Boundary-Layer Meteorol (2010) 137:477–492
DOI 10.1007/s10546-010-9539-3

ARTICLE

Analysis of Wind Roses Using Hierarchical Cluster and Multidimensional Scaling Analysis at La Plata, Argentina

Gustavo Ratto · Ricardo Maronna · Guillermo Berri

Received: 25 January 2010 / Accepted: 26 August 2010 / Published online: 16 September 2010

Abstract Knowledge of frequency wind patterns is very important for air pollution modelling, especially in a city like La Plata (approximately 850,000 inhabitants) with high vehicular and industrial activities and no air monitoring network. An hourly wind analysis was carried out on data from two local weather stations (points A and J). An initial result was that, in spite of differences in data quality, the local weather stations observations were consistent with local and regional National Meteorological Service (NMS) monthly based observations. Two non conventional multivariate statistical methods were employed to further analyse hourly data at points A and J. Hierarchical cluster resulted in a good summarising tool to visualise prevailing hourly winds. Resultant vectors emerging from the clustering process showed good similarity between sites and seasons; this allowed a further visualization of the average diurnal wind development. Multidimensional scaling (MDS) permitted a pairwise comparison of a large number of hourly wind roses. These wind roses were more similar to each other in colder seasons and at site A (the one that is closer to the river) than in warmer seasons and at site J. Most of the observed variations regarding seasons and sites revealed by cluster and MDS analysis are explained in terms of the sea-land breeze circulations. The methodology

G. Ratto (✉)
CIOp (Centro de Investigaciones Opticas), CC124, 1900 La Plata, Argentina
e-mail: gustavratto@gmail.com

G. Ratto
Facultad de Ingeniería, Universidad Nacional de La Plata, La Plata, Argentina

R. Maronna
Departamento de Matemáticas, Facultad de Ciencias Exactas, Universidad Nacional de La Plata, La Plata, Argentina

G. Berri
SMN (Servicio Meteorológico Nacional), Buenos Aires, Argentina

G. Berri
CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas), Buenos Aires, Argentina

applied proved to be of utility for simplifying the analysis of high dimensional data with numerous observations.

Keywords Cluster analysis · Multidimensional scaling · Wind rose analysis

1 Introduction

The evaluation of meteorological parameters in the context of an air pollutant monitoring program constitutes a fundamental issue for environmental modelling and planning. Several of the largest cities in Argentina have been making efforts to make visible the increasing problems related with air pollution (Olcese and Toselli 2002; Mazzeo et al. 2005; Rehwagen et al. 2005; Pulifato et al. 2007). La Plata City and surroundings (approximately 850,000 inhabitants) is one of the six most potentially polluted cities in Argentina according to Petcheneshky et al. (1998) but to date lacks a governmental network for monitoring both air pollution and meteorological parameters. The city is located in eastern Argentina (35° S 58° W) on the estuary of the La Plata river, which is one of the most important rivers in South America, and is part of the boundary between Argentina and Uruguay (see Fig. 1). The river begins at Punta Gorda with an approximate width of 1.4 km and ends at a line of around 219 km long that joins Punta Rasa (Argentina) with Punta del Este (Uruguay).

The river length is around 300 km and the complete basin covers an area of approximately 3,200,000 km^2. The geographical extent of this river creates a considerable surface temperature contrast with the continent that sets the stage for the development of a low-level circulation, with sea-land breeze characteristics. During the daytime hours the lower layers over land are warmer than over the river, creating a land-river surface temperature gradient that establishes a river-to-land wind component typical of a sea breeze. Over the northern shore of the La Plata river the surface winds increase the southerly component, while over the

Fig. 1 La Plata River region. The river begins at Punta Gorda and flows out through the line that unifies Punta del Este (Uruguay) with Punta Rasa (Argentina)

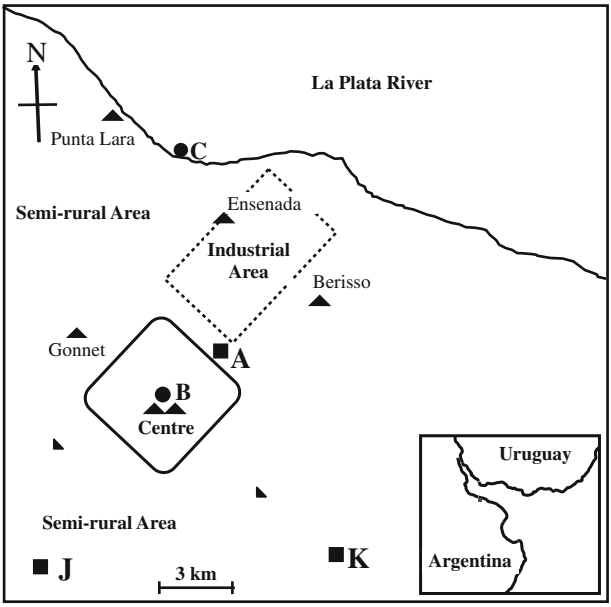

Fig. 2 La Plata City and surroundings. The *square with rounded edges* is the centre. Monitoring sites are indicated with *squares*, reference points with *circles*; the *triangles* indicate qualitatively the degree of population density (half, one and two *triangles* indicate respectively increasing population). The *rectangle in dashed lines* indicates the industrial park boundaries that mainly includes a petrochemical plant, steel processing plants and a shipyard. Point A: UTN- Reg. La Plata. Point B: city centre. Point C: river bank. Point J: Estación Agrometeorológica UNLP. Point K: La Plata's Airport (SMN local weather station)

southern shore they increase the northerly wind component. The daytime inland surface wind components create horizontal divergence and downward motion over the river and convergence and upward motion over land near the river shores. During the nighttime hours the land is cooler than the river, the land-river surface temperature gradient reverses and the winds tend to blow from land to river establishing the land breeze. The nighttime land breeze is not so well developed as the daytime sea breeze basically because of weaker low-level winds due to nocturnal stability. The daily cycle of the land-river surface temperature contrast gives rise to significant changes of the predominant wind direction across the region throughout the day (Berri et al. 2010).

Our study is intended to enrich previous work (Ratto et al. 2006, 2009) by presenting and discussing wind frequencies observed at two monitoring sites in the city. Point A belongs to UTN-Reg. La Plata and Point J (see Fig. 2) belongs to Estación Agrometeorológica Julio Hirschorn—UNLP. Neither of these monitoring sites was established to have compliance with World Meteorological Organization protocols (WMO 1983), and were installed respectively for different institutional purposes. However, as is demonstrated in this work, the supplied data allowed the derivation of useful results on the wind characteristics in the area. Both datasets were selected because the hourly data available were simultaneous within the period 1998–2003.

Two main references have been taken into account to support our observations. One lies within the city and belongs to the National Meteorological Service (NMS) while the other covers a large area of La Plata river region and consists of a network of five weather stations (named NMSN). The first one is La Plata Airport, marked as Point K in Fig. 2, with records

of period 1991–2000. The second one consists of wind records observed during 1959–1984 (Berri et al. 2010).

The main purpose of this article is to analyse seasonal hourly wind direction frequency patterns at two monitoring sites in the city in order to gain knowledge of wind characteristics that constitute a fundamental tool for environmental modelling. The observed wind patterns in the area are basically driven by the large-scale phenomena, and the sea-land breeze effects represent local perturbations of the larger-scale atmospheric patterns. At the same time, the work attempts to show the degree to which local wind direction frequency patterns are in agreement with regional ones. Nonetheless, the conditioning effect of the sea-land breeze is present when comparing the wind patterns at both sites, particularly during the warm season when this local circulation is stronger.

The comparison between observed and reference wind frequency patterns was done using squared Euclidean distance that was adopted as a similarity criterion. The comparison among observed hourly wind roses was performed by the application of two non-conventional approaches: (1) hierarchical cluster analysis, and (2) MDS analysis. Hierarchical clustering allowed us to reduce the 24-hourly wind roses representing the "day" for a given season and site to a significant lower number without loosing either their dimensionality or their main information. The obtained clusters allowed us to estimate representative daily occurrences of winds; finally, resultant vectors obtained from the clusters allowed representation of the diurnal variation in wind direction in a compact way.

Multidimensional scaling (MDS) was applied to gain insight into the underlying structure of relationships between hourly wind roses observed at Points A and J for all the seasons (represented by 192 wind roses). The implicit reduction in dimensionality (from 16 to 2) carried out with this approach allowed us to visualize hourly wind roses in a simple way and to further compare seasons and sites.

2 Data Characteristics and Instruments

The NMSN data consisted of observed winds during 1959–1984 at five weather stations in the region of the La Plata River basin. The names of the weather stations are Ezeiza (EZE), Aeroparque (AER), Martín García (MGA), Punta Indio (PIN) and Pontón Recalada (PRE) (see Fig. 1). These stations provided 8-direction wind frequency patterns grouped seasonally.

The NMS data (measured at Point K—Fig. 2) consisted of monthly averages accumulated for the decade 1991–2000 for 8-direction wind frequency patterns; the measurements were carried out at 10 m above the ground.

Point A is located in an urban area—approximately 8 km from the river bank—(see Point C in Fig. 2) and its data were registered every 15 min at a height of 12 m above the ground. Point J is located in a semi-rural area, approximately 18 km from the river bank; its data were registered every hour and measured at 5 m above the ground. The dataset belonging to site A had a deficiency in north-north-east records, a drawback that was found to be due to an obstacle that prevented a correct observation. Both monitoring sites provided complete datasets with the exception of Point J in winter 2000 whose records were very poor; missing data were replaced by the median of the four adjacent years. This approach had a minor quadratic error with respect to the average and the weighted average. Both sites were equipped with automatic weather stations providing 16-direction wind roses with an accuracy of $\pm 5°$. The detection limit for wind speeds was $0.44\,\mathrm{m\,s^{-1}}$.

Calibration and maintenance were performed according to the manufacturers' manuals. Observed wind frequencies at Point A and Point J were grouped for each of the seasons as

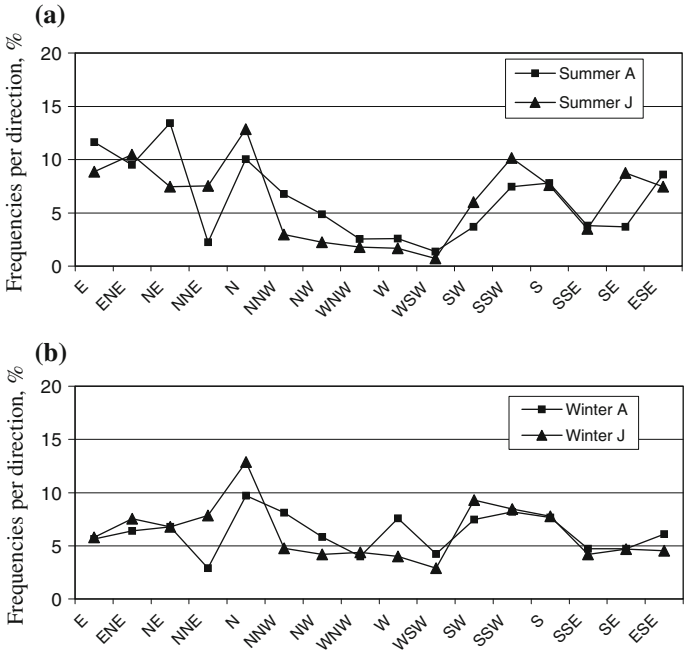

Fig. 3 16-category percentage frequency wind direction analyses for sites A and J: **a** summer **b** winter

hourly averages; for example, we called "Hour 0" the interval between 0000 and 0059 local standard time (LST).

Wind directions in the figures are expressed with their acronyms, i.e. east is E, north is N, west is W and south is S; then, north-north-east is NNE and so on.

3 Methods

3.1 Dimensionality Reduction from 16 to 8 Directions

In order to make 16-direction wind frequency patterns (such as those observed at Point A and Point J—see Fig. 3) comparable with those from Point K and the NMSN sites (8-direction wind frequency patterns) we applied the classical method proposed in Conrad and Pollak (1950). This approach takes the 16-direction wind frequency patterns and assigns half of the frequencies of the secondary directions to the two adjacent principal ones. For example, half of the frequencies observed for north-north-east is assigned to north-east and the other half to north. With this method we computed the wind direction frequency patterns for sites A and J shown in Fig. 4.

3.2 Pattern Similarities

Euclidean distance expresses how far a pattern is from another, it is indeed a dissimilarity measure because as far as the distance between two patterns increases the similarity decreases (Romesburg 2004).

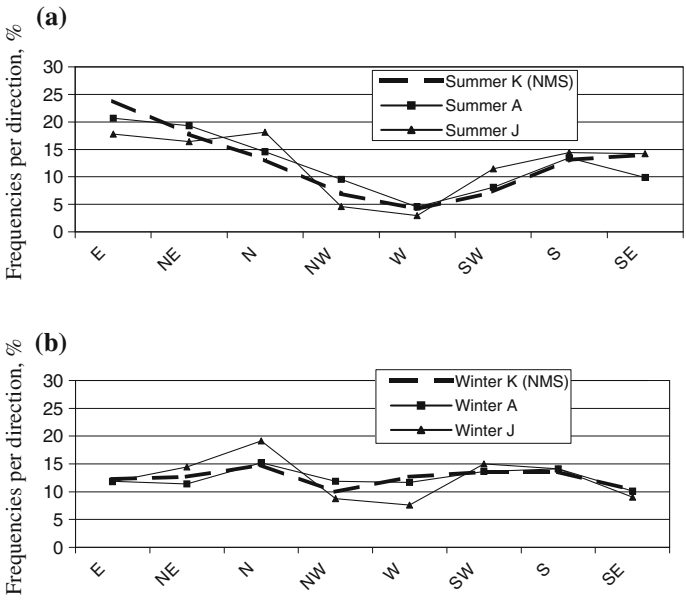

Fig. 4 8-category percentage frequency wind direction analyses involving sites A and J and for the reference site at La Plata's airport (Point K): **a** summer **b** winter

We applied the squared Euclidean distance given by Eq. 1 to analyse similarity between wind roses:

$$D^2(x, y) = \sum_{p=1}^{n} (x_p - y_p)^2 \tag{1}$$

where p is the dimension of the pattern involved (8 or 16), x_p is the frequency of the wind pattern X in the direction p, y_p is the frequency of the wind pattern Y in the direction p.

3.3 Hierarchical Cluster Approach

Details on hierarchical cluster analysis are given in Escudero (1977), Wilks (1995), Timm (2002) and Romesburg (2004). Application of different types of cluster analysis to wind patterns can be found in Kaufmann and Whiteman (1999), Darby (2005), and Beaver and Palazoglu (2006). A particular application to frequency wind roses can be found in Ratto et al. (2010). However, the procedure to carry out hierarchical clustering is outlined herein.

Cluster analysis is an exploratory technique that considers each pattern (in our case 16-dimension hourly wind roses) as a vector providing a systematic and objective way of grouping. This method is proposed as an alternative to the classical one (Alvarez and Alvares 2001), which groups consecutive hourly wind roses in fixed blocks of a predetermined number of hours. Agglomerative hierarchical clustering considers each of the vectors (or observations) as a group of only one member. Then it searches for two vectors among all these that are more similar. The found two vectors are then combined into a new group. The clustering process continues, building a hierarchy of a set of groups i.e a sequence of cluster solutions, until all the initial patterns make only one group. The optimal number of groups is somewhere between the first and the last agglomeration steps. The result tends to

minimise differences between members of a given cluster and maximize differences between members of different clusters. The criterion adopted in this work to establish the similarity between two patterns was the squared Euclidean distance, in the case of single vectors this distance can be computed using Eq. 1 (Sect. 3.2). But when the distance between a single vector and a group or between groups of vectors is required, a "method" for clustering is needed. There are a number of choices for adopting such a method. Although none of the available methods are guaranteed to be the best in any given circumstance, the average linkage (between groups) is often recommended (Wilks 1995; Escudero 1977) and is less sensitive to outliers (Figueras 2001). As an example, the corresponding expression for the "average linkage between groups" method to compute the square distance between a group and an individual is given by:

$$D^2_{(i,j),L} = \frac{n_i D^2_{i,L} + n_j D^2_{j,L}}{n_i + n_j} \tag{2}$$

where $D^2_{(i,j),L}$ is the distance between a cluster containing i and $j(i, j)$ and the individual (L), $D^2_{i,L}$ is the distance between individual (i) and individual (L), computable with Eq. 1, and $D^2_{j,L}$ is the distance between individual (j) and individual (L), computable with Eq. 1, n_i is the number of members of the cluster that contains the individual (i); $i = 1$ if there is just one member, n_j is the number of members of the cluster that contains the individual (j). In cluster (i, j) $n_i = 2$ and $n_j = 2$.

A similar expression is found when distances among groups are required. All patterns have the same dimension ($p = 16$). This method takes into account the quantity of patterns of the cluster that will be pooled with another cluster or single pattern by weighting the number of members of the cluster. The process of the hierarchical cluster analysis and its results can be observed in a dendogram (e.g. Fig. 7).

In the current study we started with a dataset of 24 hourly wind roses for each season and monitoring site. The corresponding dendograms were obtained using SPSS Version 13.0 software. The set of 24 wind roses were grouped in five clusters, this cut-off number was chosen in order to obtain a small number of representative wind roses. In cases such as those presented here, the application of hierarchical cluster analysis must be in accordance to two basic requisites: (a) the members forming the cluster have to be consecutive, and (b) when different seasons and monitoring sites are compared, the cut-off number must be common for all the cases. After trial and error, a cut-off number of 5 was find to optimise the balance between simplification and loss of information.

In order to evaluate the degree to which the dendogram (as an output of the clustering method) distorts the dissimilarity matrix between pairs of objects (input) we computed the cophenetic coefficient. This statistic is the product-moment correlation coefficient renamed for the particular application to measure cluster correlation and is explained in detail in Romesburg (2004).

3.4 Multidimensional Scaling Approach

We obtained a lower-dimensional representation of the data using non-metric MDS (Kruskal 1964a) that was considered as best suited to detect non-linear data structures. For a general reference, see Davison (1983), and for a detailed reference see Cox and Cox (2001). The term non-metric stems from the fact that this approach uses only the rank order of the dissimilarities rather than their numerical values.

Let d_{ij} $(i,\ j\ =\ 1,\ldots,n, i\ \neq\ j)$ be the dissimilarities between the objects we want to represent. In our case, the d_{ij} are Euclidean distances between the vectors representing hours of the day, but for the application of the method they may be any type of distance or a more general dissimilarity measure that is not required to satisfy the triangle inequality. For a given p the goal is to find n points in p-dimensional space, $y_i, \ldots, y_n$, such that their Euclidean distances $D_{ij} = \|y_i - y_j\|$ are a "good" representation of the original dissimilarities d_{ij}. Non-metric scaling tries to find a configuration such that the rank order D_{ij} is "as similar as possible" to that of the d_{ij}. More precisely, given a configuration $Y = \{y_1, \ldots, y_n\}$ with distances D_{ij}, let $\hat{D}_{ij}$ $(i,\ j\ =\ 1,\ldots n)$ be the set of numbers that minimises the error measure $\sum_{i<j} (d_{ij} - \hat{D}_{ij})^2$ among all sets having the same rank order as the d_{ij}. The $\hat{D}_{ij}$ are obtained by the so-called isotonic regression (Kruskal 1964b).

Then the stress corresponding to configuration Y is defined as:

$$S_{(Y)} = \sqrt{\frac{\sum_{i<j} (D_{ij} - \hat{D}_{ij})^2}{\sum_{i<j} D_{ij}^2}}. \tag{3}$$

It can be shown that the stress lies between 0 and 1. $S_{(Y)} = 0$ corresponds to a perfect fit, for in this case the distances D_{ij} of the y_i have the same rank order as the original dissimilarities d_{ij}. MDS looks for the configuration Y in p-dimensional space that minimizes $S_{(Y)}$. Kruskal (1964a) suggested the following verbal evaluation for the goodness of fit: $S_{(Y)}$ 0.2 is poor, 0.1 fair, 0.05 good, 0.025 excellent and zero perfect.

Note that the Y minimising the stress is not unique: any rigid transformation of Y (e.g., translation and rotation) yields the same D_{ij}, and multiplying Y by a scalar yields the same stress (Eq. 3). The minimization of Eq. 3 is performed by an iterative process that can be summarized in the following steps: 1. Choose the dimension p, and sort the d_{ij} in increasing order. 2. Choose an initial configuration Y (e.g. at random). 3. Compute the distances D_{ij} corresponding to Y, apply the isotonic regression algorithm to determine the $\hat{D}_{ij}$, and with them compute the stress $S_{(Y)}$. 4. Apply an iterative gradient search (e.g., steepest descent) to find a new Y with smaller stress. 5. Repeat steps 3 and 4 until the decrease in $S_{(Y)}$ is below a given threshold.

An important issue regarding the output of MDS is that the orientation of the picture (in our case two-dimensional) is arbitrary and therefore the axes, in themselves, are meaningless.

MDS allows visualization (e.g. in a two-dimensional map) of a considerable number of high dimensional patterns (e.g. 16- dimensional wind roses) and discrimination between them (e. g. seasons, sites) according to their differences (by measuring their relative distances).

4 Results and Discussion

4.1 Observed and Reference Wind Roses

So as to provide a basis to cluster and MDS analysis we first present and discuss the averaged observations of wind frequencies at Points A and J together with those taken as reference, i.e. the NMS data at Point K and those of the NMSN according to data available.

Figure 3 shows in x–y axis wind direction frequencies for summer and winter at sites A and J. These seasons were selected, due to space constraints, because they showed extreme behaviour. All seasons exhibited, in general, good similarity. Due to the deficiency in north-north-east measurements at site A (see Sect. 2) differences in this direction between sites

Table 1 Squared Euclidean distances between wind direction analyses at sites A, J and K regarding all the seasons

These distances were computed according to Eq. 1 in order to show dissimilarities between pairs of wind patterns

	A–K	J–K	A–J
Summer	38.9	88.7	87.0
Autumn	43.3	41.6	57.3
Winter	6.7	53.4	54.1
Spring	61.5	54.7	90.8
Average	37.6	59.6	72.3

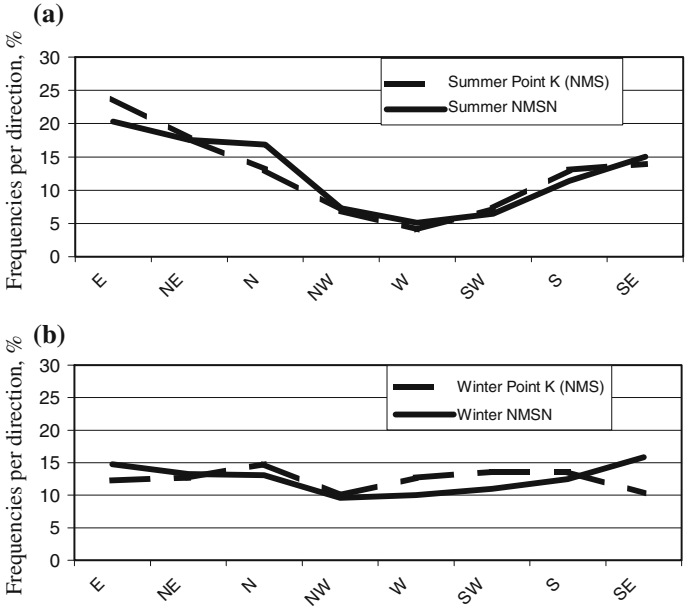

Fig. 5 Percentage frequency wind direction analyses corresponding to NMS (Point K) and NMSN data: **a** summer **b** winter

A and J were found to be among the main ones but they were all below 5.3% (see Fig. 3). Other important differences detected were north-east in summer (6.0%), north-north-west in autumn (4.5%), west in winter (3.6%) and north-east in spring (6.5%). As these data are given in 16 directions we appealed to the reducing method considered in Sect. 3.1. The resulting transformation can be appreciated in Fig. 4 where reduced observations from sites A and J are graphed together with NMS data. Summer and winter in this figure are examples of the similarity observed among the three sites over the four seasons. To assess similarity among sites in a more objective way we employed the squared Euclidean distance as stated in Sect. 3.2. Results are shown in Table 1; the second and third columns show distances between reduced wind roses at sites A and site K. While summers and winters at site A are more similar to the reference (than those corresponding to site J), they have smaller distances; autumns and springs at site J are more similar to the reference (than those of site A). From a more global perspective (involving the four seasons) site A is more similar to K than J is (see average distances in the last rows of column 2 and 3 of the table). On the other hand, sites A and J (8-directions) are more similar in colder than in warmer seasons.

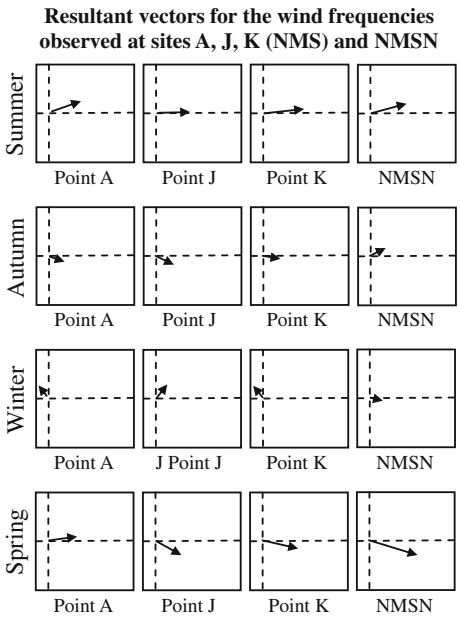

Fig. 6 Resultant vectors for the seasonal wind direction frequencies derived for sites A and J the NMS station (Point K) and for NMSN data. Note that the resultant vector is concerned solely with wind direction

Figure 5 shows seasonal wind frequencies for summer and winter corresponding to NMS and NMSN data. Both datasets revealed a similar wind pattern for all the seasons, the squared Euclidean distance between seasonal wind direction frequency patterns was 32.3 in summer, 28.3 in autumn, 55.4 in winter and 93.4 in spring. Despite differences in data quality these results reflect (as a general approximation) how local winds respond to a regional scale pattern.

By calculating the resultant vector for each of the 8-direction wind frequency patterns involving sites A, J, K and NMSN it is possible to have an overview of dominating winds. Figure 6 depicts this situation, the moduli of the resultant vector represent the amount of time that wind directions were not compensated by their corresponding opposite. The arrows point towards the direction from where the winds were blowing. Note that most of the resultant vectors belong to the first and fourth quadrants.

This figure reveals that resultant vectors (moduli and direction) are similar throughout the three sites (Points A, J and K). Warm seasons have large resultant vectors with high influence of easterly winds; besides, summer has an influence of northerly winds but spring has an influence of southerly winds instead. Autumn and winter have small resultant vectors indicating that winds tend to have an even distribution during these seasons. Considering the directions winter presents the major differences but the magnitude of the vector is very small and so is less significant. The increased easterly component during warm seasons is a direct consequence of the sea-land type of local breeze that is more intense when the land-river temperature differences are larger.

4.2 Cluster Analysis

Figure 7 is an example of the output of the clustering process. It shows the obtained dendogram for summer at Point J (selected at random). The vertical line (dashed) indicates the

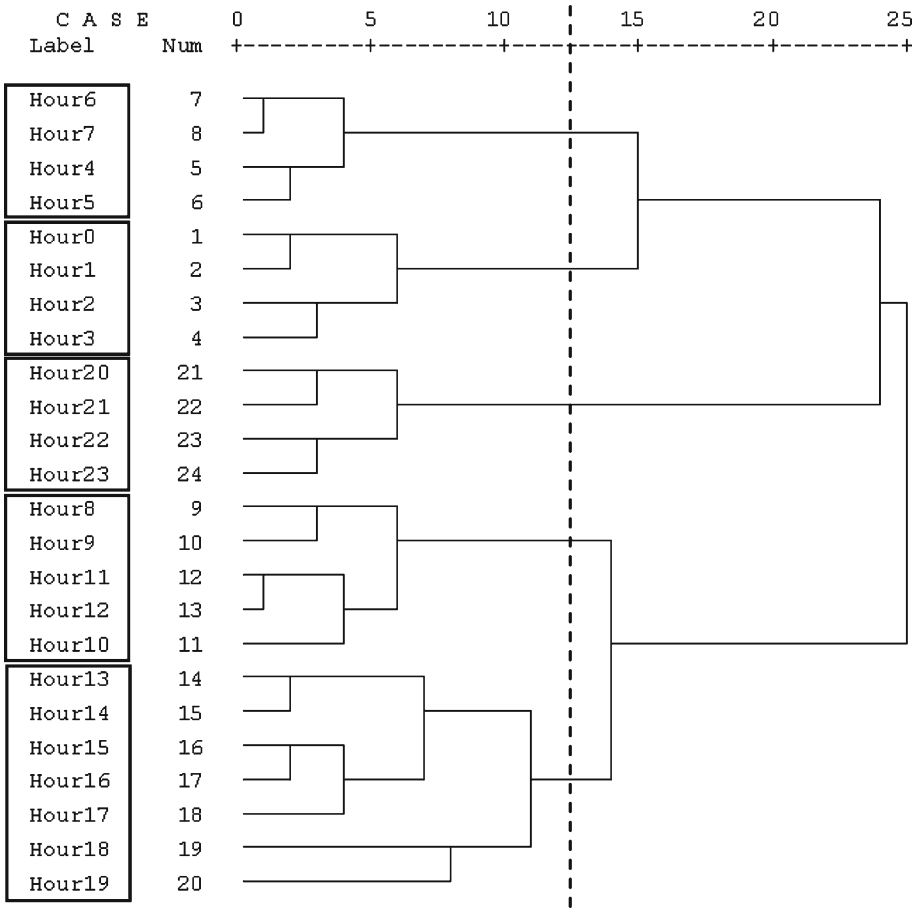

Fig. 7 Dendogram (*tree diagram*) corresponding to the cluster analysis using average linkage obtained with SPSS software for summer at Point J covering 1998–2003. The *transverse dashed line* indicates the relative distance in which the clusters formed five groups. Each cluster (embraced by a *rectangle*) denotes a group of 16-direction hourly wind roses

Table 2 Cophenetic coefficients covering all the dendograms for the seasons and sites

	Point A	Point J
Summer	0.722	0.725
Autumn	0.710	0.750
Winter	0.732	0.728
Spring	0.675	0.729

normalized distance (see the scale in the top of the graph) at which the five clusters were conformed. Including all the seasons, this distance was found to be between 12 and 14.

The cophenetic coefficients computed for all the seasons and sites are shown in Table 2. Note that all the values are lower than that of 0.8 generally proposed in the literature on

biological fields (Macedo et al. 2006) as acceptable. Although lower cophenetic coefficients have been accepted in other fields (Takahashi et al. 2007) we found no applications to wind roses, and so we leave this as an open inquiry. Values from Table 2 indicate that cluster correlations are similar for all the seasons.

By averaging the members composing each of the five clusters (of the corresponding dendogram) the wind roses shown in Fig. 8a were obtained. Each cluster is composed by a variable number of members that goes from 2 to 7, this variability is due to the similarity criterion applied. By computing the resultant vectors (following the same procedure to that for Fig. 6) for each of the wind roses of Fig. 8a and adding the rest of the seasons and sites Fig. 8b was obtained. Most of the resultant vectors, which concern only wind directions, belongs to the first and fourth quadrants showing the influence of prevailing winds. This pattern is in accordance to the general patterns obtained in Fig. 6. For example, the moduli for autumn and winter are smaller than those for summer and spring.

Although the simplification applied may be thought to be drastic the information that emerges is interesting. An overview of Fig. 8b shows that on average all seasons and sites have the same development within the day, the pattern is more visible in warmer than in colder seasons. In winters all the directions seem to be mostly compensated within the day. The largest moduli over the seasons were found between hours 20 and 23 showing agreement with the general pattern observed for hour 21 (Berri et al. 2010). For all the seasons and sites, the last three clusters show a generalized rotation of winds covering from north to south-east clockwise. From midday the moduli have in general large values for every seasons while during nighttime the moduli are much smaller and show more variability in the resultant direction. In the morning the resultant vectors shown in Fig. 8b are also in accordance with NMSN observations for hour 9 (Berri et al. 2010). At hour 15 east is added to north and north-east (prevailing during the morning) this trend can be visualized in clusters 3 and 4. The daily change of the prevailing wind directions over the La Plata river region, as discussed by Berri et al. (2010), responds to the sea-land breeze low-level atmospheric circulation that dominates the local weather and climate.

4.3 MDS Analysis

Figure 9 shows the reduced on frequency hourly wind roses obtained by the application of non-metric MDS to summer and winter datasets from Points A and J. There are 24 hourly wind roses representing each of the season at each monitoring site; finally, there are 192 wind roses to analyse. MDS provides a simple approach to compare such number of objects in a summarised way. The output of the MDS process gives points in a two-dimensional plot; each point represents an hourly wind rose given for a season and monitoring site. The main issue to consider regarding the groups of points is their relative distance that provides sounding information about their similarity. The stress coefficients (see Sect. 3.4) for all the seasons are shown in Table 3. Note that with the exception of winter at site J where the goodness of fit is poor, all the coefficients correspond to fair or good stresses according to Kruskal.

An overview of Fig. 9 shows good similarity between the pairs of the corresponding hourly wind roses for summer and winter. Not shown, summer and spring had very close patterns between hours 0 and 16 (clockwise) while autumn and winter had close patterns between hours 22 and 12 (clockwise). Warm seasons (summer and spring) presented a greater spread pattern than cold ones (autumn and winter) throughout the day. This difference is due to the sea-land breeze cycle that is more intense in warmer than in colder seasons. The same phenomenon explains the differences observed between monitoring sites; note that point A

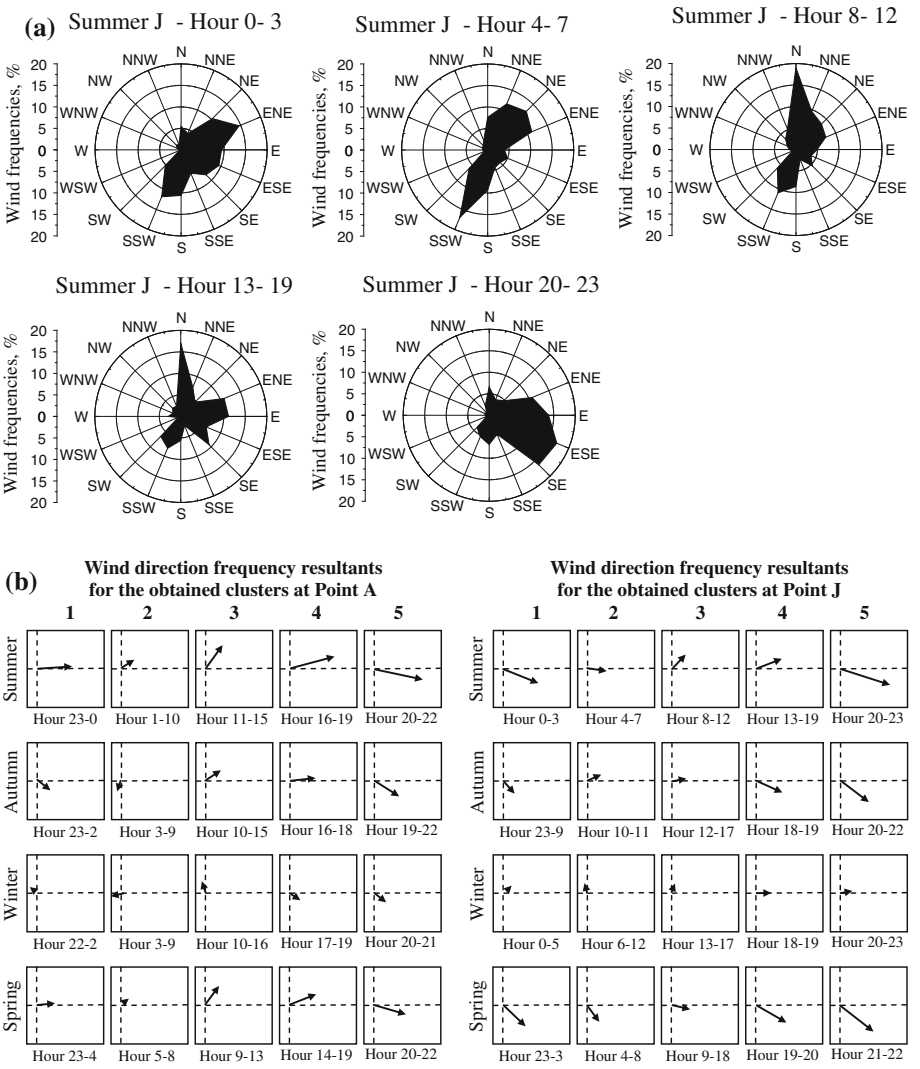

Fig. 8 **a** Resulting wind roses according to the dendogram of Fig. 7. The *y*-axis represents the percent frequencies for the corresponding wind direction. The five wind roses shown represent the 24 original hourly wind roses observed for summer at Point J. Each wind rose is the result of averaging the corresponding wind roses indicated by the formed clusters. **b** Resultant vectors of wind direction frequencies for each of the conformed clusters for the four seasons at sites A and J. Note that the resultant vector is concerned solely with wind direction

Table 3 Stress coefficients covering all the seasons and sites

	Point A	Point J
Summer	0.0364	0.0211
Autumn	0.0327	0.0614
Winter	0.0687	0.1543
Spring	0.0406	0.0440

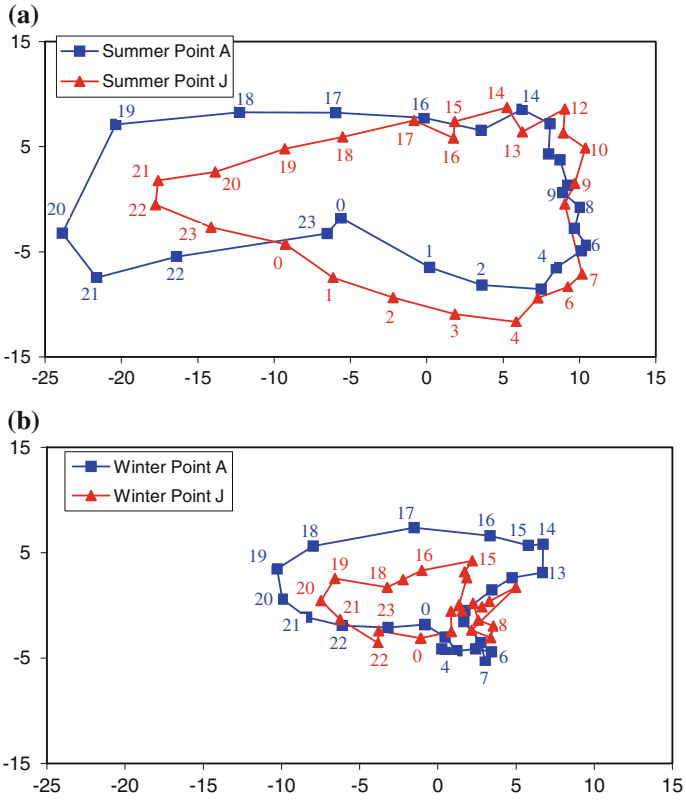

Fig. 9 Summer and winter output obtained from the application of MDS. *x* and *y* axes are given in arbitrary units. Both seasons were selected because they show extreme behaviour. Each point of the map represents an original 16-dimensional wind rose for a given hour of the day, season and site. The *number* indicates the hour of the day of the original wind rose. Some labels were omitted so as to make the map clearer. The *lines* connecting the configurations points (*blue* in the case of site A and the *red line* in the case of site J) were drawn only to help visualizing the continuity for the two-dimensional data. The contrast between seasons as well as the differences between sites, mostly attributable to sea-land breeze mechanisms, are evidenced in the map. *X* and *Y* axes have no label because the configuration of the map is arbitrary and therefore the axes themselves are meaningless

(that is close to the river) is more sensitive to sea- land mechanism compared with point J (located more inland) and so hourly wind roses are more spread.

5 Conclusions

Wind studies on an hourly basis have been rare at La Plata city. Knowledge of wind direction frequency patterns is very important for air pollution control and planning, especially in a city with high vehicular and industrial activities that has no governmental air monitoring network. Sea-land breeze circulations explained most of the observed variations between seasons and monitoring sites.

The hourly wind analysis was focused on data from two local weather stations (named Points A and J) that were not completely in agreement with WMO protocols. An initial result

of this work allows an inference that, in spite of differences in data quality, the local weather stations observations were consistent with NMS and NMSN observations. In this sense, this study shows, according to data availability, that local wind frequency patterns retain good similarity with regional patterns. Similarities among all the average seasonal observations were carried out mainly by the estimate of the squared Euclidean distance and the comparison among resultant vectors. Although non-conventional in applied meteorology, the multivariate methods used allowed us comparing complex data groups (192 hourly wind roses measured at Points A and J) in an easy way; by reducing the quantity of data to a representative number (hierarchical cluster) and by reducing dimensionality (MDS).

Cluster analysis proved to be a very helpful summarising tool to visualize prevailing hourly winds without loosing either dimensionality or representativeness. Cophenetic coefficients were very similar throughout the seasons and sites although their values were somewhat lower with respect to those considered acceptable for biological applications.

Resultant vectors (concerning solely wind directions) regarding groups of hours emerging from the clustering process showed good similarity when seasons and sites were compared; this allowed us to further visualize the average diurnal wind rotation. During the daytime prevalent averaged winds have a strong influence from the north and north-east. Around sunset winds from the east-north-east, east, east-south-east and south-east are prevalent until later at night the south-east becomes prevalent. During nighttime hours northerly winds recommence their influence, the reduced resultant moduli observed indicates that there is an even distribution among wind directions. From the seasonal point of view the vectors moduli are smaller in autumn and winter than in spring and summer.

Multidimensional scaling permitted a pairwise comparison for hourly wind roses coming from different monitoring sites on a seasonal basis. Hourly wind roses were more similar each other in colder seasons and at site A (the one that is closer to the river) than in warmer seasons and site J. These differences are in accordance to the sea-land breeze characteristics. During the warm seasons the land-river temperature contrast is larger, the sea-land breeze circulation is stronger, so that more marked spatial changes should be expected close to the river.

The variations between sites A and J as revealed by the wind direction patterns are largely attributable to the sea-land breeze effects. The area of study is a flatland and the only significant source of local atmospheric variability is the surface temperature contrast between the land and the large water surface of the La Plata river. The variability observed in the area on the seasonal time scale is driven by large-scale phenomena and the annual cycle, and the sea-land breeze effects are mainly local perturbations of the larger-scale atmospheric patterns. Nonetheless, the conditioning effect of the sea-land breeze is present when comparing the wind patterns at both sites, particularly during the warm season when this local circulation is stronger. We can conclude that the applied methodology can be of utility for simplifying the analysis of high-dimensional data with numerous observations from different places.

Acknowledgments This research was partially supported by Grants: PID 5505 from CONICET, PICTS 00899 AND 21407 from ANPCYT and X-018 from the University of Buenos Aires at Buenos Aires, Argentina. The authors are grateful to UTN for the data from the monitoring site Point A and to Agronomic Engineers Marcelo Asborno and Martín Pardi from the Experimental Station "Julio Hirschhorn" of UNLP for the data at Point J. Our particular thanks to Dr. Jorge Reyna Almandos and Engineer Fabián Videla for their contributions to this article.

References

Alvarez Escudero L, Alvares Morales R (2001) Climatología del Viento en Casablanca y sus Aplicaciones I. Climatología. Boletín de la Sociedad Cubana de Meteorología 7(2). Ciudad Habana, Cuba. http://www.met.inf.cu/sometcuba/boletin/v07_n02/espanol/art_2-1.htm. Accessed 18 Nov 2006

Beaver S, Palazoglu A (2006) Cluster analysis of hourly wind measurements to reveal synoptic regimes affecting air quality. J Appl Meteorol Clim 45:1710–1726

Berri GJ, Sraibman L, Tanco R, Bertossa G (2010) Low-level wind field climatology over the La Plata River region obtained with a mesoscale atmospheric boundary layer model forced with local weather observations. J Appl Meteorol Clim 49(6):1293–1305

Conrad V, Pollak LW (1950) Methods in climatology. Harvard University Press, Cambridge, 459 pp

Cox TF, Cox MA (2001) Multidimensional scaling, 2nd edn. Chapman and Hall/CRC, Boca Raton, 294 pp

Darby LS (2005) Cluster analysis of surface winds in Houston, TX, and the impact of wind patterns on ozone. J Appl Meteorol 44:1788–1806

Davison ML (1983) Multidimensional scaling. Wiley, New York, 242 pp

Escudero LF (1977) Reconocimiento de Patrones. Paraninfo, Madrid, 679 pp

Figueras S (2001) Análisis de conglomerados o cluster. Universidad de Zaragoza, España. http://www.5campus.org/lección/cluster. Accessed 20 Mar 2007

Kaufmann P, Whiteman CD (1999) Cluster-analysis classification of wintertime wind patterns in the Grand Canyon Region. J Appl Meteorol 38:1131–1147

Kruskal JB (1964a) Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika 29:1–28

Kruskal JB (1964b) Nonmetric multidimensional scaling: A numerical method. Psychometrika 29:115–129

Macedo IM, Pereira Masi B, Rosental Zalmon Ll (2006) Comparison of rocky intertidal community sampling methods at the northern coast of Rio de Janeiro state, Brazil. Braz J Oceanogr 54(2/3):147–154

Mazzeo NA, Venegas LE, Choren H (2005) Analysis of NO, NO_2, O_3 and NO_x concentrations measured at a green area of Buenos Aires City during wintertime. Atmos Environ 39:3055–3068

Olcese LE, Toselli BM (2002) Some aspects of air pollution in Cordoba. Argentina Atmos Environ 36:299–306

Petcheneshky T, Gravarotto MC, Benitez R, De Titto E (1998) Gestión de la Calidad de Aire Urbano-Industrial. Situación del Monitoreo de la Calidad del Aire (GEMS-AIRE) en la República Argentina. Departamento de Salud Ambiental del Ministerio de Salud y Acción Social de La Nación, AIDIS, Buenos Aires, 12 pp

Pulifato E, Rey Saravia F, Pereyra M, Pagani M (2007) Calidad del aire en el polo petroquímico de Bahía Blanca. II Congreso de PROIMCA San Nicolás, el 30 y 31 de octubre de 2007, pp 114–122. ISBN 978-950-42-0119-9—Pcia. de Buenos Aires, Argentina

Ratto G, Videla F, Reyna Almandos J, Maronna R, Schinca D (2006) Study of meteorological aspects and urban concentration of SO_2 in atmospheric environment of La Plata, Argentina. Environ Monit Assess 121:327–342

Ratto G, Videla F, Maronna R (2009) Analyzing SO_2 concentrations and wind directions during a short monitoring campaign at a site far from the industrial pole of La Plata, Argentina. Environ Monit Assess 149:229–240

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 Pollut 208:243–257

Rehwagen M, Müller A, Massolo L, Herbarth O, Ronco A (2005) Polycyclic aromatic hydrocarbons associated with particles in ambient air from urban and industrial areas. Sci Total Environ 348:199–210

Romesburg C (2004) Cluster analysis for researchers. Lulu Press, North Carolina, USA, 330 pp

SMN (2000) Estadísticas Climatológicas 1991–2000. Servicio Meteorológico Nacional, Buenos Aires, 36 pp

Takahashi K, Mirua T, Shioya I (2007) Hierarchical summarizing and evaluating for web pages. In: Arenas M, Hidders J (eds) Proceedings of the 1st workshop on emerging research opportunities for Web Data Management (EROW 2007) collocated with the 11th international conference on Database Theory (ICDT 2007) Barcelona, Spain, January 13, 2007. Pontificia Universidad Católica de Chile, Chile, University of Antwerp, Belgium

Timm, NH (2002) Applied multivariate analysis. Springer, New York, 693 pp

Wilks DS (1995) Statistical methods in the atmospheric sciences. Academic Press, San Diego, 464 pp

WMO (1983) Guide to climatological practices no. 100. WMO, Geneva, Switzerland, 198 pp