Cluster-based characterization of multi-dimensional tropospheric ozone variability in coastal regions: an analysis of lidar measurements and model results

Claudia Bernier[1], Yuxuan Wang[1], Guillaume Gronoff[2,3], Timothy Berkoff[2], K. Emma Knowland[4,5], John T. Sullivan[4], Ruben Delgado[6,7], Vanessa Caicedo[6,7], Brian Carroll[2,6]

[1] Department of Earth and Atmospheric Science, University of Houston, Houston, Texas, USA

[2] NASA Langley Research Center, Hampton, VA, 23666, USA

[3] Science Systems and Application Inc., Hampton, VA, 23666, USA

[4] NASA Goddard Space Flight Center, Global Modeling and Assimilation Office, Greenbelt, MD, 20771, USA

[5] Morgan State University, Goddard Earth Science Technology & Research (GESTAR) II, Baltimore, Maryland, USA

[6] Joint Center for Earth Systems Technology, Baltimore, MD, USA

[7] University of Maryland Baltimore County, Baltimore, MD, USA

Correspondence: Yuxuan Wang (ywang246@central.uh.edu)

Abstract. Coastal regions are susceptible to multiple complex dynamic and chemical mechanisms and emission sources that lead to frequently observed large tropospheric ozone variations. These large ozone variations occur on a meso-scale which have proven to be arduous to simulate using chemical transport models (CTMs). We present a clustering analysis of multi-dimensional measurements from ozone Light Detection And Ranging (LiDAR) in conjunction with both an offline GEOS-Chem CTM simulation and the online GEOS-Chem simulation GEOS-CF, to investigate the vertical and temporal variability of coastal ozone during three recent air quality campaigns: 2017 Ozone Water-Land Environmental Transition Study (OWLETS)-1, 2018 OWLETS-2, and 2018 Long Island Sound Tropospheric Ozone Study (LISTOS). We developed and tested a clustering method that resulted in 5 ozone profile curtain clusters. The established 5 clusters all varied significantly in ozone magnitude vertically and temporally which allowed us to characterize the coastal ozone behavior. The lidar clusters provided a simplified way to evaluate the two CTMs for their performance of diverse coastal ozone cases. An overall evaluation of the models reveals good agreement ($R \approx 0.70$) in the low-level altitude range (0 to 2000 m), with a low and unsystematic bias for GEOS-Chem and high systemic positive bias for GEOS-CF. The mid-level (2000 – 4000 m) performances show a high systematic negative bias for GEOS-Chem and an overall low unsystematic bias for GEOS-CF and a generally weak agreement to the lidar observations ($R = 0.12$ and 0.22, respectively). Evaluating cluster-by-cluster model performance reveals additional model insight that is overlooked in the overall model performance. Utilizing the full vertical and diurnal ozone distribution information specific to lidar measurements, this work provides new insights on model proficiency in complex coastal regions.

32

1. Introduction

Tropospheric ozone (O₃) is an important secondary pollutant created by multiple reactions involving sunlight, nitrogen oxides ($NO_x = NO + NO_2$), and volatile organic compounds (VOCs) which, in accumulation, can have damaging effects on human and plant health. In addition to its photochemical growth, O₃ can easily be influenced by local and regional transport mechanisms. For coastal regions, surface O₃ is highly variable in time and space due to its susceptibility to many factors such as local ship emissions, long range transport, and sea/bay breeze processes. This variability is challenging for air quality models to capture as high-resolution measurements are necessary to fully understand and simulate this O₃ behavior in coastal regions.

For example, Dreessen et al. (2019) tested the U.S. Environmental Protection Agency (EPA) Community Multiscale Air Quality (CMAQ) model's ability, configured at 12 km, to simulate O₃ exceedances at Hart Miller Island in Maryland (HMI) revealing high bias and 'false alarms' due to several reasons such as emission transport over water and the coarse model resolution's inability to capture fine-scale meteorology and transport. Multiple studies have proven the strong influence that sea/bay breeze and wind flow patterns can have on the accumulation of coastal O₃ and can often lead to poor air quality (e.g., Tucker et al., 2010; Martins et al., 2012; Stauffer et al., 2012; Li et al., 2020). Cases such as sea/bay breeze events, which directly contribute to high coastal O₃ cases, are denoted by local meteorological mechanisms such as surface wind speed deceleration, wind direction convergence and recirculation (Banta et al., 2005). Loughner et al. (2014) also highlighted the importance of understanding the ability bay breeze events have in O₃ variability not only spatially but vertically throughout the atmosphere. Air quality models with coarse horizontal and vertical resolutions are not able to capture such fine developments (Caicedo et al., 2019). Ring et al. (2018) also used CMAQ to estimate the impact of ship emissions on the air quality in eastern U.S. coastal regions indicating that an understanding of the vertical profiles of emissions was significant for improving air quality simulations. These are consistent and unanimous issues with air quality modeling in coastal regions. Since offshore sites within coastal regions are historically under sampled due to the difficulty of water-based measurements, this problem is still pertinent today.

Recently, three associated air quality campaigns set out to address this issue (https://www-air.larc.nasa.gov/index.html): 2017 & 2018 NASA Ozone Water-Land Environmental Transition Study (OWLETS-1 & OWLETS-2) and Long Island Sound Tropospheric Ozone Study (LISTOS) (e.g., Sullivan et al., 2019). These three campaigns were each conducted in highly populated coastal regions along the Chesapeake Bay in Virginia and Maryland and the Long Island Sound in the New England/Middle Atlantic region, that are vulnerable to O₃ exceedances with the goal of filling the measurement gaps in these regions. During these campaigns, a suite of detailed airborne and ground measurements were taken during the course of highly polluted summer months (end of May through August) to capture the variability of pollutants, including O₃ and its precursor species, and the distinct meteorological processes specific to land-water regions that affect them.

The three campaigns strategically placed multi-dimensional tropospheric O₃ lidar instruments on and offshore in order to capture critical land-water gradients and to fill the deficit of measurements in these under monitored areas. These measurements were supported as part of NASA's Tropospheric Ozone Lidar Network (TOLNet). Continuous profile

66 measurements from O₃ lidars highlight important regional transport and temporal variations of O_3 in the lower and middle
67 levels of the troposphere that are usually difficult to capture by most satellite-based remote-sensing instruments (Thompson et
68 al., 2014). Lidar instruments are unique in their ability to capture high resolution full O_3 2-D profile curtains over a period of
69 time that can help in understanding O_3 behavior in coastal regions. In Gronoff et al. (2019), the co-located lidar at the
70 Chesapeake Bay Tunnel Bridge (CBBT) during OWLETS-1 successfully captured a near-surface maritime ship plume
71 emission event on 01 August 2017. An ensemble of other instruments (e.g., drones, Pandora spectrometer systems, etc.)
72 launched near the shipping channel captured elevated NO_2 concentrations while the lidar instrument captured a depletion of
73 O_3 simultaneously. The lidar was able to capture the unique low range altitude O_3 concentrations which elucidated the evolution
74 of the trace-gas concentrations during this ship plume event.

75 Several studies have thoroughly evaluated the results from the air quality campaigns used in this study but were focused
76 more on specific case studies (Dacic et al., 2019; Sullivan et al., 2019; Gronoff et al., 2019). Dacic et al. (2019) used lidar
77 measurements of a high O_3 episode during OWLETS-1 to evaluate the ability of two NASA coupled chemistry-meteorology
78 models (CCMMs), the GEOS Composition Forecast ("GEOS-CF"; Keller et al., 2021) and MERRA2-GMI (Strode et al.,
79 2019), to simulate this high O_3 event. They found that the GEOS-CF model performed fairly in simulating O_3 in the lower
80 level (between 400 to 2000 m ASL) and outperformed MERRA2-GMI based on surface observations at multiple monitoring
81 sites. In the case of this event, GEOS-CF was able to simulate the 2-D O_3 profile curtains at small scales. At the time of the
82 Dacic et al. (2019) study, processed observational data was only available from OWLETS-1.

83 For this study, we took advantage of measured 2-D (vertical and diurnal) O_3 profile curtains from all three air quality
84 campaigns (Sect. 2). To characterize the different behaviors of O_3 in coastal regions, we developed a novel clustering method
85 based on the altitude and time dimensions of the lidar measurements that organized the profile curtains (Sect. 2). We used the
86 developed clusters to evaluate the ability of both offline and online GEOS-Chem and GEOS-CF simulations to reproduce the
87 coastal O_3 and wind characteristics highlighted by each cluster (Sect. 3).

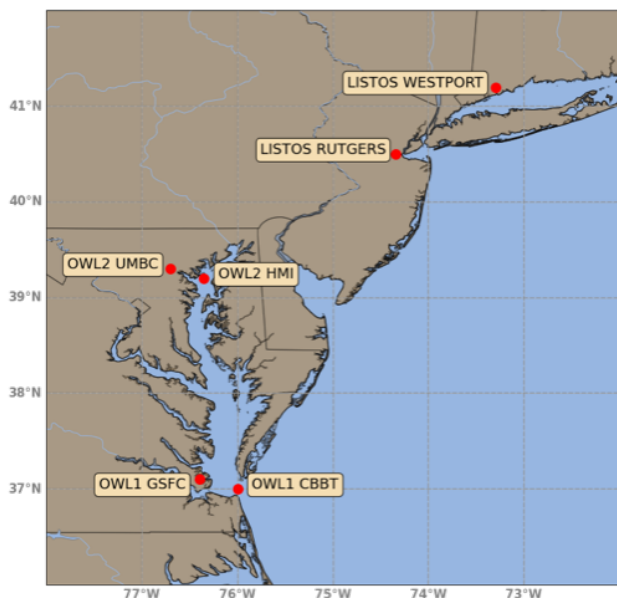
88

89 **2. Materials & Method**

90 **2.1. Air quality campaigns**

91 During the years 2017 and 2018, NASA in partnership with other U.S. national agencies and university research groups
92 orchestrated three air quality campaign studies that focused on key land and water observations: OWLETS-1, OWLETS-2,
93 and LISTOS. OWLETS-1 was conducted in 2017 from July 5 to August 3, while OWLETS-2 and LISTOS were conducted in
94 2018 from June 6 to July 6 and July 12 to August 29, respectively. All campaigns took advantage of a multitude of ground,
95 aircraft, and remote sensing measurements. For the sake of this study, we will focus on measurements from the two lidars from
96 the TOLNet: NASA Langley Mobile Ozone Lidar (LMOL) (De Young et al. 2017; Farris et al. 2018; Gronoff et al, 2019,
97 2021) and NASA Goddard Space Flight Center (GSFC) Tropospheric Ozone (TROPOZ) Differential Absorption Lidar (DIAL)
98 (Sullivan et al. 2014, 2015a), which ran simultaneously at the marked positions in Figure 1. The TOLNet data from all three

99 campaigns are available on the NASA LaRC Airborne Science Data for Atmospheric Composition archive (https://www-
100 air.larc.nasa.gov/missions.htm; accessed – 20 January 2021).

101

102 **Figure 1.** An inset map of the Chesapeake Bay airshed in Maryland, Virginia, and Long Island Sound in New York with the
103 six lidar monitoring locations used for OWLETS-1, OWLETS-2, and LISTOS highlighted and labeled.

104 The two lidars were placed strategically for each campaign (Figure 1), so that one lidar was closest to over-water
105 measurements while the other was farther inland with the goal of examining how O_3 transport and concentration is influenced
106 by specific coastal mechanisms such as the land–water breezes. For OWLETS-1, the LMOL lidar was used at the CBBT
107 [37.0366°N, 76.0767°W], depicting the real time over water O_3 measurements while the GSFC TROPOZ lidar was stationed
108 at NASA Langley Center [37.1024°N, 76.3929°W] further inland. Similarly, for OWLETS-2, the LMOL lidar was stationed
109 for the over water measurements at Hart Miller Island [39.2449° N, 76.3583° W] and GSFC TROPOZ was stationed at the
110 University of Maryland, Baltimore County (UMBC) [39.2557° N, 76.7111° W]. For LISTOS, LMOL was at the Westport site
111 [41.1415° N, 73.3579° W] and TROPOZ at Rutgers [40.2823° N, 74.2525° W]. For the sake of this study the unique benefits
112 due to the different placements (onshore versus offshore) of the co-located lidars are not specifically evaluated. Instead, the
113 study focuses on the benefits of the detail and multi-dimensionality of lidar instrument data in general.

114 Routine lidar measurements were taken for the duration of the campaigns. Both lidars retrieve data at a 5-min temporal
115 resolution and use a common processing scheme to produce a final O_3 product which was used for this study. In this study, the
116 individual profile curtains refer to the "full day", vertical and diurnal lidar measurements. In this study, 91 individual 2-D

117   profile curtains were used from both lidars from the three campaigns: 26 profile curtains from OWLETS-1, 28 profile curtains

118   from OWLETS-2, and 37 profile curtains from LISTOS.

119   To evaluate meteorological impacts on the lidar O_3 clusters and model performance we used various temperature and

120   wind measurements. Hourly observed temperature, wind speed and wind direction, and O_3 from surface monitors pertaining

121   to the study area were obtained from the Air Quality System (AQS) (data can be accessed at

122   https://aqs.epa.gov/aqsweb/airdata/). We utilized high resolution vertical and horizontal wind speed and direction data

123   monitored by Doppler wind lidar Leosphere WINDCUBE 200s instruments deployed at HMI during OWLETS-2 during

124   LISTOS (e.g., Couillard et al., 2021; Coggon et al., 2021; Wu et al., 2021).

125

126   **2.2. Clustering lidar data**

127   **2.2.1 Description of the ozone lidar measurements**

128   The lidar instrument is unique in that it provides high dimensional profile measurements of O_3, as opposed to one

129   dimensional surface measurements from air quality monitoring sites. The two TOLNet lidars used during the campaigns have

130   been evaluated for their accuracy during previous air quality campaigns (DISCOVER-AQ; https://www-

131   air.larc.nasa.gov/missions/discover-aq and FRAPPÉ; https://www2.acom.ucar.edu/frappe) and have also been compared

132   against each other (e.g., Sullivan et al., 2015; Wang et al., 2017). The two lidars have different transmitter and retrieval

133   components but produce O_3 profiles within 10 % of each other as well as compared to ozonesondes (Sullivan et al., 2015). In

134   comparison with other in situ instrument measurements, the TOLNet lidars were found to have an accuracy better than ±15 %

135   for capturing high temporal tropospheric O_3 vertically proving their capability of capturing high temporal tropospheric O_3

136   variability (Wang et al., 2017; Leblanc et al., 2018).

137   To characterize coastal O_3 during the summer months, we use a multitude of lidar profile curtains obtained during the

138   OWLETS-1, 2, and LISTOS campaigns. The two lidars used in the campaigns produced O_3 profile curtains from 0 – 6000 m

139   above ground level (AGL) with some days beginning as early as 06:00 local time (EDT) and ending measurements as late as

140   the last hour of the day. One of the challenges is that the multiple lidar datasets are not always uniform; although most of the

141   profile curtains began at or around 08:00 EDT, the lidar measurements commence and conclude at different times. At the time

142   of these campaigns, the lidar data retrieval was constrained by the availability of personnel as well as the availability of

143   electricity in remote areas. Due to this constraint, the 91 lidar curtains range from as short as a 6-hour window to a full 24-

144   hour window. Similarly, the profile curtains do not have an exact uniform altitude range either. In the processing of the lidar

145   data, some measurements may be filtered out and removed due to issues, such as clouds, which can influence and degrade the

146   retrieval leaving some blocks of empty data within the vertical altitude dimension. When the cloud conditions are perfect, the

147   limiting factor for the altitude is the solar background: the UV from the sun is a source of noise that prevents the detection of

148   the low level of backscattered photons. For LMOL, this means that the maximum altitude is about 10 km AGL at night (Gronoff

149   et al., 2021) and lowered to about 4 km AGL at solar noon (worse conditions possible for the summer in the continental U.S.

150   resulting in below 4 km AGL). This results in a general scarcity of O_3 measurements above 4000 m AGL for most of the

151 vertical profile curtains. Lidars still have limitations that prove to be a complication e.g., noise signal and manual operations.

152 At the time of writing, the operative limitation has been addressed and the lidars are now more fully automatized for use during

153 succeeding campaigns removing such constraints.

154

155 2.2.2 Clustering approach and application

156 To characterize coastal O_3, we used a cluster analysis to categorize the behavior of the tropospheric O_3 captured in the

157 profile curtains. Clustering methods are commonly used in air quality and atmospheric studies to group and characterize large

158 datasets (Darby, 2005; Alonso et al., 2006; Christiansen, 2007; Davis et al., 2010; Stauffer et al., 2018). In our previous work,

159 we have successfully used clustering methods to automatically characterize diurnal patterns of surface winds and surface O_3

160 in the Houston-Galveston-Brazoria area that proved to perform better than a rudimentary quantile method to reveal the

161 dependence of surface O_3 variability on local and synoptic circulation patterns on the Gulf Coast (Bernier et al., 2019; Li et

162 al., 2020).

163 In evaluating the structure of the lidar measurements and working within measurement limitations (described in Sect.

164 2.2.1) from the three air quality campaigns, we developed a method to cluster multi-dimensional O_3 profile curtains using K-

165 Means clustering algorithm. Input features (seed values) were rationally established to best represent the behavior of O_3

166 temporally and vertically without including an excessive amount of input features, which can weaken the results of clustering

167 (discussed in detail in Sect. S1, in Supplementary Material). With the goal of evaluating lower level tropospheric O_3 and based

168 on description of the structure and constraints of the lidar measurements, the features were tailored to the altitude range 0 –

169 4000 m AGL and time range of 06:00 EDT – 21:00 EDT.

170 Figure 2 illustrates the 8 features that represent the slabs of altitude and time used in the cluster analysis. For each O_3

171 profile curtain (total of 91), we calculated the average O_3 from the following time and altitude range: Features 1 – 4 altitudes

172 range from 0 – 2000 m; Features 5 – 8 altitudes range from 2000 – 4000 m. The two altitude ranges were determined to best

173 represent different O_3 transport events although they do not explicitly represent these layers. For Features 1 – 4, O_3 would most

174 likely primarily be affected by local production and pollution transport while for Features 5 – 8, O_3 would more likely be

175 associated with long range transport (e.g., interstate). As planetary boundary layer growth (PBL) in coastal regions do not

176 usually reach altitudes greater than 2000 m, mixing between the boundary layer and free troposphere would presumably take

177 place within the low-level altitude bin. Additional attention to the PBL in the selecting of low versus mid-level features for the

178 clustering will be investigated in future work. For clarity, we will use the terms low-level and mid-level features to address the

179 two altitude subsets e.g., Features 1 – 4 and 5 – 8, respectively. Feature 1 and 5 time range from 06:00 – 08:00 EDT; Feature

180 2 and 6 from 08:00 – 12:00 EDT; Feature 3 and 7 from 12:00 – 16:00 EDT; and Feature 4 and 8 from 16:00 – 21:00 EDT. The

181 four subset time ranges were indicated to best represent features that characterize the common diurnal behavior of O_3.

182

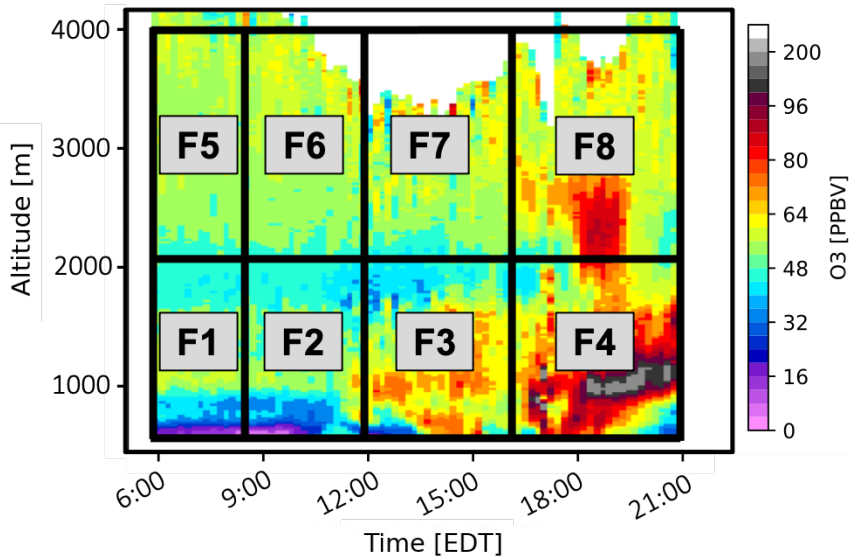



Figure 2. Clustering method developed for clustering vertical O_3 profiles taken from lidar measurements. The color coding
shows a typical day of lidar measurements of O_3 profiles on 6 August 2018, from the LMOL at Westport, CT during the
LISTOS Campaign. F1 – F8 indicate the time and altitude range of the eight features used for the clustering algorithm.

186

187   The features were evaluated for cluster tendency, essentially to confirm our dataset contained meaningful clusters
(discussed in detail in Sect. S2). Evaluating different feature options did not lead to better statistical results than with the final
chosen features. Since the choice of clustering algorithm is subjective, we chose K-means clustering for its simplicity and
widespread use. To use the K-Means clustering algorithm, the optimal number of clusters based on your dataset must be chosen
beforehand (Sect. S2). We selected six clusters as the optimal number of clusters. Since the K-Means clustering algorithm is
based on the Euclidean distance to each centroid, the input data was normalized (to a mean of zero and standard deviation of
one) to ensure each feature is given the same importance in the clustering (Aksoy & Haralick, 2001; Larose, 2005).

194   The clustering analysis initially identified six clusters (described fully in Sect. 3.2). Only one date was assigned to Cluster
6 (16 June 2018): the lidar profile curtain on this day (Figure S1) shows a large fraction of data missing, and the available data
have relatively high O_3 throughout the lowest 3 km, which is different from other clusters. Therefore, we consider Cluster 6 to
be an outlier and will not be included in the subsequent analysis.

198

2.2.3 Missing data

200   Although the input features were tailored based on the structure of the lidar measurements, the remaining data still had
missing data points. In performing a quick evaluation on the 8 input features (Figure S6), we found that Features 1, 4, 5, and
8 had the most missing data while Features 2, 3, 6, and 7 had few or zero cases of missing data. This means that the earlier
morning measurements (06:00 – 12:00 EDT) and the later evening measurements (16:00 – 21:00 EDT) had the most cases of

204 missing data points. This is plausible as the campaign teams were best able to retrieve clear measurement during
205 midday/evening hours (12:00 – 16:00 EDT). As a result, 51 out of 91 O_3 profile curtains had at least one missing data point
206 (feature) throughout the individual profile curtain.

207   A common practice for dealing with missing data is complete case analysis (CCA), in which observations with missing
208 values are completely ignored, leaving only the complete data to cluster. CCA can be inefficient as it introduces selection bias
209 since the sample data no longer retains the state of the original full dataset (Donders et al., 2006; Little & Rubin, 2014). When
210 we applied CCA, there were only 40 O_3 profile curtains of complete data, removing over half of the study profiles. Instead,
211 we used a more comprehensive solution – imputation - that yields results (Donders et al., 2006). For this study we used the
212 single imputation (SI) technique, *knnImputation,* which uses the k-nearest neighbors and searches for the most similar cases
213 and uses the weighted average of the values of those neighbors to fill the missing data (Torgo, 2010). Essentially, this method
214 selects the days that have the most similar profile curtain to any profile which has missing data points and uses those real data
215 points to calculate a weighted mean that will fill in the missing data. We acknowledge using an imputation method on the
216 dataset will possibly introduce a bias which is difficult to quantify, but this allows us to utilize all 91 O_3 profile curtains. The
217 silhouette method was used to test the quality of the newly imputed dataset and proved to be no worse, nor better, than the
218 CCA (*real data*) results. Therefore, the dataset was first imputed using SI to create a complete dataset and then the clustering
219 method described in the sect. before (2.2.2) was applied to the complete imputed dataset.

220

221 **2.3. Model simulations**

222   The offline GEOS-Chem chemical-transport model (CTM) was utilized to simulate the spatial and temporal variability
223 of coastal O_3 in the Chesapeake Bay and Long Island Sound during the time of the campaigns. The GEOS-Chem model is a
224 global 3-D CTM driven by assimilated meteorological data from the NASA Global Modeling and Assimilation Office
225 (GMAO). Our simulations were driven by reanalysis data from Modern-Era Retrospective analysis for Research and
226 Applications, Version 2 (MERRA-2; Gelaro et al., 2017). We ran a nested GEOS-Chem (v12-09) simulation at 0.5° x 0.625°
227 horizontal resolution over the eastern portion of North America and adjacent ocean (90 – 60°W, 20 – 50°N), using lateral
228 boundary conditions updated every three hours from a global simulation with 2° x 2.5° horizontal resolution. The nested
229 GEOS-Chem simulation was run with 72 vertical levels from 1013 to 0.01 hPa. Since the study focuses on the altitude range
230 0 – 4000 m, the first 20 vertical levels from GEOS-Chem were used with 14 levels within the boundary layer ($\leq$ 2000 m). The
231 nested simulation was conducted for the study periods June – September 2017 and April – August 2018. We used the standard
232 "out-of-the-box" unmodified default settings from the tropospheric chemistry chemical mechanism (tropchem) with global
233 anthropogenic emissions from the Community Emissions Data System (CEDS) inventory (McDuffie et al, 2020) and U.S.
234 Environmental Protection Agency (EPA) National Emissions Inventory (NEI) 2011 for monthly mean North American
235 regional emissions (EPA NEI, 2015).

236   We also used results from NASA's near real-time forecasting system, GEOS-CF, an online GEOS-Chem simulation (v12-
237 0-1) from GMAO (https://gmao.gsfc.nasa.gov/-weather_prediction/GEOS-CF/) with GEOS coupled to the GEOS-Chem

tropospheric-stratospheric unified chemistry extension (UCX) and run at a high spatial resolution of 0.25°, roughly 25 km (Keller et al., 2021, Knowland et al., 2021). The vertical resolution for GEOS-CF is interpolated onto 72 vertical levels from 1000 to 10 hPa. Since the study focuses on the altitude range 0 – 4000 m, the first 21 vertical levels from GEOS-CF were used with 14 levels within the boundary layer ($\leq$ 2000 m). Prior to the launch of the 12z five-day forecast, GEOS-CF produces daily global, 3-D atmospheric composition distributions using the GEOS meteorological replay technique (Orbe et al., 2017), and this study makes use of these historical estimates, made available to the public for the period since January 2018. Therefore, the GEOS-CF results shown in this study only include the dates from OWLETS-2 and LISTOS campaigns, since they both occurred in 2018.

While both model simulations use similar versions of GEOS-Chem chemistry, there are noteworthy differences to keep in mind during the analysis of the clustering. The main differences between the two models are (1) GEOS-Chem is an offline CTM using archived meteorology, while GEOS-CF simulates atmospheric composition simultaneously with meteorology (online); (2) the spatial resolution of the GEOS-CF model (0.25°) is higher than GEOS-Chem (0.5° x 0.625°); and (3) the GEOS-CF model runs with Harmonized Gridded Air Pollution (HTAP; v2.2; base year 2010) anthropogenic emissions from the Emission Database for Global Atmospheric Research (EDGAR), while GEOS-Chem was run with CEDS anthropogenic emissions (base year 2014). These imperative differences can lead to disparities in the following results.

3. Results & Discussion

3.1 Overview of the 2-D O_3 curtain clusters

The clustering results reveal distinctive characterized O_3 behavior during the three campaigns in which O_3 concentrations vary. Various O_3 and surface meteorological parameter cluster statistics for the five clusters are summarized in Table 1. With only 5 of the 2-D profile curtains assigned, Cluster 5 depicts the least common O_3 behavior during the campaigns. On the other hand, Cluster 3 is the most common O_3 behavior during the campaigns with 28 profile curtains assigned to this cluster. Following Cluster 3, Cluster 1 is the next most common cluster with 25 profile curtains. Cluster 2 and Cluster 4 fall in the middle with 14 and 18 profile curtains assigned to the cluster numbers, respectively.

Cluster #	a) No. of vertical profiles	b) O_3 Max (ppb)	c) O_3 Min (ppb)	d) T avg. (min; max) (°F)	e) WS avg. (min; max) (m s^{-1})
1	25	86.5	42.2	74.1 (67.8; 86.4)	1.5 (0.5; 2.8)
2	14	72.8	28.9	71.6 (64.0; 83.9)	1.6 (0.6; 2.9)
3	28	86.6	34.2	77.2 (67.0; 87.6)	1.3 (0.5; 2.4)
4	18	97.8	44.1	78.4 (68.0; 90.4)	1.2 (0.4; 2.3)
5	5	67.7	29.1	74.5 (66.8; 74.5)	1.2 (0.3; 3.4)

Table 1. Lidar vertical O$_3$ profile cluster statistics: a) total number of vertical profiles; b) O$_3$ maximum; c) O$_3$ minimum O$_3$;) AQS monitoring station cluster mean d) surface temperature and e) wind speed; minimum and maximums in parenthesis. The statistics and averages were derived from the total number of profile curtains assigned to each cluster.

The five clusters were distinguished by the varying O$_3$ concentrations between the low-level and mid-level as well as diurnal variations (Figure 3). In Figure 3a we separate the data by the two altitude subsets (low and mid-level) and by morning (06:00 – 12:00) and afternoon (12:00 – 21:00) to quantify the between-cluster differences. In the low-level, all five clusters exhibit the common O$_3$ diurnal pattern where surface O$_3$ is titrated overnight and reaches a minimum but then is quickly exacerbated with the increase of sunlight throughout the day and typically peaks after midday (Figure 3b). The extent of this common diurnal pattern varies by cluster.

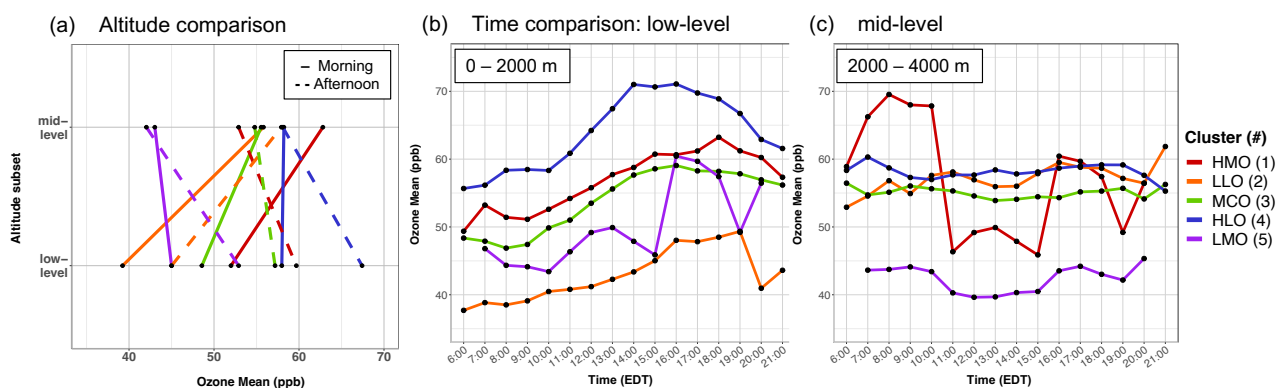



Figure 3. Lidar O$_3$ cluster average comparisons (fiv[...]ean O$_3$ averaged over time: morning hours from 6:00 – 12:00 (solid line) and afternoon hours from 12:00 – 21:00 (dashed lines). Time comparison of mean hourly O$_3$ split between the b) low-level and c) mid-level.

Cluster 1 in the low-level has the second highest morning and afternoon O$_3$ average (52 and 59 ppb) and in the mid-level the highest morning O$_3$ average (64 ppb) (Figure 3a). Cluster 1 also exhibits the most unique pattern of mid-level O$_3$ (Figure 3c), with the highest concentrations found in the early morning and an uncharacteristic plunge to lower O$_3$ concentrations from 11:00 – 15:00 EDT. This is contrary to the other clusters which do not show much O$_3$ variation temporally in the mid-level. The majority of the individual profile curtains assigned to Cluster 1 show concentrated early morning residual layers in the mid-level that diffuse after the morning, which is distinctive to the other clusters. In the low-level, Cluster 2 has the lowest morning and afternoon O$_3$ average among the clusters (39 and 45 ppb) with moderate mid-level O$_3$ concentrations. Cluster 3 has the most uniform vertical O$_3$ extent between the low and mid-level (Figure 3a), in contrast to the other clusters that differ greatly in O$_3$ concentrations between the two altitude subsets. Cluster 4 has the highest morning and afternoon O$_3$ averages (59 and 68 ppb) in the low-level, reaching > 70 ppb temporally (Figure 3b). Finally, Cluster 5 has, considerably, the lowest morning

292 and afternoon O₃ averages (42 and 43 ppb) in the mid-level, almost 10 ppb lower than the other clusters. Cluster 5 does not

293 have a smooth-evolving O₃ diurnal pattern in the lower level (Figure 3b), which can be attributed to the averaging of only five

294 different profile curtains that were assigned to this cluster (Table 1).

295   Figure 4a illustrates the mean lidar O₃ 2-D profile curtains for each of the clusters. For Cluster 1, 3, 4, and 5, higher O₃

296 concentrations in the low-level are captured during afternoon/evening time (12:00 – 21:00 EDT), with the highest low-level

297 O₃ in Cluster 4 (> 70 ppb). This behavior follows the common diurnal pattern of O₃, that was distinguishable in Figure 3b. This

298 common O₃ growth reaches vertically to approximately 1500 m for each of the clusters but is generally contained below 2000

299 m. Differing from the low-level O₃ behavior, mid-level O₃ is generally less variable in magnitude throughout the entire profile

300 curtain (except for Cluster 1; see Figure 3a). The highest O₃ concentrations for the mid-level are exhibited in Cluster 1, 2, 3,

301 and 4, with the highest mid-level O₃ in Cluster 1 during the early morning hours (≥ 70 ppb).

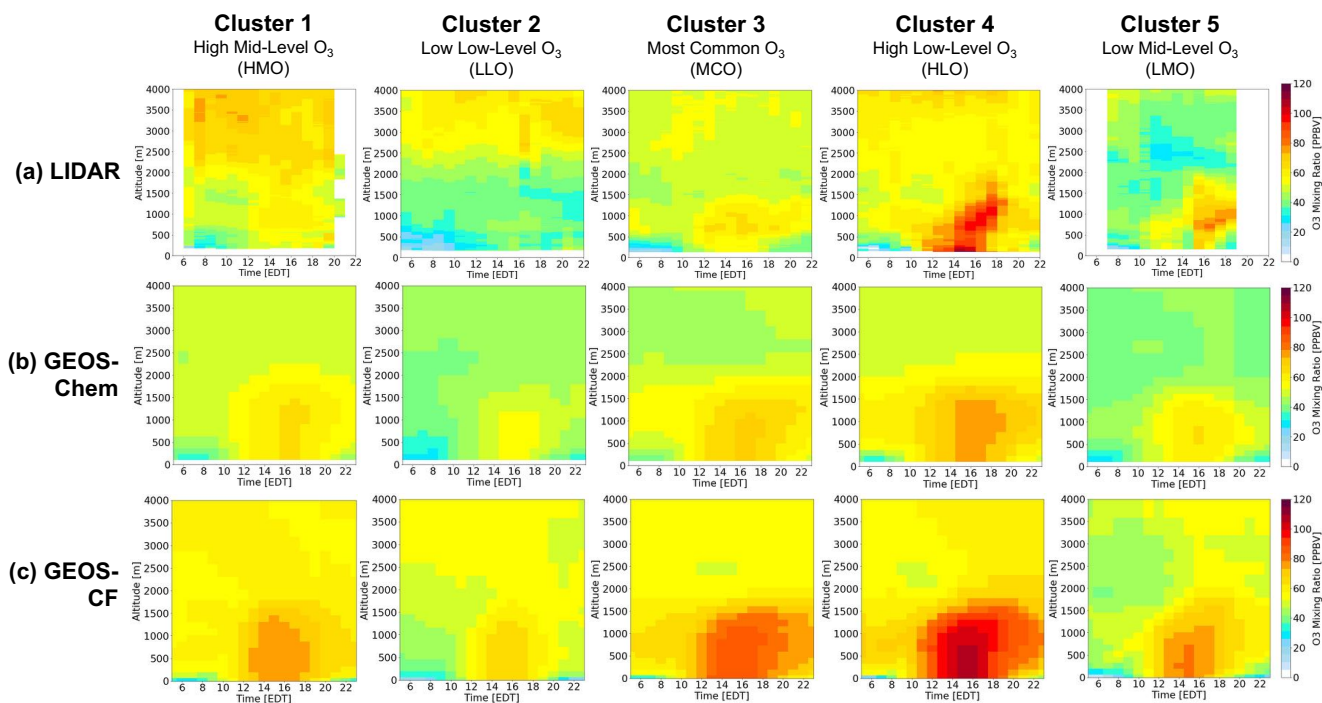
302   Following the descriptions above, each cluster is given a nomenclature according to their unique characteristics. Cluster

303 1 is termed as the highest mid-level O₃ (HMO) cluster; Cluster 2 as the lowest low-level O₃ (LLO) cluster; Cluster 3 is the

304 most common O₃ (MCO) cluster; Cluster 4 is the highest low-level O₃ (HLO); Cluster 5 is the least common and lowest mid-

305 level O₃ (LMO) cluster. The O₃ variability represented and justified above is what led to the successful clustering of the lidar

306 O₃ 2-D profile curtains.

307

309 **Figure 4.** Cluster-mean O₃ vertical profile results by cluster assignment (1- 5) and arranged: a) LIDAR; b) GEOS-Chem

310 simulation; and c) GEOS-CF simulation.

311

312   Figure 3b and 3c indicate each cluster represents a different O_3 evolution pattern, likely related to different photochemical

313 or transport regimes. This kind of evaluation is useful in that it combines O_3 information from both temporal and vertical

314 dimensions. For example, the HLO cluster reveals a unique low-level case in which high O_3 concentrations at a high elevation

315 (~1000 m) are captured early in the temporal profile that translate to the higher O_3 concentrations at the surface later in the

316 evening. The mean profile curtain indicates these cases did not have "clean air" to begin with which can allow a greater

317 accumulation in the low-level in the afternoon. In another example, several profile curtains assigned to the HMO cluster

318 indicate concentrated residual layers in the mid-level and possible entrainment to the surface as the day progressed. To prove

319 this feature, vertical velocity and vertical velocity variance data would be needed but the knowledge that a clustering approach

320 is able to highlight these features that could only be discernible through lidar measurements proves to be useful. The clustering

321 results were valuable in recognizing a significant large pollution related cluster (HLO), a total of 18 out of the 91 curtain

322 profiles which correspond with the highest daily surface maxima measured at these sites (= 97.8 ppb) (Table 1). This cluster,

323 on average, exhibited a daily surface maxima up to 10 ppb greater than any of the other clusters. Discerning these higher O_3

324 cases is imperative for mitigating severe air pollution.

325

326 **3.2. Cluster surface analysis**

327   To support the lidar clustering results, daily averaged meteorological surface observations from AQS stations nearest to

328 the lidar locations pertaining to the campaign period and GEOS-Chem surface model output were evaluated in regard to the

329 five clusters. Figure 5 shows the cluster mean surface temperature from AQS stations and GEOS-Chem model as well as the

330 simulated wind speed and direction. The average surface temperature from each station is represented as the circular markers

331 while the simulated temperatures are represented as the spatial contour and the simulated wind speed (m s^{-1}) and direction as

332 arrows. Cluster average, minimum, and maximum AQS surface temperature and wind speed can be found in Table 1d, e.

333

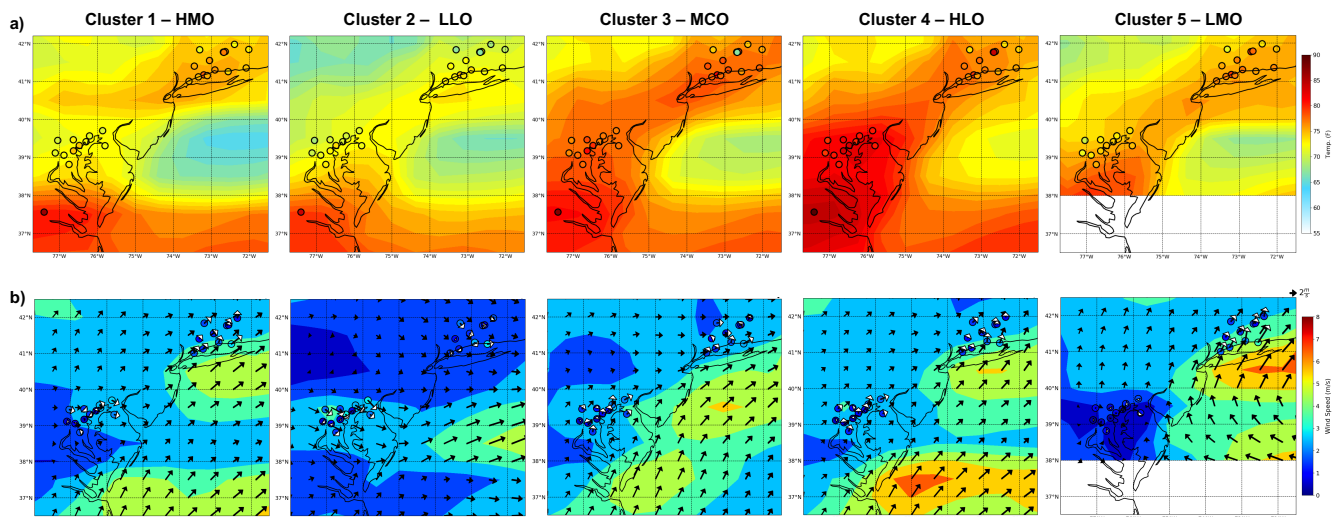



Figure 5. Cluster averaged meteorological surface AQS station observations and GEOS-Chem model results. a) Surface temperature observations represented as the circular markers and simulated surface temperatures represented as the spatial contour (top-panel). b) Surface wind speed and direction observations represented as the circular markers and white arrows and simulated wind speed and direction represented as spatial contour and black arrows (bottom-panel).

In general, the surface meteorological conditions agree with our knowledge of transport and O_3 production that would lead to each of the five clustered lidar O_3 profile curtains. It is evident that the clusters with the highest surface O_3 (HMO, MCO, and HLO) all share a predominant offshore, westerly wind. Furthermore, MCO and HLO presented higher overall observed and simulated surface temperatures compared to the other clusters (Figure 5a). These meteorological conditions are conducive to a higher production of surface O_3 concentrations which validates the higher O_3 found in the low-level results (Figure 3b, 4a).

Conversely, the lowest surface temperatures are found in LLO. Lower surface temperatures are also indicative of low vertical mixing due to less generation of convection which can reduce any possible descending O_3 from aloft. Relatively calm wind speeds, lower temperatures, and other possible meteorological factors such as high cloud cover could have contributed to the lower O_3 concentrations in LLO. Although surface O_3 concentrations in LMO reach higher levels later in the day, first at 13:00 EDT and then again at 16:00 EDT, the rest of the temporal profile stays below moderate levels. Average temperatures for LMO are moderately high but, in contrast, the average wind speed is higher (specifically over the Long Island Sound) and unique to the other clusters, wind direction is predominantly onshore (Easterly – Southerly). This prevalent onshore flow indicates a transport of cleaner marine air which corroborates the lower surface O_3 levels. LMO did not have any profile curtains assigned from OWLETS-1 which is why data for the lower Chesapeake Bay area is not shown in Figure 5.

There was only one occurrence during the dates in which the lidar instruments were operating in which there was a recorded maximum daily 8-hour average (MDA8) O_3 exceedance (> 70 ppbv). This exceedance date is 25 May 2018 in which

357 3 AQS sites in the LISTOS region measured MDA8 O_3 of 73, 72, and 72 ppbv. This curtain profile was assigned to the HMO

358 cluster (Cluster 1), the cluster with high O_3 in the mid-level and moderate O_3 in the low-level and near the surface. Since the

359 AQS stations applied here were the nearest stations to the lidar instrument placements, the MDA8 O_3 captured by the AQS

360 stations do not necessarily reflect the high O_3 concentrations capture by the lidars near the surface.

361

362 **3.3. Evaluating the GEOS-Chem and GEOS-CF model**

363 In this sect. the model results from GEOS-Chem and GEOS-CF will be compared to the lidar data using the five lidar O_3

364 profile clusters discussed in Sect. 3.1. Both model results were sampled in an equal manner, in which we extracted the same

365 cluster date assignments from the lidar clusters and created mean vertical profiles based on the model results. This allowed us

366 to evaluate the model performance based on the five characterized O_3 lidar clusters. As mentioned previously, the GEOS-CF

367 simulation data is not available for 2017. Thus, the results shown subsequently will only include GEOS-CF results from 2018

368 (only dates from the OWLETS-2 and LISTOS campaigns). The GEOS-Chem simulation results include both years thus all

369 three campaign duration periods.

370

371 **3.3.1 Overall model performance**

372 Figure 4b and 4c depict the simulated cluster-mean O_3 profile curtains from GEOS-Chem and GEOS-CF, mirroring the

373 mean lidar profile curtains in Figure 4a. For all clusters in the low-level, both models simulate a consistent accumulation of

374 O_3 near the surface after 12:00 EDT, mirroring the O_3 common diurnal pattern depicted in mean lidar profile curtains in Figure

375 4a. However, the extent the models simulate is often higher in magnitude than the observations, specifically GEOS-CF

376 consistently predicting the accumulation at a higher magnitude than GEOS-Chem. In the mid-level, both models simulate

377 much less O_3 variability than what is captured in the lidar observations. Figure 4b and 4c clearly show how the models struggle

378 to reproduce any mid-level O_3 pattern or variability that is relayed in the lidar observations.

379 We first evaluate overall correlation and biases between the model and lidar data, disregarding the specific clusters. The

380 overall correlation between the models and the lidar data is evaluated by the two altitude subsets as the performances differ

381 considerably between low-level and mid-level for both GEOS-Chem (Figure S7a) and GEOS-CF (Figure S7b) (mean

382 normalized biases found in Table S1). For both models, overall low-level O_3 correlation rounds to 0.70, signifying a strong

383 relationship between the model simulations and the lidar observations (Figure S7 - top panel). This indicates that both models

384 can simulate the development and pattern of O_3 well in the low-level. Overall, GEOS-Chem performs well in simulating low-

385 level O_3 with a lower non-systematic normalized bias ranging from -0.10 to +0.13. Thus, based on the lower bias, GEOS-

386 Chem also fairs well simulating the magnitude of low-level O_3. Overall, GEOS-CF overestimates the magnitude of low-level

387 O_3 with a systematic high positive normalized bias ranging from +0.30 to +0.67. This consistently high bias reveals that GEOS-

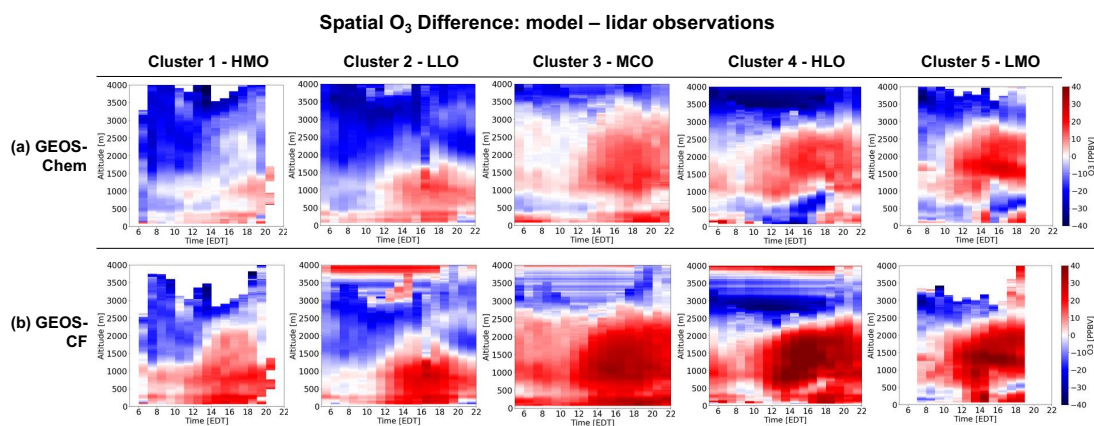
388 CF generally struggles to simulate low-level O_3 magnitude.

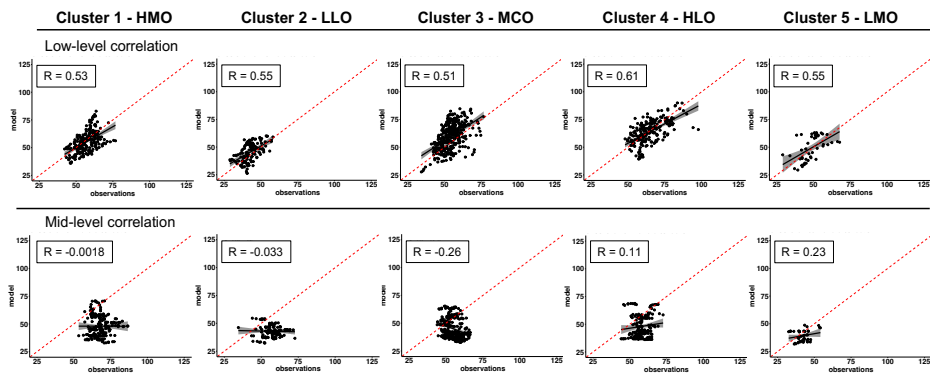
389 For the mid-level, the overall correlation reveals that GEOS-CF and GEOS-Chem both have a weak relationship with the

390 lidar (R = 0.22 and R = 0.12, respectively) (Figure S7 - bottom panel). This indicates that neither model can simulate mid-

391 level O₃ pattern well. GEOS-Chem consistently underestimates the magnitude of mid-level O₃ with a systematic high negative
392 normalized bias ranging from -0.44 to -0.18, while GEOS-CF has a lower and non-systematic normalized bias ranging from -
393 0.22 to 0.28. Overall, both models are not able to simulate the O₃ variability nor magnitude well in the mid-level. The overall
394 analysis provides a fundamental but condensed assessment of model performance.
395

3.3.2 Model evaluation based on lidar clusters

397   In this sect. we discuss significant cluster by cluster differences in model performance that are unmasked by the clustering
398 approach. To better explain the side-by-side comparison in Figure 4, spatial O₃ differences (model – lidar observations) for
399 each cluster were derived (Figure 6) as well as individual cluster correlation (Figure 7, Table S1). Subsequent mean normalized
400 biases (Table S1) were calculated from the total vertical and diurnal averages separated by low-level and mid-level.
401

403 **Figure 6**. Mean profile curtain spatial O₃ difference (model – lidar observations) for each cluster (1 – 5). GEOS-Chem
404 differences (a) and GEOS-CF differences (b).

(a) GEOS – Chem model

(b) GEOS – CF model

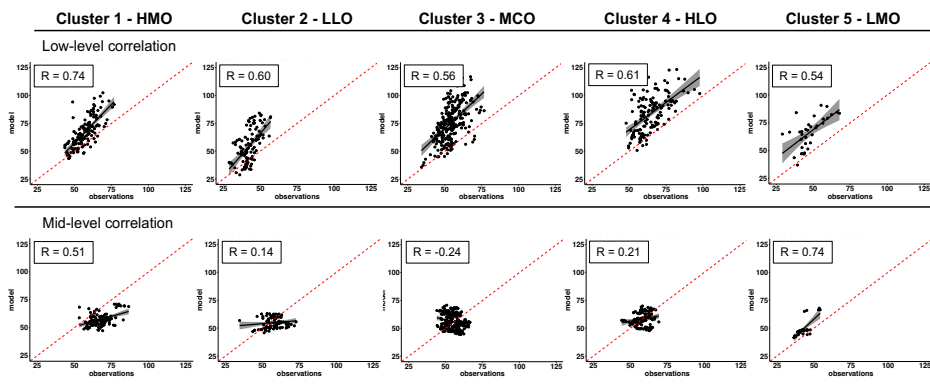



Figure 7. O_3 correlation between lidar observations and a) GEOS-Chem model simulation results and b) GEOS-CF model results by each cluster split by low-level (top panel) and mid-level (bottom panel).

In the low-level, GEOS-CF has a similar performance ability for the HMO, HLO, and LMO clusters with high positive biases at + 0.30, + 0.41, and + 0.45 respectively. These higher biases imply GEOS-CF has difficulty capturing moderate O_3 cases (HMO and LMO) as well as high O_3 cases (HLO) below 2000 m. GEOS-CF also has a high positive bias (+ 0.50) in the LLO cluster indicating the model struggles to capture the lower O_3 cases as well. This is warranted as models are intended to approximate and are not usually able to capture extremes (high or low). In the low-level, GEOS-Chem has the best performance (minimal – 0.04 bias and strong correlation, R = 0.61) in HLO, the cluster with the highest low-level O_3 accumulation and the second-best performance (minimal + 0.07 bias and fair correlation, R = 0.55) in LLO , the cluster with the lowest O_3 accumulation. These results challenge the overall assumption that models struggle to capture extreme cases. GEOS-Chem has a similar performance for the LMO and HMO clusters with low negative biases of – 0.10 and – 0.09, respectively, indicating the model is also able to capture moderate O_3 cases.

Both models perform the worst (in comparison to other clusters) in the low-level in the MCO cluster with a + 0.13 bias for GEOS-Chem and + 0.67 bias for GEOS-CF. As described in Sect. 3.1, MCO is the most common cluster with moderate -

high average O_3 concentrations in the low-level (refer to Figure 3b). Although GEOS-Chem has its worst performance in the MCO cluster, it is not necessarily a poor performance. Contrarily, the GEOS-CF performance in the MCO cluster reveals a more substantially high positive bias. This stands out as models are usually able to capture moderate levels (e.g., non-extreme cases). Evaluating the full temporal and vertical profile indicates that the higher GEOS-CF bias in the MCO cluster is additionally influenced by the greater overestimation of morning O_3, not solely the afternoon O_3. This is different to the performance in the LLO and LMO clusters where GEOS-CF also had a high positive bias in the low-level but better simulates early morning O_3. A similar conclusion can be drawn when evaluating the low-level GEOS-Chem performance. HMO, LLO, MCO, and LMO all share 'higher' biases (rounding to +/– 0.10), but the highest bias is found in the MCO cluster. This can similarly be attributed to GEOS-Chem overestimating morning O_3 in the MCO cluster in contrast to the better early morning estimation in the other clusters.

In the mid-level, GEOS-Chem underestimates O_3 magnitude to the greatest extent in the HMO and the LLO cluster (both bias = – 0.44), which are both clusters with higher mid-level O_3 concentrations (refer to Figure 3c). GEOS-Chem performs similarly in the HLO and MCO clusters, with a negative mean bias of – 0.30 and – 0.27, respectively. This indicates that GEOS-Chem most struggles to simulate higher concentrations of O_3 in the mid-level. The GEOS-Chem model actually never reaches O_3 cluster averages greater than 50 ppb, directly divulging the greater systemic negative bias in the mid-level. GEOS-Chem simulates LMO mid-level O_3 magnitude the best (– 0.18 bias), which is the cluster with the lowest O_3 average (< 45 ppb). Although for the LMO cluster GEOS-Chem has a lower bias, the correlation is still poor (R = 0.23) which indicates that the model is relatively capable of simulating mid-level O_3 only when the case devises lower concentrations but still fails to replicate any O_3 variability and pattern.

On the other hand, GEOS-CF does best simulating LLO, MCO, and HLO, which are all clusters with moderate O_3 in the mid-level (≥ 50 and ≤ 70 ppb). GEOS-CF has the highest bias in the LMO cluster (+ 0.28), the cluster with the lowest mid-level O_3 magnitude but also has the strongest correlation in the same cluster (R = 0.74). This is a unique case where although the model is not able to capture mid-level O_3 magnitude, it is able to capture the variability well. Comparing the full profile curtain, it is evident that in the LMO cluster, the GEOS-CF model simulates mid-level O_3 pattern in the morning/early afternoon fairly well. GEOS-CF also struggles to simulate mid-level O_3 in the HMO cluster, contrarily the cluster with the highest mid-level O_3 (≥ 70 ppb). This supports the previous conclusion that although GEOS-CF has a relatively lower biases in the mid-level, the model still struggles to simulate the extreme O_3 cases. Although GEOS-CF underestimates O_3 magnitude in the HMO cluster, it has a higher correlation than most of the other clusters (R = 0.51) (Figure 7, Table S1). GEOS-CF does a fair job connecting the mid-level higher O_3 pattern in the early morning that develops down to the low-level later in the afternoon (Figure 3). From this we can draw a conclusion that GEOS-CF is better able to capture mid-level O_3 patterns earlier in the temporal profile leading to better correlations with the lidar.

3.3.3 Advantages of cluster approach and derived model conclusions

454   It is warranted that models struggle simulating extreme events/cases such as seen in the low-level in the HLO cluster and

455   in the LLO cluster. However, GEOS-Chem performs best in both clusters with minimal biases and strong to fair correlations.

456   Our result suggest that GEOS-Chem does a much better job simulating extreme O_3 cases in the low-level than expected. We

457   can conclude that the non-systemic bias is not only attributed to a good simulation of afternoon O_3 but also a fair simulation

458   in morning O_3. This specific model feature is not eminent when evaluating overall performance. GEOS-CF systematically

459   overestimates low-level O_3, but the individual clusters indicate that the model has a better correlation with O_3 in the HMO

460   cluster. The higher O_3 levels measured throughout the diurnal profile from 1500 – 2000 m are well captured by the model and

461   contribute to the better low-level correlation.

462   The clustering approach also reveals more discrepancies in the models such as in the MCO cluster. Evaluating the full

463   profile curtains, we find the overestimation of early morning O_3 in the low-level in GEOS-CF adds to the systemic

464   overestimation in afternoon O_3 contributing the greater bias and poorer correlation. The same case can be found in the GEOS-

465   Chem MCO cluster performance but to a lesser extent as GEOS-Chem has a much lower positive bias. Previous studies have

466   found that excessive vertical mixing leads to overestimation of O_3 near the surface as well as underestimation of O_3 night-time

467   depletion resulting in overestimation of O_3 the next day (Dacic et al., 2020; Keller et al., 2021; Travis & Jacob, 2019).

468   Model overestimation of O_3 at night and in early morning hours is a common problem for 3-D Eulerian CTMs. Overnight, O_3

469   concentrations from the evening before can remain lingering in the residual layer. This residual layer sits at about 1000 m or

470   higher depending on the conditions of the environment. O_3 trapped in this residual layer can directly correlate with the next

471   day afternoon O_3 (e.g., Figure 3a; HLO cluster). Models struggle to resolve the shallow surface layer at night, which enhances

472   nighttime NO titration and O_3 dry deposition. If this residual layer and the titration of O_3 overnight in the shallow surface layer

473   is not resolved, next day simulated O_3 will most likely warrant even greater biases. Therefore, in the given case where there is

474   an O_3 event that lasts more than one day (at the same lidar location), the model will likely underestimate O_3 night-time

475   depletion, overpredict morning O_3, and subsequently overpredict the afternoon build-up. Given multiple cases of multi-day or

476   consecutive high O_3 events from the lidar measurements (17 total from HMO, MCO, and HLO), this is likely one of the reasons

477   for GEOS-CF overestimating early and therefore afternoon O_3 in these high O_3 cases in the low-level. In Figure 6, GEOS-CF

478   exhibits the greatest afternoon O_3 overprediction in MCO and HLO. In HLO alone, there were 4 (out of 18) of the profiles that

479   were consecutive while in MCO there were 8 (out of 28). This gives explanation for upwards of 22 – 29 % of the overestimation

480   of O_3 in the profile curtains of these clusters. These multi-day O_3 events are particularly important as they can indubitably lead

481   models to overestimations of afternoon O_3. Full vertical and temporal curtains provided by lidar instruments are essential in

482   fully understanding the development and depletion of O_3 in these cases. The mean curtain profiles in Figure 3a indicate that

483   what is captured at the surface (below 500 m) in the early morning does not represent what is captured in the residual layer

484   (1000 m) by the lidar. Therefore, surface data would not be sufficient in evaluating a multi-day event.

485   GEOS-Chem does not have such an issue overestimating low-level O_3 in the afternoon. In the other clusters, GEOS-

486   Chem actually underpredicts early morning low-level O_3 in the full vertical profile and does an overall better job than GEOS-

487   CF simulating morning low-level O_3, such as in the HLO cluster. A better estimation of early morning O_3 does not warrant the

same build-up of afternoon O_3. In these cases, GEOS-Chem handles the multi-day simulations better than GEOS-CF. This gives some explanation to why GEOS-Chem underpredicts the other clusters with higher O_3 concentrations in the low-level (HMO and HLO). GEOS-CF does best simulating morning low-level O_3 in cases of lower O_3 extent (LLO and LMO), but still overestimates the afternoon O_3. Since in these cases the afternoon does not seem to be related to early morning overestimations, other factors may be contributing. In the LLO cluster, the full curtain profile implies excessive mixing throughout the entire vertical profile could be adding to afternoon O_3 overestimation. Similarly, for the LMO cluster, mid-level O_3 seems to be at play in influencing low-level O_3 which could be adding to afternoon biases.

In the mid-level GEOS-Chem consistently underestimates O_3 but the clusters reveal a better performance in LMO. It is evident that the model is better able to capture lower magnitude O_3 cases in the mid-level. A unique case is exposed in which GEOS-CF has a strong correlation in the mid-level in the LMO cluster despite having a low correlation overall and in the other clusters. The individual cluster correlation reveals the GEOS-CF model is better able to capture the higher O_3 observations in this cluster thus capturing more of the variability. Since the version of GEOS-Chem used in this study was run with the tropchem chemistry mechanism which excludes stratospheric chemistry (now obsolete with current GEOS-Chem developments) and GEOS-CF uses the UCX chemistry mechanism that includes stratospheric chemistry, this may allude to better performance of GEOS-CF in simulating higher O_3 concentrations in the mid-level. The weak correlations in the mid-level could be due to multiple model inefficiencies such as the coarse model resolutions. Although GEOS-CF has a finer resolution than GEOS-Chem, it still may not be sufficient in horizontal and vertical grid resolution to replicate the O_3 variations captured in the 2-D lidar observations. Additionally, transport of emissions in the free troposphere (FT) is another influential factor that could contribute to the misrepresentation of mid-level O_3. In Figure S8, aircraft measurements from OWLETS-2 are used to evaluate GEOS-Chem simulated carbon monoxide (CO) in the FT (1800 – 2500 m AGL). The flight days evaluated are all curtain profiles that were assigned to the clusters with higher levels of O_3 in the mid-level (HMO, MCO, and HLO). It is evident that the model is able to capture lower levels of CO in the FT (100 – 110 ppbv) (e.g., background levels) but struggles to capture the higher levels (130 – 140 ppbv). Since increased levels of CO in the FT are indicative of possible long-range transport (Neuman et al., 2012), FT transport could be a factor contributing to the GEOS-Chem poor performance in the mid-level.

There are additional model discrepancies that can lead to underestimations of O_3 in GEOS-Chem in the mid-level that was found in all 5 clusters. One gap in the GEOS-Chem model could be the representation of tropospheric halogen chemistry which has a large effect of coastal O_3 production. Newer updates to the GEOS-Chem model (v12.9) have included updated tropospheric halogen chemistry mechanisms (iodine, bromine, and chlorine) (Wang et al., 2021) and indicate further investigation of halogen chemistry is needed for better model representation. Another study finds a similar conclusion in the proper representation of cloud uptake and tropospheric chemistry in the model (Holmes et al., 2019), warranting further testing. The role lightning plays in tropospheric oxidation is another feature that is commonly misrepresented in global models and can affect O_3 simulation (Mao et al., 2021). These are all examples of features that if not simulated correctly can lead to misestimations of O_3. The clustering approach allows us to organize the detailed lidar measurements to scope out specific

522 cases where these misrepresentations occur. These previous studies also highlight the importance of lidar measurements and
523 their ability to depict tropospheric emission development and behavior throughout the vertical profile and diurnal cycle which
524 can be used to constrain model emissions and improve simulations.

525   Although this analysis proves to be a useful technique to characterize the largely variably O_3 behavior in coastal regions
526 and evaluate the subsequent model performance, there are also limitations. In this study we are comparing single point lidar
527 versus model output, therefore we cannot simply state that the model is incorrect. We make conclusions and draw biases based
528 on the ability to subset a grid point and compare that to a single point lidar curtain to the best ability but that still leaves an
529 uncertainty.

530

531 **3.4 Cluster derived case studies to evaluate modeled wind and ozone**

532   Meteorological factors such as wind speed and direction can directly impact whether a coastal region will experience
533 clean air or O_3 exceedances. When local meteorological processes such as sea/bay breeze occur at such a fine scale, equally
534 fine resolution measurements are essential in capturing this. The Doppler wind lidar offers a focus on fine details that are only
535 revealed in the multi-dimensional data which allows for such a comprehensive evaluation of the established O_3 cluster profile
536 curtains. In this sect., we evaluate the 2-D relationship between wind and O_3 to assess model performance using lidar and
537 model derived profile curtains (Figure 8). We derived two specific case studies, each from a different cluster: MCO = 17 June
538 2018 and HLO = 30 June 2018. Utilizing the derived clusters, the case studies were chosen to focus on high low-level O_3
539 behavior cases with a goal of evaluating possible sea/bay breeze events. The two case studies are both from the HMI location
540 during the OWLETS-2 campaign. The white spaces in both the wind and O_3 lidar indicate missing data.

541

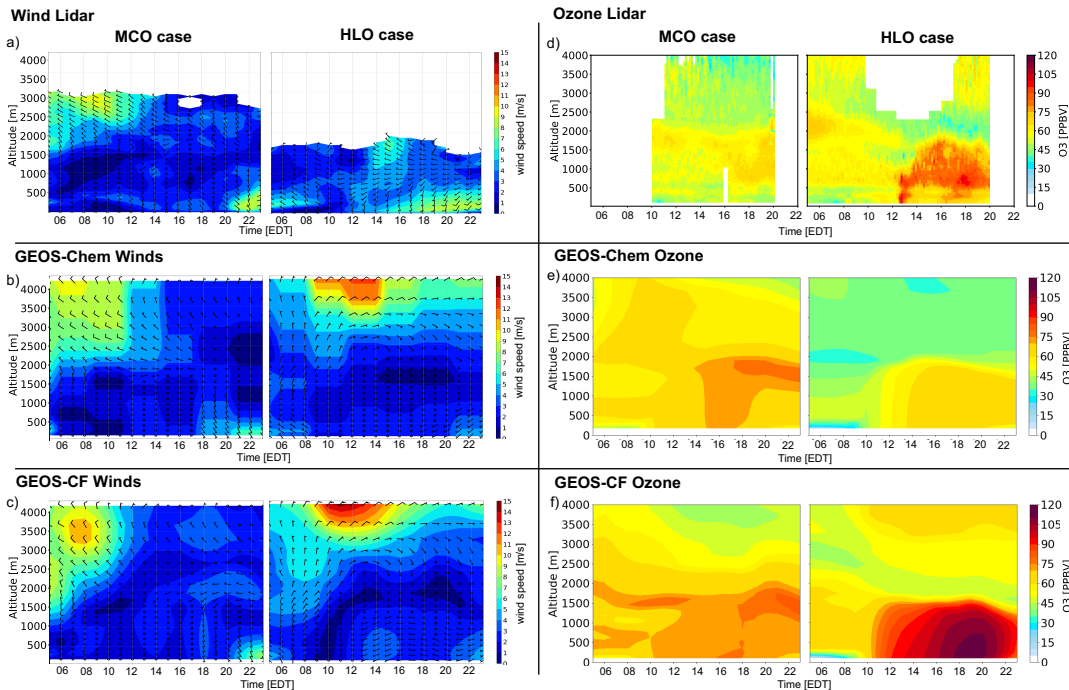



Figure 8. Profile curtains of wind speed/direction (a-c) and O₃ (d-f) from the lidar (top panel), GEOS-Chem (middle panel), and GEOS-CF (bottom panel). Results from OWLETS-2 at HMI. Wind direction is depicted by wind barbs. The white spaces indicate missing data for both the a) wind and d) O₃ lidar curtain profiles.

3.4.1 Sea breeze event interpretation

In the MCO case, the Doppler wind lidar captures a wind direction shift from westerly to easterly winds beginning at 06:00 EDT accompanied by calm winds (approximately 0 m s^{-1}) indicating an early onset sea/bay breeze event. The timing of the start of this event is simulated well but the models fail to predict an actual well-defined wind shift, instead merely simulating 0 m s^{-1} winds after 05:00 EDT. A wind direction shift is depicted in the HLO case, with westerly winds early in the morning and a shift to south-easterly winds later in the temporal profile (at about 10:00 EDT). This could also likely be a common sea breeze event which could have contributed to the high observed O₃ concentrations in the afternoon. Again, the exact timing of the start of the wind shift is captured by the models but then no defined directional shift and little to no winds are simulated after with a worse performance for the GEOS-Chem model. Based on the Doppler wind lidar curtain profiles, we can derive that the two sea/bay breeze cases are distinct. The HLO case closely mirrors a common sea/bay breeze event with a more definite wind direction shift later in the morning and winds above the surface remain consistent throughout the profile. The MCO case shows a less discernible wind shift which also begins earlier in the morning with weaker winds above the surface. These differences are not well captured by either model. It is important to note that GEOS-Chem runs with offline meteorology,

559 averaged every 3 hours. Since sea/bay breezes often happen at a finer temporal resolution, the GEOS-Chem model is at a
560 disadvantage in modeling such fine processes.

561

562 **3.4.2 Wind relation to ozone cases and clustering**

563   In this sect., the wind lidar curtains will be assessed in relation to the O_3 lidar profile curtains and the model performance.
564 We show in sect. 3.3.2 that both models have the highest bias and lowest correlation simulating low-level O_3 in the MCO
565 cluster. Mirroring those results, both models overestimate low-level O_3 in the MCO case studies (Figure 8e, f). Higher O_3
566 concentrations are captured in the lidar curtain profile throughout the day, but is constrained between 1000 – 2000 m. Both
567 models bring this high O_3 pattern down to the surface (below 500 m) which contributes to the overestimation. The models
568 predict little to no winds in the low-level simulating a stagnant environment. Simulated stagnant winds reflect lower dilution
569 rates and induce higher O_3 concentration build-up near the surface that is reproduced in both models. For the mid-level, the
570 GEOS-CF model seems to replicate O_3 pattern better, while GEOS-Chem overestimates O_3. This is a unique finding that was
571 not detected in the previous analysis where GEOS-Chem was found to consistently underestimate mid-level O_3. From the data
572 available above 2000 m, both models seem to do well replicating mid-level winds. This implies that there are more factors at
573 play such as transport or background level O_3 that may have prompted the overestimated O_3 in these cases.

574   For the HLO cluster, GEOS-CF had a high positive mean normalized bias and a reasonable relationship (R = 0.61) in the
575 low-level (sect. 3). For the individual HLO case (Figure 8f), GEOS-CF was similarly found to overestimate low-level O_3
576 magnitude while better able to capture the O_3 pattern. GEOS-CF is better able to reproduce the wind shift in HLO (Figure 8c)
577 but, like the MCO case, stagnant winds simulated earlier in the morning suggest a similar overestimation of early morning O_3.
578 This is another clear example supporting the tendency for GEOS-CF to overestimate morning O_3 which can facilitate an
579 overestimation in the afternoon. The GEOS-Chem HLO case results mirror its mean cluster performance closely by
580 underestimating both low-level and mid-level O_3. For this case, the simulated winds indicate a very different result than the
581 lidar winds, simulating no winds in the low-level for almost the entirety of the temporal profile and vertical profile. Since the
582 results reveal O_3 is underestimated, this suggests that there are more factors affecting O_3 results in this specific case. One of
583 these factors can be the simulation of the boundary layer as the sea/bay breeze develops. If the boundary layer is simulated to
584 be larger in depth, the ability for the model to simulate higher O_3 concentrations may be hindered such as found in Dacic et al.
585 (2017). Since the HLO case indicates a common sea breeze event based on the timing and shift, it appears that GEOS-Chem
586 really struggles capturing this intricate process while GEOS-CF does a better job.

587   It is evident from these cases that differences in sea/bay breeze events can lead to diverse O_3 profiles. The HLO case high
588 O_3 levels that reach down to the surface, with peaks > 75 ppb at both 12:00 and again at 16:00 EDT. Just above this extreme
589 O_3 plume at 2000 m, there is an O_3 deficit of almost 50 ppb. The MCO case differs in that the highest O_3 concentrations do not
590 reach the surface. Also, O_3 is more distributed and mixed throughout the curtain profile and the vertical gradient, although
591 present, is not as stark as the HLO case. The HLO cases also has higher O_3 captured aloft above 2500 m which is not captured
592 in the MCO case. Analyzing their full curtain profiles, it is easy to conclude why these events were not assigned to the same

593 cluster and the differences are also apparent in the individual model performance. For both cases, the models generally seem
594 to underestimate wind speed and overestimate O_3 (to different extents) but the GEOS-Chem performance in the HLO case is
595 different. The uniqueness of this case implies that GEOS-Chem struggles to simulate this sea/bay breeze based on factors other
596 than wind speed and direction.

597   It is imperative to correctly simulate coastal mechanisms in order to mitigate high O_3 events. To accurately simulate such
598 complex exchanges, high resolution vertical and horizontal simulations are needed. Because of the models' relatively coarse
599 resolutions (nominally 50 and 25 km horizonal resolution; 72 vertical levels), the fine-scale vertical wind gradients and
600 horizontal wind shifts are difficult to resolve and, in these cases, not fully able to replicate. This study also acknowledges the
601 need for an evaluation of other modeled factors, aside from model resolution, such as divulged in sect. 3.3.3, considering the
602 possible confounding effects on modeled O_3 outcome.

603

604 **4. Conclusion**

605   We developed a clustering method based on a suite of 91 multi-dimensional lidar O_3 profile curtains retrieved from
606 three recent campaigns. The K-Means clustering algorithm, driven by 8 well defined features, was applied to categorize the
607 fine resolution O_3 data, revealing five distinct O_3 behavior cases that all vary in pattern and magnitude vertically and
608 temporally. The results indicate that fine resolution data can be used to characterize highly variable vertical and temporal
609 coastal O_3 behavior and classify different cases of O_3 exploiting the multiple dimensions. Furthermore, this approach could
610 be used by states to better identify different O_3 photochemical regimes and frequency beyond just surface sampling.

611   The performance of two CTMs (GEOS-Chem and GEOS-CF) were evaluated. Overall, the models had a weak overall
612 relationship with the lidar observations in the mid-level (R = 0.12 and 0.22). GEOS-Chem had a systematic high negative bias
613 and GEOS-CF had an overall lower unsystematic bias range. In the low-level, GEOS-Chem had overall low unsystematic bias
614 range and fair relationship with the lidar observations (R = 0.66), while GEOS-CF had a systematic high positive bias but
615 overall fair relationship (R = 0.69). Utilizing the curated clusters reveals new model insight that is neglected in the overall
616 performance analysis. GEOS-Chem does best simulating extreme O_3 cases in the low-level (such as in HLO and LLO). The
617 greater underestimations of mid-level O_3 for GEOS-Chem can be alluded to multiple model discrepancies such as the
618 mechanism used (tropchem) which only considers tropospheric chemistry. Another factor inhibiting the poor simulation in the
619 mid-level is the model failing to capture long-range transport of emissions in the FT. Evaluating the full profile curtains reveal
620 that GEOS-CF low-level overestimations can be most attributed to the greater overestimation of early morning O_3. This feature
621 is affiliated to multi-day O_3 events where O_3 lingering in the residual layer overnight can contribute to higher O_3 in the afternoon
622 the next day and proves to be a challenge for CTMs. Lidar curtain profiles prove to be essential in evaluating these multi-day
623 cases as they can capture the full development and deposition of O_3 in the residual layer that is not observed at the surface.
624 Although we find the GEOS-CF model struggles to simulate O_3 magnitude in the mid-level, it can relatively emulate O_3
625 variability in some cases (LMO cluster). GEOS-CF also does fairly well in cases in which the pattern of higher mid-level O_3
626 suggests a relationship with the low-level O_3. Although GEOS-CF is run with the combined tropospheric and stratospheric

627 chemistry mechanism, has a finer grid resolution, and is an online model, we conclude there are still limitations to both models
628 which contribute to the difficulty in simulating fine-scale coastal O_3 variability.

629   We demonstrate a unique value of the clustering approach on multi-dimensional lidar data in which we use the cluster
630 results to evaluate of two cases studies from the MCO and HLO clusters. The wind speed and directional shifts (onshore to
631 offshore) illustrated in wind lidar profile curtains indicate a possible sea/bay breeze event in both case studies. The two cases
632 represent distinct sea/bay breeze events that lead to different O_3 developments that were difficult for the CTMs to reproduce,
633 due to coarse model resolution and other possible factors. With a regional model analysis being out of the scope of this study,
634 we propose to use multi-dimensional lidar measurements to evaluate finer regional modeling in our future work.

635   This work is the first time that all three associated campaign lidar data have been analyzed in conjunction. The value
636 of lidar measurements is reflected in its ability to reveal unique features within the temporal and vertical pattern of O_3 behavior.
637 Applying the clustering analysis directly to the lidar O_3 data emerges as a useful and robust approach for identifying O_3
638 regimes. Further observations using lidar instruments should be especially valuable in investigating coastal O_3 behavior as it
639 can divulge the finer-scale O_3 characteristics that remain difficult to successfully simulate in CTMs. We provide a new
640 approach that is the middle ground between looking at specific cases and summarizing overall model performance that allows
641 a synopsis of summer coastal O_3 behavior and subsequently model performance without completely muting distinct O_3
642 features. Evaluating model performance for diverse O_3 behavior in coastal regions is crucial for improving the simulation and
643 furthermore, mitigation of air quality events.

644 *Code availability.* Model code is available upon request to the first author.

645 *Data availability.* The GEOS-Chem model simulation data from this study is publicly accessible online at
646 https://doi.org/10.7910/DVN/V99LHT. The GEOS-CF model data is publicly available online at their website
647 https://gmao.gsfc.nasa.gov/-weather_prediction/GEOS-CF/. The lidar data is publicly available online at https://www-
648 air.larc.nasa.gov/missions.htm.

649 *Supplement.*

650 *Author contributions.* CB and YW conceived the research idea. CB wrote the initial draft of the paper and performed the
651 analyses and model development. All authors contributed to the interpretation of the results and the preparation of the paper.

652 *Competing interests.* The authors declare that they have no conflict of interest.

653 *Acknowledgements.* This study is supported by NASA MUREP Graduate Fellowship (80NSSC19K1680). The Ozone Water-
654 Land Environmental Transition Study (OWLETS-1, 2) and Long Island Sound Tropospheric Ozone Study (LISTOS) field
655 measurements described here were funded by the NASA's Tropospheric Composition Program and Science Innovation Fund

656 (SIF), Maryland Department of Environment, the National Oceanic and Atmospheric Administration (NOAA), the
657 Environmental Protection Agency (EPA), the Northeast States for Coordinated Air Use Management (NESCAUM), and the
658 New Jersey and Connecticut Departments of Energy and Environmental Protection. The authors acknowledge the principal
659 investigators and data operators John Sullivan, Joel Dreessen, Ruben Delgado, William Carrion, and Joseph Sparrow as well
660 as the guidance of the Tropospheric Ozone Lidar Network (TOLNet). LMOL and TROPOZ data are publicly available at
661 (https://www-air.larc.nasa.gov/missions/TOLNet/). The OWLETS and LISTOS data are available at (https://www-
662 air.larc.nasa.gov/). The Doppler wind data taken from the UMBC wind lidar and are publicly available at (https://www-
663 air.larc.nasa.gov/cgi-bin/ArcView/owlets.2018). The aircraft measurements from the UMD Cessna 402B Research Aircraft
664 are publicly available at (https://www-air.larc.nasa.gov/cgi-bin/ArcView/owlets.20180. The GEOS-CF model simulation data
665 were provided directly from the NASA Center Global Modeling and Assimilation Office (GMAO) at the Goddard Space Flight
666 Center (https://gmao.gsfc.nasa.gov/weather_prediction/GEOS-CF/).

References

668 Alonso, A. M., Berrendero, J. R., Hernández, A., and Justel, A.: Time Series Clustering Based on Forecast Densities,
669   Computational Statistics & Data Analysis, 51(2), 762–776., https://doi.org/10.1016/j.csda.2006.04.035, 2006.

670 Banta, R. M., Senff, C. J., Nielsen-Gammon, J., Darby, L. S., Ryerson, T. B., Alvarez, R. J., Sandberg, S. P., Williams, E. J.,
671   and Trainer, M: A bad air day in Houston. Bulletin of the American Meteorological Society, 86(5), 657–
672   670. https://doi.org/10.1175/BAMS-86-5-657, 2005.

673 Bernier, C., Wang, Y., Estes, M., Lei, R., Jia, B., Wang, S., and Sun, J.: Clustering Surface Ozone Diurnal Cycles to Understand
674   the Impact of Circulation Patterns in Houston, TX, Journal of Geophysical Research: Atmospheres, 124(23), 13457–
675   13474., https://doi.org/10.1029/2019jd031725, 2019.

676 Caicedo, V., Rappenglueck, B., Cuchiara, G., Flynn, J., Ferrare, R., Scarino, A. J., Berkoff, T., Senff, C., Langford, A., and
677   Lefer, B.: Bay Breeze and Sea Breeze Circulation Impacts on the Planetary Boundary Layer and Air Quality from an
678   Observed and Modeled Discover-AQ Texas Case Study, Journal of Geophysical Research: Atmospheres, 124(13),
679   7359-7378, https://doi.org/10.1029/2019jd030523, 2019.

680 Christiansen, B.: Atmospheric Circulation Regimes: Can Cluster Analysis Provide the Number?, Journal of Climate, 20(10),
681   2229–2250., https://doi.org/10.1175/jcli4107.1, 2007.

682 Coggon, M. M., Gkatzelis, G. I., McDonald, B. C., Gilman, J. B., Schwantes, R. H., Abuhassan, N., Aikin, K. C., Arend, M.
683   F., Berkoff, T. A., and Brown, S. S.: Volatile chemical product emissions enhance ozone and modulate urban
684   chemistry, Proceedings of the National Academy of Sciences, 118,32, National Academy of Sciences,
685   https://doi.org/10.1073/pnas.2026653118, 2021.

686 Couillard, M. H., Schwab, M. J., Schwab, J. J., Lu, C. H., Joseph, E., Stutsrim, B., Shrestha, B., Zhang, J., Knepp, T. N., and
687   Gronoff, G. P., Vertical Profiles of Ozone Concentrations in the Lower Troposphere Downwind of New York City

during LISTOS 2018-2019, Journal of Geophysical Research: Atmospheres, 126(23), e2021JD035108, https://doi.org/10.1029/2021JD035108, 2021.

Darby, L. S.: Cluster Analysis of Surface Winds in Houston, Texas, and the Impact of Wind Patterns on Ozone, Journal of Applied Meteorology, 44(12), 1788–1806., https://doi.org/10.1175/jam2320.1, 2005.

Davis, R. E., Normile, C. P., Sitka, L., Hondula, D. M., Knight, D. B., Gawtry, S. P., and Stenger, P. J.: A Comparison of Trajectory and Air Mass Approaches to Examine Ozone Variability, Atmospheric Environment, 44(1), 64–74., https://doi.org/10.1016/j.atmosenv.2009.09.038, 2010.

De Young, R., Carrion, W., Ganoe, R., Pliutau, D., Gronoff, G., Berkoff, T., and Kuang, S.: Langley Mobile Ozone LIDAR: Ozone and Aerosol Atmospheric Profiling for Air Quality Research, Applied Optics, 56(3), 721, https://doi.org/10.1364/ao.56.000721, 2017.

Dreessen, J., Orozco, D., Boyle, J., Szymborski, J., Lee, P., Flores, A., and Sakai, R. K.: Observed Ozone over the Chesapeake Bay Land-Water Interface: The Hart-Miller Island Pilot Project, Journal of the Air & Waste Management Association, 69, (11), 1312–1330, https://doi.org/10.1080/10962247.2019.1668497, 2019.

EPA NEI (National Emissions Inventory v1): Air Pollutant Emission Trends Data, available at: http://www.epa.gov/ttn/chief/trends/index.html last access: 23 June 2015.

Farris, B. M., Gronoff, G. P., Carrion, W., Knepp, T., Pippin, M., and Berkoff, T. A.: Demonstration of an off-Axis Parabolic Receiver for near-Range Retrieval of Lidar Ozone Profiles, Atmospheric Measurement Techniques, 12(1), 363–370, https://doi.org/10.5194/amt-12-363-2019, 2019.

Gelaro, R., Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G. Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A. M., Gu, W., Kim, G-K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J. E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S. D., Sienkiewicz, M., and Zhao, B.: The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (Merra-2), Journal of Climate, 30(14), 5419–5454., https://doi.org/10.1175/jcli-d-16-0758.1, 2017.

Gronoff, G., Robinson, J., Berkoff, T., Swap, R., Farris, B., Schroeder, J., Halliday, H.S., Knepp, T., Spinei, E., Carrion, W., Adcock, E.E., Johns, Z., Allen, D., Pippin, M.: A Method for Quantifying near Range Point Source Induced O3 Titration Events Using Co-Located Lidar and Pandora Measurements, Atmospheric Environment, 204, 43–52, https://doi.org/10.1016/j.atmosenv.2019.01.052, 2019

Gronoff, G., Berkoff, T., Knowland, K.E., Lei, L., Shook, M., Fabbri, B., Carrion, W., Langford, A. O.: Case study of stratospheric Intrusion above Hampton, Virginia: lidar-observation and modeling analysis, Atmospheric Environment, 259,118498, 1352-2310, https://doi.org/10.1016/j.atmosenv.2021.118498, 2021.

Han, J. and Kamber, M.: Data mining: concepts and techniques, San Francisco: Morgan Kaufmann Publishers, 2001.

Holmes, C. D., Bertram, T. H., Confer, K. L., Graham, K. A., Ronan, A. C., Wirks, C. K., and Shah, V.: The Role of Clouds in the Tropospheric NOx Cycle: A New Modeling Approach for Cloud Chemistry and Its Global Implications, Geophys. Res. Lett., 46, 4980–4990, https://doi.org/10.1029/2019GL081990, 2019.

Hu, L., Keller, C. A., Long, M. S., Sherwen, T., Auer, B., Da Silva, A., Nielsen, J. E., Pawson, S., Thompson, M. A., Trayanov, A. L., Travis, K. R., Grange, S. K., Evans, M. J., and Jacob, D. J.: Global simulation of tropospheric chemistry at 12.5 km resolution: performance and evaluation of the GEOS-Chem chemical module (v10-1) within the NASA GEOS Earth system model (GEOS-5 ESM), Geosci. Model Dev., 11, 4603–4620, https://doi.org/10.5194/gmd-11-4603-2018, 2018.

Keller, C.A., Knowland, K.E., Duncan, B.N., Liu, J., Anderson, D.C., Das, S., Lucchesi, R.A., Lundgren, E.W., Nicely, J.M., Nielsen, E., Ott, L.E., Saunders, E., Strode, S.A., Wales, P.A., Jacob, D.J., and Pawson, S.: Description of the NASA Geos Composition Forecast Modeling System GEOS-CF v1.0, Journal of Advances in Modeling Earth Systems, 13(4), https://doi.org/10.1029/2020ms002413, 2021.

Knowland, K.E., Keller, C.A., Lucchesi, R.: File specification for GEOS-CF products, GMAO office note No. 17 (version 1.0). available from: https://gmao.gsfc.nasa.gov/pubs/office_notes.php, 32, 2019.

Knowland, K. E., Keller, C. A., Wales, P. A., Wargan, K., Coy, L., Johnson, M. S., Liu, J., Lucchesi, R. A., Eastham, S. D., Fleming, E. L., Liang, Q., Leblanc, T., Livesey, N. J., Walker, K. A., Ott, L. E., and Pawson, S.: NASA GEOS Composition Forecast Modeling System GEOS-CF v1.0: Stratospheric Composition, 14(6), e2021MS002852, https://doi.org/10.1002/essoar.10508148.1, 2021.

Lawson, R. G., and Jurs, P. C.: New Index for Clustering Tendency and Its Application to Chemical Problems, Journal of Chemical Information and Computer Sciences, 30(1), 36–41., https://doi.org/10.1021/ci00065a010, 1990.

Leblanc, T., Brewer, M. A., Wang, P. S., Granados-Muñoz, M. J., Strawbridge, K. B., Travis, M., Firanski, B., Sullivan, J. T., McGee, T. J., Sumnicht, G. K., Twigg, L. W., Berkoff, T. A., Carrion, W., Gronoff, G., Aknan, A., Chen, G., Alvarez, R. J., Langford, A. O., Senff, C. J., Kirgis, G., Johnson, M. S., Kuang, S., and Newchurch, M. J.: Validation of the TOLNet lidars: the Southern California Ozone Observation Project (SCOOP), Atmos. Meas. Tech., 11, 6137–6162, https://doi.org/10.5194/amt-11-6137-2018, 2018.

Lei, L., Berkoff, T. A., Gronoff, G., Su, J., Nehrir, A. R., Wu, Y., Moshary, F., and Kuang, S.: Retrieval of UVB aerosol extinction profiles from the ground-based Langley Mobile Ozone Lidar (LMOL) system, Atmos. Meas. Tech., 15, 2465–2478, https://doi.org/10.5194/amt-15-2465-2022, 2022.

Li, W., Wang, Y., Bernier, C., and Estes, M.: Identification of Sea Breeze Recirculation and Its Effects on Ozone in Houston, TX, during Discover-Aq 2013, Journal of Geophysical Research: Atmospheres, 125(22), https://doi.org/10.1029/2020jd033165, 2020.

Little, R. J., A., and Rubin, D., B.: Statistical Analysis with Missing Data. Second ed., Wiley, 1987.

Loughner, C. P., Tzortziou, M., Follette-Cook, M., Pickering, K. E., Goldberg, D., Satam, C., Weinheimer, A., Crawford, J. H., Knapp, D. J., Montzka, D. D., Diskin, G. S., and Dickerson, R. R.: Impact of Bay-Breeze Circulations on Surface Air

Quality and Boundary Layer Export. Journal of Applied Meteorology and Climatology, 53, 7, 1697–1713., https://doi.org/10.1175/jamc-d-13-0323.1, 2014.

Mao, J., Zhao, T., Keller, C. A., Wang, X., McFarland, P. J., Jenkins, J. M., and Brune, W. H.: Global Impact of Lightning-Produced Oxidants, Geophys. Res. Lett., 48, https://doi.org/10.1029/2021GL095740, 2021.

McDuffie, E. E., Smith, S. J., O'Rourke, P., Tibrewal, K., Venkataraman, C., Marais, E. A., Zheng, B., Crippa, M., Brauer, M., and Martin, R. V.: A global anthropogenic emission inventory of atmospheric pollutants from sector- and fuel-specific sources (1970–2017): an application of the Community Emissions Data System (CEDS), Earth Syst. Sci. Data, 12, 3413–3442, https://doi.org/10.5194/essd-12-3413-2020, 2020.

Neuman, J. A., Trainer, M. Aikin, K., Brioude, J. Brown, S., de Gouw, J., Dube, W., Flynn, J., Graus, M., Holloway, J., Lefer, B., Nedelec P., Nowak, J., Parrish, D., Pollack, I., Roberts, J., Ryerson, T., Smit, H., Thouret, V., and Wagner, N.: Observations of Ozone Transport from the Free Troposphere to the Los Angeles Basin, Journal of Geophysical Research: Atmospheres, 117, 21, https://doi.org/10.1029/2011jd016919, 2012.

Orbe, C., Oman, L. D., Strahan, S. E., Waugh, D. W., Pawson, S., Takacs, L. L., and Molod, A. M.: Large-scale atmospheric transport in GEOS replay simulations, J. Adv. Model. Earth Syst., 9, 2545–2560, https://doi.org/10.1002/2017MS001053, 2017.

Ring, A. M., Canty, T. P., Anderson, D. C., Vinciguerra, T. P., He, H., Goldberg, D. L., Ehrman, S. H., Dickerson, R. R., and Salawitch, R. J.: Evaluating commercial marine emissions and their role in air quality policy using observations and the CMAQ model, Atmospheric Environment, 173, 96-107, https://doi.org/10.1016/j.atmosenv.2017.10.037, 2018.

Stauffer R.M., Thompson A.M., and Witte J.C.: Characterizing Global Ozonesonde Profile Variability from Surface to the UT/LS with a Clustering Technique and MERRA-2 Reanalysis, J Geophys Res Atmos. 123(11):6213-6229. https://doi.org/10.1029/2018JD028465, 2018.

Strode, S. A., Ziemke, J.R., Oman, L. D., Lamsal, L. N., Olsen, M. A., and Liu, J.: Global changes in the diurnal cycle of surface ozone, Atmospheric Environment, 199, 323-333, https://doi.org/10.1016/j.atmosenv.2018.11.028, 2019.

Sullivan, J. T., McGee, T. J., Leblanc, T., Sumnicht, G. K., and Twigg, L. W.: Optimization of the GSFC TROPOZ DIAL retrieval using synthetic lidar returns and ozonesondes – Part 1: Algorithm validation, Atmos. Meas. Tech., 8, 4133–4143, https://doi.org/10.5194/amt-8-4133-2015, 2015.

Sullivan, J.T., McGee, T.J., DeYoung, R., Twigg, L.W., Sumnicht, G.K., Pliutau, D., Knepp, T., and Carrion, W.: Results from the NASA GSFC and LaRC Ozone Lidar intercomparison: new mobile tools for atmospheric research, J. Atmos. Ocean. Technol., 32 (10), 1779-1795, https://doi.org/10.1175/JTECH-D-14-00193.1, 2015.

Thompson, A. M., Stauffer, R. M., Miller, S. K., Martins, D. K., Joseph, E., Weinheimer, A. J., and Diskin, G. S.: Ozone profiles in the Baltimore-Washington region (2006-2011): satellite comparisons and DISCOVER-AQ observations, J Atmos Chem., 72(3-4), 393-422. https://doi.org/10.1007/s10874-014-9283-z, 2015.

Torgo, L.: Data Mining with R, Chapman & Hall/CRC Data Mining and Knowledge Discovery Series, https://doi.org/10.1201/b10328, 2010.

787 Wang, L., Newchurch, M. J., Alvarez, R. J., Berkoff, T. A., Brown, S. S., Carrion, W., De Young, R. J., Johnson, B. J., Ganoe,
788   R., Gronoff, G., Kirgis, G., Kuang, S., Langford, A. O., Leblanc, T., McDuffie, E. E., McGee, T. J., Pliutau, D., Senff,
789   C. J., Sullivan, J. T., Sumnicht, G., Twigg, L. W., and Weinheimer, A. J.: Quantifying TOLNet Ozone Lidar Accuracy
790   during the 2014 DISCOVER-AQ and FRAPPÉ Campaigns. Atmos Meas Tech, 10(10), 3865-3876,
791   http://doi.org/10.5194/amt-10-3865-2017. 2017.

792 Wang, X., Jacob, D. J., Downs, W., Zhai, S., Zhu, L., Shah, V., Holmes, C. D., Sherwen, T., Alexander, B., Evans, M. J.,
793   Eastham, S. D., Neuman, J. A., Veres, P. R., Koenig, T. K., Volkamer, R., Huey, L. G., Bannan, T. J., Percival, C. J.,
794   Lee, B. H., and Thornton, J. A.: Global tropospheric halogen (Cl, Br, I) chemistry and its impact on oxidants, Atmospheric
795   Chem. Phys., 21, 13973–13996, https://doi.org/10.5194/acp-21-13973-2021, 2021.

796 Wu, Y., Nehrir, A. R., Ren, X., Dickerson, R. R., Huang, J., Stratton, P. R., Gronoff, G., Kooi, S. A., Collins, J. E., and
797   Berkoff, T. A.: Synergistic aircraft and ground observations of transported wildfire smoke and its impact on air quality
798   in New York City during the summer 2018 LISTOS campaign, Science of The Total Environment, 773,145030,
799   https://doi.org/10.1016/j.scitotenv.2021.145030, 2021.