Variability of O_3 and NO_2 Profile Shapes during DISCOVER-AQ: Implications for Satellite Observations and Comparisons to Model-Simulated Profiles

Authors: Clare Marie Flynn[1*]; Kenneth E. Pickering[1,2]; James H. Crawford[3]; Andrew J. Weinheimer[4]; Glenn Diskin[3]; K. Lee Thornhill[3]; Christopher Loughner[2,5]; Pius Lee[6]; Sarah A. Strode[7]

[1] Department of Atmospheric and Oceanic Science, University of Maryland, College Park

[2] NASA Goddard Space Flight Center

[3]NASA Langley Research Center

[4]NCAR Atmospheric Chemistry Observations and Modeling Laboratory

[5]Earth System Science Interdisciplinary Center, University of Maryland, College Park

[6]NOAA Air Resources Laboratory

[7]GESTAR, NASA/GSFC

*corresponding author, now at Department of Earth System Science, University of California, Irvine;

Tel:+13016559784, e-mail: cflynn@atmos.umd.edu

Keywords Ozone; Nitrogen Oxides; DISCOVER-AQ; Cluster analysis; aircraft measurement campaign; measurement-model comparison

Abstract

To investigate the variability of *in situ* profile shapes under a variety of meteorological and pollution conditions, results are presented of an agglomerative hierarchical cluster analysis of the *in situ* O_3 and NO_2 profiles for each of the four campaigns of the NASA DISCOVER-AQ mission. Understanding the observed profile variability for these trace gases is useful for understanding the accuracy of the assumed profile shapes used in satellite retrieval algorithms as well as for understanding the correlation between satellite column observations and surface concentrations. The four campaigns of the DISCOVER-AQ mission took place in Maryland during July 2011, the San Joaquin Valley of California during January-February 2013, the Houston, Texas, metropolitan region during September 2013, and the Denver-Front Range

region of Colorado during July-August 2014. Several distinct profile clusters emerged for the California, Texas, and Colorado campaigns for O_3, indicating significant variability of O_3 profile shapes, while the Maryland campaign presented only one distinct O_3 cluster. In contrast, very few distinct profile clusters emerged for NO_2 during any campaign for this particular clustering technique, indicating the NO_2 profile behavior was relatively uniform throughout each campaign. However, changes in NO_2 profile shape were evident as the boundary layer evolved through the day, but they were apparently not significant enough to yield more clusters. The degree of vertical mixing (as indicated by temperature lapse rate) associated with each cluster exerted an important influence on the shapes of the median cluster profiles for O_3, as well as impacted the correlations between the associated column and surface data for each cluster for O_3. The correlation analyses suggest satellites may have the best chance to relate to surface O_3 under the conditions encountered during the Maryland campaign Clusters 1 and 2, which include deep, convective boundary layers and few interruptions to this connection from complex meteorology, chemical environments, or orography. The regional CMAQ model captured the shape factors for O_3, and moderately well captured the NO_2 shape factors, for the conditions associated with the Maryland campaign, suggesting that a regional air quality model may adequately specify *a priori* profile shapes for remote sensing retrievals. CMAQ shape factor profiles were not as well represented for the other regions.

1. Introduction

Because of the global coverage, increasingly high spatial resolution, fixed temporal resolution, and improved retrievals of tropospheric column abundances, satellite observations have great potential for diagnosis of near-surface air quality conditions (Beirle et al., 2003; Boersma et al., 2008; Chatfield and Esswein, 2012; Fishman et al., 2008; Flynn et al., 2014;

Lamsal et al., 2011; Martin, 2008). Satellite observations can be especially useful for the Environmental Protection Agency (EPA) criteria pollutants (http://www.epa.gov/air/criteria.html) ozone (O_3) and nitrogen dioxide (NO_2), particularly in regions that lack a sufficient surface air quality monitoring network. Recent work has demonstrated that such an application should be possible for NO_2. Lamsal et al. (2008, 2010) observed significant correlation between *in situ* surface NO_2 mixing ratios and ground-level NO_2 observations inferred from Ozone Monitoring Instrument (OMI) column amounts, after application of local scaling factors derived from the GEOS-Chem model. Other works have instead scaled surface NO_2 values to obtain column amounts with the use of assumed NO_2 profiles, and have found good agreement between OMI tropospheric columns and these scaled columns (Boersma et al., 2009; Knepp et al., 2013; Ordóñez et al., 2006). Chatfield and Esswein (2012) examined ozonesonde data over the U.S. and observed a significant correlation between partial column (0-3 km) and near-surface O_3 (500 m) observations. Flynn et al. (2014) examined partial or full tropospheric column amounts derived from aircraft or Pandora UV/Vis spectrometer data sets for O_3 and NO_2 during July 2011 in the Baltimore-Washington metropolitan region, and found a wide range of degrees of correlation between column and surface data, with O_3 generally demonstrating a greater correlation than NO_2. These results suggest that satellite-observed ozone observations could be used to estimate surface ozone, provided there is sufficient sensitivity to the lowermost troposphere. Natraj et al. (2011) demonstrated that lower tropospheric O_3 retrievals could be greatly improved by flying both UV and thermal IR sensors together on future satellites.

These studies are encouraging that column amounts and surface concentrations can be related for trace gases, and follow in the footsteps of research relating satellite-derived aerosol optical depth (AOD) to surface particulate matter of diameter less than 2.5 microns ($PM_{2.5}$).

Recent work correlating AOD and $PM_{2.5}$ over the continental U.S. found a range of correlation coefficients, with the best correlation obtained in the northeast, where aerosol type, abundance, and boundary layer height were most uniform, and the poorest correlation over the Pacific Northwest (Engel-Cox et al., 2004; Hoff and Christopher, 2009). Wang and Christopher (2003) found a linear correlation coefficient (R) of 0.7 between AOD observed by the Moderate Resolution Imaging Spectroradiometer (MODIS) onboard the Terra and Aqua satellites and 1-hour and 24-hour averaged $PM_{2.5}$ over seven sites in Jefferson County, Alabama. The GEOS-Chem model has also been applied to the MODIS retrieval of AOD to derive local aerosol optical properties, in place of climatological values that have been averaged monthly and over large spatial regions. The use of simulated optical properties for each location of interest resulted in improved AOD values, and demonstrates the importance of spatial and temporal resolution of the input datasets to the accuracy of satellite retrievals (Drury et al., 2008; Drury et al., 2010; Wang et al., 2010). Thus, remotely-sensed columnar amounts of gaseous pollutants or optical thicknesses, often with the use of a model, may be useful for surface concentrations for both trace gases and aerosols. However, such work also highlighted the considerable difficulties inherent in understanding these relationships. Biases remain in satellite retrievals, while current satellite instruments have lower sensitivity to the lower troposphere for trace gases (Liu et al., 2010; Martin, 2008). Aerosols have additional complications, including the effects of ambient humidity on aerosol composition and optical properties, and the impacts of mixing height on aerosol extinction (Wang and Christopher, 2003). The wide range of correlations obtained and the use of model- or data-derived scaling factors demonstrate a need for better understanding of the processes connecting the column and surface.

Understanding of the variability of *in situ* profile shapes is useful for understanding the degree of correlation between column and surface data. Satellite data may be more useful for air quality applications in some parts of the day than others and under certain meteorological conditions. How well do the assumed profile shapes used in satellite retrievals capture observed conditions and ultimately what is the resulting impact on the ability of satellite-observed columns to represent surface air quality? Additionally, the assumed profile shapes used in retrievals are given not as profiles of volume mixing ratios, but as shape factors that are provided by simulations of global chemical transport models such as the NASA Global Modeling Initiative (GMI) model. The shape factor is defined as the ratio of the partial column within a vertical layer z (Ω_z) to the full tropospheric column (Ω_{trop}), and thus depends indirectly on the mixing ratio profile:

$$S(z) = \Omega_z/\Omega_{trop} \quad (1)$$

For the NASA standard NO_2 product (Bucsela et al., 2013) shape factors are used in the radiative transfer model to calculate the air mass factor (AMF), which is used to convert the slant column abundance to the vertical column abundance according to

$$\Omega_v = \Omega_s/AMF \quad (2)$$

where Ω_s is the slant column and Ω_v is the vertical column (Chance, 2002; Lamsal et al., 2014; Palmer et al., 2001). The AMF is also used to compute the O_3 vertical column from the retrieved slant column when using the differential optical absorption spectroscopy (DOAS) technique, such as used to retrieve the OMI total vertical O_3 columns (Bhartia, 2002). Liu et al. (2010) employ a similar method in their retrieval of OMI O_3 vertical columns, in which the AMF is not used but rather *a priori* partial O_3 column amounts at each model vertical layer in the retrieval computations. Accurate representation of the O_3 or NO_2 profile shapes is critical to accurate

120 representation of the shape factors used in satellite retrievals, and thus retrieval accuracy. A
121 high-resolution NO_2 retrieval from OMI has been performed by Russell et al. (2011) using NO_2
122 profiles from the WRF-Chem model at 4-km horizontal resolution. In this paper, we evaluate the
123 ability of a regional air quality model (CMAQ) to produce accurate NO_2 profiles for use in
124 satellite retrievals.

125 Previous studies have examined the observed trace gas or aerosol profile variability over
126 several regions. Diab et al. (2003) investigated O_3 aircraft profiles over Johannesburg, South
127 Africa, through application of a clustering technique. Six unique O_3 clusters were found, which
128 were further related to air mass origin through back trajectory modeling. Diab et al. (2004)
129 clustered ozonesonde profiles over Irene, South Africa, for the periods 1990-1994 and 1998-
130 2002; these clusters were also related to air mass origin and meteorological conditions. Using a
131 self-organizing maps technique, Stauffer et al. (2016) obtained nine distinct clusters of
132 tropospheric O_3 profiles form four long-term U.S. ozonesonde locations, which corresponded to
133 distinct meteorological and pollution conditions. Additionally, this work determined that O_3
134 profile climatologies greatly underrepresented O_3 profiles at these sites. Taubman et al. (2006)
135 and Hains et al. (2008) examined summertime trace gas and aerosol aircraft vertical profiles over
136 the mid-Atlantic U.S. during 1997-2003. Through a cluster analysis of back trajectories or
137 pollutant profiles, distinct pollution regimes and their associated meteorological conditions and
138 emissions were identified for the summertime Mid-Atlantic region. However, while these studies
139 examined long time record data sets, they were limited in spatial extent and occasionally also by
140 time of year examined. These studies also did not investigate NO_2 profile variability, nor
141 evaluate the ability of air quality models to replicate these profile shape clusters.

142 The ultimate goal of the DISCOVER-AQ (Deriving Information on Surface conditions
143 from Column and Vertically Resolved Observations Relevant to Air Quality) project is to provide
144 information relevant to improving our ability to relate satellite-observed column densities to
145 surface conditions for aerosols, O_3, NO_2, and formaldehyde. DISCOVER-AQ combines P-3B
146 aircraft *in situ* profiling of trace gas species, aerosol properties, and key meteorological variables,
147 UC-12 aircraft remote sensing of aerosols and trace gas columns, observations of surface
148 conditions from the existing network of surface air quality monitors, remote sensing of trace gas
149 columns and aerosols from a network of ground-based Pandora UV/vis spectrometers and a
150 network of AERONET sun photometers collocated with the air quality monitors (and additional
151 monitors at some sites), and model simulations for each of the four campaigns (Flynn et al.,
152 2014). The first campaign was conducted in the Baltimore-Washington metropolitan region of
153 Maryland during July 2011, the second in the San Joaquin Valley of California during January-
154 February 2013, the third in the Houston, TX, metropolitan area during September 2013, and the
155 last in the Front Range region of Colorado during July-August 2014. Thus, the project covered a
156 large range of meteorological and pollution conditions throughout these four campaigns.

157 In support of DISCOVER-AQ, results are presented of an agglomerative hierarchical
158 cluster analysis of P-3B *in situ* profiles for O_3 and NO_2 for each of the four campaigns. Through
159 these analyses, the variability of the *in situ* P-3B O_3 or NO_2 profiles will be characterized for each
160 campaign. Classes of profiles are identified for each trace gas and each campaign, and
161 meteorological conditions influencing these classes will be investigated as well as the associated
162 column vs. surface correlations for each cluster. Shape factors are computed from the O_3 and NO_2
163 observations. The observed cluster shape factor results are also compared to shape factors from
164 the Community Multiscale Air Quality (CMAQ) model and NASA Global Modeling Initiative

(GMI) model to assess model performance and the relevance of the model profiles for use in satellite retrievals will be evaluated. In addition, shape factors for CO are examined to better diagnose vertical mixing in the CMAQ model compared with observations.

2. Data

2.1 P-3B Aircraft Data

A complete description of the DISCOVER-AQ measurements is publicly available at http://www-air.larc.nasa.gov/missions/discover-aq/discover-aq.html. *In situ* trace gas volume mixing ratio data were collected by the P-3B aircraft over a network of six surface air quality monitoring sites during the Maryland, California, and Colorado campaigns, and over a network of eight monitoring sites during the Texas campaign. Typically, three spirals were accomplished over each spiral site during each flight day (typically morning, midday, and afternoon), with 10-15 flight days per campaign. Spiral sites for each campaign are mapped in Fig. 1a-d. The National Center for Atmospheric Research (NCAR) $NO_{xy}O_3$ instrument, a 4-channel chemiluminescence instrument for the measurement of NO, NO_2, NO_y, and O_3, provided the P-3B O_3 (uncertainty of 5%) and NO_2 (uncertainty of 10%) *in situ* observations used here. . P-3B carbon monoxide (CO) *in situ* observations were provided by the Differential Absorption Carbon Monoxide instrument (DACOM; uncertainty of 2%). The National Suborbital Education and Research Center (NSERC) P-3B data acquisition and distribution system provided the *in situ* observations of altitude and meteorological observations used in these analyses.

2.2 Surface Volume Mixing Ratio Data

2.2.1 Maryland Campaign

Surface volume mixing ratio data were provided by the Maryland Department of the Environment (MDE) for O_3 at all spiral sites from a chemiluminescence monitor, for NO_2 at

Essex from a molybdenum converter chemiluminescence monitor corrected for NO_y interferences, and for NO_y from monitors at the Aldino and Beltsville sites. The EPA provided NO_2 measurements from chemiluminescence instruments with photolytic converters at the Edgewood and Padonia spiral sites. The NASA Chemical, Optical, and Microphysical Measurements of In-situ Troposphere (COMMIT; http://smartlabs.gsfc.nasa.gov) trailer provided O_3 and photolytic converter NO_2 measurements for Fair Hill.

2.2.2 California Campaign

Surface O_3 data were provided by the San Joaquin Valley Air Pollution Control District (SJV) at all California spiral sites except Huron. The EPA provided surface photolytic converter NO_2 measurements at Bakersfield and Porterville, The Pennsylvania State University Nittany Atmospheric Trailer and Integrated Validation Experiment (NATIVE; http://ozone.met.psu.edu/Native/) provided O_3 and NO_y data at the Porterville site, and The Millersville University of Pennsylvania (MU) provided O_3 and molybdenum-converter NO_x data at the Huron site. SJV surface NO_y data were also available at the Hanford site, while EPA provided photolytic NO_2 measurements at Bakersfield. The California Air Resources Board (CARB) provided surface NO_2 at the Fresno site. Photolytic surface NO_2 data were used instead of molybdenum converter NO_2 (with interferences from NO_y species) data for the correlation analyses that follow at those sites for which it was available (Bakersfield and Porterville).

2.2.3 Texas Campaign

The Texas Commission on Environmental Quality (TCEQ) provided surface O_3 data from a chemiluminscence monitor at the Conroe, Channelview, Deer Park, Galveston, and Manvel Croix spiral sites. TCEQ also provided surface molybdenum-converter NO_2 data at the Conroe, Channelview, and Deer Park sites, and NO_y data at the Galveston and Manvel Croix

211 sites. The University of Houston (UH) provided O_3 and NO_y surface data at the Moody Tower
212 site; the surface at this site was approximately 70 m AGL, as these monitors sat on the rooftop of
213 the Moody Tower on the UH campus. NATIVE provided O_3 and NO_y at Smith Point. Almost no
214 surface data were available for the West Houston site. Lastly, NOAA provided NO_2 data from a
215 chemiluminescence monitor equipped with a photolytic converter at Galveston, and NO_2 surface
216 measurements at Manvel Croix from a cavity ring down instrument. Photolytic or cavity ring
217 down surface NO_2 data were used in place of NO_y measurements, where available.

218 **2.2.4 Colorado Campaign**

219 The Colorado Department of Health and Environment (CDPHE) provided O_3 surface
220 mixing ratio data from chemiluminescence monitors at the Golden, Chatfield Park, Fort Collins,
221 and La Casa spiral sites and molybdenum converter chemiluminescence NO_2 from the La Casa
222 site. NATIVE provided surface O_3 and NO_y data at the Platteville site, while the NASA Langley
223 Research Center Langley Aeorsol Research Group Experiment (LARGE) provided O_3 and NO_2
224 data at the Boulder Atmospheric Observatory (BAO) site from their LARGE mobile van suite of
225 instruments. EPA provided photolytic NO_2 measurements at the Golden and Fort Collins sites.
226 Neither NO_2 nor NO_y data were available for the Chatfield Park site.

227 **3. Model Simulations**

228 **3.1 HYSPLIT Simulations**

229 Back trajectories were computed for each P-3B spiral for each of the four campaigns with
230 the National Oceanic and Atmospheric Administration (NOAA) Air Resources Laboratory
231 (ARL) Hybrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT) model version 4
232 (Draxler and Rolph, 2003), to analyze air mass source regions and transport. Meteorological
233 inputs were taken from the North American Mesoscale (NAM) model 40 km Eta Data

Assimilation System (EDAS) 3-hour data archive. Back trajectories were computed for each spiral's center latitude and longitude back to three days prior to each spiral, and were computed for each 500 m increment in altitude between 500 m and 3500 m AGL. To compensate for uncertainties in the back trajectories, in this analysis, back trajectories associated with a spiral site were clustered using the clustering algorithm within the HYSPLIT model. The HYSPLIT clustering algorithm is based on the k-means clustering approach (a different clustering technique from agglomerative hierarchical clustering). This cluster analysis was performed over the entirety of the campaign to which that spiral site belonged, to determine the source regions and transport patterns that most contributed to air mass transport at the site during the campaign period.

3.2 WRF/CMAQ Simulations

The CMAQ model version 5.0 was used to simulate air quality for the Maryland campaign, as described by Loughner et al. (2014). Model output from the 12 km horizontal resolution domain was utilized in the following analyses.

CMAQ version 5.0.2 was used to simulate air quality for the Texas campaign (Christopher P. Loughner, personal communication), again driven off-line by WRF meteorology. In this case, WRF used the North American Model (NAM) 12-km analyses for initial and boundary conditions. The same WRF and CMAQ options as used for the Maryland campaign were used in this simulation, with the exception of 45 vertical layers instead of 34 layers, with 18 layers within the lowermost 2 km. However, 2012 Texas Commission on Environmental Quality (TCEQ) anthropogenic emissions were used instead of projected emissions based on the NEI.; as with the Maryland simulation, the Biogenic Emissions Inventory System (BEIS; Pouliot and Pierce, 2009) within CMAQ, and the Motor Vehicle Emissions Simulator (MOVES, specifically

MOVES2010; Kota et al., 2012) were used to compute biogenic emissions and anthropogenic mobile emissions, respectively. The Texas WRF simulations were run iteratively. This technique required WRF to be run twice: the first WRF run performed analysis nudging on all domains based on the 12 km North American Model (NAM) output, and the second WRF run performed analysis nudging on all domains based on the NAM with the exceptions of 2 m temperature and humidity for the 4 km horizontal resolution domain; the 2 m temperature and humidity from the first WRF simulation at 12 km was used to nudge these two parameters for the second WRF simulation at the 4 km resolution. The second iterative WRF run was used to drive CMAQ. The 4 km output was used for the following analyses to capture bay and sea breeze events. CMAQ output for the 4 km domain was provided in 20-minute averages for the Texas campaign, while it was given in hourly averages for the Maryland campaign.

NOAA ARL provided forecasts of O_3 and NO_2 from an experimental version of CMAQ Version 4.6 during each deployment; these simulations were examined for campaigns for which Loughner et al. simulations were unavailable (California and Colorado). The CB05 chemical mechanism was also used in the NOAA simulation. However, the NOAA model runs were driven offline by WRF (Nonhydrostatic Mesoscale Model, or NMM, core) meteorology, and used the fourth generation aerosol module (aero4), the Mellor-Yamada-Janjic (MYJ; Janjić, 1994) scheme for boundary layer mixing, the Noah land surface model, and the 2005 NEI for anthropogenic emissions; lightning NO_x emissions were not included. The available horizontal resolution was 12 km for the California campaign and 4 km for the Colorado campaign, and the available vertical resolution was 22 layers for California and 27 layers for Colorado, with 13 layers and 17 layers, respectively, within the lowest 2 km.

3.3 GMI Simulations

Profiles were obtained from the NASA GMI coupled troposphere-stratosphere chemical transport model. GMI has a horizontal resolution of 2˚ latitude by 2.5˚ longitude, with 72 vertical levels (Duncan et al., 2007), and were matched in time as closely as possible with the obsevations. Specifically, the GMI HindcastFF simulations (Strode et al., 2015) were used in the following analyses. These simulations are driven by meteorology from the Modern-Era Retrospective Analysis for Research and Applications (MERRA; Rienecker et al., 2011). Emissions inputs include year-specific fossil fuel emissions based on the Emission Database for Global Atmospheric Research (EDGAR) 2000 emissions inventory and which were scaled to particular years using annual emission factors as in van Donkelaar et al. (2008), regional anthropogenic emissions inventories for other years, anthropogenic emissions from the NEI2005, year-specific Global Fire Emissions Database v3 (GFEDv3; van der Werf et al., 2010) biomass burning emissions, Asian fossil fuel emissions from the 2006 Intercontinental Chemical Transport Experiment (INTEX-B; Zhang et al., 2009) experiment scaled to other years, and biofuels from the EPA/NEI99 over the U.S. Lightning NO_x emissions are also included (Allen et al., 2010).

4. Analysis Methods

4.1 Agglomerative Hierarchical Cluster Analysis

An agglomerative hierarchical cluster analysis was applied to the P-3B O_3 and NO_2 profiles for each campaign, following the approach of Hains et al., 2008. Any type of statistical cluster analysis seeks to group together objects with the smallest differences, such that objects (in this case, the difference D_{ij} between profiles, as defined below) within one cluster are more similar to each other than to objects within different clusters. Agglomerative hierarchical cluster analysis initially treats each object as its own cluster, and would continue to cluster until all

objects are grouped into a single cluster. A combination of a statistical criterion and manual inspection is required to determine a meaningful number of clusters.

As in Hains et al. (2008), the O_3 or NO_2 mixing ratio data for an individual profile were first averaged into altitude layers of 100 m, which were then grouped into altitude bins covering larger depths. Only the profiles that covered the full altitude range for each campaign were included in this analysis; it should be noted that all altitude data are above mean sea level (AMSL). Hains et al. (2008) employed the following equation to quantify the difference D_{ij} over the entire profile depth between profiles *i* and *j* within each profile pair:

$$D_{ij}(a) = \left[\sum_{h=1}^{h=n} abs(C_i(h) - C_j(h))\right]^2 \quad (3)$$

$$D_{ij}(b) = 1 + [1 - r] \quad (4)$$

$$D_{ij}(c) = 1 - \exp[-(s-1)^2] \quad (5)$$

$$D_{ij} = \sum_{k=1}^{k=b}[\, D_{ij}(a) * (D_{ij}(b) + D_{ij}(c))] \quad (6)$$

where *k* is the index for the b altitude bins, h is the index for the *n* altitude layers within a bin, *C* is the mixing ratio for the *ith* and *jth* profiles, *r* is the correlation coefficient for each pair of profiles within each of the *k* bins, and *s* is the slope between each profile pair within a bin. The correlation coefficient and slope were obtained from a regression analysis between each profile pair within each bin. These differences D_{ij} represent the total difference between magnitude of the mixing ratios throughout the altitude range, as well as how much the slope and correlation coefficients deviate from unity, thus accounting for how different the profile shapes are at different altitude levels (Hains et al., 2008). The objects of the cluster analysis presented here are these D_{ij} values, and were clustered with a hierarchical clustering algorithm in the Interactive Data Language (IDL). An average linkage method was used to determine the similarity between clusters, in which the distance between two clusters was defined as the average difference between the data points in the first cluster and the data points in the second cluster.

The optimal number of meaningful clusters was determined with a combination of manual inspection and a technique based on total root mean square deviation (TRMSD), as described by Taubman et al. (2006). The dendrogram (tree diagram displaying the arrangement of clusters) produced by the clustering algorithm was initially inspected to determine a reasonable maximum number of clusters to consider, before application of the TRMSD technique. In this technique, an average profile was calculated for each cluster, and then the root mean square deviation (RMSD) of each profile within the cluster from the average cluster profile was computed. These RMSD values were then summed over all clusters under consideration to give the TRMSD. This TRMSD value was computed first over all clusters initially under consideration based on inspection of the dendrogram, and then computed again each time the number of clusters was reduced. The percentage change in TRMSD as clusters were condensed was then computed, and the optimal number of clusters taken as the number of clusters immediately before a large increase (~10%) in TRMSD, with the percentage change remaining relatively high upon further agglomeration (Taubman et al., 2006). The member profiles of each cluster obtained with the TRMSD technique were also manually inspected to assess the meaningfulness of the algorithm results.

The above techniques determine the optimum number of meaningful clusters, but do not necessarily ensure that each median cluster profile obtained after application of these techniques is significantly different from the other median profiles. In other words, each median cluster profile shape and relative magnitude at each vertical level may differ from those of the other median clusters, but the distribution of magnitudes at each vertical level may overlap, for which the clustering algorithm and TRMSD computation do not test. Thus, to determine that a cluster is significantly different from the other clusters for a campaign, the cluster median profile must have

exhibited at least five consecutive altitude layers in which the 25th and 75th percentile values (error bars) did not overlap with those of any other median profile. Additionally, this technique required that each profile contain data without significant gaps in coverage over the entire altitude range of interest for each campaign, and profiles that did not meet this criterion were not included in the analysis. For example, all profiles over the Beltsville site during the Maryland campaign were disregarded, as these profiles extended to approximately 1.5 km AGL, while profiles over the other sites extended to at least 3.2 km AGL.

4.2 Column Abundance Computation

As described by Flynn et al. (2014), two P-3B column amounts were computed for O_3 and NO_2: column_air and column_ground. Column_air was computed through integration of the profile after the mixing ratio measurement at the lowest aircraft altitude (approximately 0.3 km AMSL) was held constant to the surface. Column_ground was computed in the same manner, but instead held the surface mixing ratio measurement constant to the level of the lowest aircraft altitude, when a surface measurement was available. The top of the P-3B partial columns was typically between 3-4 km AMSL.

4.3 Correlation Analysis for the P-3B Clusters

A simple linear least squares regression analysis was performed for O_3 and NO_2 between the P-3B column_air or column_ground column amounts and surface mixing ratio data for each P-3B cluster obtained from the above profile cluster analysis. Column and surface data coincident to a profile included within a P-3B profile cluster were sampled for each regression analysis. Surface data were averaged over the time of the aircraft profile spiral for use with this analysis. This was done to assess the degree of correlation between the column and surface data for each

cluster, and assess how the different conditions associated with each cluster affected the correlations.

4.4 Shape Factor Computation

Model simulated profiles were evaluated relative to the P-3B *in situ* profiles in terms of the O_3 or NO_2 shape factor, or relative vertical distribution, for each P-3B cluster. Profiles most coincident in time and space to the P-3B profiles included within a cluster were sampled, and a median simulated shape factor profile was computed for each cluster from both the model and the observations. The shape factor was defined as the ratio of the partial column within a model vertical layer to the full tropospheric column:

$$S(z) = \Omega_z/\Omega_{trop} \quad (5)$$

where Ω_z denotes partial column amount at vertical layer z and Ω_{trop} denotes the partial tropospheric column over the full depth of the P-3B spirals. For both observations and model simulations, the partial columns were computed over the depth of each CMAQ or GMI vertical layer; thus, the vertical distribution was computed over the model vertical grid, analogously to the shape factors computed from global models used in conjunction with satellite retrieval algorithms (Lamsal et al., 2014; Palmer, et al., 2001)**.** It should be noted that the CMAQ or GMI tropospheric columns used in this analysis were also computed over the P-3B spiral depths only, and so were lower tropospheric column abundances rather than full tropospheric columns, to allow for a direct comparison to the P-3B shape factors.

5. Results and Discussion

5.1 Ozone

5.1.1 P-3B Ozone Profile Clusters

Figure 2a)-d) shows the median profiles of O_3 for each of the clusters obtained for each of the four DISCOVER-AQ deployments. Clusters containing fewer than four members had too few members for a meaningful analysis and are excluded here. To determine that a cluster is significantly different from the other clusters for a campaign, the cluster median profile must have exhibited at least five consecutive altitude layers in which the 25th and 75th percentile error bars did not overlap with those of any other median profile. With few exceptions, the median profiles were significantly different only within the planetary boundary layer (PBL), and lost distinction from each other within the free troposphere. Based on this significance criterion, the Texas campaign (Fig. 2c) demonstrated the greatest number of distinct clusters (five), as well as the greatest range of mixing ratio values, of all four campaigns. Few clusters were obtained for the California and Colorado campaigns (Fig. 2b and 2d, respectively), though these clusters were statistically distinct. The Maryland campaign (Fig. 2a), however, demonstrated only one significantly independent cluster (Cluster 4), unlike the other campaigns. This is contrary to initial expectations, and is surprising, given that the largest range in O_3 mixing ratio magnitudes within the PBL was encountered during the Maryland campaign. However, the clustering algorithm considered both mixing ratio magnitude and profile shape, producing the results described in this section for Maryland.

These differences in profile variability, as indicated by the number of significant clusters, may be due to synoptic conditions favoring or inhibiting O_3 formation. For example, the California campaign took place during winter, and thus experienced less sunlight and the coldest temperatures relative to the other campaigns, which inhibited O_3 formation. Though the Colorado campaign took place during summer, cooler temperatures were also experienced here relative to the other warm season campaigns due to its elevation in addition to more convection

417 than initially expected, which also led to some inhibition of O_3 formation. Just as importantly,
418 the effects of complex terrain may have led to greater horizontal mixing within the study regions
419 during the California and Colorado campaigns, further limiting the variability of the observed O_3
420 profiles. Plots of the wind fields from the NARR display complex interactions of the synoptic
421 scale flow with local topography within the San Joaquin Valley during the California campaign
422 and within the foothills and plains at the base of the Front Range during the Colorado campaign
423 at several pressure levels. For example, along-valley flow or valley circulations were often seen
424 due to upslope or downslope flow on opposing sides of the San Joaquin Valley, leading to
425 horizontal mixing within the valley (Fig. 3). Interactions of downslope flow from the Front
426 Range often interacted with northerly or southerly winds over the plains, leading to horizontal
427 circulation patterns; these wind flow patterns also changed throughout the course of a day (Fig.
428 4). On some days, a well-defined upslope flow formed on the east side of the Front Range during
429 the daytime, which also influenced pollution profiles. In contrast, the Maryland campaign study
430 region was embedded within the synoptic scale flow, which tended to be westerly or
431 northwesterly, at most levels, and, due to this, polluted air masses from the Ohio River Valley or
432 Great Lakes Region were often transported into the study region. Ozone was thus likely the most
433 well mixed horizontally and vertically during the Maryland campaign, dampening its profile
434 variability (Fig. 5). Wind patterns had less of an impact during the Texas campaign than for the
435 other three, suggesting that local chemistry and emissions were much more important to O_3
436 profile variability than local meteorology (except for the last two flights). This is perhaps due to
437 the more complex chemical environment of the Houston metropolitan area, with large, nearby
438 sources of volatile organic compounds (VOCs).

Further, differences in profile shapes among clusters for a campaign may be partially explained by differences in the atmospheric stability encountered during the P-3B spiral sampling times. Median profiles of potential temperature (θ), an indicator of stability and degree of mixing, for each O_3 profile cluster are displayed in Figure 6a)-d). These potential temperature profiles were not themselves clustered, but θ profiles coincident to each O_3 profile included within a cluster were sampled, and the median θ profile computed for that O_3 cluster. Comparison of the median θ profile to the median O_3 profile for each cluster suggests that θ has an influence on these profile shapes for all campaigns; where the θ profile is well mixed within the PBL, the O_3 profile is also relatively well mixed, and vice versa. For example, during the Colorado campaign, Cluster 2 demonstrated a more well mixed θ median profile than Cluster 1 (Fig. 6d), corresponding to a well mixed O_3 median profile (Fig. 2d); during the California campaign, Cluster 1 displayed the most well mixed θ profile (Fig. 6b), also corresponding to a well mixed O_3 profile (Fig. 2b). However, the influence of potential temperature was somewhat weaker during the Maryland and Texas campaigns; Clusters 3 and 4 exhibited poorly mixed θ profiles (very stable) but relatively well mixed O_3 profiles during the Texas campaign (Fig. 6c), for example (Fig. 2c).

Another indicator of stability, the temperature lapse rate (Γ), defined as

$$\Gamma = {}^{-dT}/_{dz} \qquad \text{Eqn. (3.7)}$$

also emerged as an influence on profile shape for all four campaigns. The lapse rate determines the static stability of the atmosphere, and is a local property (i.e., stability is not constant for all portions of the atmosphere). Thus, the lapse rate of an air parcel determines its buoyancy, and thus whether vertical displacement of that parcel is supported or inhibited by the surrounding environment. Three stability regimes are possible: if the parcel's lapse rate is less than the moist adiabatic lapse rate (**4 K/km**), then the local atmosphere is absolutely stably stratified and vertical

462 motion is suppressed; if the lapse rate falls between the moist and dry adiabatic (9.8 K/km) lapse
463 rates, the local atmosphere is conditionally unstable, and vertical motion depends on the parcel's
464 degree of saturation; and finally, if the lapse rate is equal to or greater than the dry adiabatic lapse
465 rate, then the local atmosphere is absolutely unstable. Instability encourages vertical motions,
466 which mixes scalar quantities such as potential temperature, water vapor, and pollutants and
467 causing more uniform vertical profiles of these quantities. The lapse rate for each 100 m altitude
468 layer within each O_3 profile included within a cluster was first computed, and then these lapse
469 rates were separated into boundary layer or free tropospheric lapse rates based on the PBL height
470 associated with that O_3 profile. The median values and distributions of lapse rates within the PBL
471 and free troposphere were then compared for each O_3 cluster to determine its influence on profile
472 shape; results are displayed in Fig. 7a)-d).

473 As expected, clusters that exhibited a larger median boundary layer lapse rate value also
474 exhibited a more well mixed median O_3 profile, while smaller median PBL lapse rate values were
475 associated with less well mixed O_3 profiles; this is consistent with the influence of the potential
476 temperature profiles. For each campaign, the PBL or free tropospheric lapse rates were often
477 statistically the same among clusters (i.e.; overlap of the boxes in Fig. 7 representing the 25^{th} and
478 75^{th} percentile values), which likely dampens differences in profile shapes among clusters within
479 a campaign. Additionally, for some Texas and Colorado clusters (Fig. 7b and 7d), the lapse rates
480 were often statistically the same between the PBL and the free troposphere, resulting in smoother
481 median O_3 profiles throughout the entire profile depth than seen for the Maryland or California
482 campaigns (Fig. 7a and 7c). More interesting median O_3 profile behavior resulted when
483 differences existed between the PBL and free tropospheric lapse rate values that were larger than
484 for the other clusters within a campaign. For example, Clusters 1 and 2 of the Texas campaign

485 exhibited such large differences between the PBL and the free troposphere, resulting in O_3
486 profiles that were well mixed within the PBL and which displayed more layered behavior within
487 the free troposphere; the same profile behavior can be seen for Cluster 2 of the Maryland
488 campaign.

489 Lastly, flight date during the campaign and spiral sampling time influence the relative
490 magnitudes of the median cluster profiles for the California (Tables 1 and 2), Texas (Tables 3
491 and 4), and Colorado (Tables 5 and 6) campaigns. The most polluted median profiles were
492 associated with either O_3 pollution episodes or campaign periods and sampling times conducive
493 to O_3 photochemistry. For example, 50.8% of the profiles in Colorado Cluster 1 occurred during
494 the first campaign period (July 17-21, 2014) and were sampled mainly during Spirals 2 and 3
495 (~11am and ~4pm MDT), while Cluster 2 was sampled mostly during the last campaign period
496 (Aug. 6-10, 2014) and during Spirals 1 and 2 (~9am and ~11am MDT); this is consistent with
497 Cluster 1 demonstrating greater O_3 mixing ratios within the PBL than Cluster 2. It should be
498 noted that July often experienced larger surface O_3 concentrations than did August in Colorado
499 (Mazzuca et al., submitted). Likewise, Texas Clusters 1 and 2, the most polluted median
500 profiles, contained profiles entirely from the last campaign period (Sept. 24-26, 2013), when a
501 pollution episode occurred, and were primarily sampled during Spiral 3 (~2pm CDT). Sampling
502 time has a greater influence over the relative cluster magnitudes than campaign period for the
503 California campaign, as this campaign took place during the winter. California Cluster 2 sampled
504 primarily during Spirals 1 and 2 (~8am and ~11am PST), while the most polluted Cluster 3
505 sampled roughly evenly among all three spiral times (~8am, ~11am, and ~1pm PST). The
506 Maryland campaign also demonstrated a weak influence of spiral location for Cluster 4: only

Edgewood and Essex (the two sites nearest the Chesapeake Bay) afternoon profiles were included in this cluster, such that some influence of the bay breeze may be present.

Airmass origin, as revealed by HYSPLIT back trajectories, also influenced the relative magnitudes for the Maryland (Table 7) and Colorado campaigns (Tables 8-12). Maryland Cluster 4, the most polluted within the PBL and the only significant Maryland cluster, contained profiles whose corresponding back trajectories originated over the polluted Great Lakes region at all altitude levels between 500 m and 2500 m; the less polluted clusters included back trajectory origins over less polluted regions, such as northern Canada, at all levels. The Colorado profile back trajectories associated with Cluster 1 (most polluted) demonstrated larger percentages of air recirculated over the Denver-Boulder region, thus recirculating urban pollution, than did Cluster 2. Though airmass origin did not emerge as a significant influence on the California cluster magnitudes, it is interesting to note that the back trajectories (not shown) emphasized recirculation within the San Joaquin Valley, which is consistent with the NARR plots demonstrating significant valley recirculation and horizontal mixing during this campaign.

5.1.2 P-3B Ozone Profile Cluster Correlations

Values of R^2 from the simple linear regression analyses between surface mixing ratio and column abundance data for each profile cluster (cluster correlations) are summarized in Tables 13-14 for the Maryland and California campaigns. The Texas and Colorado campaigns did not produce significant cluster correlations and so are not included in the tables. The column-surface correlations over the full set of profiles used with the hierarchical cluster analysis (full correlations) are also included in Tables 13 and 14. Representative scatter plots are displayed in Fig. 8a)-d). It should be noted that clusters containing fewer than five members contained too few members for a meaningful analysis and are excluded here. The California, Texas, and

530 Colorado campaigns yielded very few statistically significant correlations, while the Maryland
531 campaign yielded many significant and large cluster correlations. It is also notable that, for each
532 cluster during each campaign, the column_air and column_ground correlation analyses typically
533 yielded R^2 values that were statistically the same, despite the different gap-filling methods (and
534 inherent assumptions about the strength of PBL mixing in these methods) used in the column
535 computations. This similarity in R^2 values indicates that O_3 remained generally well mixed
536 horizontally and vertically in the lower PBL during each campaign, consistent with the findings
537 of Flynn et al. (2014) for the Maryland campaign.

538 Differences in degree of correlation between the Maryland campaign and the other
539 campaigns may be due again to the season in which the campaign took place, as well as the
540 vertical wind shear of the large-scale flow over each study region. Wintertime stagnation during
541 the California campaign interfered with the column-surface connection, as O_3 could not be mixed
542 vertically as efficiently as in convective boundary layers, resulting in poor connection between
543 column and surface, while the Colorado campaign did not experience convective boundary
544 layers as deep as for Maryland or Texas, also impacting the degree of connection between
545 column and surface data. The California and Colorado campaigns also experienced changes in
546 wind direction and circulation patterns with height, as indicated by the NARR wind fields at
547 several pressure levels, discussed previously. Air was transported from different source locations
548 at different heights, rather than simply being mixed in the vertical, further interfering with the
549 column-surface connection. Again in contrast, as Maryland experienced similar wind patterns at
550 most levels and deep, convective boundary layers, O_3 was most well mixed horizontally and
551 vertically of all four campaigns, allowing for greater connection between column and surface.
552 Texas was in a complex chemical environment, such that O_3 production was much more

553 localized, as evidenced by its low correlation with CO over all profiles included in the clustering
554 analysis (Fig. 9).

555 These cluster correlations were also compared to the full correlation for the Maryland
556 (Table 13) and California (Table 14) campaigns. As the Texas and Colorado full correlations were
557 not themselves significant, the cluster correlations were statistically the same as the full
558 correlations. Neither California Cluster 2 nor Cluster 3 was found to be statistically different from
559 the full correlation at a confidence level of 95% for both column analyses (Table 14). This
560 indicates that no correlation analysis for this campaign was greater than zero. However, half of
561 the Maryland clusters (Cluster 3, 5, and 6; Table 13) were significantly smaller than the full
562 correlation for both analyses; the Cluster 1, 2, and 4 correlations were not significantly different
563 from the full correlation analyses (which also presented a high degree of correlation). It is notable
564 that the median profiles associated with Clusters 1 and 2 were well mixed, whereas those for
565 Clusters 3, 5, and 6 were less well mixed and presented more vertical variation within the PBL.
566 Thus, the column-surface correlation for these clusters degraded relative to those for the well
567 mixed Clusters 1 and 2 and the full correlations under conditions of inhibited vertical mixing.
568 This is also consistent with the results for the California campaign: the Cluster 2 median profile
569 was less well mixed than that for Cluster 3, though neither profile could be considered well
570 mixed. The Maryland Cluster 4 correlation presents an exception: despite a well mixed median
571 cluster profile, this cluster contained only five profiles, fewer than any other Maryland cluster,
572 that may have been influenced by the bay breeze, causing an interruption in the column-surface
573 connection.

574 The results presented above provide a mixed message for the ability to relate satellite O_3
575 column observations to surface mixing ratios. The complex meteorological or chemical

576 conditions encountered during the Colorado or Texas campaigns indicate that satellite
577 observations of lower tropospheric column O_3 may not be useful for estimating surface ozone
578 under such conditions. Because the correlations for each cluster and the full correlation analyses
579 were essentially null during the California campaign, this suggests that it matters little if the
580 satellite retrieval correctly approximates the O_3 lower tropospheric profile shape, as the stratified
581 conditions still prevent there being a connection between column and surface data. The Maryland
582 campaign also indicated degraded column-surface correlation under conditions of poorer vertical
583 mixing, as indicated by the median O_3 and potential temperature profiles for Clusters 3, 5, and 6.
584 However, the correlation analyses suggest that satellites may have the best chance to relate to
585 surface ozone under the conditions encountered during the Maryland campaign Clusters 1 and 2,
586 which include deep, convective boundary layers and few interruptions to this connection from
587 complex meteorology, chemical environments, or orography.

588 **5.1.3 Ozone Shape Factor Comparison**

589 As with the potential temperature median profiles for each cluster, the CMAQ- and GMI-
590 simulated O_3 mixing ratio profiles coincident to each observed profile within a cluster were
591 sampled, and the median shape factor then computed for each model and each cluster. CMAQ
592 well captured the shapes of the median P-3B shape factor profiles for most clusters in each
593 campaign. Shape factors were compared in this analysis because these are used in remote sensing
594 retrievals, rather than mixing ratio profiles; such a comparison elucidates potential implications of
595 model errors for retrievals. Representative comparison plots are displayed in Fig. 10a)-d).
596 Interestingly, CMAQ reproduced the complex vertical behavior of the Colorado median shape
597 factor profiles, though CMAQ also performed well relative to the P-3B during the other three
598 campaigns. However, the model placed the upper PBL peak shape factor value approximately 0.5

599 km above the observed peak within the PBL for many clusters, indicating that, while reproducing
600 the shapes of the median shape factor profiles, CMAQ placed a greater portion of the O_3 column
601 higher in the vertical than seen in the observations during each campaign. This is particularly
602 evident for the Texas and Colorado comparison plots in Fig. 10c) and d), and suggests that
603 vertical mixing may be overestimated within the model, as O_3 is a relatively long-lived gas within
604 the PBL. Additionally, CMAQ generally displayed a slight low bias in shape factor values within
605 the PBL and a slight high bias within the free tropospheric portions of the shape factors relative to
606 the P-3B for the Maryland, California and Texas campaigns, and bias direction varying with
607 altitude during the Colorado campaign. This further indicates that CMAQ placed a greater portion
608 of the O_3 column higher in the vertical than seen in the observations.

609 CMAQ well captured the magnitudes of P-3B O_3 shape factors for all clusters during the
610 California campaign and for Clusters 1-3 and 5 of the Maryland campaign, with R^2 values of 0.80
611 and larger (Fig. 11a)-b), when correlating CMAQ and observed shape factor profile values for
612 individual model layers across all shape factors within a cluster. However, CMAQ moderately
613 well captured the shape factor values for all clusters during the Texas and Colorado (Fig. 11c)-d)
614 campaigns, and Clusters 4 and 6 of the Maryland campaign, with R^2 values between 0.50 and
615 0.80. It is notable that Maryland Cluster 4 contained profiles only from Edgewood and Essex, and
616 that Cluster 6 contained a large percentage of profiles from Essex. These were the two most
617 polluted median cluster profiles, with PBL mixing ratio values of approximately 80 ppbv or
618 higher; in agreement with this, Texas Clusters 1 and 2, with median profile PBL mixing ratio
619 values in excess of 80 ppbv, also displayed the lowest R^2 values for that campaign. Thus, CMAQ
620 was less able to capture the magnitudes of the shape factors for very polluted profile clusters, and
621 for the two campaigns with complex pollution or meteorological conditions (Texas and Colorado,

622 respectively). This further suggests that CMAQ best captured the O_3 column relative vertical
623 distribution under conditions that were unfavorable to O_3 photochemical production (California
624 campaign), or under conditions of moderate pollution (Maryland Clusters 1-3 and 5). However,
625 the scatter plots presented in Fig. 11 also indicate that CMAQ often under- or overpredicted the
626 magnitudes of the shape factors relative to the P-3B. CMAQ tended to overpredict at many
627 vertical levels during the California campaign (Fig. 11b), with the exceptions of the 600-1200 m
628 AMSL altitude ranges, which corresponded to the upper PBL and lower free troposphere. Remote
629 sensing retrievals are more sensitive to the upper PBL/lower free troposphere altitude region than
630 near the surface (as determined by the scattering weights), such that errors in the shape factor in
631 this region relative to other altitudes lead to greater retrieval errors. Like the median cluster shape
632 factor profiles, this suggests that CMAQ placed too much of the O_3 column burden within fewer
633 vertical layers (the lower PBL and near the tops of the California P-3B spirals). In contrast,
634 CMAQ tended to underpredict in the lower PBL (300-800 m AMSL) and less often overpredicted
635 above 2000 m AMSL (free troposphere) during the Maryland campaign, and compared well to the
636 P-3B shape factor values in the middle portions of the shape factor profiles (800-2000 m AMSL;
637 Fig. 11a). Such underprediction within the lowest PBL portions of the Maryland shape factors,
638 with some compensating overprediction at higher altitudes, again suggests that CMAQ placed too
639 much of the O_3 column burden within fewer vertical layers than seen in the observations. The
640 scatter plots for the Colorado campaign (Fig. 11d) are consistent with the model biases suggested
641 by the median shape profiles: CMAQ demonstrated underprediction within the lower PBL (below
642 2500 m AMSL, or below 1000 AGL) and within the highest spiral altitudes (above 4000 m
643 AMSL or 3500 m AGL), and overprediction for altitudes between 2500-4000 m AMSL. These
644 results and those comparing the median shape factors profiles are encouraging that a regional air

645 quality model such as CMAQ may be able to replicate the shape factors during winter, when
646 photochemical O_3 production is inhibited (California), and the moderate pollution conditions
647 associated with the Maryland campaign Clusters 1-3 and 5.

648 Additionally, the CMAQ and P-3B CO shape factor magnitudes associated with each O_3
649 profile cluster were compared as a preliminary investigation of the errors in simulated vertical
650 mixing and the impact of those errors on the simulated O_3 shape factors (Fig. 11e)-g). The CO
651 comparisons for the Maryland campaign demonstrated the clearest case of overestimated vertical
652 mixing of all four deployments (Fig. 11e). As with the O_3 shape factors, the CO shapes factors
653 were underestimated within the lowermost portion of the profiles (below 1200 m AMSL), with
654 the bias direction transitioning through the middle altitude range of the profiles, leading to general
655 overestimation of the CO shape factor values in the upper PBL and lower free troposphere (1600
656 m AMSL and above). As CO is a long-lived chemical species (much longer than the time scale of
657 turbulent mixing in the PBL; Zhang et al., 2016), with emissions dominated by surface sources,
658 underestimation in the lower simulated shape factor profile and overestimation aloft suggests that
659 the model mixed too much CO into the upper portions of the profiles, and thus likely over-mixed
660 other pollutants such as O_3. The California clusters also suggested overly vigorous vertical mixing
661 relative to the observed CO shape factors. The simulated CO shape factor values tended to be
662 overestimated throughout most of the PBL (50-900 m AMSL), transitioning to a lack of preferred
663 bias direction above 900 m. This is consistent with the overestimation of the O_3 shape factors
664 below 600 m. However, as the depth to which the CO shape factors were overestimated extends
665 further than the depth to which the O_3 values were overestimated, this suggests that, while errors
666 in vertical mixing likely caused errors in the O_3 column vertical distribution, vertical mixing may
667 have played less of a role than for the Maryland campaign. In contrast, the CO comparisons for

668 the Colorado campaign suggest that vertical mixing may have been underestimated relative to the
669 observations, and that CO emissions may have been overestimated; simulated CO shape factor
670 values were overestimated below 2500 m AMSL (1000 m AGL), and tended to be underestimated
671 throughout the remainder of the profile depth. This is does not correspond to the underestimated
672 O_3 shape factor values below 2500 m and above 4500 m AMSL, with overpredicted O_3 between
673 these altitude ranges. Vertical mixing may have exerted the least influence on column values
674 during the Colorado campaign, relative to the other three campaigns, due to complex circulations
675 such as the Denver cyclone or the mountain solenoidal circulations that models struggle to
676 replicate.

677 The GMI median shape factors were also computed for each cluster of the Maryland and
678 California campaigns (Fig. 12a)-b). GMI well captured the altitude variation of the O_3 shape
679 factor profiles for both campaigns, indicating that this global model captured the relative O_3
680 vertical distributions. However, unlike CMAQ, GMI tended to display a low bias in shape factor
681 values within the lowermost PBL and a high bias within the mid-PBL, with the preferred bias
682 direction changing for each cluster within the free tropospheric portion of the shape factor
683 profiles, indicating that GMI did not mix O_3 as high in the vertical as in the observations and
684 placed too much of the O_3 column burden within the PBL and lower portions of the profiles.
685 Comparisons of observed and GMI-simulated CO shape factors (Fig. 13) typically display
686 overestimates in CO shape factor magnitudes between 800 m and 1600 m for the Maryland
687 campaign, and 300-900 m for the California campaign (the middle portion of the PBL for each
688 campaign), with underestimates in the lowermost PBL. This suggests that mixing into the middle
689 portion of the PBL from the lowermost profile is overestimated in GMI for both campaigns,
690 though not as severely as for CMAQ. GMI performed similarly well as CMAQ for each campaign

691 relative to the P-3B O_3 cluster median shape factors, while CMAQ was also able to capture the
692 vertical structure of the Colorado campaign shape factors to some extent. This implies that a
693 regional model may be able to estimate O_3 shape factors accurately enough for use in remote
694 sensing retrieval algorithms.

695 GMI moderately well captured the P-3B O_3 shape factor values for all clusters during the
696 California campaign, and for Clusters 1-3 during the Maryland campaign (example O_3 and CO
697 scatter plots shown in Fig. 13a)-b). Values of R^2 for the correlation between P-3B and GMI
698 shape factor values ranged between 0.60 and 0.90. However, GMI compared more poorly for
699 Clusters 4-6 of the Maryland campaign, with R^2 values less than 0.60 (Fig. 13a). The contrast in
700 performance for these two groups of clusters is again explained by the difference in degree of
701 pollution for the cluster median *in situ* profiles: Maryland Clusters 4-6 were more polluted than
702 Maryland Clusters 1-3 and all California clusters, with median PBL mixing ratios greater than 65
703 ppbv. This suggests that GMI performed better for clean to moderately polluted conditions. The
704 scatter plots for GMI vs. P-3B shape factor values demonstrate that GMI often underpredicted
705 the shape factor values within the lowermost portions of the profiles, while above this altitude
706 range, GMI well predicted or slightly overpredicted the shape factor values relative to the P-3B;
707 GMI again demonstrated a tendency to underpredict in the lower free troposphere during the
708 Maryland campaign and overpredict during the California campaign It is these underpredictions
709 of small shape factor values that likely drive down the correlations between simulated and
710 observed shape factors. These altitude biases coupled with the smooth median shape factor
711 profiles indicate that the GMI relative O_3 vertical distributions were too homogeneous relative to
712 the observations, as GMI placed more of the O_3 column burden approximately evenly within the
713 PBL and the lower free troposphere than seen in the observations, causing GMI to underpredict

714 the O_3 distribution within most of the free tropospheric portions of the profiles. This demonstates
715 a disadvantage of a global model such as GMI relative to the regional CMAQ model to
716 accurately simulate the lower tropospheric O_3 shape factors.

717 **5.2 Nitrogen Dioxide**

718 **5.2.1 P-3B NO_2 Profile Clusters**

719 The median profiles for the P-3B NO_2 clusters are displayed in Fig. 14a)-d). The
720 Maryland campaign produced the greatest number of NO_2 clusters, though none of these were
721 significant (i.e.; all median profiles were statistically the same; Fig 14a). The Texas campaign
722 also produced a large number of clusters yet only two significant clusters, each containing only
723 one profile; all other clusters were statistically the same (Fig. 14c). These results indicate that the
724 NO_2 profiles displayed relatively uniform behavior during these two campaigns, which may be
725 expected given that these were warm season campaigns and flights were generally conducted on
726 sunny days with convective boundary layers and with conditions conducive for O_3 production.
727 On the other hand, all three clusters obtained for the California and Colorado campaigns were
728 significant (Figs. 14b and 14d). Distinction among clusters was again found only within the PBL
729 rather than the free troposphere, suggesting influence of both chemistry and vertical mixing on
730 NO_2 profile shapes. However, as Cluster 1 contained the vast majority of profiles, the Colorado
731 results again indicate that the NO_2 profiles displayed relatively uniform behavior throughout the
732 campaign period; only the California campaign indicated any NO_2 profile variability. The largest
733 ranges of mixing ratio values were also encountered during the California campaign. The
734 uniqueness of the California NO_2 profile clusters may have been due to the inhibited O_3
735 formation during this campaign; less NO_2 was converted to O_3 during the day, allowing a greater
736 variation in profile shapes than the other three summer campaigns, in which NO_2 was more

737 rapidly converted to O_3 during the day. The significant Texas and Colorado clusters may be
738 considered as outliers, as they each contained few profile members. No clear meteorological
739 conditions emerged as influences on the NO_2 clusters, further indicating that NO_2 profile
740 variability is influenced by photochemical loss of NO_2 as well as vertical mixing, as may be
741 expected given its short lifetime; this result is consistent with Zhang et al. (2016), which found
742 that the observed NO_x vertical gradient during the Maryland campaign is sensitive to both
743 chemistry and boundary layer mixing. However, it should also be noted that the agglomerative
744 hierarchical clustering technique used in this study produced clusters with little segregation
745 among the NO_2 profiles even though these two processes led to changes in profile shape as the
746 the boundary layer evolved through the day Other clustering techniques (which are not evaluated
747 here) may not have produced such few clusters for NO_2.

748 **5.2.2 P-3B NO_2 Profile Cluster Correlations**

749 Values of R^2 from the simple linear regression analyses between surface mixing ratio and
750 column abundance for each profile cluster during the California and Texas campaigns that
751 presented a statistically significant correlation are summarized in Table 15. The remainder of the
752 Texas and California clusters, as well as Maryland and Colorado clusters, did not produce
753 significant correlations, and are not summarized in the tables. Representative scatter plots are
754 shown in Fig. 15a)-d). Again, clusters containing fewer than four members are excluded here.
755 Flynn et al. (2014) demonstrated that, for NO_2, column_air better represented the "true" lower
756 tropospheric column than did column_ground during the Maryland campaign; thus, only the
757 column_air correlations will be analyzed here. Many fewer NO_2 cluster correlations were
758 significant than for the O_3 clusters across all campaigns, which is not unexpected given that NO_2
759 is generally less well mixed than O_3. No cluster correlations were significantly different from the

760 full correlations for the Maryland, Texas, and Colorado campaigns; only the California Cluster 3
761 correlation was statistically significantly smaller than the full correlation (as it was not a
762 significant correlation).

763 It is also notable that California Cluster 2 presented a significant correlation, and was not
764 significantly different from the full correlation for California (which was itself statistically
765 significant). The key differences here between California Cluster 2 and Cluster 3 discussed
766 above may lie in the differences in the shapes of the median cluster profiles. California Cluster 2
767 displayed a more well mixed NO_2 profile, in which NO_2 mixing ratios decreased less sharply in
768 the PBL, than did Cluster 3. In contrast, Cluster 3 displayed a much smoother, exponential decay
769 profile shape, indicative of a less well mixed profile and more stable atmospheric conditions.
770 Cluster 3 also displayed an increase in NO_2 mixing ratios within the lowermost portion of the
771 profile not present in the Cluster 2 profile. This comparison in median profile shapes is
772 consistent with the greater correlation between column and surface data obtained for Cluster 2.
773 However, as Cluster 2 contained the majority of NO_2 profiles during the California campaign,
774 the correlations for this cluster were not statistically significantly different from the full
775 correlation for California. Additionally, the Cluster 4 column_air correlation of the Texas
776 campaign was significant, contained the vast majority of NO_2 profiles, and displayed the most
777 well mixed median profile shape (NO_2 mixing ratios did not display as great a vertical gradient
778 as for the other clusters. These comparisons further suggest that the vertical distribution, and, by
779 proxy, vertical mixing, has an influence on NO_2 profile variability, as well as photochemical loss
780 of NO_x and production of O_3. Furthermore, as a small number of clusters contained the majority
781 of profiles for each campaign, these results again indicate that NO_2 behavior did not change
782 much over the course of each of these campaigns, further supporting the lack of NO_2 profile

783 variability, and that the ability to relate column observations to surface concentrations depends
784 upon NO_2 exhibiting a more uniform profile vertical gradient in mixing ratio.

785 **5.2.3 NO_2 Shape Factor Comparison**

786 Potential errors in satellite NO_2 retrievals are greater when there is bias in the shape
787 factor profile at altitudes where the instrument is more sensitive. Instruments such as OMI are
788 more sensitive to NO_2 in the upper PBL and lower free troposphere than near the surface (as
789 determined by profiles of the scattering weights used in the retrieval), so that errors in the shape
790 factors will increase retrieval error should they occur in the upper PBL/lower free troposphere.
791 The median CMAQ NO_2 shape factor profiles for each cluster within each campaign compared
792 poorly to the median P-3B shape factor profiles, indicating that CMAQ did not capture the NO_2
793 relative vertical distribution on the CMAQ vertical grid (Fig. 16a)-d). While CMAQ often
794 displayed a peak in shape factor values within the PBL when the corresponding P-3B cluster also
795 displayed a peak, the model commonly misplaced the vertical location of that peak, as seen for
796 Colorado Cluster 1 (Fig. 16d) and California Cluster 2 (Fig. 16b). Additionally, these simulated
797 peaks were often too broad relative to the P-3B shape factor peaks, as seen for California and
798 Texas (Figures 16b, c). These peaks often extended through the upper PBL and the lower free
799 troposphere, the altitude regions to which instruments such as OMI are more sensitive for NO_2
800 than near the surface, and thus where retrieval errors would be greater. Indeed, for the Maryland,
801 California, and Texas campaigns, CMAQ often greatly overestimated the NO_2 shape factors
802 within the lower free troposphere for most clusters. The CMAQ shape factors missed the PBL
803 peaks entirely and were too smooth relative to the P-3B for each cluster of the Maryland
804 campaign (Fig. 16a). These issues further indicate a general over-mixing within CMAQ, as the

805 NO_2 relative vertical distribution was spread too evenly among the vertical levels relative to the
806 observations.

807 CMAQ typically struggled to simulate the shape factor values for most clusters during
808 the Maryland, Texas, and Colorado campaigns, with R^2 values below 0.40 (Fig. 17a)-d). CMAQ
809 failed to capture the California shape factors, demonstrating no correlation between the
810 simulated and observed values for any cluster (R^2 values between 0.01 and 0.05); the scatter
811 plots of the simulated shape factor values plotted against the P-3B values demonstrated both
812 over- and underprediction with no consistent bias. These plots also demonstrate that CMAQ
813 often randomly and severely over- or underpredicted relative to the P-3B throughout the depths
814 of the shape factor profiles for the other campaigns. This likely drove down the correlations
815 between simulated and observed values for these campaigns, and further indicates that CMAQ
816 struggled to reproduce the NO_2 vertical distributions under the conditions associated with the
817 California, Texas, and Colorado campaigns. Though CMAQ struggled to capture the shape
818 factor values for the Maryland campaign, CMAQ did perform best relative to the observations
819 for this campaign. In fact, CMAQ compared moderately well to the P-3B for Clusters 4 and 5
820 (R^2 values of 0.44 and 0.55, respectively). The scatter plots for the Maryland campaign also
821 demonstrate the least scatter of all four campaigns, as well as consistent, severe overestimation
822 of the NO_2 shape factors above 1200 m (upper PBL/lower free troposphere) and underprediction
823 below this altitude (lower PBL). These patterns of NO_2 shape factor bias are consistent with the
824 patterns of bias in the associated CO shape factor comparison plots (Fig. 18), suggesting that
825 overestimated vertical mixing contributes to the errors in the simulated Maryland NO_2 shape
826 factors. This NO_2 comparison, in addition to the comparisons for O_3, further suggests that
827 simulated vertical mixing was most vigorous for the Maryland campaign within CMAQ relative

828 to the other three campaigns. As the greatest correlations between observed and simulated
829 values, and smallest scatter, occurred for the Maryland campaign, this suggests that CMAQ was
830 best able to capture the NO_2 relative vertical distributions under the conditions conducive to O_3
831 photochemical production and NO_2 photochemical loss that occurred during this campaign. Use
832 of CMAQ NO_2 profiles in satellite retrievals over Maryland would produce less error than in the
833 other DISCOVER-AQ deployment regions. However, it should be noted that the slopes for the
834 comparison between simulated and observed NO_2 shape factor values were far from unity for
835 each cluster during each campaign, indicating that, while producing the least error for the
836 Maryalnd campaign, CMAQ still produced large errors in the simulated shape factors.
837 The GMI NO_2 shape factors compared well to the observations for Maryland, but
838 compared poorly for the California campaign (Fig. 19a)-b). GMI displayed approximately the
839 same smoothly decaying shape factors as seen for the P-3B during the Maryland campaign.
840 However, GMI often also displayed a high bias in NO_2 shape factor values within the upper
841 portion of the PBL, indicating that GMI placed a greater column burden higher in the PBL than
842 was observed. However, during the California campaign, the GMI NO_2 shape factors resembled
843 the GMI O_3 shape factors much more closely than they did the P-3B NO_2 shape factors. As with
844 O_3, much of the good comparison between GMI and the P-3B during the Maryland campaign
845 may be due to the model smoothing out the vertical variability when computing the shape factors
846 on the GMI vertical grid. This grid is likely too coarse within the PBL to capture the stagnant
847 conditions in the San Joaquin Valley, and thus to capture the very stable, stratified, and poorly
848 mixed NO_2 profile shapes and shape factors. Additionally, these median shape factor profiles
849 indicate that the NO_2 column burden was distributed too evenly within GMI relative to the
850 observations, as with O_3.

GMI poorly captured the shape factor values for all clusters of the California campaign, with R^2 values between 0.10 and 0.30 (Figure 20a)-b), as with CMAQ for this campaign; GMI compared moderately well for most clusters of the Maryland campaign, with R^2 values above 0.30. GMI compared exceptionally poorly to the P-3B shape factor values for Maryland Cluster 2 (R^2 = 0.18), but compared exceptionally well for Maryland Clusters 4 and 7 (R^2 = 0.82 and 0.63, respectively). The poor comparison of the GMI shape factor values to the observed California shape factors is particularly evident in the scatter plots, as they display both extreme over- and underprediction for all three clusters; most simulated shape factor values fell between approximately 0.05 and 0.09, regardless of the coincident observed values. This indicates that GMI is both under- and overpredicting. Much of this error in the GMI-simulated NO_2 shape factors may be due to overestimated vertical mixing, as the associated CO shape factor comparison plots display a consistent underprediction in CO in the lower PBL, with overprediction in the upper PBL/lower free troposphere (Fig. 20c). This further highlights the influence of atmospheric stability and mixing on NO_2 mixing ratio profiles and shape factors. The scatter plots for the Maryland clusters indicate that a few extreme over- and underpredicted values typically undermined the correlation between the simulated values and the observed shape factor values. The CO shape factor comparison plots for Maryland do not indicate the same errors in vertical mixing as for California. These results suggest that GMI performs better for the conditions associated with the Maryland campaign, similarly to CMAQ, in which NO_2 was more rapidly converted to O_3, than for the California campaign, where NO_2 concentrations were able to build up and the PBL was stably stratified. As the correlations between simulated and observed shape factor values were similar between CMAQ and GMI, and both models captured the Maryland median shape factors better than those for California, this suggests that neither type

874 of model presents an advantage for remote sensing retrievals in terms of profile shape for a short-
875 lived trace gas such as NO_2.

876 **6. Conclusions**

877 All *in situ* O_3 profile clusters produced by the agglomerative hierarchical cluster analysis
878 were significant for the California, Texas, and Colorado DISCOVER-AQ deployments, with
879 Texas producing the greatest number of distinct clusters (five). Only one distinct cluster emerged
880 for the Maryland deployment. The season in which the campaign took place, as well as the
881 relative degrees of vertical mixing and horizontal mixing, drove the differences in profile
882 variability, as denoted by the number of significant clusters that emerged, among campaigns.
883 Further, atmospheric stability, as indicated by the lapse rate and potential temperature profiles,
884 played a key role in modulating the differences in median O_3 profile shape among clusters within
885 each campaign, while time period during the campaign, spiral sampling time of day, and airmass
886 origin influenced the relative magnitudes of the clusters. In contrast, very few *in situ* NO_2 profile
887 clusters were distinct for each campaign, and those that were distinct contained very few profiles,
888 indicating that NO_2 profile behavior remained relatively uniform throughout the course of each
889 campaign; the interplay of NO_2 photochemical loss and vertical mixing likely played the key role
890 in determination of the amount of NO_2 profile variability. Future work could include a
891 comparison between the number of distinct clusters obtained and influences on these clusters
892 between the results of the agglomerative hierarchical clustering technique presented here and
893 another clustering technique, to determine if such a difference in technique is significant; this
894 could be especially interesting for the NO_2 profiles, as meteorological and chemical influences
895 on the clustering from different techniques could be examined.

When significant, the correlations between the corresponding P-3B column and surface data for each cluster were generally larger for O_3 than NO_2, which is not surprising given that O_3 is more well mixed in the vertical and horizontal than NO_2. However, many cluster correlations and correlations over the full set of profiles used in the clustering analysis were not significant for both O_3 and NO_2, suggesting that, even if the model profiles used in a satellite retrieval algorithm were to correctly represent the actual lower tropospheric profiles for these trace gases, meteorological or chemical conditions may prevent a strong column-surface connection. For O_3, column observations may be most representative of surface concentrations under the conditions of deep, convective boundary layers, reduced wind shear, and few terrain influences associated with the Maryland deployment. The degree of vertical mixing was also found to have an influence on the ability to relate column and surface quantities for NO_2, consistent with the results of Flynn et al. (2014), as greater column-surface correlation was found for clusters with more well mixed median NO_2 cluster profiles than for those with less well mixed median profiles. However, a typically relatively low degree of correlation was associated with even these well mixed median cluster profiles, indicating that accurate representation of the lower tropospheric NO_2 profile in a satellite retrieval does not guarantee the ability to connect column and surface.

Lastly, the regional CMAQ model and global GMI model simulated lower tropospheric O_3 shape factors compared moderately well to the P-3B shape factors corresponding to most clusters for the Maryland and California campaigns. These results suggest that a regional air quality model that captures the observed shape factor variability would be potentially useful in remote sensing retrievals for O_3. Both models performed better relative to the observations for the Maryland campaign than for any other campaign. These results demonstrate that models may

be best able to capture trace gas profiles under the conditions of convective boundary layers and O_3 photochemical formation associated with the Maryland campaign. These results also demonstrate the necessity of sufficient vertical resolution within the PBL and lower tropospohere, where the majority of the NO_2 burden resides, and for which both models struggled to capture both the shape and magnitude of the lower tropospheric shape factors.

Acknowledgements

Funding for this work was provided by the NASA Earth Venture-1 DISCOVER-AQ project. The authors thank the MDE, the SJV ACPD, the CARB, the TCEQ, the CDPHE, Millersville University, the Penn State NATIVE team, the NASA COMMIT team, the NASA LARGE group, the EPA, and NOAA for providing the surface O_3 and NO_2 datasets used in this work. The authors also thank Donald Lenschow for providing PBL height estimates for the Maryland campaign.

References

Allen, D., Pickering, K., Duncan, B., Damon, M. (2010). Impact of lightning NO emissions on North American photochemistry as determined using the Global Modeling Initiative (GMI) model. *Journal of Geophysical Research Atmospheres*, *115*(D22), 10.1029/2010JD014062.

Beirle, S., Platt, U., Wenig, M., Wagner, T. (2003). Weekly cycle of NO2 by GOME measurements : a signature of anthropogenic sources. *Atmospheric Chemistry and Physics*, *3*(2), 2225–2232.

Bhartia, P. K. (2002). OMI Algorithm Theoretical Basis Document, Volume II, OMI Ozone Products. *NASA-OMI, Washington, DC, ATBD-OMI-02, Version 2.0*, *II*(August).

Boersma, K. F., Jacob, D. J., Eskes, H. J., Pinder, R. W., Wang, J., & van der A, R. J. (2008). Intercomparison of SCIAMACHY and OMI tropospheric NO2 columns: Observing the diurnal evolution of chemistry and emissions from space. *Journal of Geophysical Research: Atmospheres*, *113*(16). http://doi.org/10.1029/2007JD008816

Boersma, K. F., Jacob, D. J., Trainic, M., Rudich, Y., DeSmedt, I., Dirksen, R., & Eskes, H. J. (2009). Validation of urban NO2 concentrations and their diurnal and seasonal variations observed from the SCIAMACHY and OMI sensors using in situ surface measurements in Israeli cities. *Atmospheric Chemistry and Physics*, *9*(12), 3867–3879. http://doi.org/10.5194/acp-9-3867-2009

Bucsela, E. J., Krotkov, N. A., Celarier, E. A., Lamsal, L. N., Swartz, W. H., Bhartia, P. K., Boersma, K.F., Veefkind, J.P., Gleason, J. F., & Pickering, K. E. (2013). A new stratospheric and tropospheric NO2 retrieval algorithm for nadir-viewing satellite

952 instruments: Applications to OMI. *Atmospheric Measurement Techniques*, *6*(10), 2607–
953 2626. http://doi.org/10.5194/amt-6-2607-2013
954 Chance, K. V. (2002). OMI Algorithm Theoretical Basis Document Volume IV, OMI Trace Gas
955 Algorithms. *NASA-OMI, Washington, DC, ATBD-OMI-04, Version 2.0*, *IV*(August), 1–50.
956 Retrieved from http://ozoneaq.gsfc.nasa.gov/media/docs/ATBD-OMI-04.pdf
957 Chatfield, R. B., & Esswein, R. F. (2012). Estimation of surface O 3 from lower-troposphere
958 partial-column information: Vertical correlations and covariances in ozonesonde profiles.
959 *Atmospheric Environment*, *61*, 103–113. http://doi.org/10.1016/j.atmosenv.2012.06.033
960 Diab, R. D., Raghunandan, A., Thompson, A. M., & Thouret, V. (2003). Classification of
961 tropospheric ozone profiles over Johannesburg based on MOZAIC aircraft data.
962 *Atmospheric Chemistry and Physics Discussions*, *3*(1), 705-732.
963 Diab, R. D., Thompson, A. M., Mari, K., Ramsay, L., & Coetzee, G. J. R. (2004). Tropospheric
964 ozone climatology over Irene, South Africa, from 1990 to 1994 and 1998 to 2002. *Journal*
965 *of Geophysical Research D: Atmospheres*, *109*(20).
966 Drury, E., D. J. Jacob, J. Wang, R. J. D. Spurrs, and K. Chance (2008). Improved algorithm for
967 MODIS satellite retrievals of aerosol optical depths over land, J. Geophys. Res., 113,
968 D16204, doi:10.1029/2007JD009573.
969 Drury, E., D. J. Jacob, R. J. D. Spurr, J. Wang, Y. Shinozuka, B. E. Anderson, A. D. Clarke, J.
970 Dibb, C. McNaughton, and R. Weber (2010). Synthesis of satellite (MODIS), aircraft
971 (ICARTT), and surface (IMPROVE, EPA-AQS, AERONET) aerosol observations over
972 North America to improve MODIS aerosol retrievals and constrain surface aerosol
973 concentration and sources, J. Geophys. Res., D14204.
974 Duncan, B. N., Strahan, S. E., & Yoshida, Y. (2007). Model study of the cross-tropopause
975 transport of biomass burning pollution. *Atmospheric Chemistry and Physics Discussions*,
976 *7*(1), 2197–2248. http://doi.org/10.5194/acpd-7-2197-2007
977 Engel-Cox, J.A.; Holloman, C.H.; Coutant, B.W.; Hoff, R.M. (2004). Qualitative and
978 Quantitative Evaluation of MODIS Satellite Sensor Data for Regional and Urban Scale Air
979 Quality; Atmos. Environ., 38, 2495-2509.
980 Fishman, J., Bowman, K. W., Burrows, J. P., Richter, A., Chance, K. V., Edwards, D. P., Martin,
981 R. V., Morris, G. A., Pierce, R. B., Ziemke, J. R., Schaack, T. K., and Thompson, A. M
982 (2008). Remote sensing of tropospheric pollution from space. *Bulletin of the American*
983 *Meteorological Society*, *89*(6), 805–821. http://doi.org/10.1175/2008BAMS2526.1
984 Flynn, C. M., Pickering, K. E., Crawford, J. H., Lamsal, L., Krotkov, N., Herman, J.,
985 Weinheimer, A., Chen, G., Liu, X., Szykman, J., Tsay, S., Loughner, C., Hains, J., Lee, P.,
986 Dickerson, R. R., Stehr, J. W., & Brent, L. (2014). Relationship between column-density
987 and surface mixing ratio: Statistical analysis of O3 and NO2 data from the July 2011
988 Maryland DISCOVER-AQ mission. *Atmospheric Environment*, *92*, 429–441.
989 http://doi.org/10.1016/j.atmosenv.2014.04.041
990 Hains, J. C., Taubman, B. F., Thompson, A. M., Stehr, J. W., Marufu, L. T., Doddridge, B. G., &
991 Dickerson, R. R. (2008). Origins of chemical pollution derived from Mid-Atlantic aircraft
992 profiles using a clustering technique. *Atmospheric Environment*, *42*(8), 1727–1741.
993 http://doi.org/10.1016/j.atmosenv.2007.11.052
994 Hoff, R., S.A. Christopher, (2009). Remote Sensing of Particulate Matter Air Pollution from
995 Space: Have we reached the promised land?, *J. Air & Waste Manage. Assoc.*, 59:642-675.
996 Janjić, Z. I. (1994). The Step-Mountain Eta Coordinate Model: Further Developments of the
997 Convection, Viscous Sublayer, and Turbulence Closure Schemes. *Monthly Weather Review*.

998 http://doi.org/10.1175/1520-0493(1994)122<0927:TSMECM>2.0.CO;2
999 Knepp, T., Pippin, M., Crawford, J., Chen, G., Szykman, J., Long, R., Cowen, L., Cede, A.,
1000 Abuhassan, N., Herman, J., Delgado, R., Compton, J., Berkoff, T., Fishman, J., Martins, D.,
1001 Stauffer, R., Thompson, A. M., Weinheimer, A., Knapp, D., Montzka, D., Lenschow, D., &
1002 Neil, D. (2013). Estimating surface NO2 and SO2 mixing ratios from fast-response total
1003 column observations and potential application to geostationary missions. *Journal of*
1004 *Atmospheric Chemistry*, (2), 1–26. http://doi.org/10.1007/s10874-013-9257-6
1005 Kota, S. H., G. Schade, M. Estes, D. Boyer, and Q. Ying (2015), Evaluation of MEGAN
1006 predicted biogenic isoprene emissions at urban locations in Southeast Texas, *Atmos.*
1007 *Environ.,* 110,54–64.

1008 Lamsal, L. N., Krotkov, N. A., Celarier, E. A., Swartz, W. H., Pickering, K. E., Bucsela, E. J.,
1009 Gleason, J. F., Martin, R. V., Philip, S., Irie, H., Cede, A., Herman, J., Weinheimer, A.,
1010 Szykman, J. J., & Knepp, T. N.. (2014). Evaluation of OMI operational standard NO2
1011 column retrievals using in situ and surface-based NO2 observations. *Atmospheric Chemistry*
1012 *and Physics Discussions*, *14*(10), 14519–14573. http://doi.org/10.5194/acpd-14-14519-2014
1013 Lamsal, L. N., Martin, R. V., Padmanabhan, A., van Donkelaar, A., Zhang, Q., Sioris, C. E.,
1014 Chance, K., Kurosu, T. P., & Newchurch, M. J. (2011). Application of satellite observations
1015 for timely updates to global anthropogenic NOx emission inventories. *Geophysical*
1016 *Research Letters*, *38*(5), 1–5. http://doi.org/10.1029/2010GL046476.
1017 Lamsal, L. N., Martin, R. V., van Donkelaar, A., Steinbacher, M., Celarier, E. A., Bucsela, E.,
1018 Dunlea, E. J., and Pinto, J. P. (2008), Ground-level nitrogen dioxide concentrations inferred
1019 from satellite-borne Ozone Monitoring Instrument, *J. Geophys. Res., 113*, D16308, doi:
1020 10.1029/2007JD009235.
1021 Lamsal, L. N., Martin, R. V., van Donkelaar, A., Celarier, E. A., Bucsela, E. J., Boersma, K. F.,
1022 Dirksen, R., Luo, C., Wang, Y. (2010), Indirect validation of tropospheric nitrogen dioxide
1023 retrieved from the OMI satellite instrument: Insight into the seasonal variation of nitrogen
1024 oxides at northern midlatitudes, *J. Geophys. Res., 115*, D05302,
1025 doi:10.1029/2009JD013351.
1026 Lee, C. J., Brook, J. R., Evans, G. J., Martin, R. V., & Mihele, C. (2011). Novel application of
1027 satellite and in-situ measurements to map surface-level NO2 in the Great Lakes region.
1028 *Atmospheric Chemistry and Physics*, *11*(22), 11761–11775. http://doi.org/10.5194/acp-11-
1029 11761-2011
1030 Liu, X., Bhartia, P. K., Chance, K., Spurr, R. J. D., & Kurosu, T. P. (2010). Ozone profile
1031 retrievals from the Ozone Monitoring Instrument. *Atmos. Chem. Phys. Atmospheric*
1032 *Chemistry and Physics*, *10*, 2521-2537. Retrieved from www.atmos-chem-
1033 phys.net/10/2521/2010/
1034 Loughner, C. P., Tzortziou, M., Follette-Cook, M., Pickering, K. E., Goldberg, D., Satam, C.,
1035 Weinheimer, A., Crawford, J. H., Knapp, D. J., Montzka, D. D., Diskin, G. S., & Dickerson,
1036 R. R. (2014). Impact of bay-breeze circulations on surface air quality and boundary layer
1037 export. *Journal of Applied Meteorology and Climatology*, *53*(7), 1697–1713.
1038 http://doi.org/10.1175/JAMC-D-13-0323.1
1039 Martin, R. V. (2008). Satellite remote sensing of surface air quality. *Atmospheric Environment*,
1040 *42*(34), 7823–7843. http://doi.org/10.1016/j.atmosenv.2008.07.018
1041 Mazzuca, G. M., K. E. Pickering, R. D. Clark, C. P. Loughner, A. Fried, R. R. Dickerson, D. C.
1042 Stein-Zweers, A. J. Weinheimer, (2016) Use of tethersonde and aircraft profiles to study the

impact of mesoscale and microscale meteorology on air quality, Atmos. Environ., submitted.

Natraj, V., Liu, X., Kulawik, S.S., Chance, K., Chatfield, R., Edwards, D.P., Eldering, A., Francis, G., Kurosu, T.P., Pickering, K., Spurr, R., Worden, H., (2011). Multispectral sensitivity studies for the retrieval of tropospheric and lowermost tropospheric ozone from simulated clear sky GEO-CAPE measurements. Atmospheric Envi- ronment 45, 7151e7165. http://dx.doi.org/10.1016/j.atmosenv.2011.09.014.

Ordóñez, C., Richter, A., Steinbacher, M., Zellweger, C., Nüß, H., Burrows, J. P., & Prévôt, A. S. H. (2006). Comparison of 7 years of satellite-borne and ground-based tropospheric NO2 measurements around Milan, Italy. *Journal of Geophysical Research: Atmospheres*, *111*(5), 1–12. http://doi.org/10.1029/2005JD006305

Palmer, P. I., Jacob, D. J., Chance, K., Martin, R. V., Spurr, R. J. D., Kurosu, T. P., Bey, I., Yantosca, R., Fiore, A., & Li, Q. (2001). Air mass factor forumlation for spectroscopic measurements from satellites: Application to formaldehyde retrievals form the Global Ozone Monitoring Experiment. *Journal of Geophysical Research*, *106*(D13), 14539–14550.

Pleim, J. (2006). A simple, efficient solution of flux-profile relationships in the atmospheric surface layer. *Journal of Applied Meteorology and Climatology*, (1), 341–347. http://doi.org/10.1175/JAM2339.1

Pleim, J. E. (2007). A combined local and nonlocal closure model for the atmospheric boundary layer. Part II: Application and evaluation in a mesoscale meteorological model. *Journal of Applied Meteorology and Climatology*, *46*(9), 1396–1409. http://doi.org/10.1175/JAM2534.1

Pouliot, G., and T. E. Pierce (2009), Integration of the Model of Emissions of Gases and Aerosols from Nature (MEGAN) into the CMAQ modeling system. http://www3.epa.gov/ttnchie1/conference/ei18/session3/pouliot.pdf

Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M. G., Schubert, S. D., Takacs, L., Kim, G., Bloom, S., Chen, J., Collins, D., Conaty, A., da Silva, A., Gu, W., Joiner, J., Koster, R. D., Lucchesi, R., Molod, A., Owens, T., Pawson, S., Pegion, P. Redder, C. R., Reichle, R., Robertson, F. R., Ruddick, A. G., Sienkiewicz, M., & Woollen, J. (2011). MERRA: NASA's modern-era retrospective analysis for research and applications. *Journal of Climate*, *24*(14), 3624-3648.

Russell, A. R., Perring, A. E., Valin, L. C., Bucsela, E. J., Browne, E. C., Wooldridge, P. J., & Cohen, R. C. (2011). A high spatial resolution retrieval of NO2 column densities from OMI: Method and evaluation. *Atmospheric Chemistry and Physics*, *11*(16), 8543-8554.

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, X.Y. Huang, W. Wang, and J. G. Powers, (2008): A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 113 pp.

Stauffer, R. M., A. M. Thompson, and G. S. Young (2016), Tropospheric ozonesonde profiles at long-term U.S. monitoring sites: 1. A climatology based on self-organizing maps, J. Geophys. Res. Atmos., 121, 1320–1339, doi:10.1002/2015JD023641.

Strode, S. A., Duncan, B. N., & Yegorova, E. A. (2015). Implications of model bias in carbon monoxide for methane lifetime. *Atmospheric Chemistry and Physics Discussions*, *15*, 20305–20348. http://doi.org/10.5194/acpd-15-20305-2015

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: Atmospheres, *111*(10), 1–14. http://doi.org/10.1029/2005JD006196

van Donkelaar, A., Martin, R. V., Leaitch, W. R., Macdonald, A. M., Walker, T. W., Streets, D. G., Zhang, Q., Dunlea, E. J., Jimenez, J. L., Dibb, J. E., Huey, L. G., Weber, R., and Andreae, M. O.: Analysis of aircraft and satellite measurements from the Intercontinental Chemical Transport Experiment (INTEX-B) to quantify long-range transport of East Asian sulfur to Canada, Atmos. Chem. Phys., 8, 2999-3014, doi:10.5194/acp-8-2999-2008, 2008.

van der Werf, G. R., Randerson, J. T., Giglio, L., Collatz, G. J., Mu, M., Kasibhatla, P. S., Morton, D. C., DeFries, R. S., Jin, Y., and van Leeuwen, T. T.: Global fire emissions and the contribution of deforestation, savanna, forest, agricultural, and peat fires (1997–2009), Atmos. Chem. Phys., 10, 11707-11735, doi:10.5194/acp-10-11707-2010, 2010.

Wang, J., and S.A. Christopher (2003). Intercomparison between satellite-derived aerosol optical thickness and PM2.5 mass: Implication for air quality studies, *Geophys. Res. Lett.*, 30, 2095, doi:10.1029/2003GL018174.

Wang, J., X. Xu, R. Spurr, Y. Wang, and E. Drury, (2010). Improved algorithm for MODIS satellite retrievals of aerosol optical thickness over land in dusty atmosphere: Implications for air quality monitoring in China, Remote Sensing of Environment, 114, 2575-2583.

Xiu, a, & Pleim, J. (2001). Development of a land surface model. Part I: Application in a mesoscale meteorological model. *Journal of Applied Meteorology*, 192–209. http://doi.org/10.1175/1520-0450(2001)040<0192:DOALSM>2.0.CO;2

Yarwood, G., Rao, S., Yocke, M., & Whitten, G. Z. (2005). Updates to the Carbon Bond Chemical Mechanism: CB05. *Camx.Com*. Retrieved from http://www.camx.com/files/cb05_final_report_120805.aspx\npapers2://publication/uuid/200E81B3-7B33-4353-82C0-76D262E2A5E4

Zhang, Q., Streets, D. G., Carmichael, G. R., He, K. B., Huo, H., Kannari, A., Klimont, Z., Park, I. S., Reddy, S., Fu, J. S., Chen, D., Duan, L., Lei, Y., Wang, L. T., & Yao, Z. L. (2009). Asian emissions in 2006 for the NASA INTEX-B mission. *Atmospheric Chemistry and Physics*, *9*(14), 5131-5153.

Zhang, Y., Wang, Y., Chen, G., Smeltzer, C., Crawford, J., Olson, J., Szykman, J., et al. (2016). Large vertical gradient of reactive nitrogen oxides in the boundary layer: Modeling analysis of DISCOVER-AQ 2011 observations. *Journal of Geophysical Research: Atmospheres*, *121*(4), 1922-1934.

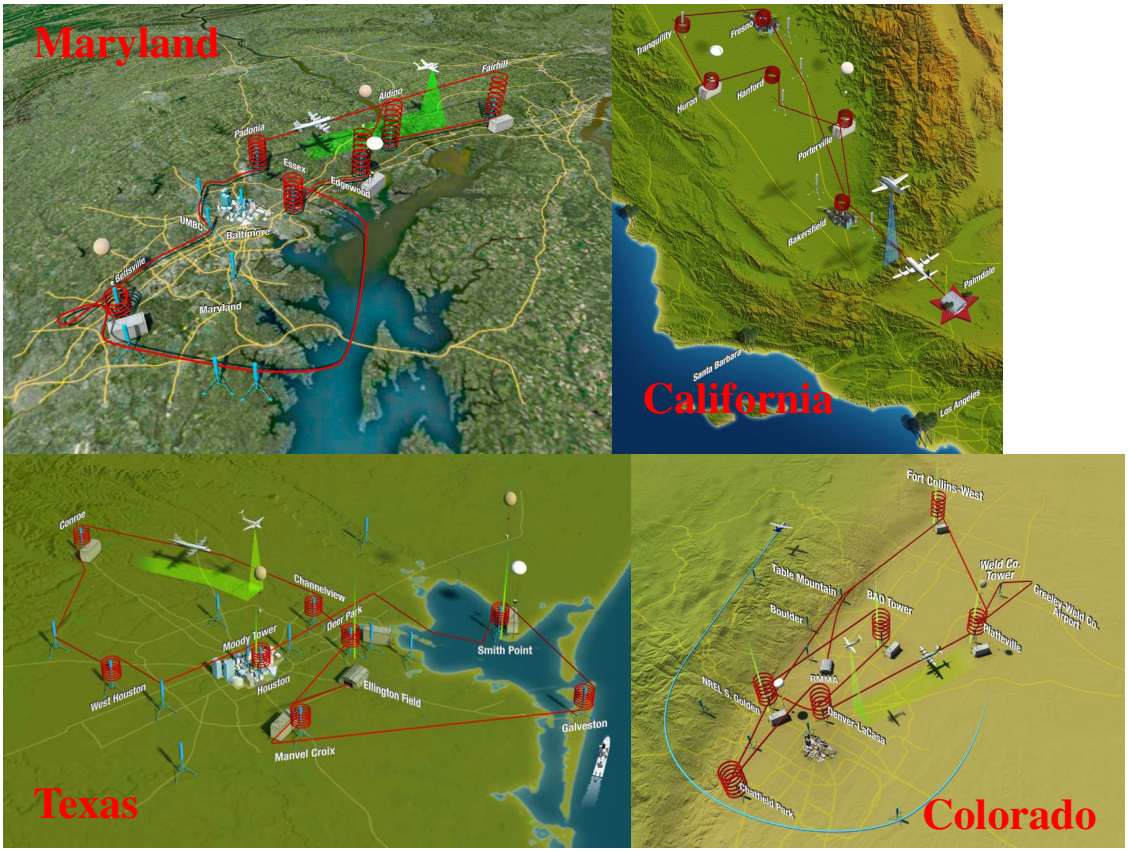


Fig. 1. a) The six spiral sites for the Maryland campaign; b) the six spiral sites for the California campaign; c) the eight spiral sites for the Texas campaign; and d) the six spiral sites for the Colorado campaign. Spiral sites named in white font.

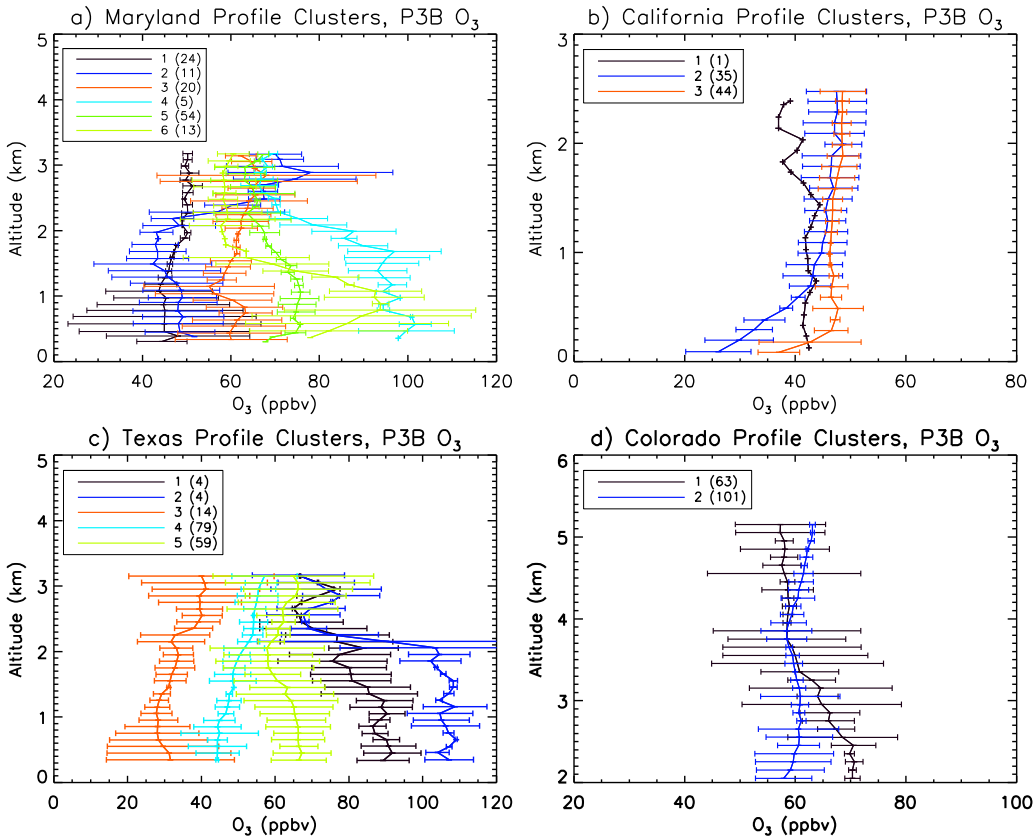
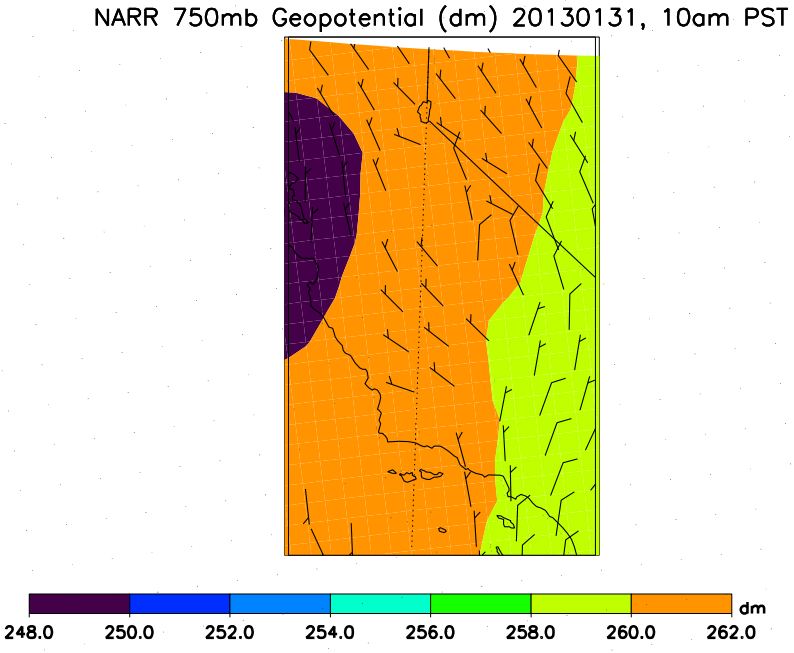


Fig. 2. The median profiles for each *in situ* O_3 profile cluster a) for the Maryland campaign; b) for the California campaign c) for the Texas campaign; and d) for the Colorado campaign. Cluster numbers displayed in legend, with number of profiles in each cluster given in parentheses. Error bars represent 25th and 75th percentile values.


NARR 750mb Geopotential (dm) 20130131, 10am PST
248.0
250.0
252.0
254.0
256.0
258.0
260.0
262.0
dm

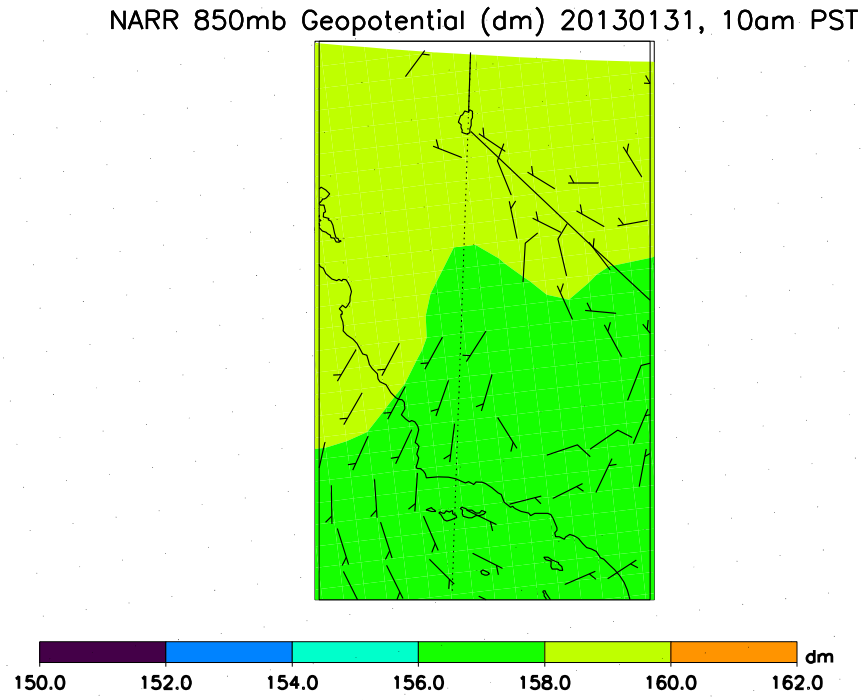
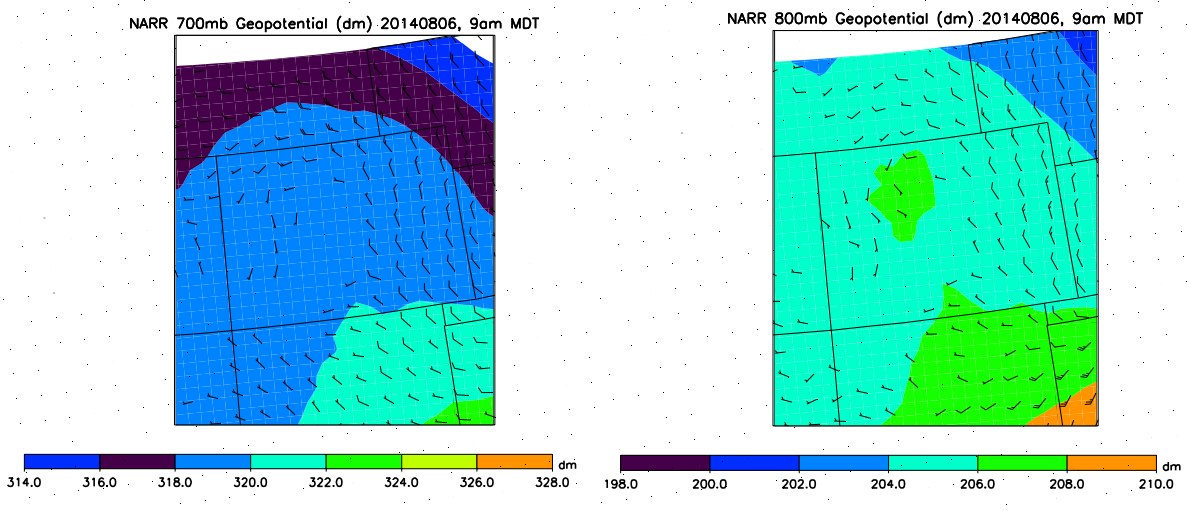


Fig. 3. NARR wind fields and geopotential height at the 750 mb and 850 mb at 10am PST on January 31, 2013, displaying typical circulation patterns within the San Joaquin Valley, California.



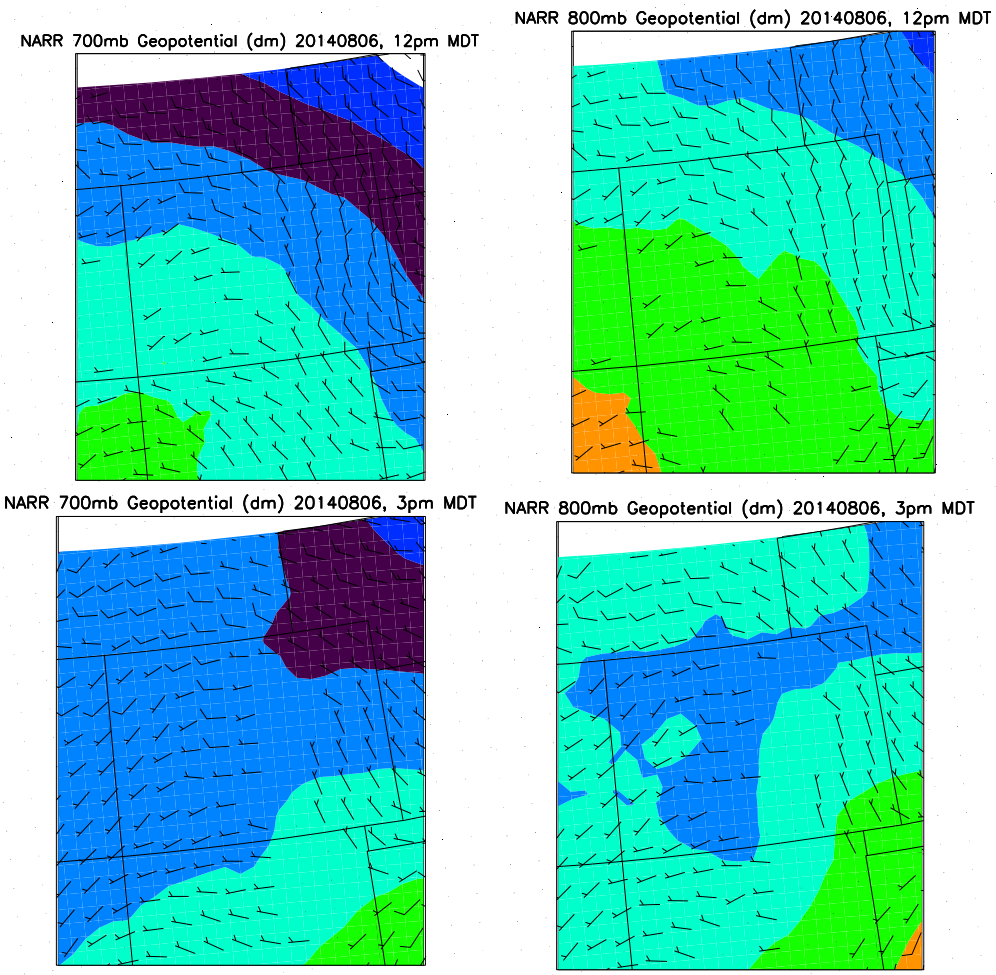


Fig. 4. NARR wind fields and geopotential height at the 700 mb and 800 mb at 9am, 12pm, and 3pm MDT on August 6, 2014, displaying typical circulation patterns and changes in wind flow over Colorado.

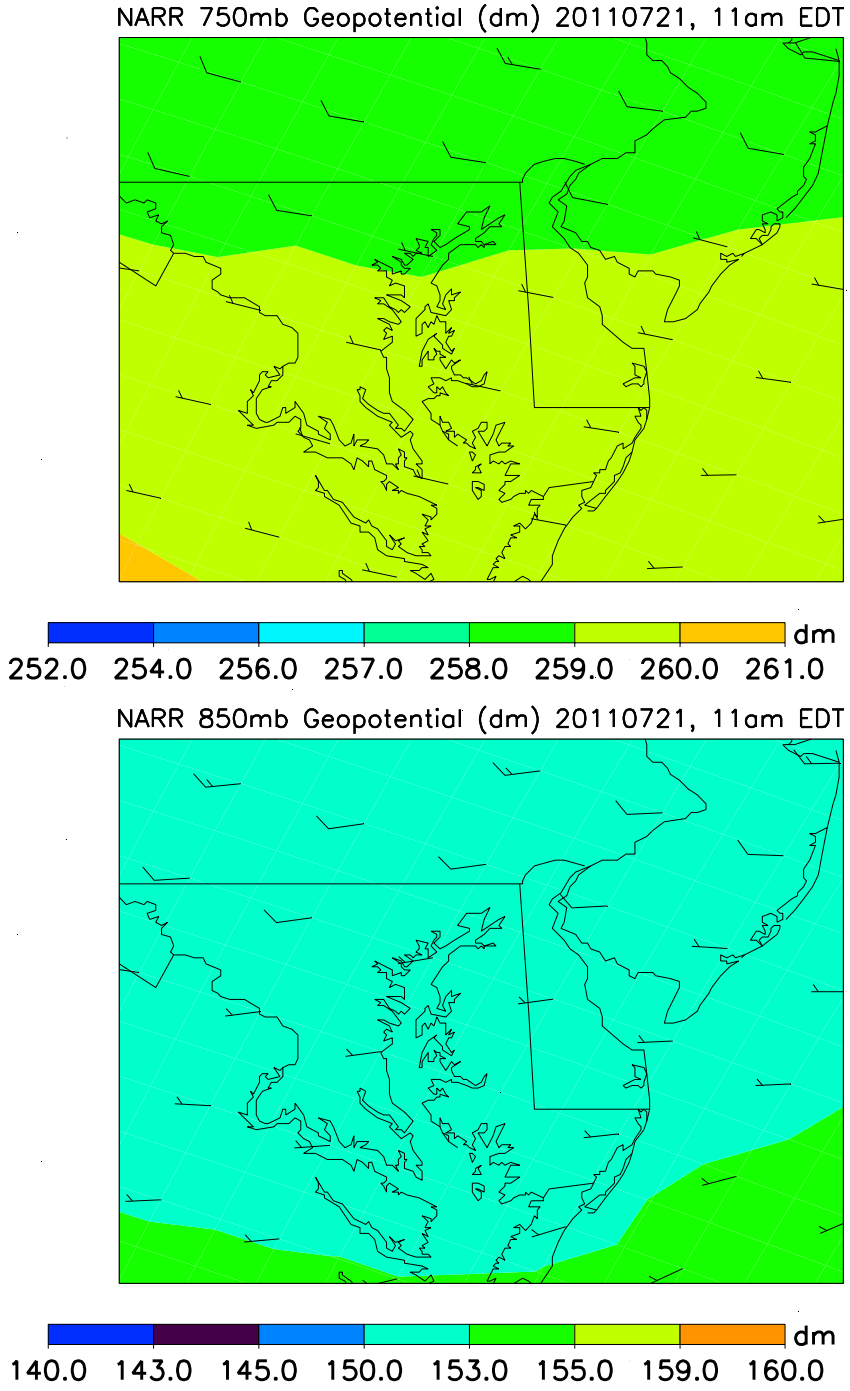


Fig. 5. NARR wind fields and geopotential height at the 750 mb and 850 mb at 11am EDT on July 21, 2011, displaying typical westerly flow patterns over Maryland.

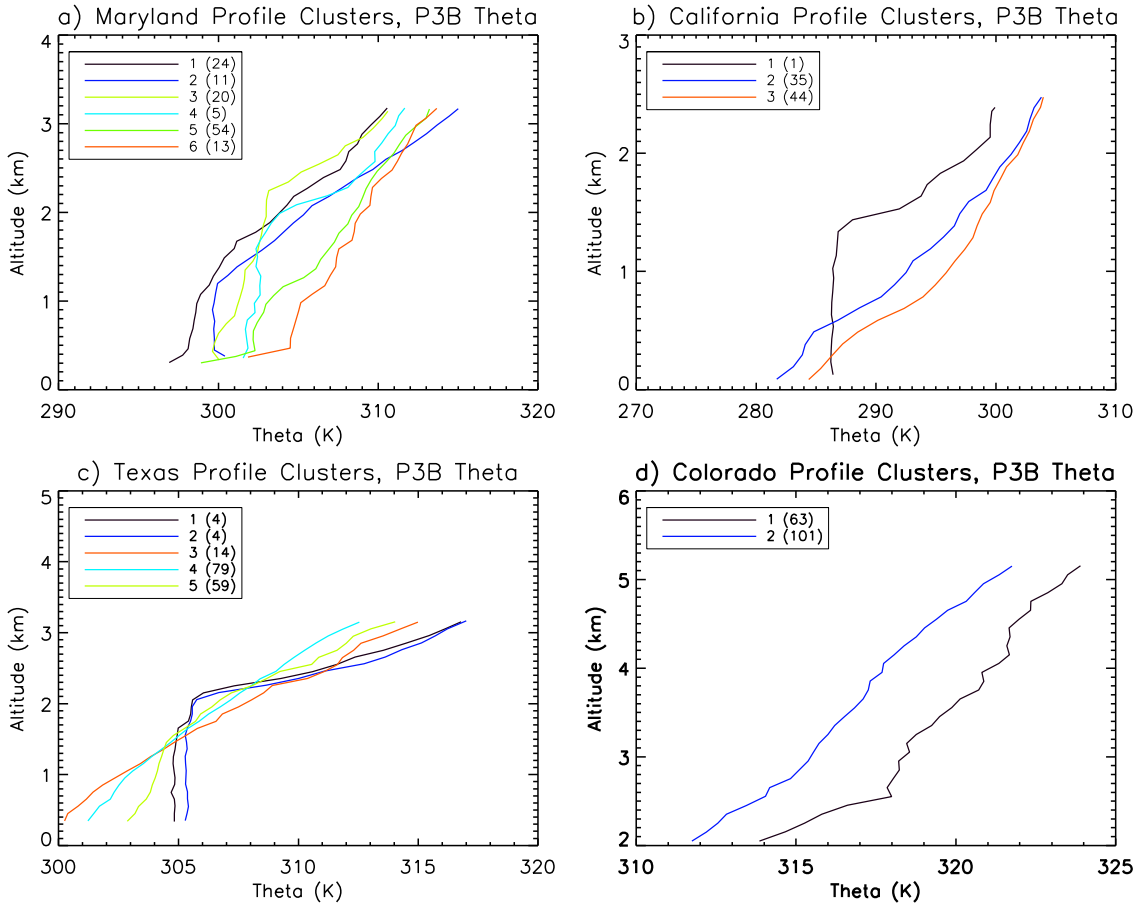


Fig. 6. The median potential temperature profiles associated with each *in situ* O_3 profile cluster a) for the Maryland campaign; b) for the California campaign c) for the Texas campaign; and d) for the Colorado campaign. Cluster numbers displayed in legend, with number of profiles in each cluster given in parentheses.

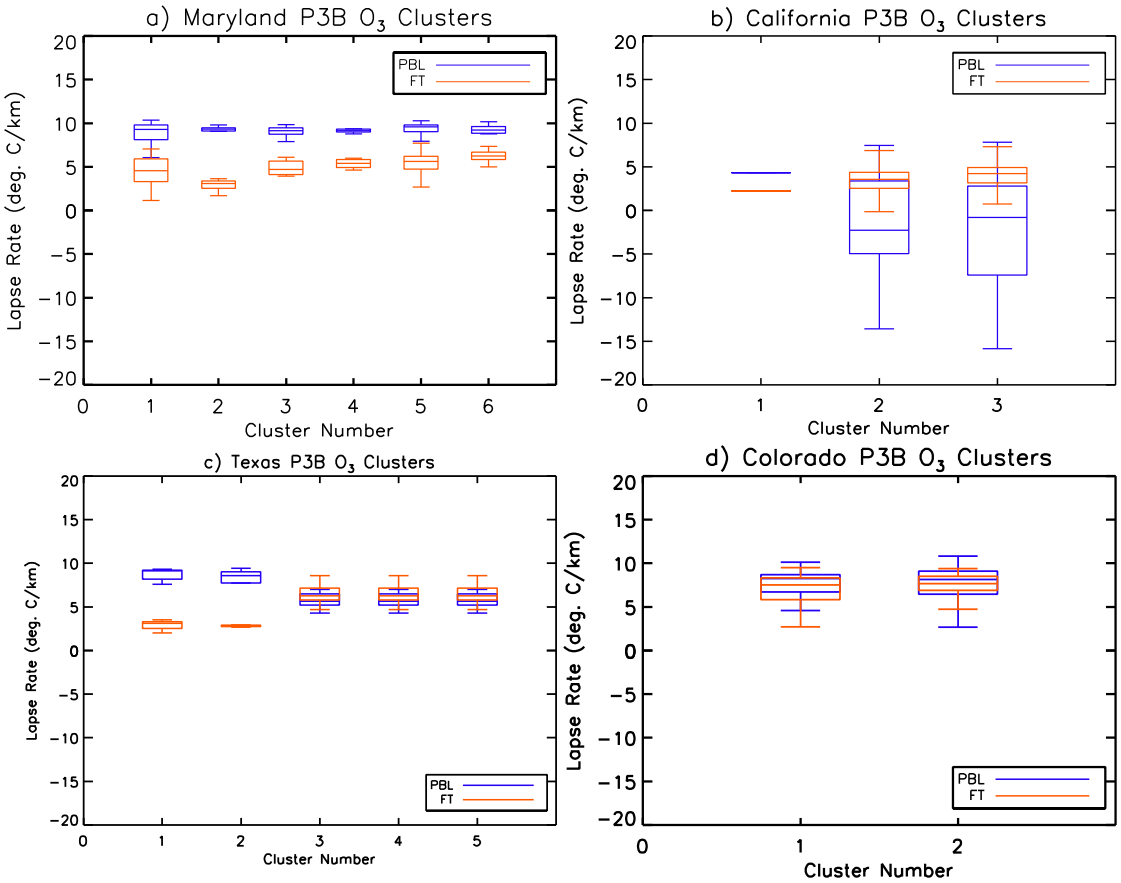


Fig. 7. The median lapse rates and distributions associated with each *in situ* O_3 profile cluster, computed separately for the PBL and free troposphere (FT), a) for the Maryland campaign; b) for the California campaign c) for the Texas campaign; and d) for the Colorado campaign.

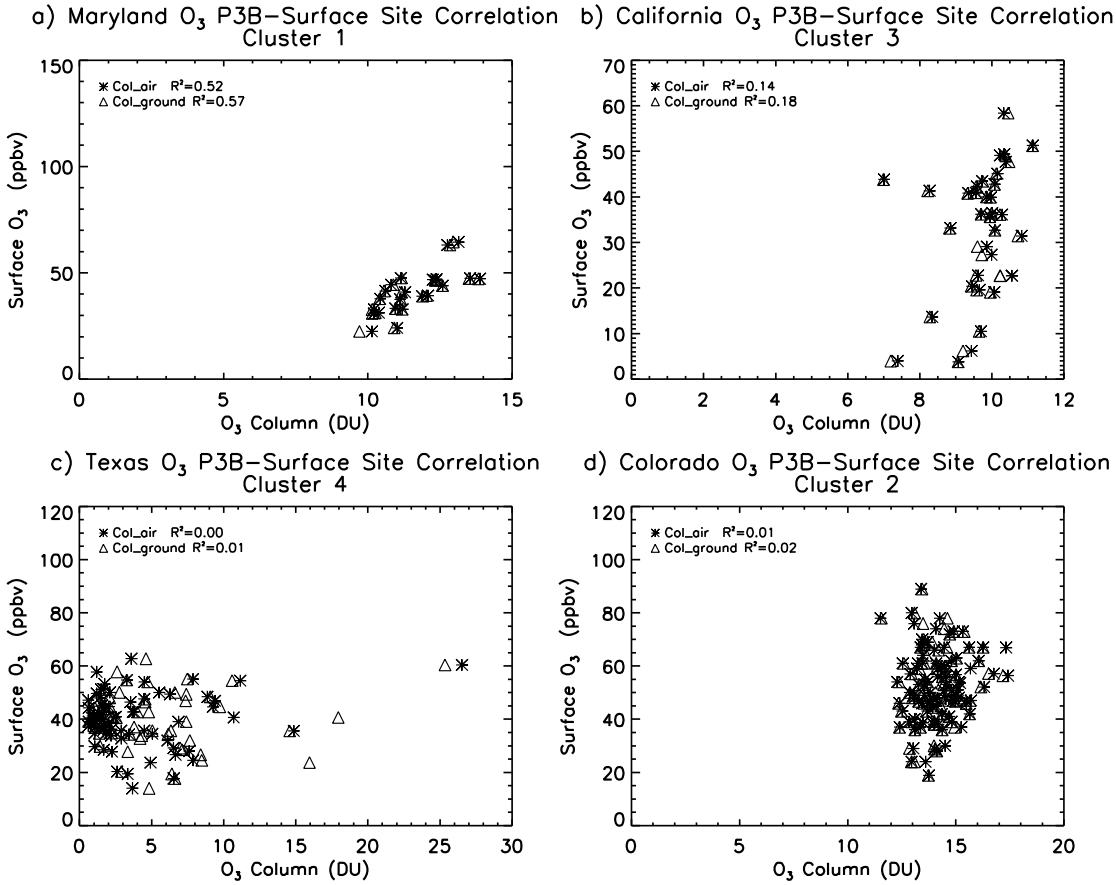


Fig. 8. Representative scatter plots for the O_3 column-surface correlations a) for the Maryland campaign; b) for the California campaign c) for the Texas campaign; and d) for the Colorado campaign. R^2 values for the column_air and column_ground correlations displayed in the legends.

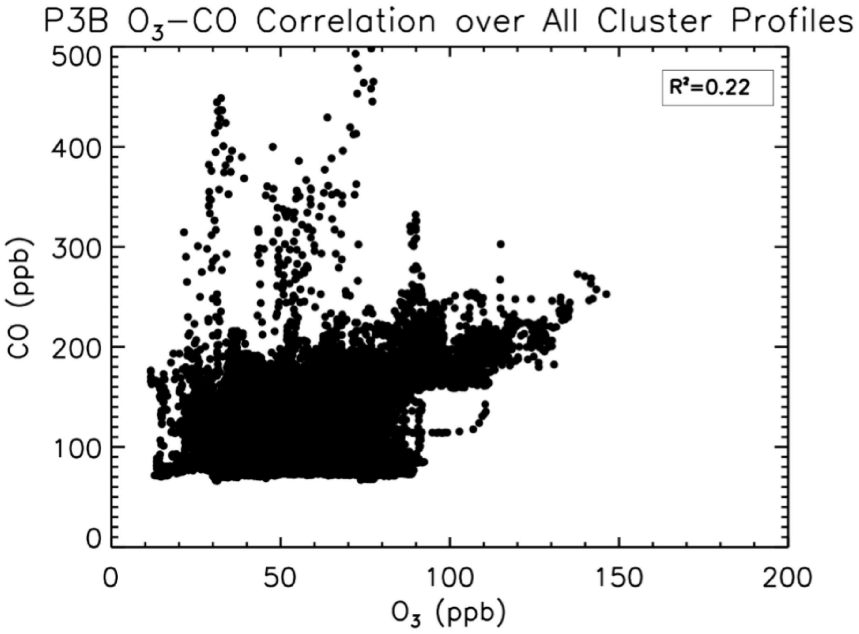


Fig. 9. Correlation plot of *in situ* CO vs. O_3 over the spiral depths of the profiles included in the clustering analysis for the Texas campaign.

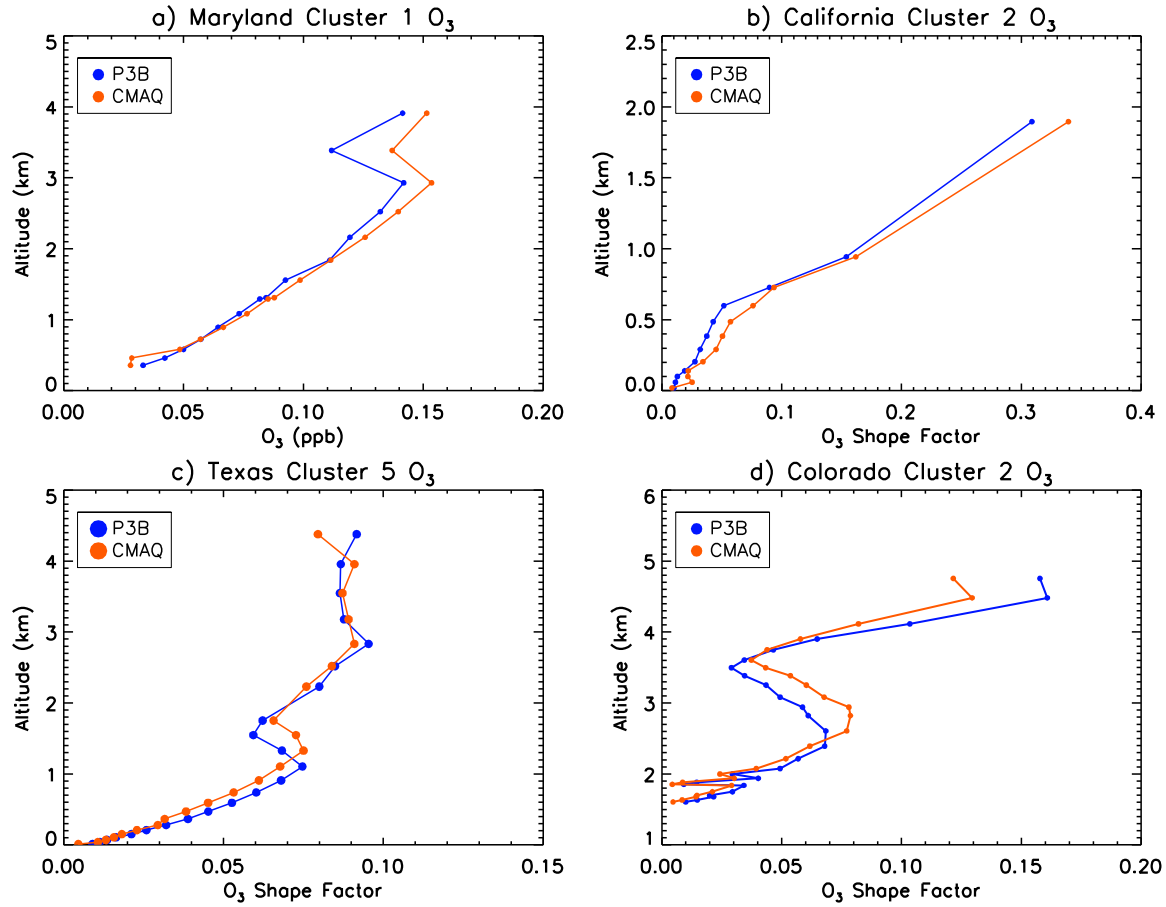


Fig.10. Representative shape factor comparison plots for CMAQ vs. P-3B for O_3 a) for the Maryland campaign; b) for the California campaign, c) for the Texas campaign; d) for the Colorado campaign. Shape factors computed on CMAQ vertical grid.

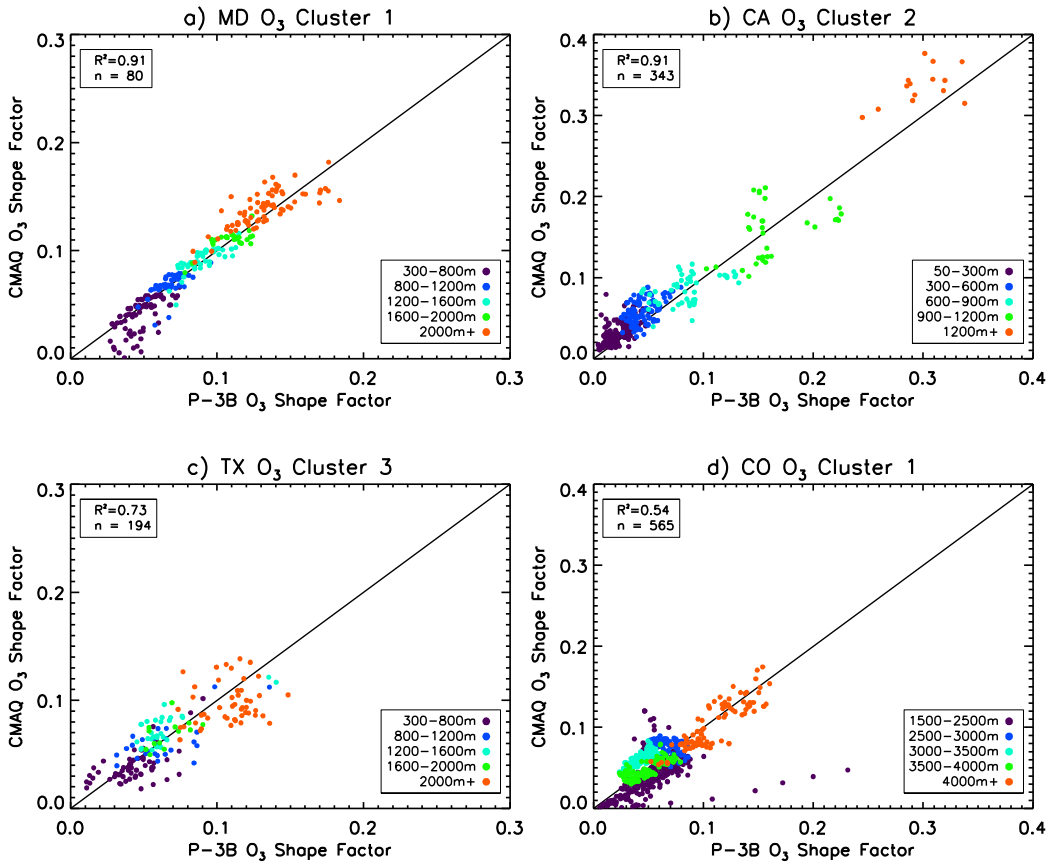


Fig. 11. Comparison scatter plots of CMAQ vs. observed shape factor values for O_3 a) for the Maryland campaign Cluster 1; b) for the California campaign Cluster 2, c) for the Texas campaign Cluster 3; d) for the Colorado campaign Cluster 1. Shape factors computed on CMAQ vertical grid. Scatter plots colored by altitude layers with legend in bottom right displaying the altitude layer ranges for each campaign. R^2 value and number of data points for the correlation between simulated and observed shape factor values displayed in upper left.

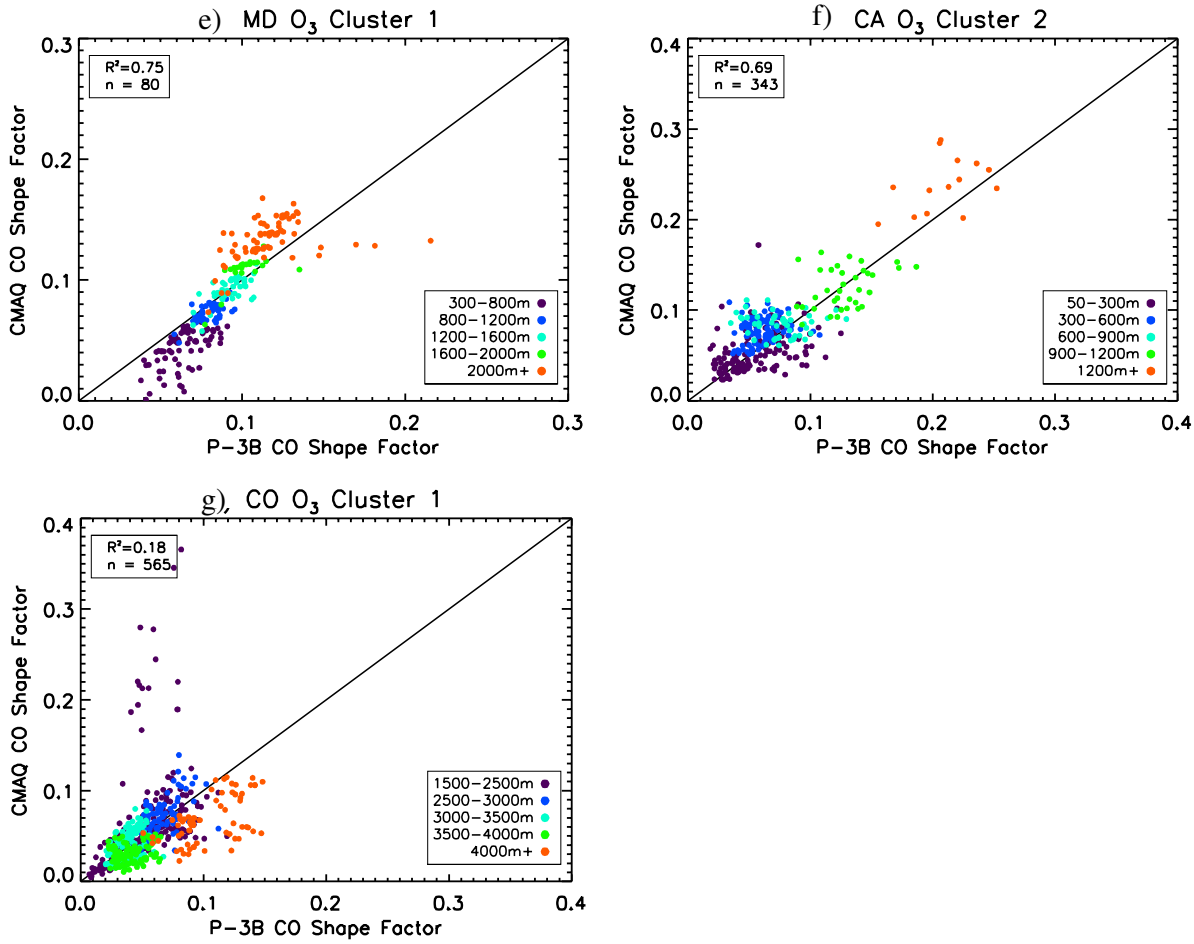


Fig. 11. Comparison scatter plots of CMAQ vs. observed shape factor values for CO associated with the O_3 clusters e) for the Maryland campaign Cluster 1; f) for the California campaign Cluster 2, g) for the Colorado campaign Cluster 1. Shape factors computed on CMAQ vertical grid. Scatter plots colored by altitude layers with legend in bottom right displaying the altitude layer ranges for each campaign. R^2 value and number of data points for the correlation between simulated and observed shape factor values displayed in upper left.

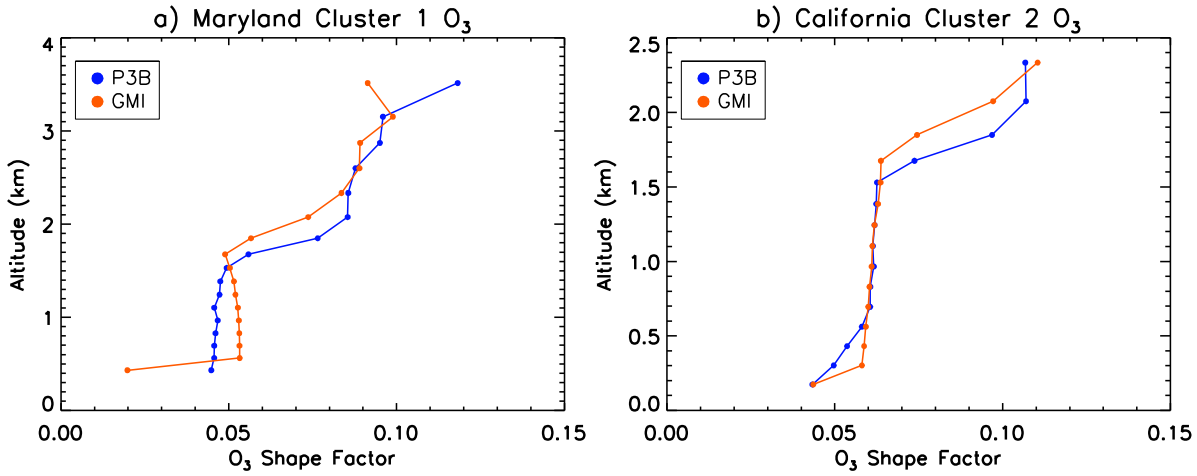


Fig. 12. Representative shape factor comparison plots for GMI vs. P-3B for O_3 a) for the Maryland campaign; b) for the California campaign. Shape factors computed on GMI vertical grid.

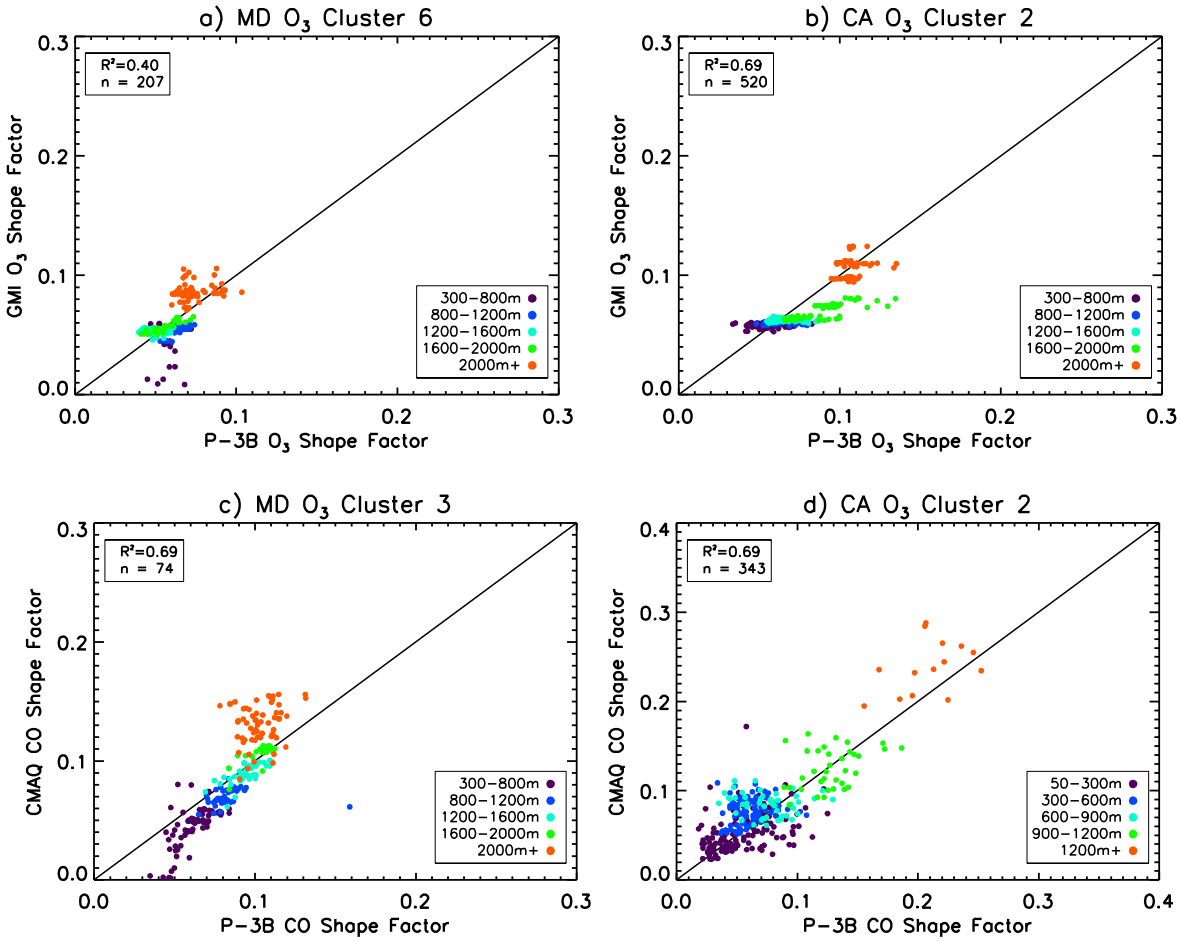


Fig. 13. Comparison scatter plots of GMI vs. observed shape factor values for O_3 and CO a) scatter plot for the Maryland campaign Cluster 2; b) scatter plot for the Maryland campaign Cluster 6; and for CO associated with each cluster for c) the Maryland campaign Cluster 3; and d) for the California campaign Cluster 2. Shape factors computed on GMI vertical grid. Scatter plots colored by altitude layers with legend in bottom right displaying the altitude layer ranges for each campaign. R^2 value and number of data points for the correlation between simulated and observed shape factor values displayed in upper left.

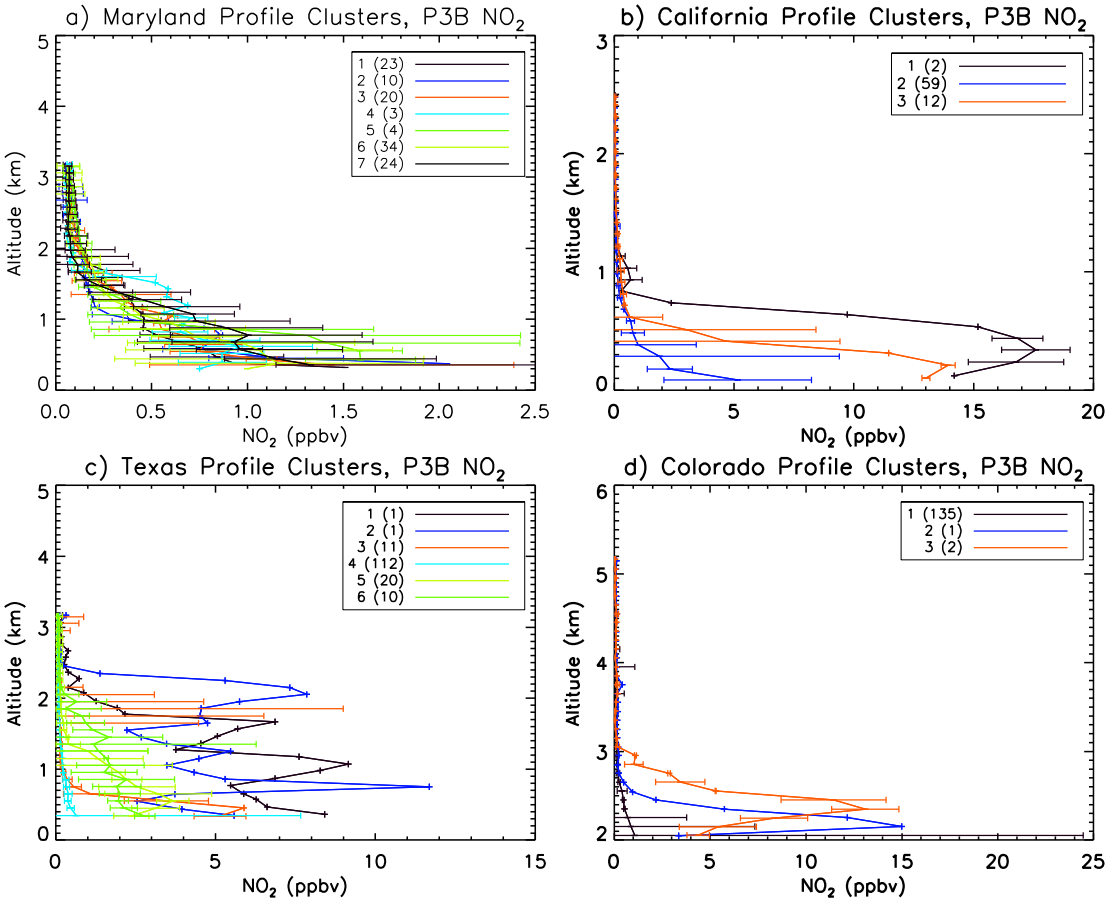


Fig. 14. The median profiles for each *in situ* NO_2 profile cluster a) for the Maryland campaign; b) for the California campaign c) for the Texas campaign; and d) for the Colorado campaign. Cluster numbers displayed in legend, with number of profiles in each cluster given in parentheses. Error bars represent 25th and 75th percentile values.

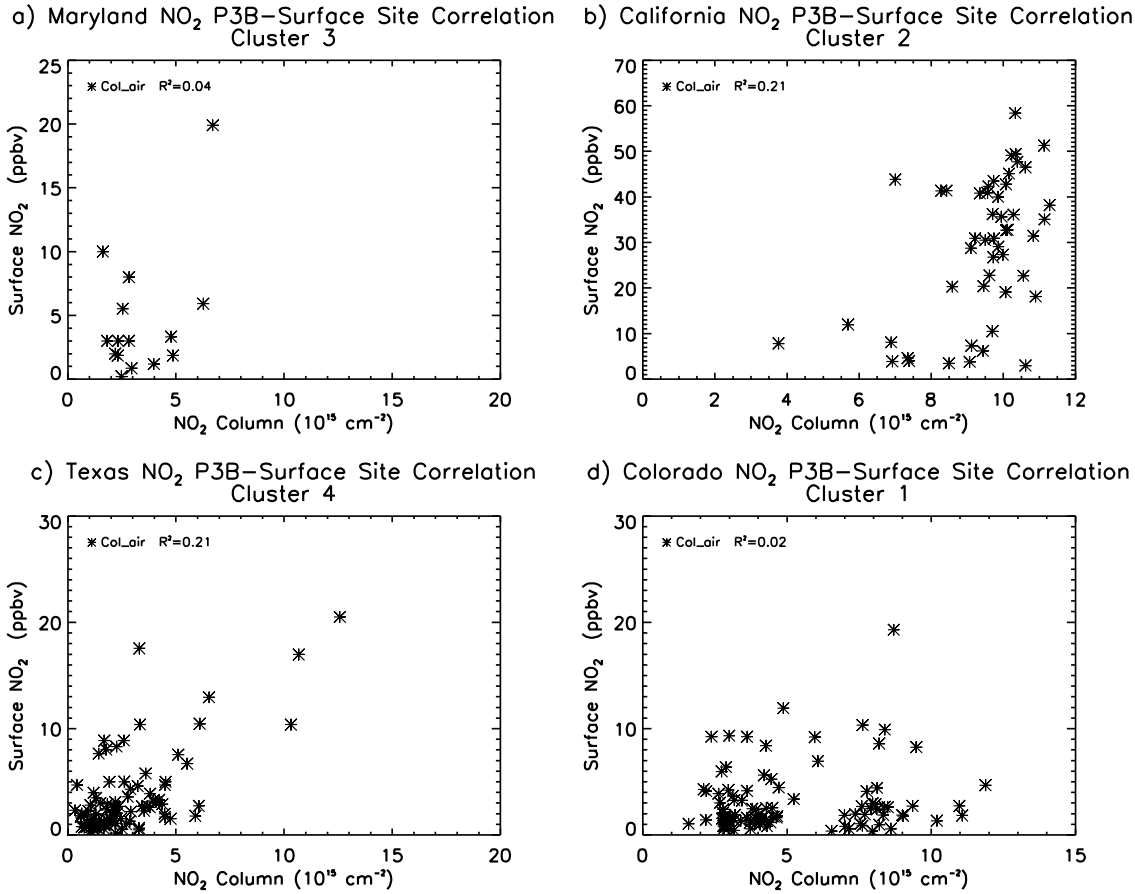


Fig. 15. Representative scatter plots for the NO_2 column-surface correlations a) for the Maryland campaign; b) for the California campaign c) for the Texas campaign; and d) for the Colorado campaign. R^2 values for the column_air and column_ground correlations displayed in the legends.

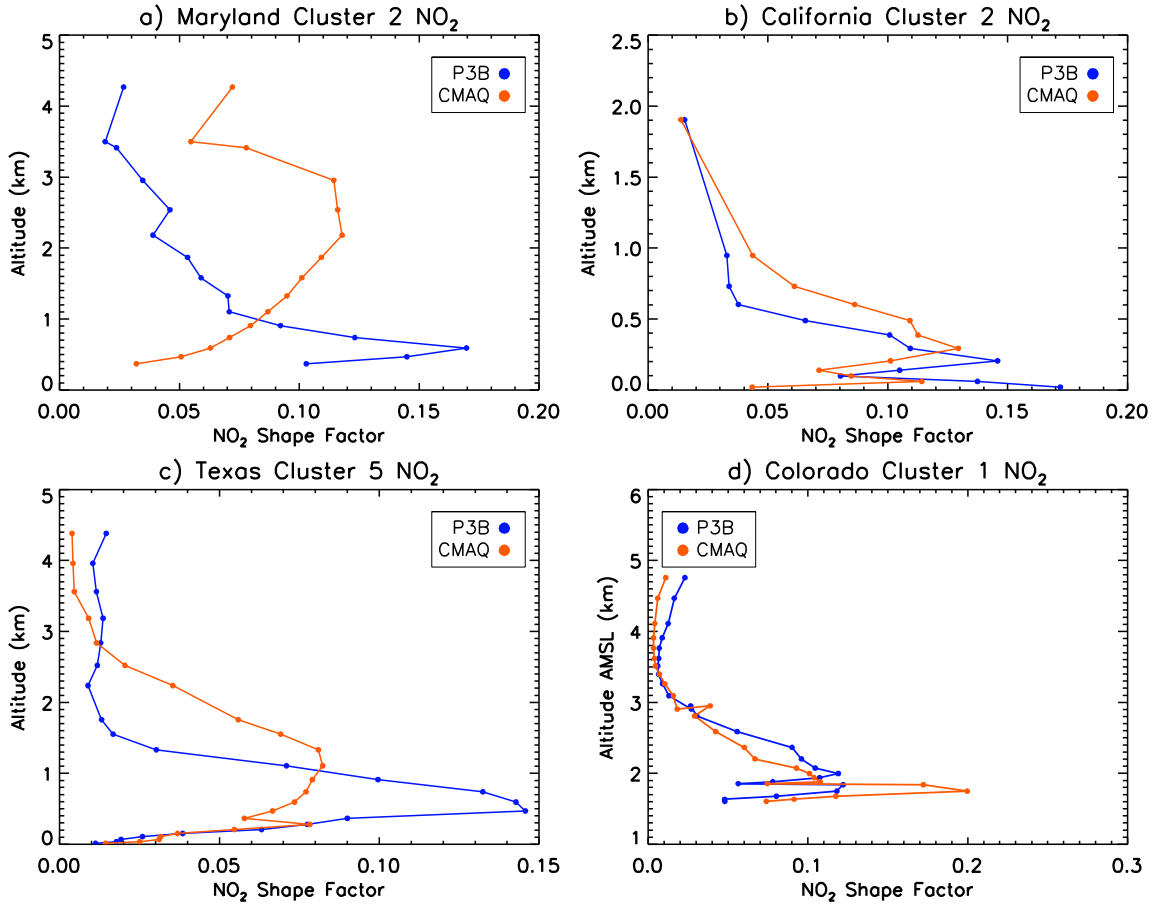


Fig. 16. Representative shape factor comparison plots for CMAQ vs. P-3B for NO_2 a) for the Maryland campaign; b) for the California campaign; c) for the Texas campaign; d) for the Colorado campaign. Shape factors computed on CMAQ vertical grid.

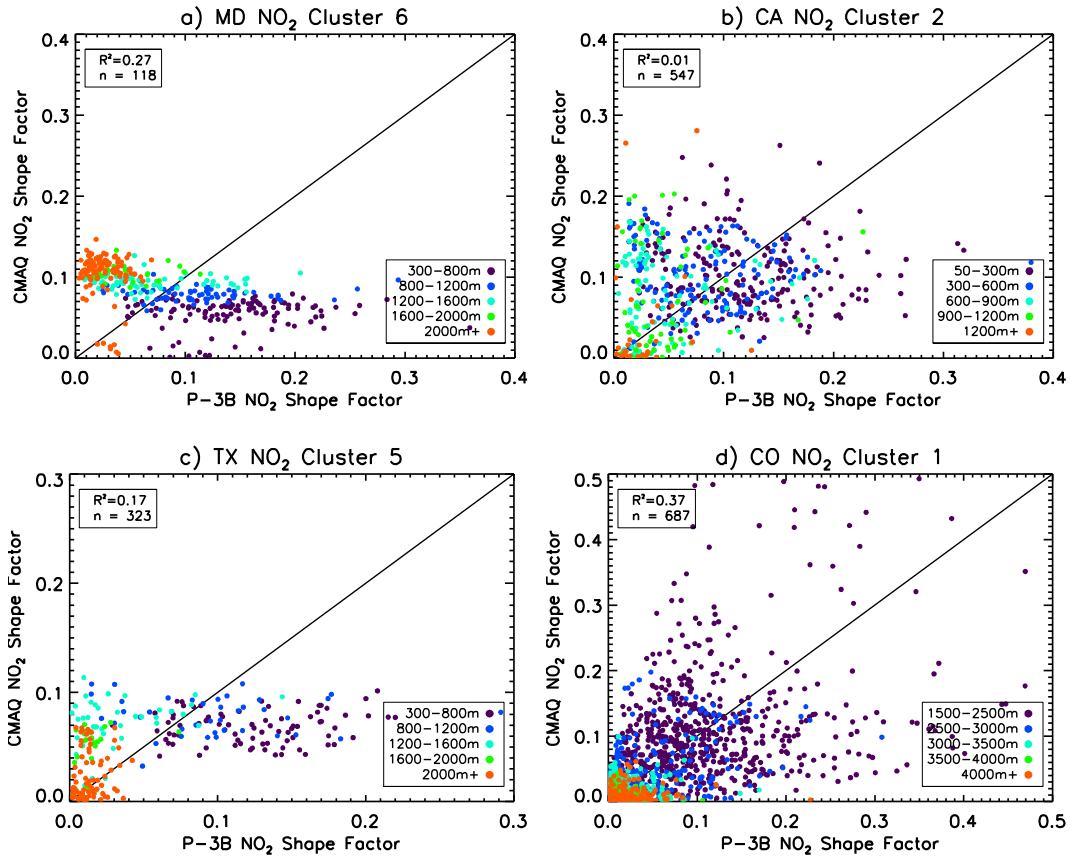


Fig. 17. Comparison scatter plots of CMAQ vs. observed shape factor values for NO_2 a) for the Maryland campaign Cluster 6; b) for the California campaign Cluster 2; c) for the Texas campaign Cluster 5; d) for the Colorado campaign Cluster 1. Shape factors computed on CMAQ vertical grid. Scatter plots colored by altitude layers with legend in bottom right displaying the altitude layer ranges for each campaign. R^2 value and number of data points for the correlation between simulated and observed shape factor values displayed in upper left.

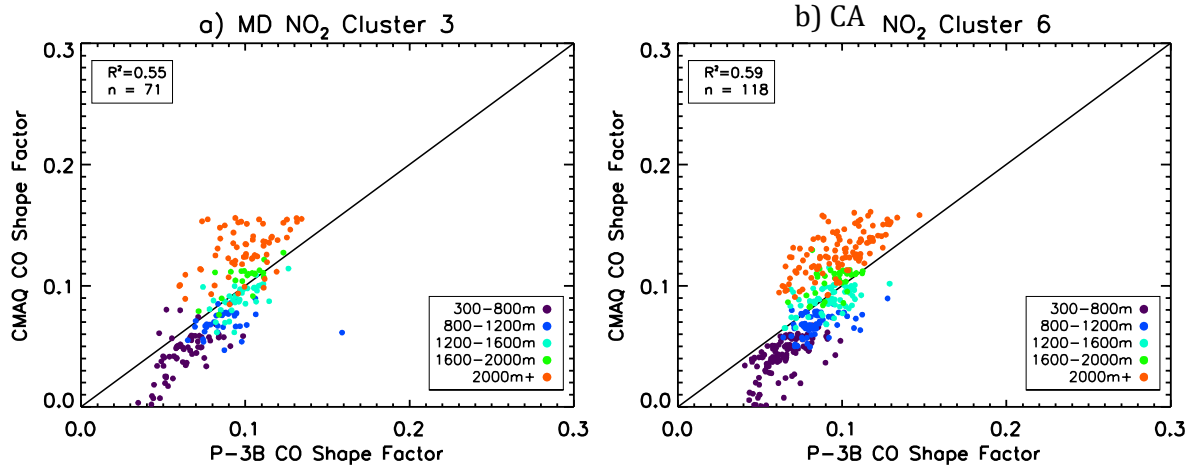
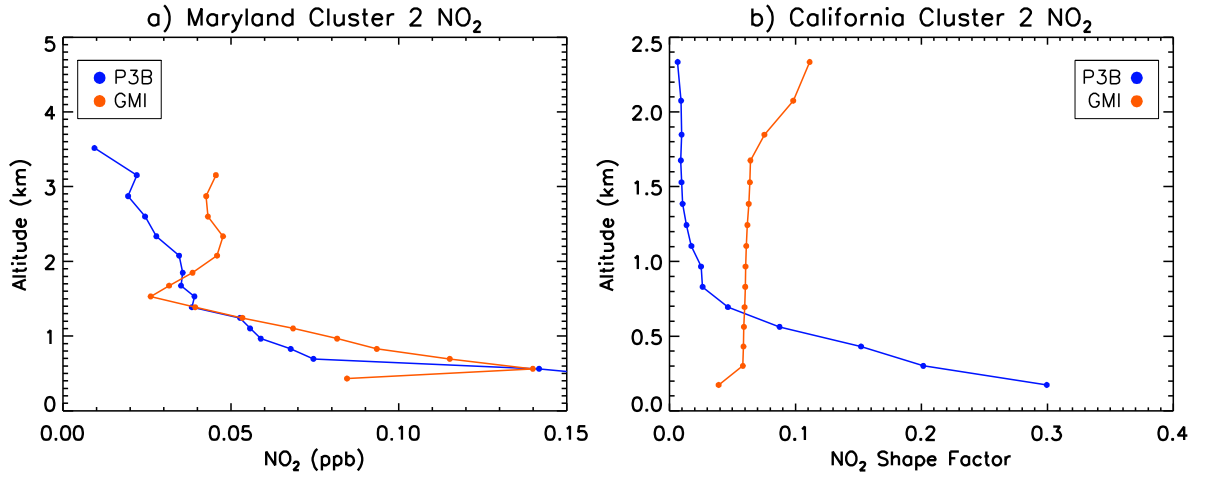


Fig. 18. Comparison scatter plots of CMAQ vs. P-3B shape factor values for CO associated with the NO_2 clusters a) for the Maryland campaign Cluster 3; b) for the Maryland campaign Cluster 6. Shape factors computed on CMAQ vertical grid. Scatter plots colored by altitude layers with legend in bottom right displaying the altitude layer ranges for each campaign. R^2 value and number of data points for the correlation between simulated and observed shape factor values displayed in upper left.



z

Fig. 19. Representative shape factor comparison plots for GMI vs. P-3B for NO_2 a) for the Maryland campaign; b) for the California campaign. Shape factors computed on GMI vertical grid.

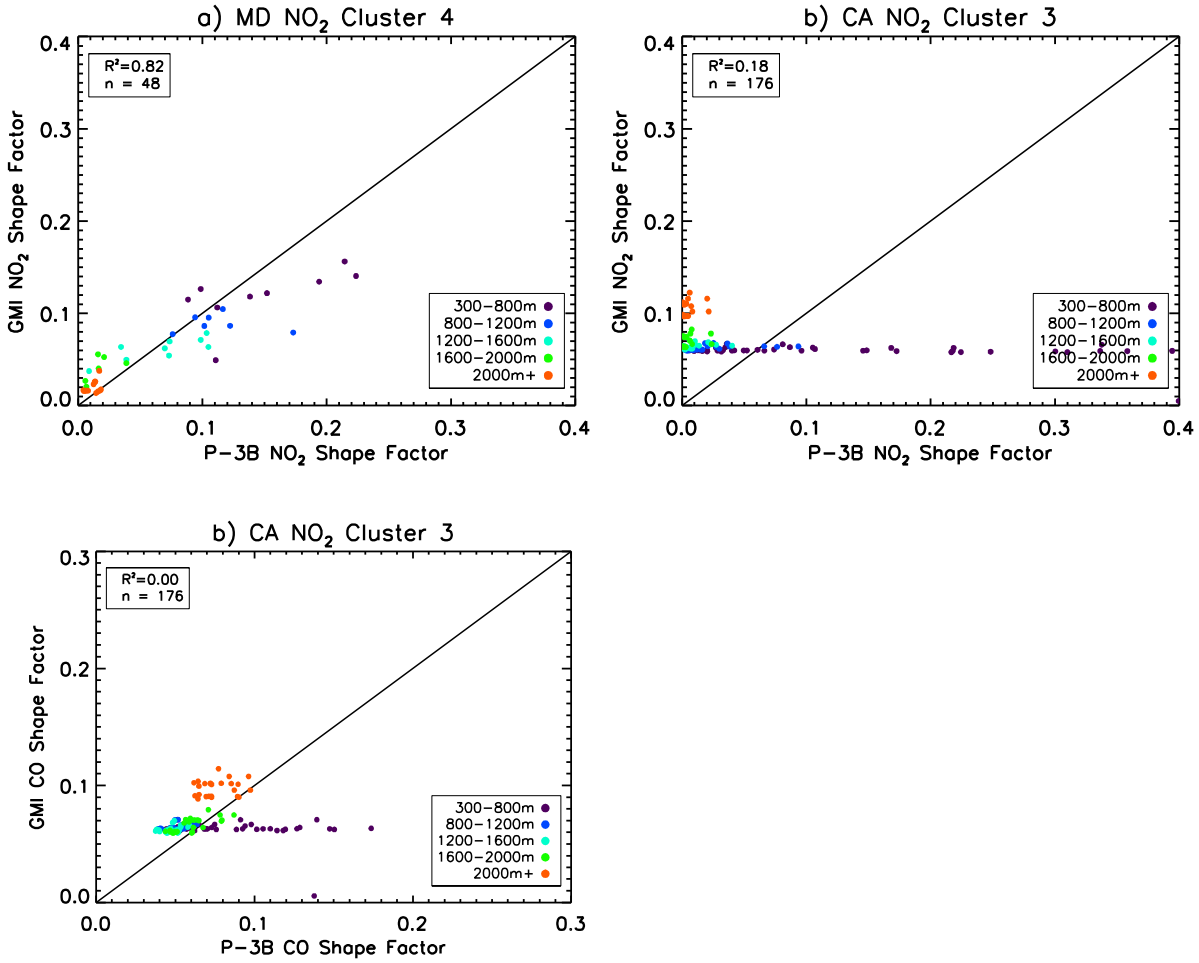


Fig. 20. Shape factor comparison scatter plots for GMI vs. P-3B shape factor values for NO_2 a) for the Maryland campaign Cluster 4; b) for the California campaign Cluster 3, and for CO for the California campaign Cluster 3. Shape factors computed on GMI vertical grid. Scatter plots colored by altitude layers with legend in bottom right displaying the altitude layer ranges for each campaign. R^2 value and number of data points for the correlation between simulated and observed shape factor values displayed in upper left.

California Campaign Campaign Period			
Dates	**Cluster 2**	**Cluster 1**	**Cluster 3**
20130116- 20130122	48.6%	0%	56.8%
20130130- 20130206	51.4%	100%	43.2%

Table 1: Percentages of profiles within each cluster that fell within each campaign period, as denoted by the Dates column, for the California campaign O_3 clusters. Clusters listed from left to right in order of least to most polluted.

California Campaign Spiral Sampling Time			
Spiral	**Cluster 2**	**Cluster 1**	**Cluster 3**
Spiral 1 (~8am PST)	42.9%	0%	31.8%
Spiral 2 (~11am PST)	40.0%	0 %	31.8%
Spiral 3 (~1pm PST)	17.1%	100%	36.4%

Table 2: Percentages of profiles within each cluster that fell within each spiral sampling time, as denoted by the Spiral column, for the California campaign O_3 clusters. Clusters listed from left to right in order of least to most polluted.

Texas Campaign Campaign Period					
Dates	**Cluster 3**	**Cluster 4**	**Cluster 5**	**Cluster 1**	**Cluster 2**
20130904-20130906	28.6%	21.5%	3.4%	0%	0%
20130911-20130914	0%	65.8%	49.2%	0%	0%
20130924-20130926	71.4%	12.7%	47.5%	100%	100%

Table 3: Percentages of profiles within each cluster that fell within each campaign period, as denoted by the Dates column, for the Texas campaign O_3 clusters. Clusters listed from left to right in order of least to most polluted.

Texas Campaign Spiral Sampling Time					
Spiral	**Cluster 3**	**Cluster 4**	**Cluster 5**	**Cluster 1**	**Cluster 2**
Spiral 1 (~9am CDT)	71.4%	43.0%	23.7%	25%	25%
Spiral 2 (~12pm CDT)	21.4%	32.9%	40.7%	25%	0%
Spiral 3 (~2pm CDT)	7.1%	22.8%	35.6%	50%	75%
Spiral 4 (~3pm CDT)	0%	1.3%	0%	0%	0%

Table 4: Percentages of profiles within each cluster that fell within each spiral sampling time, as denoted by the Spiral column, for the Texas campaign O_3 clusters. Clusters listed from left to right in order of least to most polluted.

Colorado Campaign Campaign Period		
Dates	**Cluster 2**	**Cluster 1**
20140717-20140721	28.7%	50.8%
20140728-20140803	30.7%	42.9%
20140806-20140810	40.6%	6.3%

Table 5: Percentages of profiles within each cluster that fell within each campaign period, as denoted by the Dates column, for the Colorado campaign O_3 clusters. Clusters listed from left to right in order of least to most polluted.

Colorado Campaign Spiral Sampling Time		
Spiral	**Cluster 2**	**Cluster 1**
Spiral 1 (~9am MDT)	55.4%	26.9%
Spiral 2 (~11am MDT)	39.6%	33.3%
Spiral 3 (~4pm MDT)	4.9%	39.7%

Table 6: Percentages of profiles within each cluster that fell within each spiral sampling time, as denoted by the Spiral column, for the Colorado campaign O_3 clusters. Clusters listed from left to right in order of least to most polluted.

Maryland Campaign HYSPLIT Back Trajectories						
Airmass Origin	**Cluster 1**	**Cluster 2**	**Cluster 3**	**Cluster 5**	**Cluster 6**	**Cluster 4**
N. Canada	54.2%	27.3%	20%	11.1%	7.7%	0%
NW Canada/Great Lakes	0%	0%	10%	25.9%	0%	100%
Long Range Transport	16.7%	27.3%	0%	7.4%	7.7%	0%
Ohio River Valley/Westerly Flow	25%	45.5%	70%	46.3%	84.7%	0%

Table 7: Percentages of profiles within each cluster that fell within each Airmass Origin Category, based on HYSPLIT back trajectory clusters, for the Maryland campaign O_3 clusters. Clusters listed from left to right in order of least to most polluted. HYSPLIT back trajectories initiated at all vertical levels included.

Colorado Campaign HYSPLIT Back Trajectories at 500 m		
Airmass Origin	**Cluster 2**	**Cluster 1**
Northerly Flow	12.9%	19.0%
Southwesterly Flow	29.7%	23.8%
Local Recirculation	22.7%	52.4%
Westerly Flow	1.9%	0%

Table 8: Percentages of profiles within each cluster that fell within each Airmass Origin Category, based on HYSPLIT back trajectory clusters, for the Colorado campaign O_3 clusters. Clusters listed from left to right in order of least to most polluted. HYSPLIT back trajectories initiated at 500 m vertical level only.

Colorado Campaign HYSPLIT Back Trajectories at 1000 m		
Airmass Origin	**Cluster 2**	**Cluster 1**
Northerly Flow	22.8%	15.9%
Southwesterly Flow	37.6%	42.9%
Local Recirculation	11.9%	38.1%
Westerly Flow	1.9%	0%

Table 9: Percentages of profiles within each cluster that fell within each Airmass Origin Category, based on HYSPLIT back trajectory clusters, for the Colorado campaign O_3 clusters. Clusters listed from left to right in order of least to most polluted. HYSPLIT back trajectories initiated at 1000 m vertical level only.

Colorado Campaign HYSPLIT Back Trajectories at 1500 m		
Airmass Origin	**Cluster 2**	**Cluster 1**
Northerly Flow	9.9%	11.1%
Southwesterly Flow	32.7%	23.8%
Westerly Flow	3.9%	11.1%
Local Recirculation	25.7%	44.4%

Table 10: Percentages of profiles within each cluster that fell within each Airmass Origin Category, based on HYSPLIT back trajectory clusters, for the Colorado campaign O_3 clusters. Clusters listed from left to right in order of least to most polluted. HYSPLIT back trajectories initiated at 1500 m vertical level only.

Colorado Campaign HYSPLIT Back Trajectories at 2000 m		
Airmass Origin	**Cluster 2**	**Cluster 1**
Northerly Flow	28.7%	12.7%
West-northwesterly Flow	23.8%	36.5%
Local Recirculation	16.8%	42.9%

Table 11: Percentages of profiles within each cluster that fell within each Airmass Origin Category, based on HYSPLIT back trajectory clusters, for the Colorado campaign O_3 clusters. Clusters listed from left to right in order of least to most polluted. HYSPLIT back trajectories initiated at 2000 m vertical level only.

Colorado Campaign HYSPLIT Back Trajectories at 2500 m		
Airmass Origin	**Cluster 2**	**Cluster 1**
Northerly Flow	26.7%	12.7%
West-northwesterly Flow	24.8%	46.0%
Local Recirculation	19.8%	38.1%

Table 12: Percentages of profiles within each cluster that fell within each Airmass Origin Category, based on HYSPLIT back trajectory clusters, for the Colorado campaign O_3 clusters. Clusters listed from left to right in order of least to most polluted. HYSPLIT back trajectories initiated at 2500 m vertical level only.

Maryland Campaign O_3 Correlations		
Cluster Number	**Column_Air R^2**	**Column_Ground R^2**
1	0.53	0.57
2	0.87	0.88
3	*0.24*	*0.34*
4	N.S.	N.S.
5	*0.14*	*0.23*
6	*N.S.*	*0.36*
Full	0.71	0.74

Table 13: R^2 values for the correlation between column and surface data for each O_3 profile cluster of the Maryland campaign. Cluster correlations denoted in italic font indicate clusters that were statistically significantly different from the full correlation.

California Campaign O_3 Correlations		
Cluster Number	**Column_Air R^2**	**Column_Ground R^2**
1	--	--
2	N.S.	N.S.
3	0.15	0.19
Full	0.08	0.11

Table 14: R^2 values for the correlation between column and surface data for each O_3 profile cluster of the California campaign. Cluster correlations denoted in red font indicate clusters that were statistically significantly different from the full correlation.

California And Texas Campaign NO_2 Correlations	
Cluster Number	**Column_Air R^2**
California #2	0.21
California Full	0.11
Texas #4	0.21
Texas Full	0.53

Table 15: R^2 values for the correlation between column and surface data for each NO_2 profile cluster of the California campaign. Cluster correlations denoted in red font indicate clusters that were statistically significantly different from the full correlation.