Atmospheric Environment 70 (2013) 64–74

Contents lists available at SciVerse ScienceDirect

Atmospheric Environment

journal homepage: www.elsevier.com/locate/atmosenv

Cluster analysis of roadside ultrafine particle size distributions

Kelly Sabaliauskas, Cheol-Heon Jeong, Xiaohong Yao, Yun-Seok Jun, Greg Evans*

Southern Ontario Centre for Atmospheric Aerosol Research, Department of Chemical Engineering & Applied Chemistry, University of Toronto, 200 College St., Toronto, Ontario M5S 3E5, Canada

HIGHLIGHTS

► Ultrafine particles were measured for five years near a major roadway.
► The k-means clustering algorithm was applied to particle size distributions (PSD).
► Eight PSD types were found to show different temporal patterns.
► The PSD types had differing correlations with NO, NO_2, $PM_{2.5}$ and wind speed.
► Physical interpretations of the conditions yielding each PSD type were elucidated.

ARTICLE INFO

Article history:
Received 16 August 2012
Received in revised form
28 November 2012
Accepted 15 December 2012

Keywords:
Clustering
K-means
Ultrafine particles
Particle size distribution

ABSTRACT

This study reports the diurnal, seasonal, and annual variation of ultrafine particle size distributions in downtown Toronto. The k-means clustering algorithm was applied to five years of size-resolved data for particles with diameters less than 100 nm. Continuous particle number concentrations were measured 16 m from a major arterial roadway between March 2006 and May 2011 using a Fast Mobility Particle Sizer. Eight particle size distribution (PSD) types were identified. The PSD types exhibited distinct weekday–weekend and diurnal patterns. The relative frequency that each PSD occurred varied with season and wind direction and was correlated with other pollutants. These temporal patterns and correlation helped in elucidating the sources and processes that each of the eight PSD represent. Finally, similar PSD types were observed in residential areas located 6 and 15 km away from the central monitoring site suggesting that these PSD types may be generalizable to other sites. Identification of PSD types was found to be a valuable tool to support the interpretation of PSD data so as to elucidate the sources and processes contributing to ultrafine particle concentrations.

1. Introduction

The proliferation of high-time resolution particle sizing instruments has greatly increased the rate at which long-term data on ultrafine particles (UFP: particles with diameters less than 100 nm) is being generated around the world (Mejía et al., 2007; Sabaliauskas et al., 2012; Wåhlin, 2009; Wang et al., 2011). Studies examining the behaviour of UFP have focused on reporting the particle number (PN) concentration as a function of particle size, correlation of PN with gas and particle phase species, and back trajectory analysis to identify possible origins (Ketzel et al., 2004; Mejía et al., 2008; Wang et al., 2011). Although useful, the above mentioned analysis techniques do not fully capitalize on the wealth of information contained within the data. A complimentary representation is worth exploring as the health effects associated with exposure to particulate matter (PM) are widely believed to be dependent on particle size (Sioutas et al., 2005). The particle size distribution (PSD) of UFP present at any point in time and space arises from a combination of the contributing sources and the subsequent transformation processes. UFP are emitted in significant quantities from vehicles (Kittelson, 1998; Seigneur, 2009) and form as a result of interactions between sulphuric acid, water and ammonia (Korhonen et al., 1999). UFP in vehicular emissions differ in size and chemical composition by engine type (Kittelson, 1998), driving conditions (Ketzel et al., 2003), sulphur content of the fuel (Wåhlin, 2009), fuel composition (Turrio-Baldassarri et al., 2004) and emission control technology (Seigneur, 2009). Finally, UFP can shrink or grow rapidly in the atmosphere through evaporation and condensation processes (Yao et al., 2010).

The presence of certain PSD types on a given day may be indicative of the presence of combustion sources, occurrence of

* Corresponding author. Tel.: +1 416 978 1821.
E-mail address: greg.evans@utoronto.ca (G. Evans).

1352-2310/$ – see front matter
http://dx.doi.org/10.1016/j.atmosenv.2012.12.025

nucleation and growth events, or influence of long-range transport, photochemistry and meteorology. Factor analysis, Positive Matrix Factorization (PMF) and cluster analysis have previously been applied to PSD data (Zhou et al., 2004; Kim et al., 2004; Ogulei et al., 2007; Beddows et al., 2009; Kasumba et al., 2009). For factor analysis and PMF, the size bins within the PSD are treated as variables and are modelled as the linear combination of the potential factors plus a residual error term. Although a PSD may consist of a linear combination of multiple sources, interdependencies between the sources and subsequent transformation through condensation, evaporation, and coagulation processes can alter the shape of the PSD in a non-linear fashion. For example, coagulation could be a source of larger particles and a sink for smaller particles. However, particle sinks are not accounted for in the PMF model. Studies that have applied PMF have used monitoring sites located away from major traffic sources (Ketzel and Berkowicz, 2004). Zhou et al. (2004) applied PMF to two months of data from Pittsburgh, USA. However, the occurrence of numerous particle nucleation and growth events that interfered with the PMF analysis led to the removal of all but 17 days from the analysis. Kim et al. (2004) applied the Environmental Protection Agency's Unmix and PMF to two months of PSD data from Seattle, USA. The monitoring site was located far from traffic sources and no nucleation and growth events occurred during the measurement period. Thus, the majority of the data could be retained in the study. In both studies, factors associated with vehicular emissions, long-range transport and photochemical processes were identified.

In areas near traffic, or during nucleation and growth events, the PSD can be altered through coagulation, condensation, and evaporation (Ketzel and Berkowicz, 2004). In these cases, cluster analysis is appropriate because there are no assumptions that the PSD will behave as a linear combination of multiple factors without being impacted by non-linear transformative processes. Beddows et al. (2009) applied the k-means clustering algorithm to PSD data collected at multiple sites in London and at a rural site in the United Kingdom, so as to identify clusters representing the different PSD types at these sites. At the rural site, the clusters had distinctive diurnal behaviour and exhibited correlation with wind speed and combustion-related gas pollutants. At the urban sites in London, some clusters differed from those observed at the rural site and in some cases were unique to each sampling location. In addition to being location specific, some clusters appeared at certain times of the day (i.e., at night or midday). Therefore, cluster-based descriptions are likely to be site specific to an extent and only some PSD types may be common across large or distinctive geographical areas.

Clustering may offer a simple and efficient way to separate PSD shapes into common PSD types or clusters. The variability of these PSD types over time may provide insight into the temporal variability in underlying sources and processes contributing to the formation, growth and removal of UFP. Specifically, the type of PSD present at a site should change over time due to changing sources (gasoline vehicles, diesel vehicles, nucleation events, other regional combustion sources) and processes (condensational growth, evaporation, coagulation, removal). Furthermore, different PSD types could potentially arise from the same source, due to differing amounts of processing, while other PSD types could be from different sources after undergoing the same processing. This study applies cluster analysis to five years of UFP data collected at a roadside site in downtown Toronto, Canada. Previous studies investigating PSD factors or clusters have typically relied on UFP data collected over relatively short-time horizons (1 year or less). This paper examines seasonal and diurnal trends, the relationship between the PSD types and more traditional UFP metrics such as the PN concentration or GMD, the geographic range of the clustering solution, and two case studies exploring long-range transport and seasonal diurnal trends.

2. Methodology

2.1. Study location

The Greater Toronto Area is the most populous region in Canada with a population of 5.5 million people (Statistics Canada, 2012). Downtown Toronto experiences an increase in vehicular traffic between 7:00 and 9:00 in the morning on weekdays that remains constant throughout the day and gradually decreases between 19:00 and 21:00. On weekends, the traffic volume is lower, and the daily maximum occurs later in the day (11:00–13:00). Due to the presence of numerous restaurants, bars, and clubs in the downtown core, overnight traffic volumes are significantly higher on weekends than on weekdays. The traffic fleet is dominated by gasoline-powered spark ignition passenger vehicles. The sulphur content in gasoline and diesel is less than 30 ppm and 15 ppm, respectively.

2.2. Data sources

Ultrafine particle measurements were collected at the Southern Ontario Centre for Atmospheric Aerosol Research (SOCAAR) located on a four-lane major arterial roadway in downtown Toronto (Fig. S1, upper panel). The SOCAAR sampling inlet is located 16 m from the roadway and 6 m aboveground. Traffic volumes range from 15,000 to 20,000 vehicles per day with speeds up to 50 km h^{-1}. The UFP data consisted of 36,252 hourly-averaged UFP size distributions measured by a Fast Mobility Particle Sizer (FMPS: Model 3091, TSI Incorporated, Shoreview, MN, USA). Hourly averages were calculated from 1-min averages for hours that had a minimum of 70% of data reporting (42 min). The resulting hourly-averaged size distributions used as the input data encompassed 83% of the five year period and the missing hours were uniformly distributed across years and seasons. The FMPS reported a number-based size distribution of $dN/d\log D_p$ across 32 size bins between 5.6 and 560 nm. The FMPS had variable counting efficiency for particles less than 8 nm and greater than 300 nm throughout the measurement period. To ensure the reported number concentration was consistent, only size bins between 8 and 300 nm (26 size bins) were included in this study. The PSD was corrected using the procedure described by Jeong and Evans (2009). The resulting size distributions were similar to those obtained using a Scanning Mobility Particle Sizer.

Hourly average concentrations for nitric oxide (NO), nitrogen dioxide (NO_2), ozone (O_3), carbon monoxide (CO), sulphur dioxide (SO_2) and fine particulate matter ($PM_{2.5}$) were available from the Ontario Ministry of the Environment's Downtown Toronto monitoring station located 1 km from the SOCAAR monitoring site (MOE, 2012). Hourly temperature, relative humidity, wind speed and wind direction data was obtained from Pearson International Airport's meteorological station operated by Environment Canada (Environment Canada, 2011). The National Air Pollution Surveillance Network (NAPS) provided speciation data for PM_{10}, $PM_{2.5}$, gas and particle phase polycyclic aromatic hydrocarbons (PAHs), and volatile organic compounds (VOCs) on a 24 h sampling interval (Environment Canada, 2012). The detailed chemical sampling methodology for each species is reported by Dabek-Zlotorzynska et al. (2011).

2.3. Data treatment

Particle size distribution data has two components: the total number concentration and its shape. To normalize for this day-to-