	**0,7058** 0,4228 0,8634	**0,9354** 0,8544 0,9720	**0,9003**
K	**0,6773** 0,3769 0,8488	**0,6341** 0,3100 0,8262	**0,8929** 0,7651 0,9530	**0,7273** 0,4584 0,8742	**0,7329**
Average	**0.8190**	**0.7849**	**0.6913**	**0.8072**	**0.7756**

Table 2: Robust correlation coefficients -with their corresponding confidence intervals (CI)- involving seasonal average wind roses and the corresponding OOCWR at a given site. Lower case letters indicate: [a] lower limit of the CI and [b] upper limit of the CI. The averages are simply displayed to help synthesizing the results.

Table 3		
	Averaged Wind velocities	Outgoing of Calms Wind velocities
Summer	10.7	5.5
Autumn	10.0	4.8
Winter	10.4	5.3
Spring	11.2	5.1
Average	**10.6**	**5.2**

Table 3: Seasonal wind velocities (km h^{-1}) for the averaged wind roses and for the OOCWR.

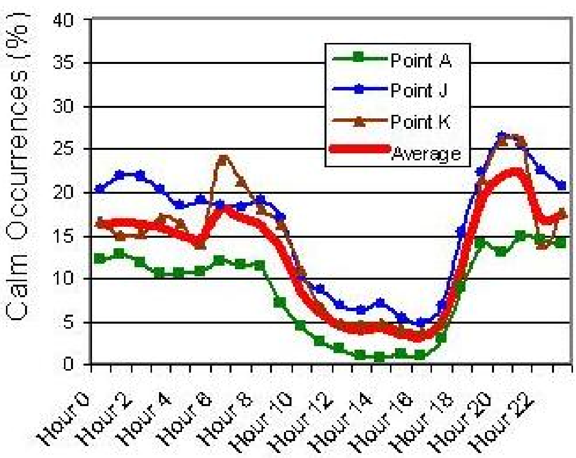

Figure 2: Hourly calm distribution for summer at different sites. Y axis represents the accumulated occurrences of calms respect to the total of occurrences (calms and winds with velocities ≥ 1.6 km h^{-1}) expressed as percentage. X axis represents the hours throughout the day in Local Time. The general average is 14.7%.

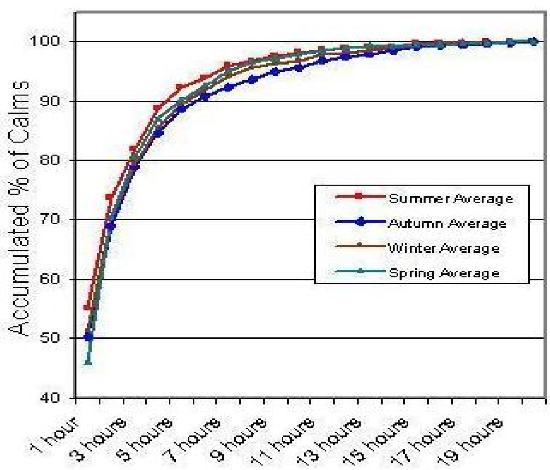

Figure 3: Accumulated percent calms averaged per season in one- hour intervals covering the data from points A, J and K. The percentages are expressed respect to all the durations and hours.

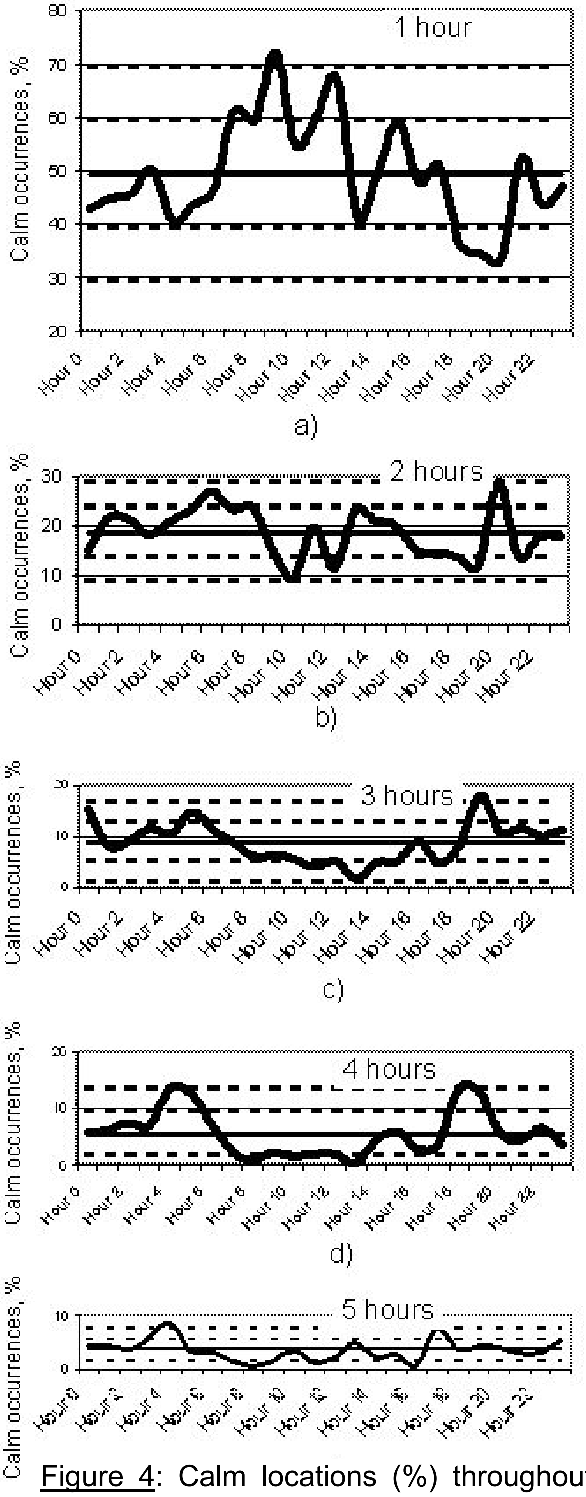

Figure 4: Calm locations (%) throughout the day for different durations. (a) distribution of calms lasting 1 hour (b) lasting 2 hours (c) lasting 3 hours (d) lasting 4 hours (e) lasting 5 hours. The percentages are expressed respect to all the durations (up to 20 hours) along one specific hour. The central straight line (bolded) represents the average for the specified duration. The two dashed lines above and below the central line represent the first and second standard

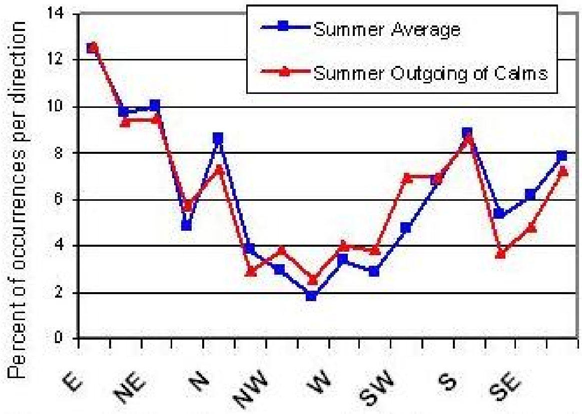

Figure 5: For the averaged wind rose Y axis represents the accumulated occurrences of winds (velocities ≥ 1.6 km h^{-1}) averaged for the three monitoring sites per season expressed as percentages respect to all the directions. Y axis for the OOCWR represents the firsts accumulated occurrences of winds (velocities ≥ 1.6 km h^{-1}) averaged for the three monitoring sites per season expressed as percentage respect to all the directions. X axis are the 16 directions of the wind rose.

References

Alvarez Morales, R. and Alvarez Escudero, L., 2000. *El efecto de acumulación y su influencia en el patrón de dispersión de contaminantes*. Revista Brasileira de Meteorologia. 15 (A1), 103- 111.

EPA, 2009. *Scout 2008 Version 1.0 User Guide*, EPA/600/R-08/038, US EPA.

Maronna, R. 1976. Robust M-estimators of multivariate location and scatter. Ann. of Stat. 4, 51-67.

Maronna, R., Martin, R., Yohai, V. 2006. *Robust Statistics. Theory and Methods.* John Wiley and Sons Ltd. London.

Ratto, G., Videla, F., , Maronna, R., Flores, A., De Pablo, F., 2010. *Air pollutant transport analysis based on hourly winds in the city of La Plata and surroundings, Argentina.* Water Air Soil Poll 208, 243-257.

Rousseeuw, P.J., Leroy, A.M., 1987. *Robust Regression and Outlier Detection*. Wiley, New York.

WHO, 1988. *La Salud en las Américas*, Vol. 2, Washington.

Acknowledgments

The authors are grateful to Technological National University- UTN, to Dr. Christian Weber and to the National Meteorological Service for the data at points A, J and K respectively. Also our particular thanks to the Department of Geodesic of the Ministry of Dwelling of Buenos Aires Province at La Plata (Map) to Dr. Andres Porta for his contributions and to Arq. Alex Rodriguez for the revision of the paper.

Keywords: air pollutants accumulation, calm analysis, La Plata, robust correlation