Available online at www.sciencedirect.com
ScienceDirect
Atmospheric Environment 42 (2008) 1727–1741
ATMOSPHERIC ENVIRONMENT
www.elsevier.com/locate/atmosenv

Origins of chemical pollution derived from Mid-Atlantic aircraft profiles using a clustering technique

Jennifer C. Hains[a,*,1], Brett F. Taubman[b], Anne M. Thompson[c], Jeffrey W. Stehr[d], Lackson T. Marufu[e], Bruce G. Doddridge[e], Russell R. Dickerson[a,c]

[a]*Department of Chemistry, University of Maryland, MD, USA*
[b]*Department of Chemistry, Appalachian State University, USA*
[c]*Department of Meteorology, Pennsylvania State University, PA, USA*
[d]*Department of Atmospheric and Oceanic Science, University of Maryland, MD, USA*
[e]*NASA Langley Research Center, USA*

Received 24 July 2007; received in revised form 18 November 2007; accepted 18 November 2007

Abstract

Upwind sources of NO_x and SO_2 play a crucial role in the amount of O_3 and aerosols in the lower troposphere in the Mid-Atlantic US. This paper describes a novel method of clustering trace gas and aerosol profiles allowing for the quantification of the relationship between point sources and pollution levels. This improves our understanding of pollution origins and has the potential for prediction of episodes of poor air quality. A hierarchical clustering method was used to classify distinct chemical and meteorological events from over 200 aircraft vertical profiles in the lower troposphere. Profile measurements included O_3, SO_2, CO and particle scattering from June to August 1997–2003, in the Mid-Atlantic US (mostly in Maryland, Pennsylvania and Virginia). The clustering technique could discriminate distinct profile shapes including measurements made during the 2002 Canadian forest fires. Forty-eight-hour back trajectories were run for each profile and the integrated NO_x and SO_2 point source emissions encountered by each trajectory were calculated using data from the EPA Clean Air Market Division's emissions database. There was a strong correlation between integrated NO_x emissions and O_3 profiles, indicating that O_3 profiles are strongly influenced by and can be predicted with point source emissions. There is a prevalent concentration of SO_2 over the eastern US with mixing ratios decreasing smoothly from about 3.5 ppb near the surface to 0.2 ppb at 2400 m.

Keywords: Ozone; SO_2; Clustering; Emissions

1. Introduction

Understanding meteorological and emission influences on the vertical distribution of trace gases and aerosols can improve the modeling and prediction of pollution events. The goal of this paper is to cluster profiles of O_3, SO_2 and aerosol scattering (based on the profile shape and mixing ratio or scattering coefficient) and describe how meteorological variables and emissions may influence the different clusters. Previous studies (Dorling et al., 1992a, b; Lee et al., 1994; Moy et al., 1994; Dorling

*Corresponding author.
E-mail address: hains@knmi.nl (J.C. Hains).
[1]Currently at KNMI De Bilt, The Netherlands.

1352-2310/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.atmosenv.2007.11.052

and Davies, 1995; Moody et al., 1995, 1998; Harris and Oltmans, 1997; Brankov et al., 1998; Cape et al., 2000; Eneroth et al., 2003; Berto et al., 2004; Jorba et al., 2004; Russell et al., 2004; Taubman et al., 2006) were devoted to clustering back trajectories to describe meteorological patterns associated with different trace gas and aerosol values. Moy et al. (1994), Brankov et al. (1998), and Taubman et al. (2006) were able to use back trajectory clusters to describe meteorological patterns associated with smog events. The converse of this method, clustering O_3 profiles to identify different transport patterns, has been applied to ozonesonde and aircraft data (Diab et al., 2003, 2004; Colette and Gerard, 2005; Colette et al., 2005).

Models used to predict O_3 and $PM_{2.5}$ (the mass concentration of fine aerosol with aerodynamic diameter $<2.5\,\mu m$) levels have limited ability to describe lower tropospheric transport within the planetary boundary layer (Seigneur, 2001; Mebust et al., 2003; Zhang et al., 2004; Hodzic et al., 2005). There is inadequate information on the planetary boundary layer distribution of trace gases and aerosols to improve the models. The University of Maryland has conducted summertime aircraft measurement campaigns over the Mid-Atlantic region since 1996 to gain a better understanding of the chemistry and dynamics of the lower troposphere. Species measured aboard the University of Maryland research aircraft include O_3, SO_2, CO, particle absorption at 565 nm, and total particle scattering at 450, 550, and 700 nm as well as geographic position, static pressure, temperature and relative humidity. For brevity, this paper presents results for O_3, SO_2 and particle scattering measurements. Results from CO and absorption measurements are presented in Hains (2007). Measurements of O_3, pressure, temperature and relative humidity have been made aboard the aircraft since 1997, CO measurements have been made since 1999, SO_2 and particle absorption measurements have been made since 2000 and particle scattering measurements have been made since 2001.

Previous studies (Hennigan et al., 2006; Taubman et al., 2006) found that aircraft profiles with large values of O_3, SO_2 and sulfate were associated with back trajectories that crossed over regions of the Ohio River Valley (where numerous coal fired power plants are located). Other studies have found significant reductions in O_3 resulting from the EPA's NO_x SIP (State Implementation Plan) call (Gego et al., 2008) and from the North American electrical blackout on 14 August 2003 (Marufu et al., 2004). Hu et al. (2006) performed a modeling study of the North American electrical blackout but did not find significant reductions in O_3.

To understand the influence of meteorology and emissions on trace gases and aerosols in the lower troposphere, vertical profiles of the species have been clustered by shape and absolute value (calculated from differences in mixing ratio or scattering coefficients). Taubman et al. (2006) grouped 48 h back trajectories associated with 550 of the University of Maryland profiles into eight distinct meteorological regimes and used these clusters of back trajectories to describe differences among morning and afternoon profiles of O_3, SO_2, CO, particle scattering, Ångström exponent, and particle absorption. The Ångström exponent (α) represents the relative size of particles and was calculated using:

$$a = \frac{[\log(\sigma_{450}) - \log(\sigma_{700})]}{[\log(450) - \log(700)]}. \quad (1)$$

Here σ_{450} and σ_{700} are the scattering coefficients at 450 and 700 nm, respectively. The present study introduces a method for clustering profiles by similarities among the mixing ratios or scattering coefficients. This allows for separation of profiles based on small-scale structure and these differences can be partially attributed to emissions. A quantitative method to relate pollution profiles to their sources will be presented in Section 3. A subset of the profiles analyzed by Taubman et al. (2006) is analyzed in this study. Clustering results from O_3, SO_2, and scattering are presented in this paper and clustering results from CO, absorption and Ångström exponent are presented in Hains (2007).

To explain the effects of emissions on particle scattering it is necessary to account for the effects of relative humidity. Day et al. (2000), Day and Malm (2001), Kotchenruther et al. (1999), Malm et al. (2000a, b, 2003) and Malm and Day (2001), have described some of the effects of relative humidity on particle scattering.

Statistics of SO_2 profiles analyzed in this paper have been presented in Taubman et al. (2006). The shape of the SO_2 profile shows a peak near the surface that decreases smoothly aloft. This shape can be explained by a source near the surface and a major sink near the top of the planetary boundary layer (reaction with peroxides in fair weather cumulus clouds) as described in Marufu et al.

(2005). We compared profiles measured during the INTEX-NA aircraft intensive (Hennigan et al., 2006) with our SO_2 measurements for similar locations and found similar SO_2 profile shapes and mixing ratios.

2. Observations

2.1. Aircraft instrumentation and back trajectory analysis

Details of the University of Maryland research aircraft and instruments onboard the aircraft are described in Taubman et al. (2004a, b) and will be briefly summarized here. The research aircraft is a twin engine Piper Aztec outfitted for atmospheric research. O_3 was measured with a commercial ultraviolet photometric analyzer (TEI Model 49). CO was measured with a commercial non-dispersive infrared gas filter correlation analyzer (TEI Model 48) with modifications described by Dickerson and Delany (1988). SO_2 was measured with a modified (Luke, 1997) pulsed fluorescence detector (TEI Model 43C). Trace gas measurements were recorded every 10 s and 1 min running averages were used in this study. Particle light scattering was measured with a three wavelength (450, 550, and 700 nm) integrating nephelometer (TSI Model 3563), with corrections subsequently made to the raw measurements (Anderson et al., 1996, 1999; Anderson and Ogren, 1998; Taubman et al., 2004a, b). Particle light scattering measurements were recorded every 30 s. Descriptions of the instrument and corrections made to measurements are described in Taubman et al. (2004a, b). The scattering measurements were made after the sample airflow was dried from ambient conditions to an RH of $<20\%$. A growth factor $F(\mathrm{RH})$ was used to account for hydroscopic particle growth. The growth factor $F(\mathrm{RH})$ is the ratio of ambient light scattering σ_{sp} (λ, RH) to dry light scattering $\sigma_{sp}(\mathrm{ref})$:

$$F(\mathrm{RH}) = \frac{\sigma_{sp}(\lambda, \mathrm{RH})}{\sigma_{sp}(\mathrm{ref})}. \quad (2)$$

The growth factor was calculated using the following relationship between particle scattering coefficients at two values of RH:

$$\frac{\sigma_{sp}(\lambda, \mathrm{RH})}{\sigma_{sp}(\mathrm{ref})} = \left[\frac{1 - \mathrm{RH}_{amb}}{1 - \mathrm{RH}_{ref}}\right]^{-\gamma}. \quad (3)$$

Here RH_{amb} is the ambient RH, RH_{ref} is the RH inside the nephelometer and γ is an empirically derived constant. The constant γ was estimated to be 0.35 as per Remer et al. (1997) due to similarities of the sampling platforms and regions of the studies.

Latitude and longitude were measured with a Garmin GPS-90, static pressure with a Rosemount Model 2008 pressure transducer, and temperature and relative humidity were measured with a Rustrak RR2-252 probe using a thermistor and capacitive thin film, respectively.

Flights analyzed in this paper were conducted from June to August 1997–2003 in the Mid-Atlantic region from 35.2°N to 44.5°N and 68.4°W to 81.6°W. Fig. 1 shows the locations of spirals as well as the number of profiles made at each location. Over 50% of spirals were made in Maryland, Virginia and Pennsylvania. All flights were made on days when smog events were forecast. Flights were made between small regional airports with spirals made over the airports from the surface to roughly 3 km above sea level. Spirals were completed within 30 min at a vertical climb rate of 100 m min^{-1}. The SO_2 instrument was zeroed at the bottom and top of every spiral made. The SO_2 data was zero corrected during data processing. Flight patterns were generally chosen to capture transport of pollutants to locations downwind of urban locations, thus flights conducted in the morning (before 12 noon EST) were typically upwind of urban areas in the Mid-Atlantic, while flights conducted in the afternoon (after 12 noon EST) were typically downwind of urban locations (mainly the Baltimore, Washington and Philadelphia metropolitan areas). Most of the flights were conducted in mid morning or mid afternoon, with 42% of the spirals made within ± 2 h of 9:30 h EST and 38% of flights made within ± 2 h of 14:30 h EST.

Fig. 2a–c shows the temperature, potential temperature and water vapor mixing ratio measured during all flights analyzed in this paper. These profiles were sorted into morning and afternoon flights, where morning flights were made before 12 noon EST and afternoon flights were made after 12 noon EST. The morning potential temperature profile shows instability from the surface to 300 m but general stability up to 800 m. The water vapor mixing ratio decreases constantly throughout the morning profile. The profiles of potential temperature and water vapor mixing ratio suggest that the morning median boundary layer top is around 300 m. The afternoon temperature profile shows evidence of subsidence inversions at 2400 and 2800 m. The afternoon potential temperature shows

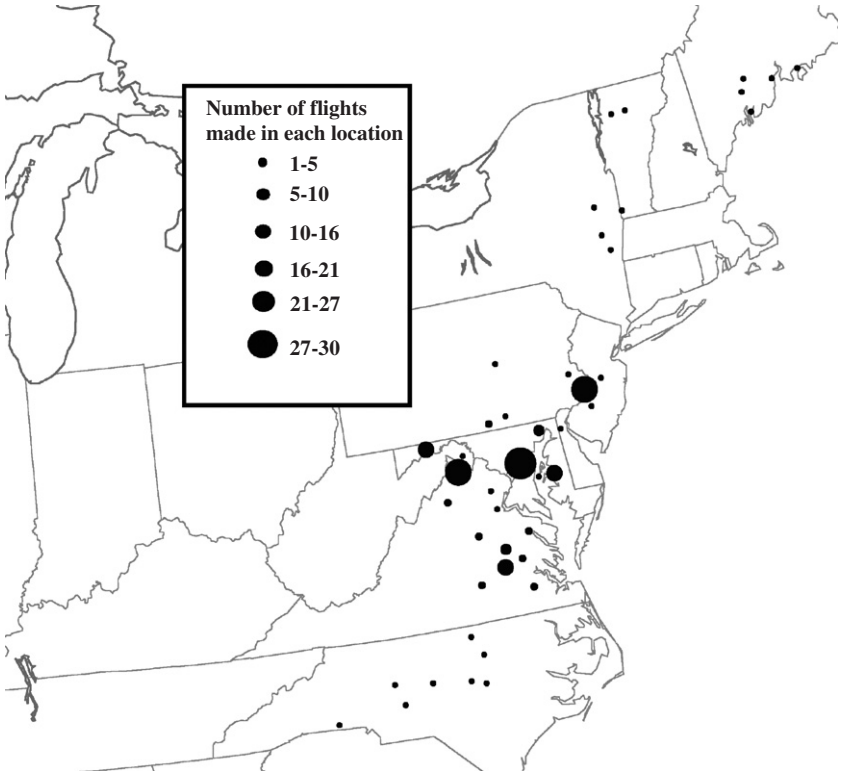

Fig. 1. Location of all flights made in this paper. The size of the circles represents the number of flights made in a given location. Over 50% of the flights were made in Maryland, Virginia and Pennsylvania.

greater instability than the morning profile up to 1100 m. The median water vapor mixing ratio reflects the well-mixed boundary layer that rose to 800–1600 m in the afternoon.

Back trajectories were calculated using the NOAA Air Resources Laboratory (ARL) HYbrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT) model (Version 4) (Draxler and Rolph, 2003, http://www.arl.noaa.gov/ready/hysplit4.html) and 80 km Eta Data Assimilation System (EDAS) 3-hourly archive data. Two-day back trajectories were calculated for each spiral, ending at an altitude of 250, 500, 750 and 1000 m, at the latitude and longitude of the spiral and at the time the spiral was made.

2.2. Cluster analysis

Statistical cluster analysis involves determining the differences between the objects being analyzed, and clustering those objects with the smallest differences. Taubman et al. (2006) examined statistics for the entire dataset and showed median profiles and the statistical spread around those profiles for O_3, SO_2, CO, single scattering albedo, and α. The trace gases showed distinctive profiles; for example, most of the SO_2 was found below 500 m throughout the day, while O_3 was most concentrated above 500 m in the morning and was more uniform from the surface to 2000 m in the afternoon (the largest values in the profile were found near the 1100 m level).

For this work, the raw data were averaged into altitude layers of 100 m (gases) or 200 m (aerosols) and then the layers were grouped into bins shown in Table 1. Typically six trace gas measurements were averaged in each 100 m layer and six aerosol scattering measurements were averaged in each 200 m layer (the number of measurements averaged in a layer depended on the sampling rate and aircraft ascent or descent speed, described in Section 2.1). Not all of the profiles described by Taubman et al. (2006) have been used for this analysis. For this study it was necessary to use the profiles that covered the full altitude range between 150 and 2450 m. Therefore, only profiles that met this criterion were analyzed. Typically, when back trajectories are clustered the Euclidian distance is

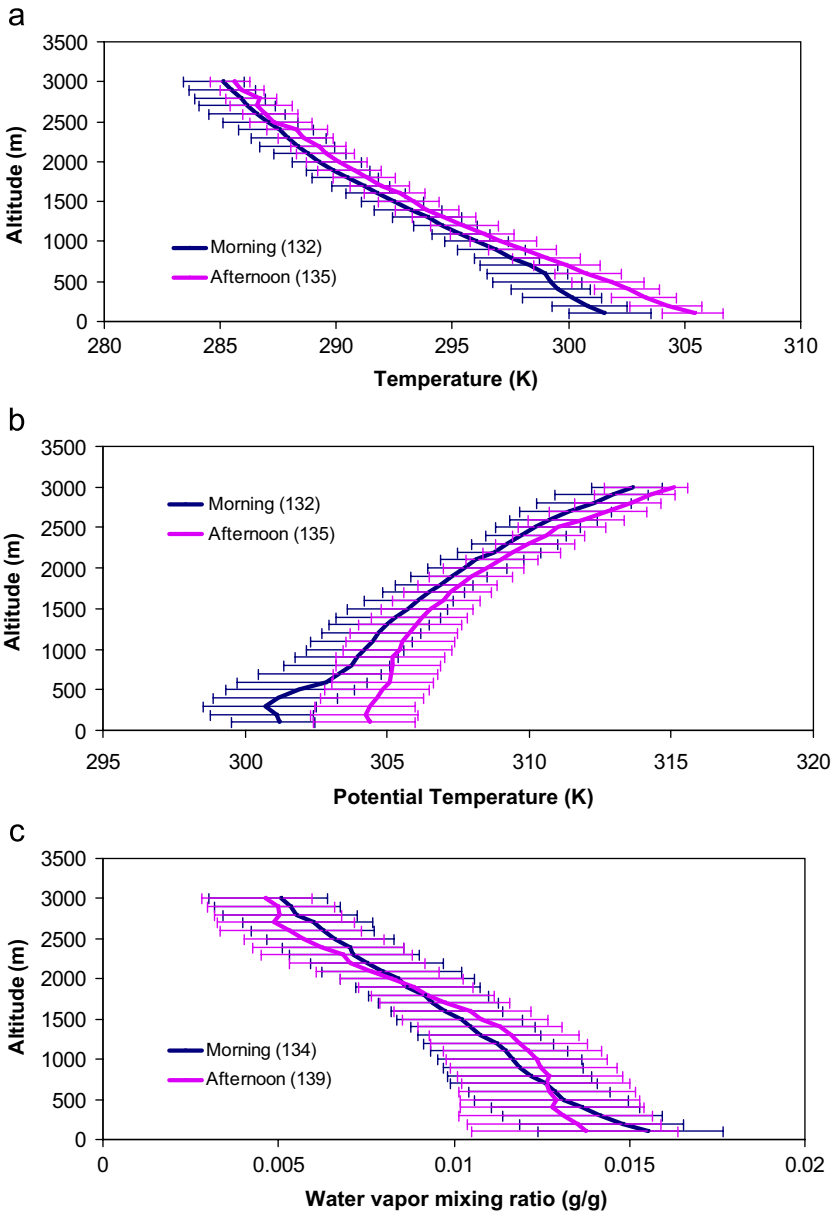

Fig. 2. The median (a) temperature, (b) potential temperature and (c) water mixing ratio for morning (blue) and afternoon (pink) profiles. The error bars represent the 25th and 75th percentiles. The number in parentheses is the number of profiles used to obtain the median.

Table 1
Altitude bins for trace gases and aerosols used in Eq. (1)

k bin	Altitude bins for trace gases (m)	No. of layers in bin (n_k)	Altitude bins for aerosols (m)	No. of layers in bin (n_k)
1	151–650	5	100–900	4
2	651–1150	5	901–1700	4
3	1151–1650	5	1701–2500	4
4	1651–2450	8		

used to determine the difference between each pair of back trajectories (Cape et al., 2000; Taubman et al., 2006). In this analysis we calculate the difference between pairs of chemical profiles so we must account for differences in shape and mixing ratios (or scattering coefficients). The slope and correlation of the points in each pair of profiles under comparison were considered as well as the total difference in mixing ratio (or scattering coefficient) between the two profiles within each

altitude bin. The following equation was used to calculate the differences among the aircraft profiles:

$$D_{ij} = \sum_{k=1}^{k=4} \left\{ \left(\sum_{a=1}^{a=n_k} abs(c_{ia} - c_{ja}) \right)^2 (1 + [1 - r] + [1 - \exp(-(s-1)^2)]) \right\}. \quad (4)$$

Here k is the index for the four different altitude bins for the profiles and a is the index for the n_k layers within the kth bin (Table 1). The species mixing ratio (or scattering coefficient) is represented by c for the ith and jth profile. In each altitude bin, k, there are at least four layers of trace gas or aerosol data. A regression was made to obtain the slope, s, and the correlation coefficient, r, for each pair of profiles using the mixing ratio (or scattering coefficient) within the k bins. The difference, D_{ij}, between two pairs of profiles represents the total difference between the mixing ratios (or scattering coefficients) at each altitude of the profile, as well as how much the slope and correlation coefficient deviate from unity. The first part of Eq. (4) determines the square of the sum of the differences between values at each altitude bin, k. The second part of the equation multiplies the difference by one plus differences associated with the correlation and slope. When the correlation is small or negative the profiles are very different and the $1-r$ portion increases, which increases the total difference D_{ij}. As the correlation coefficient approaches unity the $1-r$ portion approaches zero, and this decreases the total difference D_{ij}. The exponent of the slope portion is used to account for the slope between the pairs of profiles. A slope near unity suggests that the profiles are similar and thus should add little to the total difference. The exponent of the slope was used to guarantee that slopes much different from unity will make the exponential term small and thus increase the total difference. Once the difference between each pair of profiles has been calculated, the profiles with the smallest differences are clustered. These profile differences (D_{ij}) were clustered with a hierarchical clustering algorithm in Matlab that generates cluster trees using an average linkage method. The optimal number of clusters was chosen using techniques described by Taubman et al. (2006). For this analysis profiles made in the morning and afternoon were analyzed together, so that clusters contain both morning and afternoon profiles.

3. Results

3.1. Ozone

Fig. 3 shows the median profiles for each of the six O_3 clusters calculated in the above manner

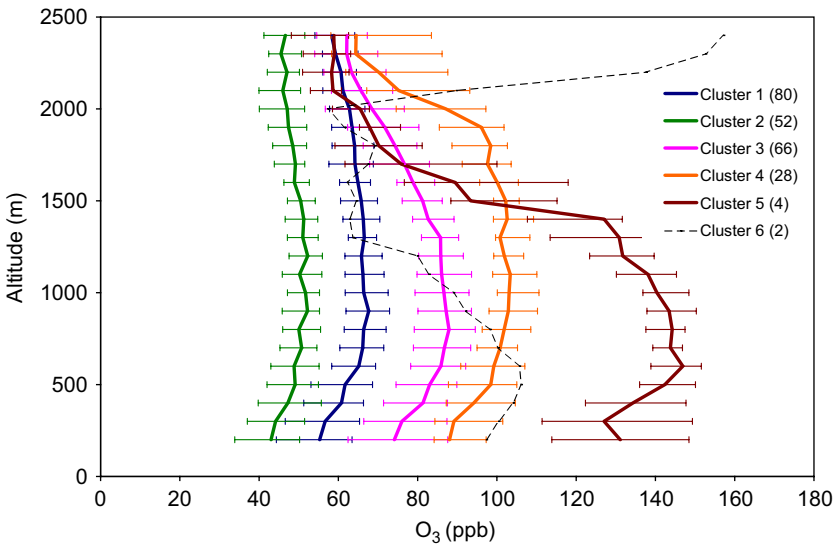

Fig. 3. O_3 profiles for each cluster. The median is shown with the solid line and error bars represent the 25th and 75th percentiles. The number of profiles in each cluster is shown in parentheses in the key. Clusters 1 and 2 show the smallest O_3 values, while Clusters 3–5 show the largest. Cluster 6 profiles were made when air parcels from Canadian forest fires traveled to the Mid-Atlantic region and the peak above 2 km shows their influence.

(a total of 232 O_3 profiles were analyzed). Cluster 6 has very large values of O_3 measured above 2000 m. There are only two profiles in this cluster and they were made on 8 and 9 July 2002 when smoke from Canadian forest fires entered the Mid-Atlantic region (Colarco et al., 2004; Taubman et al., 2004a). Transported O_3 can be seen in the peak (of over 150 ppb) above 2 km.

Back trajectories ending at 1000 m were examined to determine source regions and transport patterns for each cluster. Hains (2007) provides details on this analysis. Profiles from Clusters 3–5 had large O_3 values and their associated back trajectories show passage over the Northern Ohio River Valley, where there is a higher concentration of NO_x sources. Taubman et al. (2006) found a similar relationship between back trajectories concentrated over the Northern Ohio River Valley and large mixing ratios of O_3 and suggested that the large concentration of power plants in this region contributes to the O_3 in the Mid-Atlantic region. The back trajectories for Clusters 3–5 also show larger densities around the I-95 corridor, which is suggestive of stagnation events that lead to higher O_3 values. Cluster 2 has back trajectories that pass over the Atlantic Ocean, which may explain the smaller O_3 values associated with this cluster. Cluster 1, with the second smallest O_3 values, has a column content that is 18% smaller than that of Cluster 3, even though the back trajectories associated with Cluster 1 are similar to those of Cluster 3. To address this discrepancy, the integrated NO_x point source emissions along the back trajectories were examined.

The NO_x emissions were integrated along each back trajectory to examine the impact of upwind emissions on O_3 mixing ratios. Emissions from the EPA Clean Air Market unit level emissions database (http://cfpub.epa.gov/gdm/index.cfm?fuseaction=emissions.wizard) were used in this study. The daily reported emissions were used. The emission data were collected with EPA's Continuous Emission Monitoring (CEM) program. A circle was drawn, centered at each back trajectory position for each hour of the 2 day back trajectory (Fig. 4). The radius of the circle was 80 km to account for uncertainties associated with the back trajectory position and the influence of eddy diffusion and mixing processes. The radius of 80 km was also chosen because the HYSPLIT back trajectories were calculated from the 80 km Eta Data Assimilation System. The emissions within each circle were summed. Emissions from the day

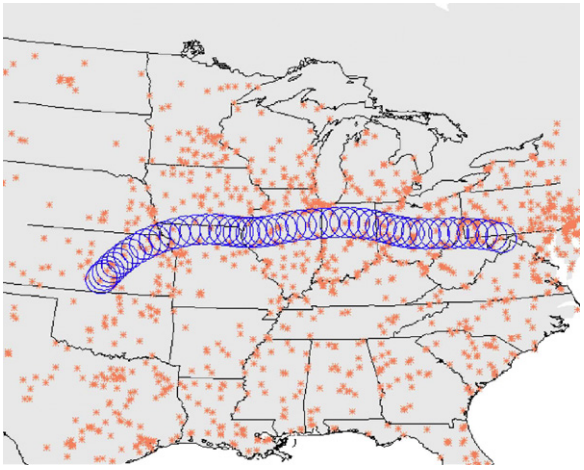

Fig. 4. Circles drawn around an example back trajectory. The emissions contained in each circle were summed. Then emissions from each circle were summed. The pink "*" represents point source locations.

on which the back trajectory crossed a circle were used for the date of each back trajectory position. The summed emissions will be referred to as integrated emissions. The integrated emissions were calculated using back trajectories ending at 250, 500, 750 and 1000 m. Statistics (median, 5th, 25th, 75th and 95th percentiles) for the integrated NO_x emissions (calculated using back trajectories ending at 750 m) associated with O_3 Clusters 1–5 are shown in Fig. 5.

Clusters 3–5 have the largest O_3 column contents and the largest NO_x emissions, whereas Clusters 1 and 2 have the smallest O_3 column contents and the smallest NO_x emissions. Even though back trajectories associated with Clusters 1 and 3 are similar, Cluster 1 is associated with 18% less integrated NO_x emissions, explaining the 18% smaller O_3 values. The median O_3 column content and integrated NO_x emissions for Clusters 1–5 are highly correlated (Fig. 6), suggesting that NO_x emissions from point sources play an important role in downwind O_3 production. The strong correlation is independent of the ending altitude of back trajectories used to calculate the integrated NO_x emissions (and the r^2 is >0.94). This strong correlation between NO_x sources and O_3 profiles agrees with work presented by Gego et al. (2008) who found significant reductions in O_3 resulting from NO_x reductions from the EPA's NO_x SIP call. Marufu et al. (2004) also found similar results, namely a 50% reduction in O_3 profiles during the North American electrical blackout in August 2003. NO_x point source emissions are

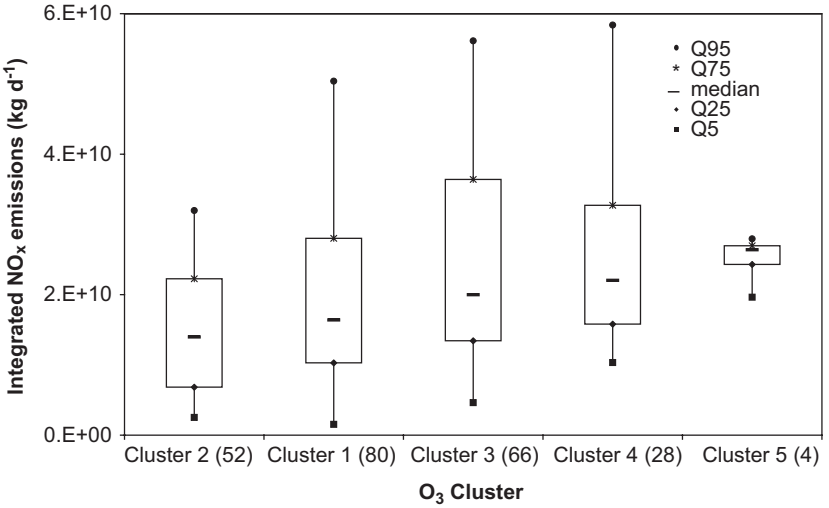

Fig. 5. Statistics for NO_x emissions encountered by back trajectories for each O_3 cluster. The number of profiles in each cluster is shown in parentheses on the x-axis. The NO_x emissions are sums of all emissions ($kg\,d^{-1}$) encountered by a back trajectory (ending at 750 m). Clusters 1 and 2, with the smallest O_3 values are associated with the smallest NO_x emissions. Clusters 3–5, with the largest O_3 values are associated with the largest NO_x emissions.

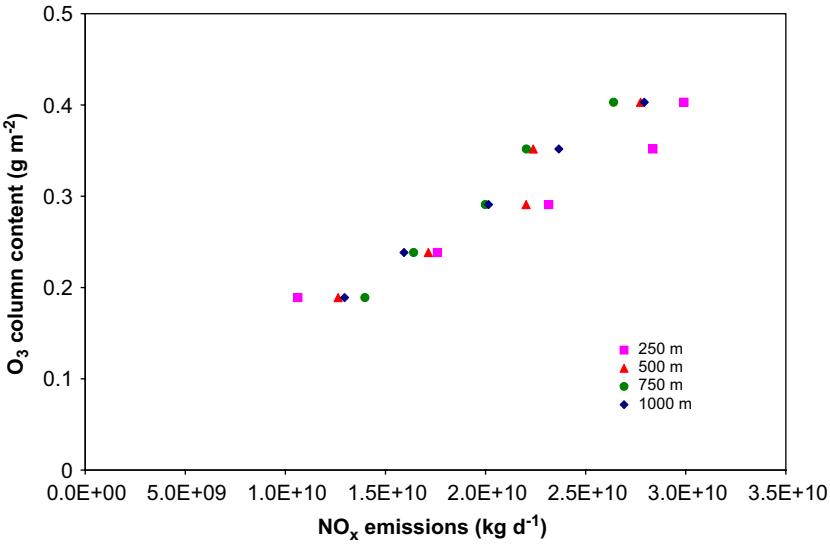

Fig. 6. Median O_3 column content and median integrated NO_x emissions (from point sources) for O_3 Clusters 1–5. Integrated NO_x emissions were calculated using back trajectories ending at 250 m (pink squares), 500 m (red triangles), 750 m (green circles) and 1000 m (blue diamonds). The O_3 column content is strongly correlated with integrated NO_x emissions and this strong correlation is independent of the ending altitude of back trajectories used to calculate the integrated NO_x emissions.

positively correlated with mobile source emissions and this must be taken into consideration when reading Fig. 6.

Profiles were also analyzed by time of day, where morning profiles are defined as those made before 12 noon EST and afternoon profiles are those made at or after 12 noon EST. Sixty-one percent and 68% of the profiles in Clusters 3 and 4 were measured in the afternoon, whereas only 38% and 46% of the profiles in Clusters 1 and 2 were measured in the afternoon. Greater O_3 values in Clusters 3 and 4 may be partly explained by the increased number of afternoon profiles which were generally made downwind of urban centers, and had more time for O_3 production. The morning profiles were clustered separately (as were the afternoon profiles)

and there was still a positive correlation between the O_3 column content and the integrated NO_x emissions.

3.2. SO_2

The clustering technique produced three distinct SO_2 profile clusters (Fig. 7). Because the methodology grouped species profiles, these clusters are not related to the O_3 clusters. Of the 192 profiles analyzed, 170 (89%) fell into the relatively clean Cluster 3. Back trajectories associated with Cluster 3 show a broader area of origin than the more heavily polluted Clusters 1 and 2. The median SO_2 profile from Cluster 2 shows very large values near the surface that decrease above 500 m. Profiles in Cluster 1 show large values from the surface to ~1000 m that do not drop off as rapidly as those in Cluster 2, indicating better mixing in the lower troposphere.

The integrated SO_2 emissions along each back trajectory and statistics for each SO_2 cluster were calculated but the differences between clusters were small. The number of profiles in Clusters 1 and 2 are small and therefore the difference in the integrated emissions has limited statistical significance. SO_2 emissions integrated over 24 and 12 h showed similar relationships with SO_2 profiles as SO_2 emissions integrated over 48 h. The weak relationship between emissions and SO_2 profiles, and the small number of meaningful SO_2 clusters generated, suggests profiles with larger values result from chance encounters with fresh plumes, and that the lifetime of SO_2 in the summer is shorter than 48 h.

3.3. Aerosol scattering

Fig. 8 shows the median scattering for all flights conducted between 2001 and 2003 (June–August) for the 176 profiles analyzed. The clustering methodology produced four scattering clusters, but two are sparsely populated (Clusters 3 and 4) and associated with the Canadian forest fire episode (Fig. 8). The median scattering profile for Cluster 1 is similar to the median profile of all flights. Cluster 1, with 139 profiles, has smaller values than Cluster 2, with 34 profiles (Fig. 8). Back trajectories associated with profiles from Cluster 2 show winds from the Northern Ohio River Valley, whereas Cluster 1 has back trajectories with more variable origins and greater mean wind speeds. Slower wind speeds and stagnant conditions allow for the conversion of SO_2 to sulfate. RH also affects particle scattering. Fig. 9 shows the median RH profiles for Clusters 1 and 2. The median RH associated with Cluster 1 is around 53% while the median RH associated with Cluster 2 is around 65%. Our technique for correcting for particle growth suggests that the RH differences contribute only 10% to the scattering differences. Day and

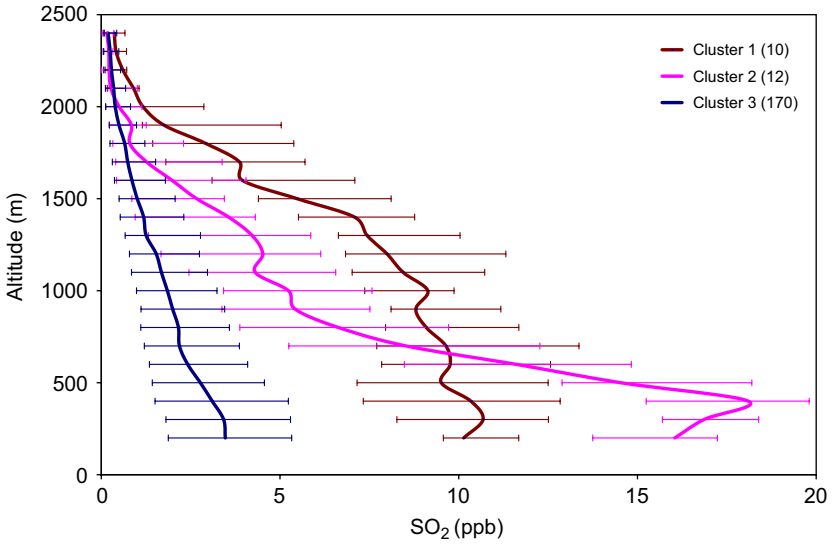

Fig. 7. SO_2 profiles for each cluster. The median is shown with a solid line and the error bars represent the 25th and 75th percentiles. The number of profiles in each cluster is shown in parentheses in the key. Cluster 3 is the background Mid-Atlantic summertime SO_2 profile, representing the majority of SO_2 profiles measured. Clusters 1 and 2 represent profiles impacted by chance plume encounters.

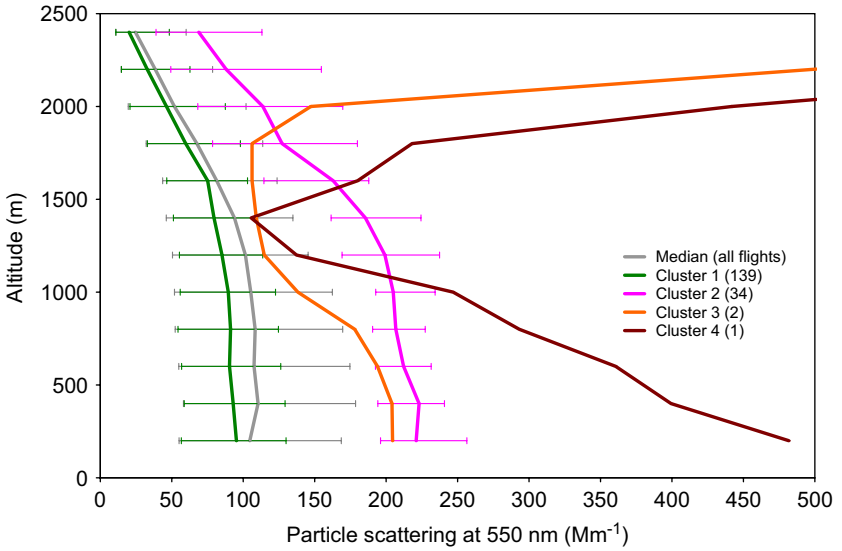

Fig. 8. Particle scattering profiles for each cluster. The median is shown with a solid line and error bars represent the 25th and 75th percentiles. The number of profiles in each cluster is shown in parentheses in the key. Cluster 2 has profiles with twice the scattering value as Cluster 1. Profiles from Clusters 3 and 4 were measured when the Canadian forest fires impacted the region.

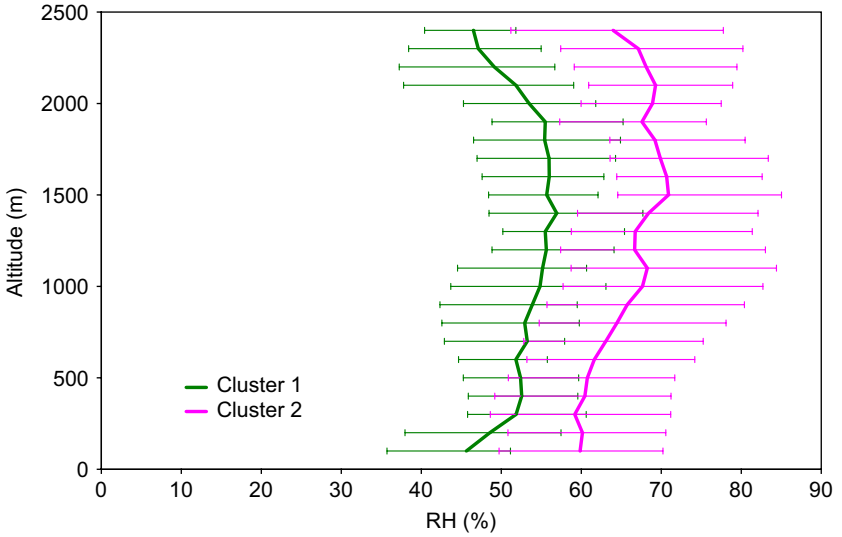

Fig. 9. RH profiles associated with aerosol scattering Clusters 1 and 2. The median is shown with a solid line and the error bars represent the 25th and 75th percentiles. Cluster 2 with larger scattering profiles than Cluster 1, has larger RH values than Cluster 1.

Malm (2001) have measured aerosol scattering at 530 nm at ambient RH (b_{spw}) as well as dry scattering (b_{spd}), with RH < 30%, in the Great Smoky Mountains national park. They calculated the scattering ratio (b_{spw}/b_{spd}) for various RH ranges. At RH values between 50% and 55% they found the average scattering ratio to be 1.33 and at RH values between 60% and 65% they found the average scattering ratio to be 1.45. This suggests that the differences between RH in Clusters 1 and 2 may account for an average of 9% of the scattering differences between Clusters 1 and 2. When the ambient RH is >85% the hydroscopic growth rate is nonlinear. Only 10% of the profiles were made when the RH was >85%.

Fig. 10 shows statistics of integrated SO_2 emissions for Clusters 1 and 2. The median SO_2 emissions for Cluster 2 are almost a factor of two greater than

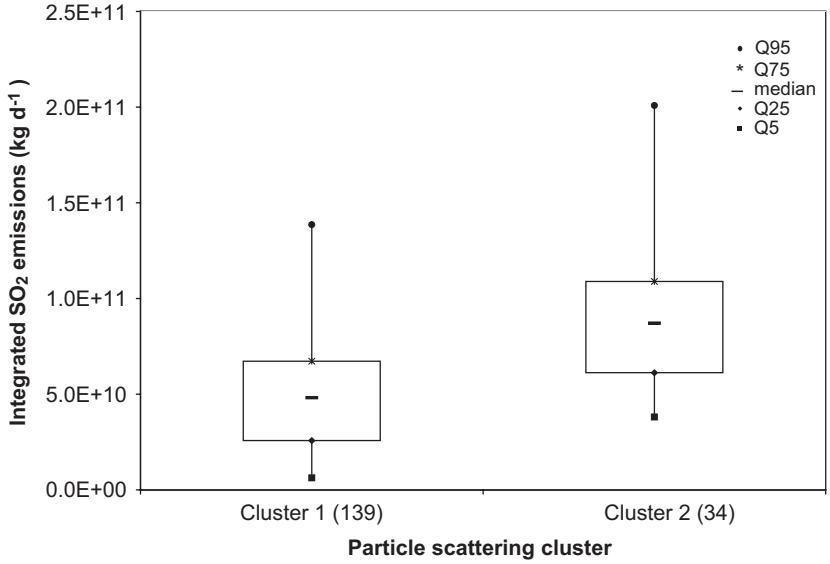

Fig. 10. Statistics for SO_2 emissions encountered by back trajectories for each scattering cluster. The number of profiles in each cluster is shown in parentheses on the x-axis. The SO_2 emissions are sums of all emissions ($kg\,d^{-1}$) encountered by a back trajectory (ending at 750 m). Cluster 2 is associated with almost twice the emissions of Cluster 1, partially explaining why Cluster 2 profiles have twice the scattering values as Cluster 1 (Fig. 8).

those for Cluster 1 and the 200 m averaged scattering in Cluster 2 is also about twice as large as that in Cluster 1 (Fig. 8). This analysis brings up the following question: Why is there a strong positive relationship between scattering profiles and SO_2 emissions, but a weak relationship between SO_2 profiles and SO_2 emissions? This question may partly be explained by the rapid conversion of SO_2 to sulfate in the summer, resulting in minimal SO_2 that can be measured downwind of power plants. Typically, large SO_2 mixing ratios are only measured near a power plant. The strong positive relationship between SO_2 emissions in the Northern Ohio River Valley and particle scattering in the Mid-Atlantic indicates an important source of sulfate aerosols is Northern Ohio River Valley coal fired power plants (e.g., Taubman et al., 2006). This agrees with findings from Hennigan et al. (2006) who found larger sulfate concentrations associated with air parcels passing over the Ohio River Valley.

3.4. Discussion

In order to better understand the chemistry associated with each O_3 cluster, the median profiles for SO_2, particle scattering and CO measured simultaneously with the O_3 profiles from each of the clusters (herein referred to as matching species profiles) were examined (Fig. 11). The clusters with the least O_3 (Clusters 1 and 2) are associated with the least SO_2 and scattering particles, while the clusters with more O_3 (Clusters 3 and 4) are associated with the most SO_2 and scattering particles (measurements of scattering and SO_2 were not made for Cluster 5). The SO_2/CO ratio (Table 2) was also used to determine whether mobile or point source pollution was most influential on the O_3 values. The larger SO_2/CO ratio for Clusters 3 and 4, with large O_3 values, suggests that these clusters were strongly associated with point source emissions. CO profiles were similar for Clusters 1–4, but very large CO was measured in Cluster 5. The larger CO suggests that localized pollution from mobile sources may have affected the Cluster 5 O_3 profiles.

Median matching profiles of SO_2, O_3 and Angstrom exponent (α) were examined for the scattering clusters (Fig. 12). Cluster 2, with the most scattering particles, was also associated with the most O_3 and SO_2, as well as the largest particles (small α values). This suggests that days with more aerosol pollution are often associated with conditions conducive to O_3 production.

The strong correlation between clustered profiles and matching species profiles suggests that the species have similar sources or source regions. The positive relationship between O_3 and SO_2 profiles reaffirms the relationship between point sources and O_3 production presented in Section 3.1. The positive

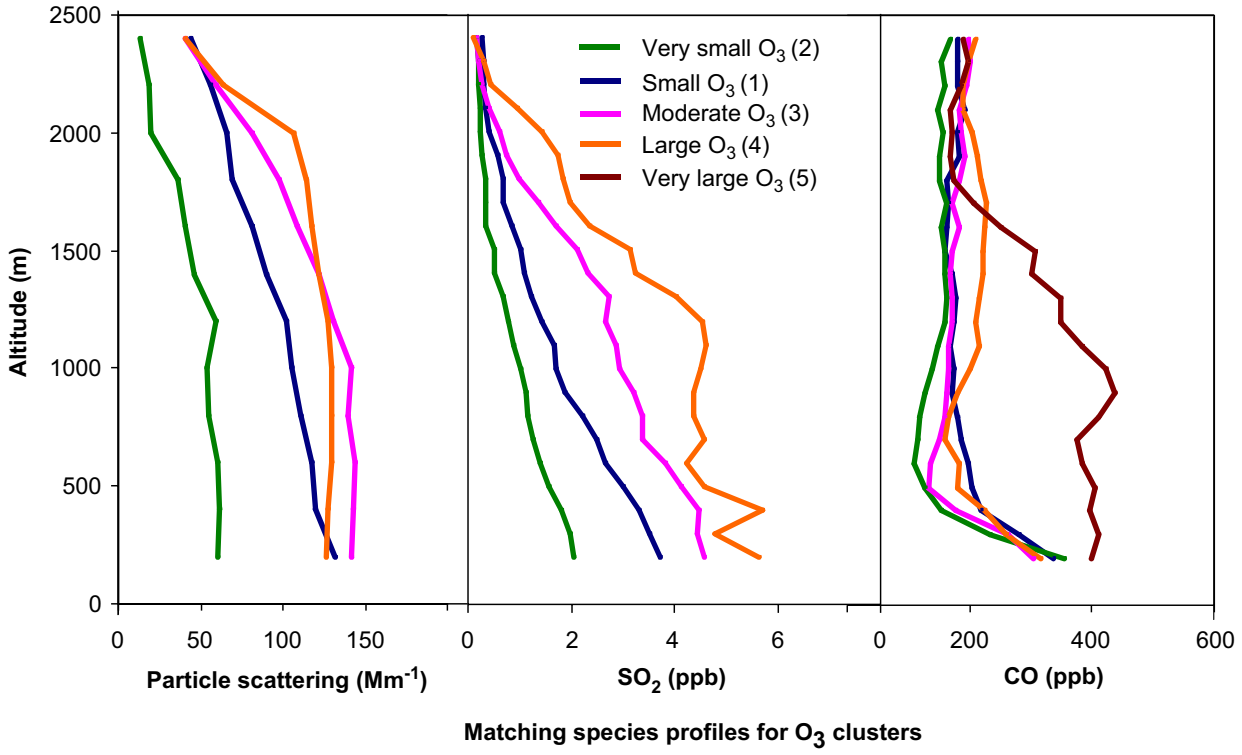

Fig. 11. Matching species profiles for O_3 clusters. Profiles with the smallest O_3 values are associated with the smallest scattering and SO_2 values, while profiles with larger O_3 values are associated with more SO_2 and scattering. The profile with the largest O_3 is associated with the most CO, suggesting that this cluster was influenced by mobile sources. The O_3 cluster number is shown in parentheses in the key.

Table 2
SO_2/CO ratios for O_3 Clusters 1–4

Cluster	SO_2/CO ratio	S.E.	*n*
1	0.008	0.002	38
2	0.005	0.001	18
3	0.013	0.001	25
4	0.019	0.004	10

Clusters 3 and 4, with large O_3 values also have large SO_2/CO ratios, suggesting that they are most heavily influenced by point source emissions. The standard error (S.E.) is the standard deviation divided by the number of profiles (*n*). The difference between the SO_2/CO ratios is larger than the S.E.

relationship between scattering and SO_2 profiles reaffirms the relationship between SO_2 emissions and scattering presented in Section 3.3.

4. Conclusions

Clustering vertical profiles of species allows for separation of distinct pollution events from a large collection of profiles, enabling a better understanding of how meteorology and chemistry affect the shape and mixing ratio (or scattering coefficient) of the profiles. This study resulted in the determination of a quantitative relationship between pollution profiles and sources. The clustering methodology was able to produce six distinct O_3 clusters. One of the clusters was associated with profiles affected by Canadian forest fires. The other five clusters were positively correlated with integrated NO_x emissions from point sources. There is a positive relationship between point source and mobile source NO_x emissions and both point and mobile source emissions affect O_3 production. The linear relationship and strong correlation suggests that reductions of NO_x emissions (from point sources) should reduce ozone levels. This study has laid the groundwork for a more rigorous study relating both point and mobile sources of NO_x to vertical ozone profiles. The EPA's NO_x SIP call required NO_x reductions by 2004 for 21 states. This study will provide a baseline for looking at future impacts of the NO_x SIP call on O_3 values.

SO_2 profiles were less influenced by regional emissions than local emissions. The amount of SO_2

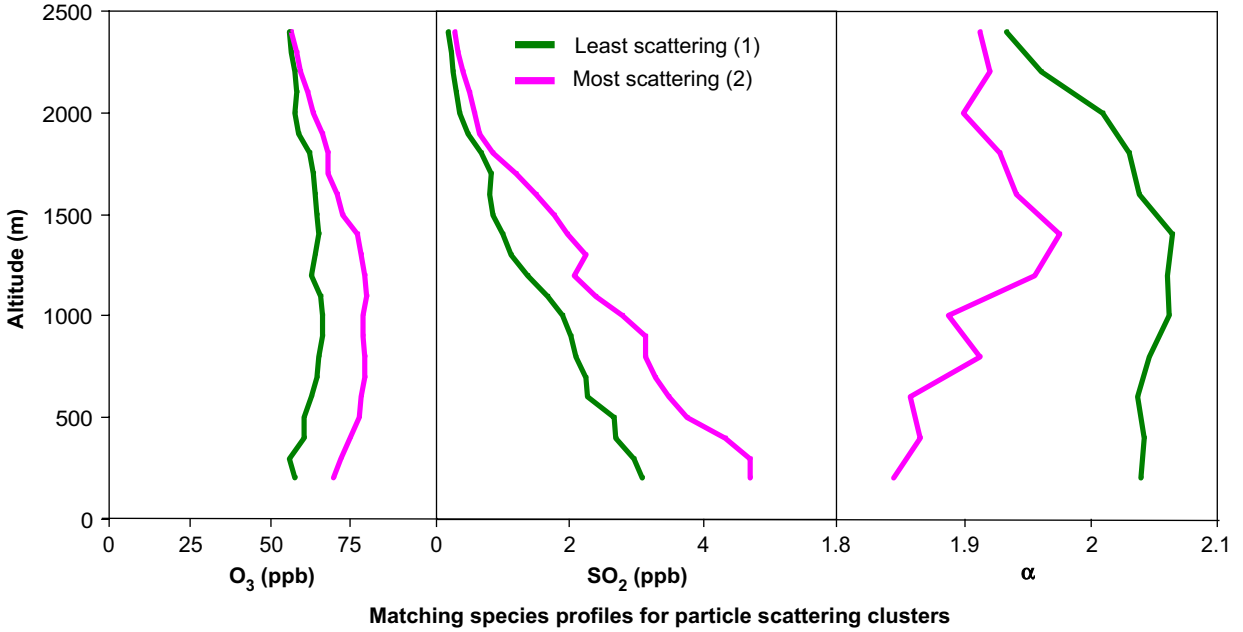

Fig. 12. Matching species profiles for particle scattering clusters. Profiles with the most particle scattering are associated with the most O_3 and SO_2 as well as the largest particles (smallest α values). The scattering cluster number is shown in parentheses in the key.

emitted into an air parcel over the previous 48 h did not have a major impact on observed SO_2 values. The product of SO_2 oxidation, sulfate (using particle light scattering as a proxy), does correlate with SO_2 emissions integrated over the previous 48 h. The characterization of the planetary boundary layer and lower free tropospheric composition of trace gases and aerosols increases our understanding of the relationship between the meteorology and chemistry of the lower troposphere. This will improve chemical transport modeling of these species and our ability to accurately forecast pollution events.

Acknowledgments

This work was supported in part by the Maryland Department of the Environment. The authors would like to thank Tad Aburn, Diane Franks and Matthew Seybold for their support of this project. The authors would also like to thank Charles Piety for forecasting.

References

Anderson, T.L., Ogren, J.A., 1998. Determining aerosol radiative properties using the TSI 3563 Integrating Nephelometer. Aerosol Science and Technology 29, 57–69.

Anderson, T.L., Covert, D.S., Marshall, S.F., Laucks, M.L., Charlson, R.J., Waggoner, A.P., Ogren, J.A., Caldow, R., Holm, R.L., Quant, F.R., Sem, G.J., Wiedensohler, A., Ahlquist, N.A., Bates, T.S., 1996. Performance characteristics of a high-sensitivity, three-wavelength, total scatter/backscatter nephelometer. Journal of Atmospheric and Oceanic Technologies 13, 967–986.

Anderson, T.L., Covert, D.S., Wheeler, J.D., Harris, J.M., Perry, K.D., Trost, B.E., Jaffe, D.J., Ogren, J.A., 1999. Aerosol backscatter fraction and single scattering albedo: measured values and uncertainties at a coastal station in the Pacific NW. Journal of Geophysical Research 104, 26,793–26,807.

Berto, A., Buzzi, A., Zardi, D., 2004. Back-tracking water vapour contributing to a precipitation event over Trentino: a case study. Meteorology Z 13 (3), 189–200.

Brankov, E., Rao, S.T., Porter, P.S., 1998. A trajectory-clustering-correlation methodology for examining the long-range transport of air pollutants. Atmospheric Environment 32 (9), 1525–1534.

Cape, J.N., Methven, J., Hudson, L.E., 2000. The use of trajectory cluster analysis to interpret trace gas measurements at Mace Head, Ireland. Atmospheric Environment 34, 3551–3663.

Colarco, P.R., Schoeberl, M.R., Doddridge, B.G., Marufu, L.T., Torres, O., Welton, E.J., 2004. Transport of smoke from Canadian forest fires to the surface near Washington, DC: Injection height, entrainment, and optical properties. Journal of Geophysical Research 109, D06203.

Colette, A., Gerard, A., 2005. Impact of vertical transport processes on the tropospheric ozone layering above Europe. Part II: Climatological analysis of the past 30 years. Atmospheric Environment 39, 5423–5435.

Colette, A., Gerard, A., Borchi, F., 2005. Impact of vertical transport processes on the tropospheric ozone layering above Europe. Part I: Study of air mass origin using multivariate analysis, clustering and trajectories. Atmospheric Environment 39, 5409–5422.

Day, D.E., Malm, W.C., 2001. Aerosol light scattering measurements as a function of relative humidity: a comparison between measurements made at three different sites. Atmospheric Environment 35, 5169–5176.

Day, D.E., Malm, W.C., Kreidenweis, S.M., 2000. Aerosol light scattering measurements as a function of relative humidity. Journal of Air and Waste Management Association 50, 710–716.

Diab, R.D., Raghunandan, A., Thompson, A.M., Thouret, V., 2003. Classification of tropospheric ozone profiles over Johannesburg based on mosaic aircraft data. Atmospheric Chemistry and Physics 3, 713–723.

Diab, R.D., Thompson, A.M., Mari, K., Ramsay, L., Coetzee, G.J.R., 2004. Tropospheric ozone climatology over Irene, South Africa, from 1990 to 1994 and 1998 to 2002. Journal of Geophysical Research 109, D20301.

Dickerson, R.R., Delany, A.C., 1988. Modification of a commercial gas filter correlation CO detector for increased sensitivity. Journal of Atmospheric and Oceanic Technologies 5 (3), 424–431.

Dorling, S.R., Davies, T.D., 1995. Extending cluster analysis—synoptic meteorology links to characterise chemical climates at six northwest European monitoring stations. Atmospheric Environment 29 (2), 145–167.

Dorling, S.R., Davies, T.D., Pierce, C.E., 1992a. Cluster analysis: a technique for estimating the synoptic meteorological controls on air and precipitation chemistry—method and applications. Atmospheric Environment 26A (14), 2575–2581.

Dorling, S.R., Davies, T.D., Pierce, C.E., 1992b. Cluster analysis: a technique for estimating the synoptic meteorological controls on air and precipitation chemistry—results from Eskdalemuir, S. Scotland. Atmospheric Environment 26A (14), 2583–2602.

Draxler, R.R., Rolph, G.D., 2003. HYSPLIT (HYbrid Single-Particle Lagrangian Integrated Trajectory) Model access via NOAA ARL READY website 〈http://www.arl.noaa.gov/ready/hysplit4.html〉, NOAA Air Resources Laboratory, Silver Spring, MD.

Eneroth, K., Kjellstrom, E., Holmen, K., 2003. A trajectory climatology for svalbard. Investigating how atmospheric flow patterns influence observed tracer concentrations. Physics and Chemistry of the Earth 28, 1191–1203.

Gego, E., Porter, P.S., Gilliland, A., Rao, S.T., 2008. Observation-based assessment of the impact of nitrogen oxides emissions reductions on ozone air quality over the eastern United States. Journal of Applied Meteorology and Climatology 46 (7), 994–1008.

Hains, J.C., 2007. A chemical climatology of lower tropospheric trace gases and aerosols over the US Mid-Atlantic. Ph.D. Dissertation, Department of Chemistry, University of Maryland, College Park, MD.

Harris, J.M., Oltmans, S.J., 1997. Variations in tropospheric ozone related to transport at American Samoa. Journal of Geophysical Research 102 (D7), 8781–8791.

Hennigan, C.J., Sandholm, S., Kim, S., Stickel, R.E., Huey, L.G., Weber, R.J., 2006. Influence of Ohio River valley emissions on fine particle sulfate measured from aircraft over large regions of the eastern United States and Canada during INTEX-NA. Journal of Geophysical Research 111, D24S04.

Hodzic, A., Vautard, R., Bessagnet, B., Lattuati, M., Moreto, F., 2005. Long-term urban aerosol simulation versus routine particulate matter observations. Atmospheric Environment 39, 5851–5864.

Hu, Y., Odman, M.T., Russell, A.G., 2006. Re-examination of the 2003 North American electrical blackout impacts on regional air quality. Geophysical Research Letter 33 (22), L22810.

Jorba, O., Perez, C., Rocadenbosch, F., Baldasano, J.M., 2004. Cluster analysis of 4-day back trajectories arriving in the Barcelona area, Spain, from 1997 to 2002. Journal of Applied Meteorology 43 (6), 887–901.

Kotchenruther, R.A., Hobbs, P.V., Hegg, D.A., 1999. Humidification factors for atmospheric aerosols off the mid-Atlantic coast of the United States. Journal of Geophysical Research 104, 2239–2251.

Lee, G., Merrill, J.T., Huebert, B.J., 1994. Variation of free tropospheric total nitrate at Mauna Loa Observatory, Hawaii. Journal of Geophysical Research 99 (D6), 12821–12831.

Luke, W.T., 1997. Evaluation of a commercial pulsed fluorescence detector for the measurement of low-level SO_2 concentrations during the gas-phase sulfur intercomparison experiment. Journal of Geophysical Research 102 (D13), 16255–16265.

Malm, W.C., Day, D.E., 2001. Estimates of aerosol species scattering characteristics as a function of relative humidity. Atmospheric Environment 35, 2845–2860.

Malm, W.C., Day, D.E., Kreidenweis, S.M., 2000a. Light scattering characteristics of aerosols as a function of relative humidity: Part I—A comparison of measured scattering and aerosol concentrations using theoretical models. Journal of Air & Waste Management Association 50, 686–700.

Malm, W.C., Day, D.E., Kreidenweis, S.M., 2000b. Light scattering characteristics of aerosols as a function of relative humidity: Part II—A comparison of measured scattering and aerosol concentrations using statistical models. Journal of Air and Waste Management Association 50, 701–709.

Malm, W.C., Day, D.E., Kreidenweis, S.M., Collett, J.L., Lee, T., 2003. Humidity-dependent optical properties of fine particles during the Big Bend Regional Aerosol and Visibility Observational Study. Journal of Geophysical Research, 108(D9), doi:10.1029/2002JD002998.

Marufu, L.T., Taubman, B.F., Bloomer, B., Piety, C.A., Doddridge, B.G., Stehr, J.W., Dickerson, R.R., 2004. The 2003 North American electrical blackout: an accidental experiment in atmospheric chemistry. Geophysical Research Letter 31, L13106.

Marufu, L.T., Taubman, B.F., Bloomer, B., Piety, C.A., Doddridge, B.G., Stehr, J.W., Dickerson, R.R., 2005. Reply to comment by D.A. Hansen et al. on "The 2003 North American electrical blackout: an accidental experiment in atmospheric chemistry". Geophysical Research Letter 32, L10813.

Mebust, M., Eder, B., Binhowski, F., Roselle, S., 2003. Models-3 community multiscale air quality (CMAQ) model aerosol component 2. Model evaluation. Journal of Geophysical Research 108 (D6), 4184.

Moody, J.L., Oltmans, S.J., Levy II, H., Merrill, J.T., 1995. Transport climatology of tropospheric ozone: Bermuda,

1988–1991. Journal of Geophysical Research 100 (D4), 7179–7194.

Moody, J.L., Munger, J.W., Goldstein, A.H., Jacob, D.J., Wofsy, S.C., 1998. Harvard forest regional-scale air mass composition by patterns in atmospheric transport history (PATH). Journal of Geophysical Research 103 (D11), 13181–13194.

Moy, L.A., Dickerson, R.R., Ryan, W.F., 1994. Relationship between back trajectories and tropospheric trace gas concentrations in rural Virginia. Atmospheric Environment 28 (17), 2789–2800.

Remer, L.A., Gasso, S., Hegg, D.A., Kaufmann, Y.J., Holben, B.N., 1997. Urban/industrial aerosol: ground based sun/sky radiometer and airborne in situ measurements. Journal of Geophysical Research 102, 16849–16859.

Russell, A., McGregor, G.R., Marshall, G.J., 2004. An examination of the precipitation delivery mechanisms for Dolleman Island, eastern Antarctic Peninsula. Tellus 56A, 501–513.

Seigneur, C., 2001. Current status of air quality models for particulate matter. Journal of Air and Waste Management Association 51, 1508–1521.

Taubman, B.F., Marufu, L.T., Vant-Hull, B.L., Piety, C.A., Doddridge, B.G., Dickerson, R.R., Li, Z., 2004a. Smoke over haze: aircraft observations of chemical and optical properties and the effects on heating rates and stability. Journal of Geophysical Research 109 (D2), D02206.

Taubman, B.F., Marufu, L.T., Piety, C.A., Doddridge, B.G., Stehr, J.W., Dickerson, R.R., 2004b. Airborne characterization of the chemical, optical, and meteorological properties, and origins of a combined ozone/haze episode over the eastern US. Journal of Atmospheric Science 61 (14), 1781–1793.

Taubman, B.F., Hains, J.C., Thompson, A.M., Marufu, L.T., Doddridge, B.G., Stehr, J.W., Piety, C.A., Dickerson, R.R., 2006. Aircraft vertical profiles of trace gas and aerosol pollution over the mid-Atlantic United States: statistics and meteorological cluster analysis. Journal of Geophysical Research 111 (D10), D10S07.

Zhang, Y., Pun, B., Wu, S., Vijayaraghavan, K., Seigneur, C., 2004. Application and evaluation of two air quality models for particulate matter for a Southeastern US episode. Journal of Air and Waste Management Association 54, 1478–1493.