Cluster analysis of urban ultrafine particles size distributions

Dayana M. Agudelo-Castañeda\textsuperscript{a}, Elba C. Teixeirab,c, Marcel Braga\textsuperscript{c}, Silvia B.A. Rolima\textsuperscript{a}, Luis F.O. Silvad, David C.S. Beddows\textsuperscript{e}, Roy M. Harrison\textsuperscript{e,f}, Xavier Querol\textsuperscript{b}

\begin{itemize}
\item \textsuperscript{a} Department of Civil and Environmental Engineering, Universidad del Norte, Km 5 – Vía Puerto, Barranquilla, Atlántico, 081007, Colombia
\item \textsuperscript{b} Research Department, Fundação Estadual de Proteção Ambiental Henrique Luis Roesler, Av. Borges de Medeiros, 261, Porto Alegre, RS, 90020-021, Brazil
\item \textsuperscript{c} Postgraduate Program in Remote Sensing, Geosciences Institute, Universidade Federal do Rio Grande do Sul, Av. Bento Gonçalves, 9500, Porto Alegre, RS, 91501-970, Brazil
\item \textsuperscript{d} Civil and Environmental Department, Universidad De La Costa, Calle 58 #45-66, Barranquilla, Atlántico, 080002, Colombia
\item \textsuperscript{e} National Centre for Atmospheric Science, Division of Environmental Health & Risk Management, School of Geography, Earth & Environmental Sciences, University of Birmingham, Edgbaston, Birmingham, B15 2TT, United Kingdom
\item \textsuperscript{f} Department of Environmental Sciences/Center of Excellence in Environmental Studies, King Abdulaziz University, Jeddah, 21589, Saudi Arabia
\item \textsuperscript{g} Institute for Environmental Assessment and Water Research (IDAEA-CSIC), C/ Jordi Girona 18-26, 08034, Barcelona, Spain
\end{itemize}

\textbf{ARTICLE INFO}

\textbf{Keywords:}
Nanoparticles
Clusters analysis
Particle number concentration
Particle size distribution
Ultrafine particles

\textbf{ABSTRACT}

Measurements of particle size distribution were made in one location of an urban area in the period January–September/2015 in order to understand the sources and mechanisms influencing ultrafine particle (UFP) number concentrations (PNC\textsubscript{2.5-250}) using a Scanning Mobility Particle Sizer Spectrometer (SMPS). k-means cluster analysis was applied to interpret the sources, temporal and spatial trends of UFP. Eight clusters were obtained. Main PSD patterns of each cluster, mean concentration of other air pollutants tracing specific sources and processes, and that of meteorological variables, as well as the hourly and seasonal frequencies of occurrence were used to support the interpretation of their origin. Thus, clusters were attributed to traffic rush hours, midday summer new particle formation, diurnal new particle formation and growth, growth of nucleated and other urban particles, urban background, regional and urban background and regional and urban background on cold nights. Many PSDs of the clusters were dominated by nucleation mode particles: midday nucleated fresh particles, photochemically induced (NPF); diurnal nucleation episodes (NPF2); growth of nucleated particles in nocturnal aging (GNPF). Origins of the clusters were related to local/regional sources (mostly traffic and biomass burning), atmospheric processes (photochemical formation and growth) and urban/regional background. Results clearly shows that traffic is a major UFP source in nucleation mode and occurred in higher concentrations in winter (08:00 to 12:00 h) during traffic rush hours, and at night. Photochemical nucleation occurred with a relatively low frequency but yielding very high PNC.

1. Introduction

Atmospheric ultrafine particles (UFP, or particles < 100 nm) can affect atmospheric chemistry, human health and climate (Atkinson et al., 2014; Kulmala et al., 2016a, 2004; among others). Therefore, a number of studies have reported the importance of studying, and some of them of minimizing concentrations, of atmospheric UFP (Horvath et al., 1996; Kulmala et al., 2004; Murr and Garza, 2009). The atmospheric particle number concentration (PNC) is dominated by that of the UFP size range (Cheung et al., 2013), and accordingly we will use in this paper the terms PNC and UFP interchangeably. These UFP contribute to an increase in the health impact of aerosols because their very fine size allows UFP to penetrate and deposit in the deepest parts of the respiratory system, or even penetrate the pulmonary epithelium and olfactory nerve (HEI, 2013, and papers therein) and reach the cardiovascular system and hence other body organs.

The origin of UFP can be related to specific emission sources (mostly combustion sources), such as road traffic, shipping, airports, industrial sources (Buonanno and Morawska, 2015; Charron and Harrison, 2003; Cheung et al., 2013; Johnson et al., 2014; Keuken et al., 2015a, 2015b), or newly produced within the atmosphere by nucleation processes (Jamriska et al., 2008; Kumar et al., 2010; Morawska et al., 2008). Thus, UFP may be of both primary and secondary origin.

In urban areas evidence suggest that road traffic is the main source...
of UFP in cities (Dall’Osto et al., 2012; Kumar et al., 2014; Ma and Birmili, 2015; Morawska et al., 2008; Pey et al., 2009; Salma et al., 2014) and these arise from primary UFP exhaust emissions (Charron and Harrison, 2003; Shi et al., 2001; Shi and Harrison, 1999; Uhrner et al., 2007), including those from new particle formation (NPF) from semi-volatile phases which condense to new UFP during dilution and cooling of the exhaust emissions very close to the source point (Charron and Harrison, 2003; Kittelson et al., 2006; Robinson et al., 2007). Some previous studies showed that the airport/aerial liner is also one of the main UFP sources in cities. However, its affecting area may be limited to rural-urban/near airport areas (Ren et al., 2016).

PSD of particles emitted from diesel vehicle engines fall mainly in the range of 20–130 nm, while for petrol these are in the 40–80 nm range (Morawska et al., 2008; Ristovski et al., 2006). Brines et al. (2015) using a long time series of data on ambient PSD of urban UFP found major modes in the 20–40 nm (traffic-related nucleated particles) and another at 70–130 nm (soot particles) for periods of high ‘fresh’ traffic pollution from a number of cities. In addition, they found to this freshly emitted traffic PSD, two other traffic-related size distributions. One of them, with a minor 20–40 nm mode and a dominant mode at 70–90 nm, interpreted as the result of growth (by condensation and coagulation) during evening and night of the fresh traffic particles; and the second with similar modes, but shifted to 10–20 nm and a main peak at 50–90 nm throughout the day, with a peak during daytime. They attributed the shift to smaller sizes of the 20–40 nm peak of freshly emitted particles due to particle evaporation.

On the other hand, photochemical NPF events are common in less polluted environments. These new particles are formed from photochemical nucleation of H$_2$SO$_4$ and H$_2$SO$_4$-NH$_3$ followed by growth by condensation of the same gaseous species and oxidation products of volatile organic compounds (VOCs) (Kulmala et al., 2016a, 2004). Thus, high insolation (favouring photochemical transformations), high wind speed (favouring nucleation by decreasing the condensation sink when UFP are dispersed), low relative humidity (favouring nucleation instead of condensation), and available SO$_2$ (supplying H$_2$SO$_4$ for nucleation) may produce intensive NPF episodes (Kulmala et al., 2016a, 2004; Kulmala and Kerminen, 2008), characterized by a marked increase in PNC in the nucleation mode, and a subsequent particle growth yielding the ‘banana like’ nucleation bursts. Cheung et al. (2013) and Brines et al. (2015) demonstrated that in the urban atmosphere most NPF events are not followed by a large condensation/growth stage, but are only nucleation bursts, probably due to the delay of nucleation due to the high particle concentrations in the morning traffic rush hours, and the increase of the condensation sink by the afternoon traffic emissions. Thus the high condensation sink caused by traffic emissions gives only a relatively short period for nucleation in the middle of the day.

While at remote and low-pollution sites PSD can be similar over longer periods of time, excluding the nucleation and growth episodes, this is not generally the case at more polluted urban sites given the variations of emission fluxes and the influence of meteorology on these, (Wegner et al., 2012). Studies performed in Rochester, US showed that policy measures to abate primary particles and precursors and recession effects caused a clear decrease of UFP-PNC (in this case at a rate of $-4.5\%$/year, Masiol et al., 2018). Thus, in spite of the complexity and multiple factors affecting PNC, policy measures might have a clear influence.

Studies on UFP, PSD and NPF have been conducted in different environments and numerous locations around the world, principally in temperate regions and developed countries (mostly Europe, Canada and US). Though, in developing countries such as Brazil, no studies have been done at urban/industrial regions. This subject requires research to get better information on the sources and aerosol processes governing concentrations and variability of UFP. This evaluation can be achieved by applying cluster analysis to long time series of measurements of PSDs (Beddows et al., 2009). This clustering technique classifies aerosol size spectra into a reduced number of categories or clusters that can be characterized considering their size peaks, temporal trends and meteorological and gaseous pollutants average values (Brines et al., 2015).

Thus, the aim of this study is to identify possible sources and aerosol processes related with the emission, formation and transformations of UFPs in the urban area of Porto Alegre (South Brazil). To this end, cluster analysis was applied to a long data set of PSDs. The interpretation of the origin of the PSDs of each cluster was supported by calculating the averaged concentrations of PNCs and other pollutants, as well as the means of meteorological parameters, and their frequency of occurrence at hourly and seasonal scales.

2. Study area

The study area is the city of Canoas, located in the Metropolitan Area of Porto Alegre (MAPA), in the central-eastern region of the state of Rio Grande do Sul, Southern Brazil. This area has its limits within 29°54’ to 29°20’ S and 51°17’ to 50°15’ W. The meteorological conditions have been described in previous studies (Teixeira et al., 2012).

According to Koppen’s international system of climate classification, the climate type of the study area is a humid subtropical climate (Cfa) with well distributed rain over the year. Due to its location, the area of study shows well-defined seasons and a climate strongly influenced by cold air masses migrating from the polar regions. The historical average rainfall is 1300–1400 mm yr$^{-1}$ (INPE-CPTEC, 2012). The area of study is located in a subtropical, temperate climate with four well-defined seasons: summer (January–March), autumn (April–June), winter (July–September), and spring (October–December) with means temperature of 24.4 °C, 17.7 °C, 15.5 °C and 21.4 °C, respectively (Table S1). The wind direction shows marked seasonal variations. During summer and spring, the prevailing direction is E-SE, while in fall and winter, besides E-SE winds, winds from W and NW also occur. During the day, the wind reaches its lowest speed at dawn and early morning, with highest speeds in the late afternoon, between 17:00–18:00. The prevailing wind results from interactions of mesoscale phenomena, especially sea/land breezes (from the Atlantic Ocean and the Patos Lagoon) and valley/mountain breezes (from the nearby Serra Geral Mountains located to the north of the MAPA).

According to the Brazilian Institute of Geography and Statistics (IBGE, 2013), this region comprises an area of 9653 km$^2$, representing 3.76% of the total area of the state, and has a population of 4.12 million inhabitants, i.e., 37.7% of the total population of Rio Grande do Sul. The MAPA is the most urbanized area in Rio Grande do Sul, and it also contains different types industries, coal power plants and significant influence of mobile sources as: BR-116 highway located 450 m East, Air Base located $\approx 3000$ m East/Northeast; oil refinery REFAP located $\approx 6800$ m · North/Northeast, steel mill GERDAU located $\approx 12700$ m – North and the Petrochemical Complex at $\approx 21000$ m-West/Northwest. In addition to the industrial emissions, it is estimated that the most significant contribution to the MAPA emission inventory of air pollutants is due to the large number of vehicles in circulation in the region, 1.96 million (DETRAN, 2013). In 2009, the distribution of the fleet by fuel type in MAPA was 69% gasoline, 16% gasoline motorcycles, 11% diesel, and 4% alcohol (Teixeira et al., 2011). Canoas is under a strong influence of road traffic emissions, with daily traffic congestion.

3. Material and methods

24-hour continuous sampling was carried out using a TSI Scanning Mobility Particle Sizer Spectrometer (SMPS) 3936NL88 for atmospheric particles with diameters from 2.5 to 250 nm and for PNC up to > 1000000 $\#/cm^3$. SMPS included a neutralizer before entering the Differential Mobility Analyzer. SMPS is a high resolution nanoparticle sizer that uses an electrostatic classifier. This method is based on the physical principle that the ability of a particle to traverse an electric field is related to particle size. This equipment operates in conjunction
with a particle counter (CPC). CPC 3788 is a particle counter that uses water-based condensation technology. Particles that are too small to scatter enough to be detected by conventional optics are grown to a larger size by condensation to be measured. Those equipment were calibrated by the manufacturer (TSI) using NIST traceable analytical tools. This calibration was valid during the time they were used in this research.

Measurements of meteorological variables (temperature, relative humidity, wind and solar radiation) and concentrations of other pollutants (PM10, NO, NO2, O3, SO2) were carried out simultaneously to those of the SMPS from January to September 2015. The SMPS and meteorological station were located in the city of Canoas (Military Air Base), in the metropolitan area of Porto Alegre, in the air quality monitoring station. This area has its limits within 29°54′ to 29°20′ S and 51°17′ to 50°15′ W. The monitoring station is located at 20 m.a.s.l. at the following location: 29°55′50.0″ S 51°10′56.5″ W at 370 m from the highway (BR-116).

4. Data analysis

Given the amount of data to be analyzed and the complexity of the study a statistical analysis was applied to the SMPS data set using k-means cluster analysis that classifies PSD spectra with the highest degree of similarity into the same category or cluster, therefore reducing the number of spectra to interpret (Beddows et al., 2009). The cluster analysis was performed using hourly averaged PSD data (39 size bins and 4760 h). Cluster validation indices were used to choose the optimum number of spectra to divide the data as described elsewhere (Beddows et al., 2009; Dall’Osto et al., 2013a). This is solely a statistical analysis based on the clustering of the shape of the spectra. The use of cluster analysis was justified in this research using a Cluster Tendency test, providing a calculated Hopkins Index of 0.20 and implying the presence of structures in the form of cluster in a dataset. The choice of k-means clustering was made from a selection of the partitional cluster packages (Beddows et al., 2009). The k-means method aims to minimize the sum of squared distances between all points and the cluster center. K-means clustering identifies homogeneous groups by minimizing the clustering error defined as the sum of the squared Euclidean distances between each dataset point and the corresponding cluster center. The complexity of the dataset is reduced allowing characterization of the data according to the temporal and spatial trends of the clusters. In order to choose the optimum number of clusters the Dunn-Index (DI) was used, which aims to identify dense and well-separated clusters. DI is defined as the ratio between the minimal intercluster distance to maximal intra-cluster distance. Since internal criteria seek clusters with high intra-cluster similarity and low inter-cluster similarity, algorithms that produce clusters with high DI are more desirable. In other words, for Dunn’s index we wanted to find the clustering which maximizes this index. The Dunn-Index for the results of the k-means cluster analysis for different cluster numbers showed a clear maximum for 8 clusters, some of which belonged only to specific times of the day, specific mechanisms as well as specific seasons.

The interpretation of the origin of each cluster was based on the dominant size modes, their seasonal and hourly frequency of occurrence and the average values obtained for other air pollutants and meteorological variables for each cluster.

5. Results and discussion

5.1. Averaged particle number concentrations

PNC obtained in this study was compared with other researches done around the world (Table S2). Differences between these sites might be due to differences in experimental methods, meteorology, cutoff size for the PNC measurements, distance to emission sources, topographic settings, seasonal effects, among others (Kumar et al., 2014), but also to the vehicle fleet composition. PNC obtained in the present study was lower than Los Angeles (US), although higher than those reported in Europe studies. Lower concentrations of PNC in Brisbane are, probably, due to lower diesel fleet proportion and higher precipitation rates (Brines et al., 2015). Similar PNC values were observed in Latin America, i.e. Santiago de Chile (Kumar et al., 2014). Probably, the higher values obtained in Canoas (Brazil), compared with other cities, is due to the diverse industries and mobile sources, that may led to the high levels of SO2 and UFP. Especially, the presence of the
emission sources of the oil refinery (REFAP) and the heavy duty vehicle fleet (uses diesel), and the increasing SO2 trend in the MAPA (Landim et al., 2018). Highest UFP values (> 5.0E+5) were presented during two winter days (9/06/2015 and 10/08/2015) at 10:00 h local time in the 50–80 nm, typical of traffic related particles, especially diesel-soot particles, as will be discussed further on.

5.2. Clusters identification and interpretation

The average PSD in dN/dlog(Dp) for each of the eight clusters is presented in Fig. 1. Statistical optimization was used for the cluster validation index to choose the optimum number of spectra to divide the clusters for the time periods of specific cluster occurrence allowed the interpretation of their possible origin (Table 2).

5.2.1. Traffic clusters

5.2.1.1. Cluster 1 – fresh vehicle exhaust during traffic rush-hours (TR1). Cluster 1 is the most frequent cluster (16.8% of the total hourly PNDs). As explained above, the diesel vehicles fleet (uses diesel), and the increasing SO2 trend in the MAPA (Landim et al., 2018). The diesel vehicles fleet is characterized by an occurrence peak during traffic rush hours in the morning (09:00 h local time - UTC-3 in winter and UTC-2 in summer) and evening (20:00 h - UTC-3 in winter and UTC-2 in summer) (Fig. 1), and the highest NO2 concentrations (Table 2), thus indicating a high influence of vehicle exhaust emissions (Brines et al., 2014; Dall’Osto et al., 2013b). As discussed in the introductory section, Brines et al. (2015) found a major mode for the traffic cluster representing freshly emitted UFP in the 20–40 nm (traffic-related nucleated particles) and another at 70–130 nm (diesel soot particles) for a number of cities. Studies done in Brisbane, Australia, Toronto, Canada, and Los Angeles, US, with a much lower proportion of diesel vehicles in their respective car fleets the traffic clusters were characterized by modes at 14–20, 22–26 and 15–30 nm, respectively (Kim et al., 2002; Sabaliauskas et al., 2013; Brines et al., 2015).

As explained above, the diesel vehicles fleet distribution in the MAPA is 11%. In addition to the timing and intensity of traffic emissions, the morning peak of this cluster is favoured, too, by an undeveloped mixed boundary layer, so that all emissions are accumulated (Agudelo-Castañeda et al., 2013; Wang et al., 2010). There is a synergistic effect between three factors: wind speed, height of the mixing layer and vehicular traffic. Reduced PNCs at midday may be favoured by atmospheric dispersion processes due to an increase of the mixing layer height (MLH) and to higher wind speed (Charron and Harrison, 2003). Wind speed tends to increase from the afternoon (the sampling site is downstream from the prevailing wind), and decreases early in the evening. The MLH decreases abruptly in the late afternoon, due to the extinction of the solar radiation, initiating the formation of the nocturnal layer, which is more stable. Following this, there is an increase of the accumulation of PNC favoured by the evening traffic rush hour, the wind carrying vehicular emissions towards the sampling site and by the decrease of the MLH, the nocturnal boundary layer, more stable (Agudelo-Castañeda et al., 2013). In the morning there is an inverse phenomenon, despite the increase of the vehicular emissions (rush hour), the winds are weaker and of variable direction (affecting the transport of pollutants to the site of sampling). Moreover, the incidence of solar radiation increases the height of the mixing layer (beginning of the convective layer).

5.2.1.2. Cluster 4 – aged vehicle exhaust, morning traffic (TR2). Cluster 4 contributed with 13.7% of total hourly PNDs, with moderate-high PNC (peak of ~20,000) of a unimodal size distribution spectrum with a peak centered in the 20–30 nm range (slightly shifted towards coarser sizes than TR1). This cluster has a marked maximum frequency of occurrence slightly delayed to the morning traffic rush hours. The late afternoon traffic rush hour is less evident in terms of frequency of occurrence of

<table>
<thead>
<tr>
<th>Cluster number</th>
<th>Contribution to hours of PNC measurements</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>16.8%</td>
<td>Fresh vehicle exhaust during traffic rush hours</td>
</tr>
<tr>
<td>2</td>
<td>11.3%</td>
<td>Regional and urban background</td>
</tr>
<tr>
<td>3</td>
<td>14%</td>
<td>Growth of nucleated and other urban particles</td>
</tr>
<tr>
<td>4</td>
<td>13.7%</td>
<td>Traffic morning rush hour</td>
</tr>
<tr>
<td>5</td>
<td>3.9%</td>
<td>Midday summer new particle (photochemical) formation</td>
</tr>
<tr>
<td>6</td>
<td>18%</td>
<td>Diurnal new particle (photochemical) formation and growth</td>
</tr>
<tr>
<td>7</td>
<td>11.9%</td>
<td>Urban background</td>
</tr>
<tr>
<td>8</td>
<td>10.4%</td>
<td>Regional and urban background in cold nights</td>
</tr>
</tbody>
</table>

### Table 2

Mean value of traffic pollutants and meteorological variables for the time periods of specific cluster occurrence.

<table>
<thead>
<tr>
<th>Cluster</th>
<th>PM&lt;sub&gt;10&lt;/sub&gt; μg·m&lt;sup&gt;-3&lt;/sup&gt;</th>
<th>NO&lt;sub&gt;x&lt;/sub&gt; μg·m&lt;sup&gt;-3&lt;/sup&gt;</th>
<th>NO&lt;sub&gt;2&lt;/sub&gt; μg·m&lt;sup&gt;-3&lt;/sup&gt;</th>
<th>O&lt;sub&gt;3&lt;/sub&gt; μg·m&lt;sup&gt;-3&lt;/sup&gt;</th>
<th>SO&lt;sub&gt;2&lt;/sub&gt; μg·m&lt;sup&gt;-3&lt;/sup&gt;</th>
<th>Humidity, %</th>
<th>Wind speed, m·s&lt;sup&gt;-1&lt;/sup&gt;</th>
<th>Solar radiation, W·m&lt;sup&gt;-2&lt;/sup&gt;</th>
<th>Temperature, °C</th>
<th>Precipitation, mm</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>19</td>
<td>12</td>
<td>34</td>
<td>13</td>
<td>5.1</td>
<td>80</td>
<td>1.4</td>
<td>557</td>
<td>18</td>
<td>0.30</td>
</tr>
<tr>
<td>2</td>
<td>34</td>
<td>24</td>
<td>44</td>
<td>17</td>
<td>4.8</td>
<td>81</td>
<td>0.9</td>
<td>414</td>
<td>18</td>
<td>0.13</td>
</tr>
<tr>
<td>3</td>
<td>25</td>
<td>13</td>
<td>28</td>
<td>18</td>
<td>4.7</td>
<td>80</td>
<td>1.4</td>
<td>401</td>
<td>20</td>
<td>0.30</td>
</tr>
<tr>
<td>4</td>
<td>26</td>
<td>24</td>
<td>44</td>
<td>20</td>
<td>4.3</td>
<td>76</td>
<td>1.0</td>
<td>665</td>
<td>20</td>
<td>0.18</td>
</tr>
<tr>
<td>5</td>
<td>16</td>
<td>10</td>
<td>22</td>
<td>17</td>
<td>6.2</td>
<td>71</td>
<td>4</td>
<td>1046</td>
<td>21</td>
<td>0.19</td>
</tr>
<tr>
<td>6</td>
<td>17</td>
<td>9</td>
<td>20</td>
<td>27</td>
<td>6.4</td>
<td>80</td>
<td>4</td>
<td>583</td>
<td>19</td>
<td>0.37</td>
</tr>
<tr>
<td>7</td>
<td>35</td>
<td>9</td>
<td>12</td>
<td>20</td>
<td>6.2</td>
<td>80</td>
<td>4</td>
<td>388</td>
<td>20</td>
<td>0.37</td>
</tr>
<tr>
<td>8</td>
<td>57</td>
<td>18</td>
<td>16</td>
<td>20</td>
<td>6.2</td>
<td>85</td>
<td>4</td>
<td>151</td>
<td>17</td>
<td>0.11</td>
</tr>
</tbody>
</table>
this cluster. Cluster 4 is associated with high levels of pollutants from vehicle exhaust such as NOx, NO and PM10. (Fig. 1 and Table 2). Brines et al. (2015) reported the occurrence of clusters with coarser PSD representing the growth (aging) of freshly emitted traffic UFP. Seasonal maxima are recorded for TR2 in May-July, whereas for TR1 it is in July-September.

To support the interpretation of the traffic origin of these clusters we calculated the frequency of occurrence of TR1 (Cluster 1) and TR2 (Cluster 4) for the weekend and week days and we normalized the frequencies for the number of weekend and weekdays with measurements. Results evidenced that frequencies were distributed similarly for weekdays, while for weekend the frequency was reduced (Table S3).

5.2.2. Background clusters
5.2.2.1. Cluster 8 nocturnal urban background aerosols on cold nights (NRU). Cluster 8 accounts for 10.4% of the total hourly PSDs, with peak in the Aitken mode at 90 nm reaching around 18,000 #/cm³ (Fig. 1). This cluster is characterized by the coarsest PSD, nocturnal higher frequency of occurrence, mostly in June and August (winter), the lowest temperature, wind speed, insolation, precipitation and O₃ concentrations, as well as the highest PM₁₀, NO, NOx and humidity (Table 2). UFP concentration peaks were present at diameters of 50–200 nm. These cluster patterns have been attributed mainly to regional background aerosols during nights and cold days (Krecl et al., 2017; Salimi et al., 2014). Ammonium nitrate, and in general secondary aerosols, formation is favoured at night in cold and high humidity periods, with thermal inversions that favours also the increase of PM and UFP concentrations by a decrease of the MLH and the prevalence of stagnant conditions. Since we did not measure nitrate, the contribution of this specie to this cluster cannot be ensured.

5.2.2.2. Cluster 2 regional-urban background particles – biomass burning (R-UB). Cluster 2 reached a frequency of occurrence of 11.3% of the total hourly PSDs, and presented, again, a unimodal size distribution with a mode centered at 50–60 nm, in the Aitken mode, with PNC (~15,000 #/cm³). The diurnal pattern is associated with a minimum during morning rush hours and late afternoon, and a frequency peak in the late night. This cluster is characterized by a coarser mode than the ‘fresh’ traffic emission cluster. Coagulation and condensation processes affecting traffic particulate emissions lead to particle growth and to the shift to coarser modes. This cluster has a very similar PSD to that of the urban background cluster reported by Brines et al. (2015) for the city of Brisbane (60 nm and unimodal), with a similarly low proportion of diesel vehicles in the urban fleet. Moreover, biomass burning is one of the largest sources of accumulation mode particles globally (Kumar et al., 2013); consequently these particles with a regional and local origin, may also influence urban background UFPs of this cluster. This cluster is highly associated with NOx and PM₁₀, confirming that incomplete combustion of biomass also may produce both gaseous and particulate air pollution (Table 2). The highest occurrence was observed in May–July and at night, implying a low MLH so UFP accumulated (Águedo-Castañeda et al., 2017). One of the lowest wind speeds also points to nocturnal accumulation of urban background particles.

5.2.2.3. Cluster 7 - urban background (UB). Cluster 7 accounted 11.9% of the total hourly PSDs (Fig. 1). This cluster shows a bimodal size distribution, with a nucleation mode at 20 nm, and a broader Aitken mode with peak at 70 nm (both reaching moderate concentrations ~8000 #/cm³). Aitken nuclei mode (20–90 nm) refers to an overlapping fraction of the nucleation and accumulation mode, that arise from the growth or coagulation of nucleation mode particles as well as by production in high numbers by primary combustion sources such as vehicles (Kulmala et al., 2004; Kumar et al., 2010).

This cluster is characterized by high concentrations of traffic and industrial pollutants (Table 2) such as PM₁₀ (35 μg/m³), NOx (36 μg/m³) and SO₂ (5.2 μg/m³) concentrations. The relatively low PNC and the nocturnal higher frequency of occurrence, very similar to the R-UB cluster, but with higher diurnal frequencies points also to urban background UFPs occurring as a link between traffic particles and the coarser regional-urban background clusters in such a way that it has a bi-modal PSD because the diurnal PSD (from 10:00–12:00 h) still contains particles in the early stages of aging. Maxima of occurrence are sporadically recorded in May and August.

MAPA region is characterized by different types of industries that influence air quality, including several fixed sources. Sources include an oil refinery, steel mills, steelworks; petrochemical Industrial Complex (Pole); and two thermoelectric plants: Thermoelectric Plant of Charqueadas (TERMOCHAR) and of São Jerônimo (TERMOSJ). The thermoelectric plants use coal and the steelworks fuel oil. PNC for 50–100 nm particles (Table 2) are similar to cluster 8 (Regional and urban background in cold nights) and cluster 2 (Regional and urban background).

5.2.3. Clusters of nucleation
5.2.3.1. Cluster 5 – midday nucleated fresh particles, photochemically induced (NPF). Cluster 5 represents only 3.9% of the total hourly PSD with peak concentrations in the nucleation mode reaching ~80,000 #/cm³. Cluster 5 showed the highest normalized concentration at <5 nm (Fig. 1). New particle formation by nucleation grew to reach the instrument detection size limit before growing further and gradually shifting to coarser fractions. This cluster occurred mainly at midday (Fig. 1) in January (warm season) and May (atypial events of warm days), and was associated with the highest insolation, temperature, concentrations of O₃ and SO₂, wind velocity, and with the lowest humidity and NOx, NO and PM₁₀ concentrations (Table 2). All these conditions favour photochemical nucleation according to Kulmala et al., (Kulmala et al., 2016a, 2004), and were found similarly associated with urban PSD clusters attributed to photochemical nucleation in a number of cities by Salimi et al. (2014), Shi et al. (2001), Dall’Osto et al. (2013a, 2012), and Brines et al. (2015, 2014), among others. The diurnal profile showed a peak frequency of occurrence of the NPF cluster at 15:00 h local time (UTC-3 in winter and UTC-2 in summer), characteristic of high insolation and low condensation sink (due to PM dilution) and a minimum during night, when growth processes prevail (see R-UB and RB clusters). As described before, photooxidation of SO₂ to H₂SO₄ is usually the cause of nucleation (Kulmala et al., 2016b, 2004). The higher SO₂ levels associated with the NPF cluster are evidenced, and this pollutant is attributed to diverse industrial plants being operating at MAPA.

5.2.3.2. Cluster 6 – diurnal nucleation episodes (NPF2). Cluster 6 (Fig. 1) represented 18% of the hours of measurement data. Interestingly, this cluster showed a relationship with low concentrations of traffic generated primary pollutants (PM₁₀, NOx) associated with high levels of radiation and wind speed. This cluster presented a unimodal size distribution with the position of the mode at around 15 nm likely due to the contribution of nucleation processes, for low PNC (~15000 #/cm³). Further analysis showed a relationship with the highest precipitation value and wind speed (Table 2). It appears that several types of atmospheric processes lead to particle formation, a possible regional nucleation event. During such events the growth of nucleated particles continues throughout the day, as observed in the daily profile (Fig. 1). Several studies have shown that such events can occur more or less uniformly in air masses that extend over distances of hundreds of kilometers, in outflows of mid-latitude convective storms (Kulmala et al., 2004). Moreover, wind has a strong influence on the particle number, such that a stronger wind speed could reduce twofold the total number counts of particles at diameters ranging from 30 to 450 nm, but had no effect on the small particles (11–30 nm) (Vu et al., 2015). This is likely to be due to the reduction in condensation sink due to the accumulation and coarse mode particles reflected in a low PM₁₀.
concentration.

5.2.3.3. Cluster 3 – growth of nucleated particles in nocturnal aging (GNPF). Cluster 3 represents 14% of the total number particle distribution (Fig. 1), and showed a mode for particles with a diameter less than 20 nm with low number concentration (~8000 #/cm³). The difference with UB (CL7) is the large proportion of the nucleation mode UFP presented in CL3. The diurnal pattern of occurrence is not strongly marked but peaking at nocturnal and early morning hours. Its diurnal pattern of occurrence and association with a low concentration of traffic generated primary pollutants (NOₓ) indicated that it had nontraffic related sources. Given the nocturnal higher frequency, this cluster presents lower levels of wind speed and insolation than clusters 5 and 6 (Table 2). Consequently, this cluster was attributed mainly to the growth of nucleated particles newly formed particles from cluster 6 in nocturnal aging. Thus, the cluster proximity diagram analysis (Fig. 2) shows that the NPF (clusters 5 and 6) are at one end of the diagram and inter-related because of the finest PSD and the diurnal occurrence. Also, cluster 3 is directly related with 6 and 4 because it contains particles from NPF2 and TR2 that undergo growth processes during aging, that at the end will yield to UB PSD (cluster 7), with a bi-modal PSD; this size and time connection is also evidenced with a direct relation between clusters 3 and 7.

Clusters 1 and 4 are associated with most of the clusters that represent UFP generated by traffic, whose impact on urban UFP accounts for these two PSDs dominating the frequency of occurrence in the study area. In the opposite side of the diagram is the background cluster 2 presenting large modal diameters, due to the fact that the spectrum of the cluster showed aged UFP. Background clusters (2, 7, 8) are associated with traffic (1 and 4), nucleation and growth (5, 6 and 3). It is interesting to note that cluster 1 and cluster 2 are associated. Also, cluster 7 and 8 are associated, too (Fig. 2). This suggests that the sources/processes generating cluster 1 and cluster 7 (linking traffic and background PSDs) develop and contribute to generation of wholly background clusters (cluster 2 and 8).

5.3. Particle number concentrations on cold and warm days

Fig. 3 show average PNCs for each size mode for warm and cold days. Average PNC show that particles in the nucleation mode (<10 nm) had a maximum value of ~17,000 #/cm³ on warm days, the highest mean PNC being in summer. As explained before in the results obtained for cluster 5, the increased nanoparticle concentration in the nucleation mode may be due to photochemical nucleation which depends strongly on the intensity of solar radiation. On the other hand, the mean PNC for cold days started to increase up to 20 nm with a predominance of #/cm³, in the Aitken mode. Probably, lower temperatures promote nucleation processes and atmospheric lifetime may increase, associated with increased traffic exhaust emissions (Ripamonti et al., 2013). In the 10–20 nm range, other studies revealed that maximum particle concentrations occurred for both cold and warm days (Morawska et al., 2008; Wehner and Wiedensohler, 2003). Fig. 4 compares average hourly PNC for summer and winter. PNC revealed higher levels in winter time, ranging from 20,245 to 21,945 #/cm³, in the morning from 8:00 to 12:00 during the periods of higher vehicle flow, and at night starting at 18:00, in a range of 21,281–21,071 #/cm³. This reflects the results discussed above, which for cluster 1 a peak was observed during traffic rush hours in the morning and evening. This pattern may be attributed to particles generated from vehicle exhaust emissions. In addition, on cold days, increased nucleation of combustion exhaust emitted from motor vehicles may occur particularly during morning rush hours, as stated before (Morawska et al., 2008). These authors reported that on colder days the greater atmospheric stability (less dispersion) and lower mixing layer height probably contribute to the increase in PNC. Evidence suggests that average PNC was highest in cold days affected by traffic emissions (Fujitani et al., 2012; Pirjola et al., 2006).

Several other studies have been cited on seasonal variations most of which reveal lower concentration in summer and higher concentrations in winter (Wehner and Wiedensohler, 2003; Wu et al., 2008). These authors reported, too, that particle formation is enhanced in periods of lower temperature, from traffic exhaust, especially during the rush hours and with higher temperatures from photochemical particle formation, with levels ranging from 14,028 to 13,841 #/cm³ between 12:00 and 15:00.

6. Conclusion

The measurements of PSD were made in an urban area, MAPA-Brazil, in the period of January 2015 to September 2015, employing a SMPS. K-means clustering analysis was performed on the data collected, resulting in eight size distributions that described the aerosol population.

The present technique has evaluated the particle size distributions in relation to predominant sources; however variations in particles from a source or contributions from multiple sources can make it difficult to identify the origin. Eight clusters and the associated sources were identified in the urban area. A total of 30.5% of the nanoparticles were contributed by two clusters representing the traffic emissions.

The nanoparticles associated with traffic emissions characterized in clusters TR1 and TR2, showed a diurnal variation, with peaks during traffic rush hours in the morning, and evening, characterized by a nucleation mode with a broader Aitken mode. This study clearly showed that the traffic conditions influenced the particle measurements and clearly identified the traffic sources. Background clusters showed similar characteristics with UFPs associated with traffic particles and coarser modes. One of them occurring during colder nights is probably associated to the condensation of secondary species such as nitrate. These three clusters represent 33.6% the total of particle count in the study.

The other three clusters observed were due to nucleation, and they represent 35.9% of the nanoparticles. A nucleation mode characterized the nanoparticles in clusters reflecting diurnal nucleation episodes (NPF2), growth of nucleated particles in nocturnal aging (GNPF) and midday nucleated fresh particles (NPF). Photochemically induced nucleation made a large contribution to nanoparticles in the nucleation mode (<10 nm). The nucleation clusters showed associations with different meteorological variables and air pollutants. This may be explained mainly by aged particles formed from growth of photochemically induced nucleation particles in the ambient air. The nanoparticle concentration was affected by environmental conditions and depended upon the emission type and meteorological conditions, showing seasonal variation. This study showed the contribution of meteorological variables including wind speed and especially precipitation that contributed to nucleation particle formation. The high SO₂ concentration associated with cluster NPF also associated with high solar radiation contributed to nucleation events. Also, an increased PNC
for particles < 10 nm (nucleation mode) was observed on warmer days due to photochemical nucleation.

These results are also important in terms of exposure assessment in public health research. The highest particle number concentrations typically occur in urban areas, and these nanoparticles also have the greatest effect on human health due to deposition of particles in the respiratory system. Consequently, the health impact of nanoparticles may vary from cluster to cluster due to the differing regional deposition of particles of different sizes (Vu et al., 2015).

Acknowledgments

Authors thanks to CNPq for the financial support.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.apr.2018.06.006.

References


