

JGR Atmospheres

RESEARCH ARTICLE

10.1029/2019JD031725

Key Points:

- We developed a new method clustering diurnal ozone patterns that better characterizes ozone variability in Houston than peak ozone alone
- Clustering was heavily influenced by early morning ozone values and showed a strong relationship to peak afternoon ozone
- Uneven distributions of clusters between years elucidate interannual ozone variability and link to synoptic meteorological processes

Supporting Information:

- Supporting Information S1

Correspondence to:

Y. Wang,
ywang246@central.uh.edu

Citation:

Bernier, C., Wang, Y., Estes, M., Lei, R., Jia, B., Wang, S.-C., & Sun, J. (2019). Clustering surface ozone diurnal cycles to understand the impact of circulation patterns in Houston, TX. *Journal of Geophysical Research: Atmospheres*, 124. <https://doi.org/10.1029/2019JD031725>

Received 24 SEP 2019

Accepted 15 NOV 2019

Accepted article online 19 NOV 2019

Clustering Surface Ozone Diurnal Cycles to Understand the Impact of Circulation Patterns in Houston, TX

Claudia Bernier¹ , Yuxuan Wang¹ , Mark Estes², Ruixue Lei^{1,3}, Beixi Jia⁴, Sing-Chun Wang¹ , and Jiajia Sun¹ 

¹Department of Earth and Atmospheric Science, University of Houston, Houston, TX, USA, ²Texas Commission on Environmental Quality, Austin, TX, USA, ³Now at Department of Meteorology and Atmospheric Science, Pennsylvania State University, University Park, Pennsylvania, USA, ⁴CMA Public Meteorological Service Center, China Meteorological Administration, Beijing, China

Abstract The diurnal cycle of surface ozone is directly influenced by the chemistry and meteorological processes which affect a region. This study uniquely employs a clustering methodology to examine the complete diurnal pattern of surface ozone for the Houston-Galveston-Brazoria region and links the identified patterns to meteorological regimes for June, July, and August of 3 years (2011, 2014, and 2015). Four features were implemented into the clustering algorithm: ozone rate of decrease at night, daily minimum before sunrise, rate of increase after sunrise, and an average of afternoon ozone. Four clusters were chosen, ranging from a mostly flat diurnal pattern with low mixing ratios (~20 ppbv) throughout the day (Cluster 1), to a more variable diurnal cycle with very high mixing ratios (>70 ppbv) in the afternoon (Cluster 4). The clusters are found to associate with distinctive circulation patterns and well-known regional meteorological processes, such as the low-level jet and Bermuda High. The uneven distribution of the clusters between the years helps elucidate ozone interannual variability due to meteorology: Cluster 4 had 0 days assigned from 2014 due to the greater influence of circulation patterns bringing clean air from the Gulf of Mexico. We show that the clustering method better characterizes ozone variability than the simplistic method of dividing peak ozone into quantiles. With the clustering analysis, we demonstrate that the ozone diurnal pattern holds more value than just peak ozone hours of the day in providing a clearer understanding of ozone variability and associated meteorological processes.

1. Introduction

Tropospheric ozone is a secondary pollutant produced by the photochemical oxidation of its precursor gases (e.g., volatile organic compounds and nitrogen oxides) that has been proven to have a great impact on vegetation and human health. Ozone variability depends on the photo-oxidation of its precursor gases and meteorology (e.g., Jacob & Winner, 2009; Mazzuca et al., 2016). At the surface, the lowest ozone during the course of a day typically occurs at night. This is due to the shutdown of photochemical ozone production and shallow nighttime surface layer which leads to enhanced ozone deposition and accumulation of NO_x emissions that titrate ozone (via NO + O₃ → NO₂ + O₂). As the sun rises, the planetary boundary layer (PBL) height in the morning increases due to the solar heating of the surface which initiates turbulent mixing in the lower level of the boundary layer. Solar radiation also leads to the rapid increase of surface ozone in the morning due to the photolysis of ozone's precursor gases. After this rapid increase, the PBL height and ozone mixing ratios stay consistently high in the afternoon. When the sun sets, convection reduces and mixing ceases causing the PBL height and ozone concentrations to decrease overnight until the sun rises again the following morning. Such diel patterns of surface ozone are typical for low-elevation sites influenced by anthropogenic emissions, whereas high-elevation sites or those affected by complex terrains may exhibit different diurnal patterns (e.g., Baumgardner & Edgerton, 1998; Seguel et al., 2013). While the ozone diurnal cycle over the course of a given day is largely determined by the diurnal cycle of solar radiation and PBL dynamics, changing weather conditions can cause the diurnal pattern of ozone to vary on a day-to-day basis due to the strong dependence of both local ozone production/loss and regional transport on meteorological factors (e.g., temperature, wind, and cloudiness).

To determine the severity of the pollution for a region, the predominant focus has been on the peak ozone level of a day (e.g., maximum daily average 8-hr ozone or MDA8 ozone). This can be partially attributed to the fact that the U.S. National Ambient Air Quality Standard (NAAQS) for ozone is based on the MDA8

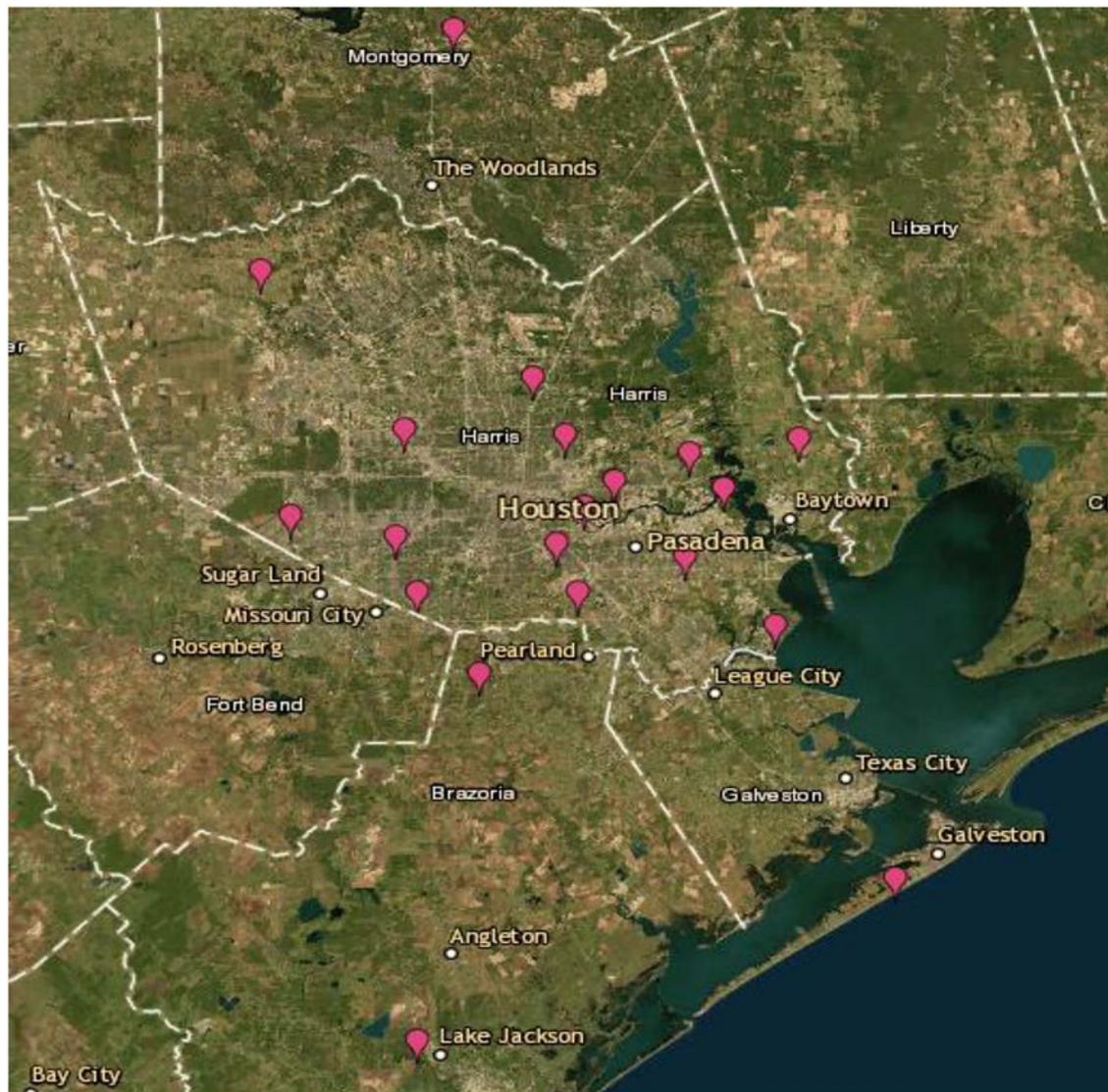


Figure 1. Location of 19 active ozone monitoring sites (pink symbols) in the HGB region from EPA AirData database (<https://epa.maps.arcgis.com/apps/webappviewer>).

ozone metric. The use of MDA8 ozone in research is a good approach but could potentially overlook certain details in the diurnal pattern of ozone which may contain useful information on how much and when the accumulation of ozone is occurring. Thus, the focus of this study is to analyze the diurnal variation of ozone concentration and understand if and how ozone concentrations early in the morning could foreshadow how ozone will behave as the day proceeds. The study region is the Houston-Galveston-Brazoria (HGB) region, denoted as 94.5°W to 96.0°W and 28.5°N to 30.5°N (Figure 1). With a population of more than 6 million people, the region's air quality status has remained in nonattainment for exceeding the NAAQS for ozone (TCEQ, 2012). Downtown Houston is only about 80 km from the Gulf of Mexico and 60 km from the shores of Galveston Bay, placing the city in an interesting spot for ozone accumulation due to the unique local meteorological air circulation patterns that can occur.

The synoptic meteorological processes that affect ozone production in the HGB region during the summer include the Bermuda High and the low-level jet (LLJ) (e.g., L. Shen et al., 2015; Wang et al., 2016; Zhu & Liang, 2013). The LLJ is defined as a relative maximum in the vertical profile of wind speed occurring at low levels (Markowski & Richardson, 2010). For the HGB region, the LLJ is predominantly a nighttime phenomenon where a strong gradient wind flows from the HGB region to the Great Plains due to differential

heating and a sloping terrain. The Bermuda High is a semipermanent high pressure that usually sits off the East Coast of the United States in the North Atlantic Ocean bringing strong southerly winds through the Gulf States, especially during midsummer (Li et al., 2011). The Bermuda High begins its westward extension in June and is usually strongest and located farthest west in July. It brings clean air from the Gulf of Mexico along with moisture that helps produce cloud cover, which are conditions that typically link to low ozone in the HGB region (Wang et al., 2016). The Bermuda High begins to retreat back eastward in August, coinciding with an increasing trend in surface ozone from July to August and September/October in the HGB region. Zhu and Liang (2013) found that the westward extension of the Bermuda High acts to amplify the LLJ on the west flank of the high by creating a stronger pressure gradient. On one hand, the Bermuda High and LLJ bring clean Gulf air through southeastern Texas, lowering ozone concentrations (Wang et al., 2016). On the other hand, for the Central Gulf States (CGS; denoted as 30 to 35°N and 85 to 95°W), the extension of the Bermuda High coupled with the LLJ can create subsidence over the area while suppressing rainfall, raising temperatures, and propagating high ozone concentrations (Pu et al., 2016). As the HGB region falls right on the edge of the CGS, it is susceptible to both regimes (Shen et al., 2015).

Previous studies have demonstrated how these synoptic air circulations patterns can alter weather conditions over the Gulf States and influence the ozone concentration levels (e.g., Pu et al., 2016; Shen et al., 2015; Tucker et al., 2010; Wang et al., 2016; Weaver & Nigam, 2008; Zhu & Liang, 2013). For the HGB region, when the strong southerly flow from the Bermuda High is not as present, this allows for other systems to enter the region, such as midlatitude cyclones which can bring dirtier continental air elevating the regional background ozone and promoting local ozone formation in postfrontal conditions (Lei et al., 2018). On the local scale, the HGB region is affected by sea breeze circulations which causes the diurnal temperature cycle to produce a 24-hr wind vector clockwise rotation, due to the temperature contrast at the coast. Generally, when the Bermuda High flow is dominant, the sea breeze is enhanced by the strong southerly flow (Tucker et al., 2010), but without this dominant flow, the sea breeze circulation can produce a midday wind direction reversal causing the polluted air to recirculate through the region, leading to high ozone (Banta et al., 2005; Nielsen-Gammon et al., 2005; Tucker et al., 2010; Zhong et al., 2007; Haman et al., 2014). The sea breeze effect is a local process that affects specific areas (sites) around the HGB region. Another local process is PBL dynamics, which affects individual sites but is subject to synoptic scale changes. All these factors provide the region with a complex regional air pollution problem. As an initial attempt, we set the scope of this study to be on the mean regional diurnal surface ozone in the HGB region. Thus, we leave the local effects of sea breeze or PBL dynamics to future studies though we acknowledge their impacts on local ozone in this paper.

Considering the well-established dependence of ozone variability on synoptic meteorology and the presence of distinct circulation patterns in the HGB region as described above, we hypothesize that the clustering method will reveal interesting and unique findings within the diurnal pattern of ozone. This hypothesis will be tested by analyzing the diurnal ozone concentration patterns during the summer months of June, July, August (JJA) of three meteorologically and chemically diverse years (2011, 2014, and 2015) for the HGB region through the K-means clustering method.

The rest of the paper is organized as follows. Section 2 presents the data used for the analysis and method used for the clustering analysis. We also describe features of the diurnal patterns used in the clustering method. In section 3, the results of the four established clusters are explained in depth. In section 4, we quantify the impact of the Bermuda High and LLJ to the diurnal surface ozone clusters. Concluding remarks are given in section 5.

2. Data and Methods

2.1. Ozone Observations

The HGB region has one of the nation's densest ozone monitoring networks, offering a rich observational data set to test the hypothesis. Hourly surface ozone concentrations from 19 active monitoring sites in the HGB area were obtained from the Environmental Protection Agency (EPA) Air Quality Station AirData website (http://www3.epa.gov/airquality/airdata/ad_data_daily.html) (Figure 1). With the hourly data from each site, a diurnal ozone concentration pattern for any given day could be produced as shown in Figure 2.

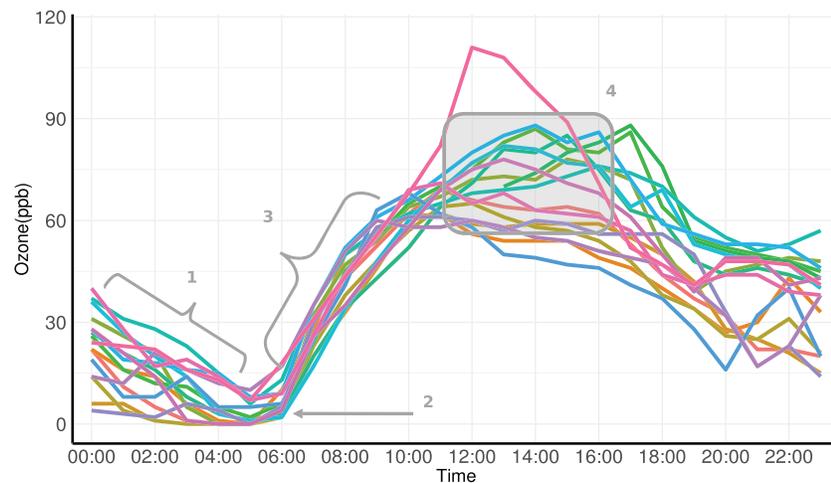


Figure 2. Conceptual model of diurnal ozone behavior for the HGB region, showing ozone mixing ratios from each monitoring site at every hour for 1 day (June 2, 2015) (thin multicolored lines). The four features chosen for clustering are indicated (gray). Feature 1: rate of change in ozone (i.e., slope) from 00:00 AM–05:00 AM. Feature 2: minimum ozone mixing ratio averaged at 05:00 AM and 06:00 AM. Feature 3: rate of change in ozone from 06:00 AM–09:00 AM. Feature 4: ozone mixing ratio at its maximum averaged from 12:00 PM–04:00 PM.

Ozone mixing ratios from all monitoring sites were averaged by the hour to achieve a regional-mean diurnal ozone concentration variation for each day in JJA.

The study focused on the summer months JJA for the years 2011, 2014, and 2015. These 3 years were specifically chosen because they were dissimilar from each other. The year 2015 was a typical high ozone year, with regional mean daily ozone concentrations reaching 80 ppbv and MDA8 ozone reaching near 100 ppbv on six different days. 2011 was an abnormal year with severe drought conditions for the HGB region. For this year, MDA8 ozone reached higher than 100 ppbv on 5 different days, potentially linked with drought conditions (Wang et al., 2015, 2017). 2014 was a relatively “clean” year with regional mean daily ozone ranging from less than 20 ppbv to about 80 ppbv. The clustering analysis was applied on the 3 years as a whole dataset. In combining all 3 years, we can use the differences in ozone diurnal variability to explain the interannual variability of ozone between the 3 years.

2.2. Meteorological Data

For the LLJ analysis, daily percentages or frequencies for the LLJ over the HGB region for each day were extracted from the North American Regional Reanalysis (NARR) vertical wind profiles by specific criteria adopted from Doubler et al. (2015), defined as follows: (1) the elevation of maximum wind speed had to be located at or below 3,000 m above ground level (AGL), (2) wind speeds were $\geq 12 \text{ ms}^{-1}$, (3) wind speed decreased by $\geq 6 \text{ ms}^{-1}$ from the level of maximum wind speed to the next minimum or to 5,000 m AGL (whichever was lower), and (4) the wind speed decreased by $\geq 6 \text{ ms}^{-1}$ below the level of maximum wind speed. The LLJ% refers to the percentage of 3-hourly winds identified as LLJ by the criteria during a 24-hr interval at each grid point, averaged over the HGB region.

For circulation patterns, we extracted wind fields and geopotential heights at 850 hPa from European Centre for Medium-Range Weather Forecasts European Reanalysis-Interim to produce wind field plots for each day. Our previous work (Wang et al., 2016) suggested that the westward extension of the Bermuda High is an important synoptic-scale feature that determines the interannual variability of monthly-mean MDA8 ozone in JJA for the HGB region and the Gulf coast in general. To present this feature, we use the Bermuda High Longitudinal index (BH-Lon) adapted from Li et al. (2011) as used in Wang et al. (2016) to determine the westward extension of the Bermuda High relative to the HGB region. The BH-Lon is defined as the longitude at which the 1,560 geopotential meter (gpm) isoline and the 850 hPa wind ridgeline intersect. This ridgeline is roughly the zonal line that separates the easterly trade winds in the south from the westerly winds in the north. The 1,560 gpm isoline is linearly interpolated to the precision of 0.1° from the gridded reanalysis data, and the 850 hPa wind ridgeline is

mathematically written as $u = 0$ and $\partial u / \partial y \geq 0$, where u is the zonal wind component, and y is the meridional coordinate. The HGB region (Figure 1) falls directly under the strong, southerly winds from the Bermuda High during most days of the summer.

2.3. Clustering

Clustering methods have been widely used in many scientific studies and have proven to be effective in discovering complex and sometimes hidden patterns in large data sets (e.g., Bezdek, 1981; Dunn, 1973; MacQueen, 1967; Sáenz & Durán-Quesada, 2015). The technique of clustering works by distributing data objects into specific groups, with the goal that all data objects assigned to the same group, or cluster, have common characteristics while each cluster has distinct characteristics (Darby, 2005). Clustering is useful in finding the underlying structure of the data set that might be complicated and not obvious to the human eye. For example, clustering analysis and pattern recognition have been used to establish air circulation regimes in a particular region, which can be linked to pollution concentrations (e.g., Balashov et al., 2017). Darby (2005) used the clustering method to identify different circulation patterns based on hourly average winds and then linked the results with MDA8 ozone. The majority of previous studies exercised clustering methods by focusing on meteorology, such as wind circulation data, as the input data to determine the clusters (e.g., Darby, 2005; Pakalapati et al., 2009; Sourì et al., 2016). In this study, we use features within the diurnal surface ozone pattern as input data for the clustering, similar to the ozone profiles used in Stauffer et al. (2017). The choice of which clustering algorithm to use is largely subjective. In this study, we chose K-means clustering because of its conceptual and numerical simplicity as well as its popularity in data science. For K-means, the features of interest and optimal number of clusters must be determined prior to performing the clustering.

The first step is to determine the features of interest, which are the specific variables that will be applied to the clustering algorithm. By inspecting the mean diurnal ozone pattern for each day, we were able to establish four definite features in the pattern that seem to characterize the diurnal ozone behavior (Figure 2). The first feature chosen is the rate of change in ozone (i.e., slope) from 00:00 AM to 05:00 AM. During this 5-hr period, ozone levels usually decrease. The second feature is the minimum ozone-mixing ratio for each day, which we found to occur usually at 05:00 AM and 06:00 AM for each day and as such, the ozone mixing ratios at these times were averaged to produce one mean minimum value. The next feature is the usual rapid increase of ozone concentration between 06:00 AM and 09:00 AM that occurs due to photo-oxidation processes as well as the contributions from higher ozone concentrations aloft in the residual layer which can be mixed down at this time. This is typically the time period when ozone is increasing with time at the fastest rate, thus it represents the greatest positive slope. Over the next few hours, 12:00 PM to 04:00 PM, ozone stays consistently high as daylight persists, reaching its maximum of the day. These 4 hours were averaged to obtain the mean maximum ozone for each day as the fourth feature for the clustering.

Before the four features were applied to the K-means clustering algorithm, the data for the features were normalized to make the values of each feature have zero-mean and unit variance. Since K-means clustering algorithm organizes the clusters based on the Euclidean distance to the centroid, normalizing the features ensures that each feature is given the same importance to the clustering (Aksoy & Haralick, 2001; Larose, 2005). The K-means clustering method also requires one to choose the optimal number of clusters based on the data; the “k” value must be preset. There are multiple methods one can use to determine the optimal number of clusters. For this study, we used seven different methods (Figures S1a and S1b): Dindex, Hubert Index, Bayesian Inference Criterion, Elbow Method, Average Silhouette Approach, Gap Statistic, and the NbClust package for various indices (Charrad et al., 2014; Kaufman & Rousseeuw, 1990; Tibshirani et al., 2001). All methods, except one (silhouette approach), selected four as the optimal number of clusters for the data. The “k” value of four was then implemented and the K-means clustering method was applied to the normalized features for the three summers all together. Figure S2 shows the correlation matrix of the four features (normalized) by cluster.

3. Clustering Results

3.1. Mean Diurnal Variations and Cluster Profiles

After completing the clustering for the three summers combined, a cluster number (1–4) was assigned to each date of the study period. The number of days assigned to each cluster is summarized in Table 1. June

Table 1
Number of days in each cluster for summers 2011, 2014, and 2015

Number of Days in each cluster	2011	2014	2015
Cluster 1	14	28	17
Cluster 2	55	48	47
Cluster 3	15	16	20
Cluster 4	7	0	8

3, 2011 was removed from the data before the clustering because of its very rare diurnal pattern (Figure S3), so in total 275 days from the three summers were used for the clustering. The number of days in each of the clusters for all 3 years is similar. Cluster 4 is the diurnal pattern of extreme ozone pollution with the fewest number of days in each summer, while Cluster 2 contains the most days. 2014 had 0 days assigned to Cluster 4, which is the high ozone/extreme cluster as shown below. This is consistent with the fact that 2014 is the lowest ozone year among the 3 years

analyzed. The clustering results suggest that the reason for 2014 being a low ozone year is due to lack of extreme, high ozone patterns and the associated meteorological conditions for that year. We will elaborate on this point in the meteorological analysis (section 4).

The mean diurnal variations of ozone differ between the four clusters identified (Figure 3a, Table S1). For all three summers, Cluster 1 has the lowest average ozone concentration in the afternoon (Feature 4), the lowest average increase slope from 06:00 AM to 09:00 AM (Feature 3), and the lowest average decreasing slope from 00:00 AM to 05:00 AM (Feature 1). Opposed to the other three clusters, Cluster 1 has the highest average minimum ozone concentration (Feature 2). These characteristics make Cluster 1 the least varying diurnal pattern of all the clusters, with mean ozone mixing ratios ranging from only 10 to 26 ppbv. We will explain in section 4 that a reason for this less variable diurnal ozone pattern is due to meteorological factors such as strong southeasterly winds bringing clean Gulf coast air through the HGB region throughout the day

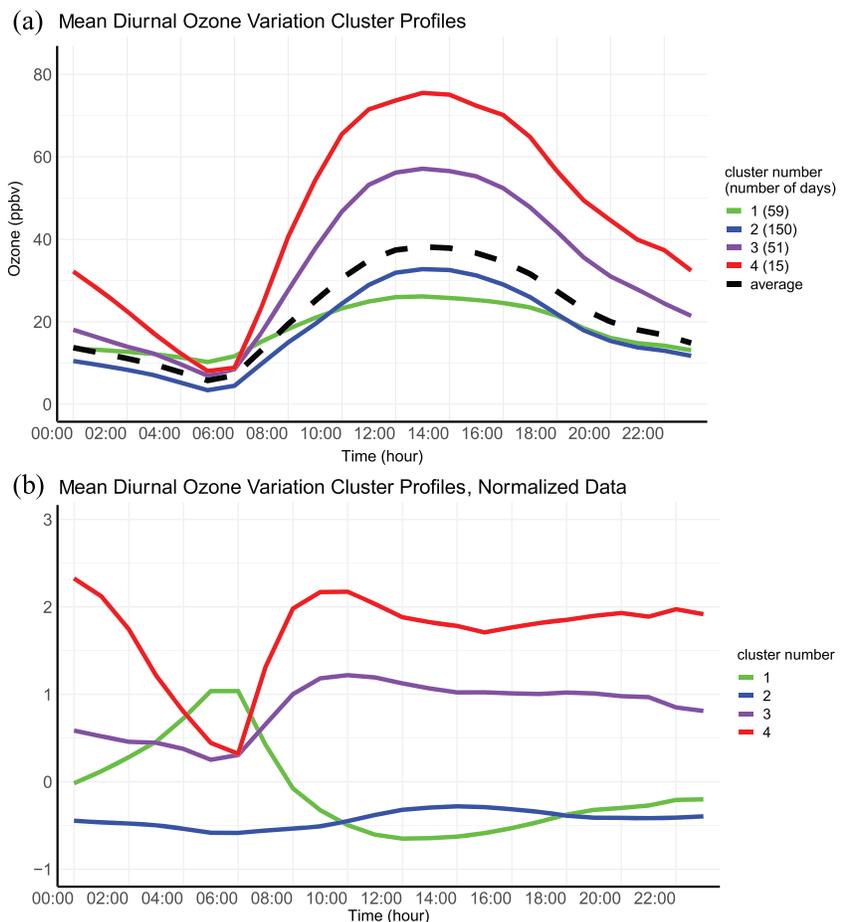


Figure 3. (a) Mean diurnal variations of surface ozone for each cluster for all three summers combined. Black dashed line represents the average mean diurnal pattern based on all summers. (b) Mean diurnal variations of surface ozone for all three summers with normalized data.

which transports pollutants away. This also prevents the formation of a shallow, stable boundary layer at the surface which typically facilitates efficient NO titration and dry deposition of ozone. These meteorological factors impede the local production of ozone which causes the diurnal pattern for the days in Cluster 1 to have smaller amplitude when compared with other clusters.

Conversely, Cluster 4 exhibits the highest average ozone concentration in the afternoon (Feature 4), and the greatest average increase/slope from 06:00 AM to 09:00 AM (Feature 3) and mean ozone mixing ratios reaching up to 80 ppbv by noon. Cluster 4 days are generally sunny and stagnant in the HGB region. These conditions promote ozone production and accumulation, and rapidly increase ozone concentrations by the hour. While Cluster 3 seems to be similar to Cluster 4, the two clusters are each unique. For Cluster 3, Feature 1 (i.e., rate of decrease in ozone concentrations from 00:00 AM to 05:00 AM) has a lower slope than in Cluster 4; that is, nighttime ozone concentrations in Cluster 3 not only start lower but also decrease at a slower pace which could be an indication of stagnation in Cluster 4. Minimum morning ozone (Feature 2) in Clusters 3 and 4 is identical, but differing rates of increase during the day (Feature 3) lead to much higher afternoon ozone (Feature 4) in Cluster 4. This makes Cluster 4 most likely associated with days of high ozone and possible exceedances. Cluster 2 has the most days, accounting for 54% of the total days, and its diurnal pattern most closely follows that of the average diurnal pattern for all three summers combined (Figure 3a). Figure 3a also demonstrates how the diurnal patterns of Clusters 1 and 4 are the least similar to the average.

To further illuminate how the four features discussed in section 2.3 act to characterize the clusters, Figure 3b shows the mean diurnal ozone variation for each cluster using normalized hourly ozone data. The normalization was conducted by every hour for the days in all 3 years so that each hour has zero mean and unit variance of ozone. In the normalized plot (Figure 3b), Cluster 2 is almost a straight line with hourly values close to zero, which again demonstrates it is the common, base diurnal pattern. The normalized data show more clearly where Clusters 3 and 4 are alike and different. The greatest similarity between them is in Feature 2, the minimum ozone. Cluster 3 has a slight negative slope (Feature 1) compared to Cluster 4 which has a greater negative slope. Feature 1 in Cluster 3 actually more closely imitates that of Cluster 2, just at an overall higher ozone level. Clusters 3 and 4 differ in Feature 3 with Cluster 4 having a greater positive slope. Feature 4 for Clusters 2, 3, and 4 are all parallel but at different levels of ozone, ranging with Cluster 2 at the lowest and Cluster 4 at the highest. Cluster 1 is the most distinct diurnal pattern of all, with a positive slope (Feature 1), the highest minimum (Feature 2), a negative slope (Feature 3), and the lowest maximum (Feature 4). This is completely contrasting Cluster 4. The normalized diurnal patterns demonstrate that Clusters 1 and 4 deviate most from the common diurnal pattern, Cluster 2, and are thus the more extreme clusters as mentioned before. They also illustrate the clusters are unique not just in terms of the maximum ozone of the day (i.e., Feature 4) but also the other features which depict the overall diel cycle.

3.2. Advantages of Clustering Ozone Diurnal Patterns

The clustering based on ozone diel pattern can be connected to MDA8 ozone but has its unique advantages. To illustrate the connection, we averaged MDA8 ozone data by day from each of the 19 HGB sites and analyzed by cluster and year. Table S2b also shows the MDA8 ozone values for each year and cluster but is averaged over the 19 HGB sites. In Figures 4a–4c, the already established cluster numbers are assigned to the MDA8 ozone data by each day for all three summers analyzed. For all the summers, Cluster 1 proves to be the cluster with the lowest MDA8 ozone and Cluster 4 with the highest MDA8 ozone, which aligns with the diurnal surface ozone cluster results. All the Cluster 4 days (15 days total) had MDA8 ozone occurrences above the 70 ppbv exceedance threshold of the NAAQS. Cluster 3 follows with about 69% of the days (33 days total) and Cluster 2 with just about 9% of the days (14 days total) exceeding 70 ppbv. Cluster 1 only has 1 day exceeding 70 ppbv and that day occurred in 2011, the extreme drought summer. Figures 4d–4f summarize the MDA8 ozone results from the 19 HGB sites averaged by cluster for all three summers. The MDA8 ozone for all three summers increases by cluster number as expected. The spread of the MDA8 ozone is also unique by summer. 2014 (Figure 4e) has the greatest spread of days for Cluster 1 (ranging from 30–49 ppbv) and for Cluster 2 (ranging from 40–58 ppbv). Conversely, 2011 (Figure 4d) and 2015 (Figure 4f) have less fluctuating MDA8 ozone for Clusters 1 and 2. Compared to 2011 and 2015, the MDA8 ozone averaged over all sites (Table S2b) for 2014 is about 10 ppbv lower for Cluster 3 and below the exceedance threshold, although some site specific MDA8 ozone exceedances did occur (Figures 4b and 4e). The mean MDA8 ozone for 2011 is slightly higher than 2015, except for Cluster 3. The cluster averages for Cluster 1 is where 2011 and 2015

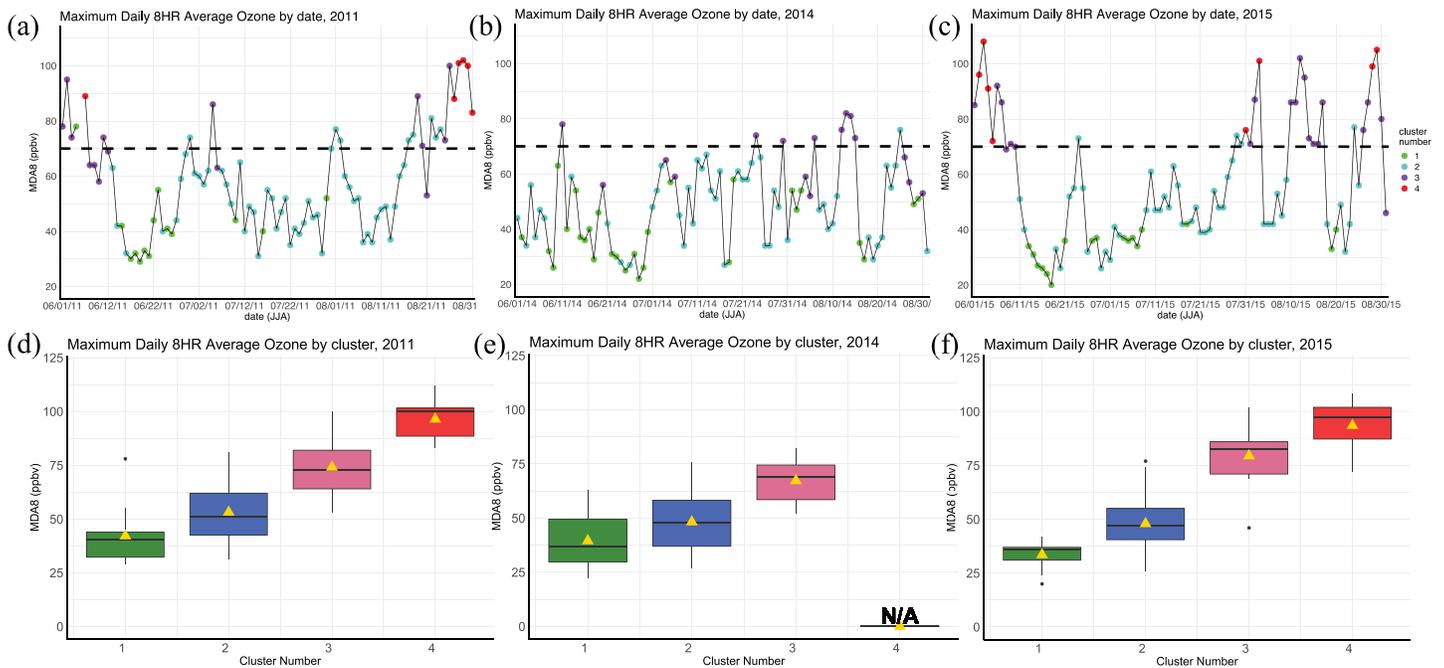


Figure 4. Time series of single-site MDA8 ozone for each day labeled by cluster number for each year: (a) 2011, (b) 2014, and (c) 2015. Black dashed line represents ozone exceedance threshold (MDA8 ozone >70 ppbv). Single-site MDA8 ozone boxplot distribution by cluster for (d) 2011, (e) 2014, and (f) 2015. Mean depicted by gold triangle.

differ with 2011 having a much higher MDA8 ozone average (Table S2b). We will analyze the differences of the years and possible meteorological influences more in sections 3.3 and 4.

As the four clusters appear to follow the rankings of MDA8 ozone, one could argue that splitting MDA8 or daily-peak ozone into quantiles ranging from low to high ozone would divide the days into similar clusters. To emphasize the uniqueness of clustering based on the diurnal pattern of ozone, we divided MDA8 ozone from the same 275 days of data into four quantiles (highest ozone, moderate ozone, low ozone, and lowest ozone) while ignoring the other hours of the day. In Figures S4a–S4b, we show the mean diurnal patterns for the four MDA8 ozone quantiles and the normalized patterns. There are several differences when comparing to the original clustering analysis. First, using the four quantiles of MDA8 ozone, the morning hours are barely distinctive between the groups, with Quantiles 2 and 3 following the exact same shape, Quantile 1 only slightly different, and Quantile 4 only different by less than 10 ppbv. Second, while the quantile approach separates the afternoon hours from low to high ozone, the highest ozone quantile has almost 15 ppbv lower ozone concentrations in the afternoon than that of Cluster 4 from the cluster analysis. This indicates that the inclusion of ozone behavior during other hours of the day can better characterize high ozone days than using the MDA8 ozone alone. This is due to the fact that the cluster analysis splits the days into groups and does not consider the amount in each; the higher ozone clusters are better represented than in the divided quantiles. Third, the divided quantiles assigned 9 days in 2014 to the highest quantile (4) unlike the clustering analysis which did not. This can mask that 2014 was the cleaner year. Overall, the quantiles based on MDA8 ozone (or similarly defined peak ozone metrics) will mute the dissimilarities within the different ozone groups, especially in the morning hours, and weakens the representation of the afternoon high ozone hours.

To further emphasize the importance of using the entire diurnal pattern to cluster ozone, we conducted a clustering analysis using only the first two features concerning the hours before sunrise (Features 1 and 2). Out of the 275 days total, only 44 of those days from the new clusters did not match the original clusters. Specifically, 87% of the days assigned in the new Cluster 1 matched the original Cluster 1, 90% from Cluster 2, 58% from Cluster 3, and 77% from Cluster 4.

We applied the same clustering process using the last two features after sunrise (Features 3 and 4). The results showed that 48% of the days assigned to the after-sunrise Cluster 1 matched the original Cluster 1, 44% from Cluster 2, 73% from Cluster 3, and 72% from Cluster 4. The match is overall weaker for Clusters

1 and 2 than using the before-sunrise features but better or similar for Clusters 3 and 4. This is because Cluster 3 and Cluster 4 are more distinguished in the afternoon hours than Cluster 1 and Cluster 2. The results from the after-sunrise clusters somewhat echo the argument made from the equally divided peak ozone quantiles in that the after sunrise/peak ozone hours are not enough to clearly capture the large variability of ozone. We conclude that the entire diurnal pattern of ozone has more value than just the peak afternoon hours in characterizing ozone variability for the region, and there is a clear linkage from the before-sunrise ozone behavior to that of after-sunrise hours and MDA8 ozone.

3.3. Interannual Differences by Cluster

As the clustering was conducted to the three summers as a whole, the clustering profiles for specific years can reveal the year-to-year differences in ozone. Figure 5 shows cluster-mean diurnal ozone variations for each year. We have already pointed out that one of the major disparities among the years is that 2014 does not have any days fall under Cluster 4. The highest ozone days in 2014 (Figure 5b) are in Cluster 3 where the mean diurnal pattern reaches just to about 50 ppbv. This is lower than the Cluster 3 average for the other 2 years (Figures 5a and 5c) which both reach up to about 58 and 64 ppbv. All 3 years have a very similar Cluster 2, the most common cluster. The shape of the Cluster 1 diurnal pattern is also consistent across the 3 years, although Cluster 1 in 2015 has a generally lower average ozone increase (Feature 3) and maximum (Feature 4) than the other 2 years. Thus, the main reason why 2014 was a low ozone year is due to lack of Cluster 4 and lower ozone in Cluster 3. Where 2011 and 2015 deviate the most is between Cluster 3 and 4. In Table S1, Cluster 4 in 2011 reaches a slightly higher average maximum (Feature 4, 75 ppbv) than 2015 (73 ppbv), whereas Cluster 3 in 2011 reaches a lower average maximum (Feature 4, 55 ppbv) than 2015 (63 ppbv). This creates a difference between Clusters 3 and 4 of 20 ppbv for 2011 as opposed to 10 ppbv for 2015. 2011 also has a lesser difference between Clusters 2 and 3, 22 ppbv, than 2015, 33 ppbv. Therefore, a larger gap exists (Figures 5a and 5c) between Clusters 2 and Cluster 3 in 2015 than in 2011. Cluster 3 in 2011 has a larger negative slope (Feature 1, Table S1a) and reaches a lower minimum (Feature 2, Table S1a) than 2015. This could explain the difference later in the day (Features 3 and 4) between the 2 years.

In terms of when each cluster occurs more frequently, notable dissimilarities and some similarities exist amid the summers. 2011 (Figure 4a) and 2015 (Figure 4c) are most similar in that Clusters 3 and 4 days occur primarily during June and August and Clusters 1 and 2 days occur primarily during July with few MDA8 ozone exceedances occurring in July. The HGB region usually experiences the highest summer surface ozone concentrations during June and August (Nielsen-Gammon et al., 2005; Wang et al., 2016). In July, the effects of the Bermuda High (Wang et al., 2016) and LLJ are the strongest (Doubler et al., 2015), and background ozone reaches a minimum due to these strong circulation processes (Nielsen-Gammon et al., 2005). As a result, surface ozone concentrations usually decrease during the month of July which is demonstrated in Figures 4a and 4c. As these meteorological processes are less influential in August, 2011 and 2015 had few or zero Cluster 1 days in that month. 2014 (Figure 4b) was the generally clean year so the clusters are more distributed throughout the 3 months. For 2014, we expect the meteorological processes to not vary much and be more consistently impactful during all 3 months which could explain this uncommon MDA8 distribution within the clusters and the months (Figures 4b and 4e).

4. Synoptic Meteorological Processes by Cluster

To fully understand the underlying differences between the clusters, we describe the synoptic meteorological conditions for each cluster in this section. We also evaluate the impact of certain weather conditions brought on by strong or weak air circulation patterns on daily ozone concentrations based on previous studies and cases. We focus on the LLJ and Bermuda High, which are known to affect daily ozone variations in the HGB region.

4.1. Cluster Circulation Patterns

The cluster-mean air circulation results extracted from ERA-Interim reanalysis are presented in Figures 6a–6k for 2011, 2014, and 2015. It is clear from the plots that each cluster and summer corresponds to expected meteorological processes and weather conditions specific to the HGB region. Clusters 1 and 2 (Figures 6a–6f) have strong winds, which extend in a clockwise rotation from a high pressure off the east coast of the United States (the Bermuda High). These strong winds travel right through the HGB region and reduces the rate of

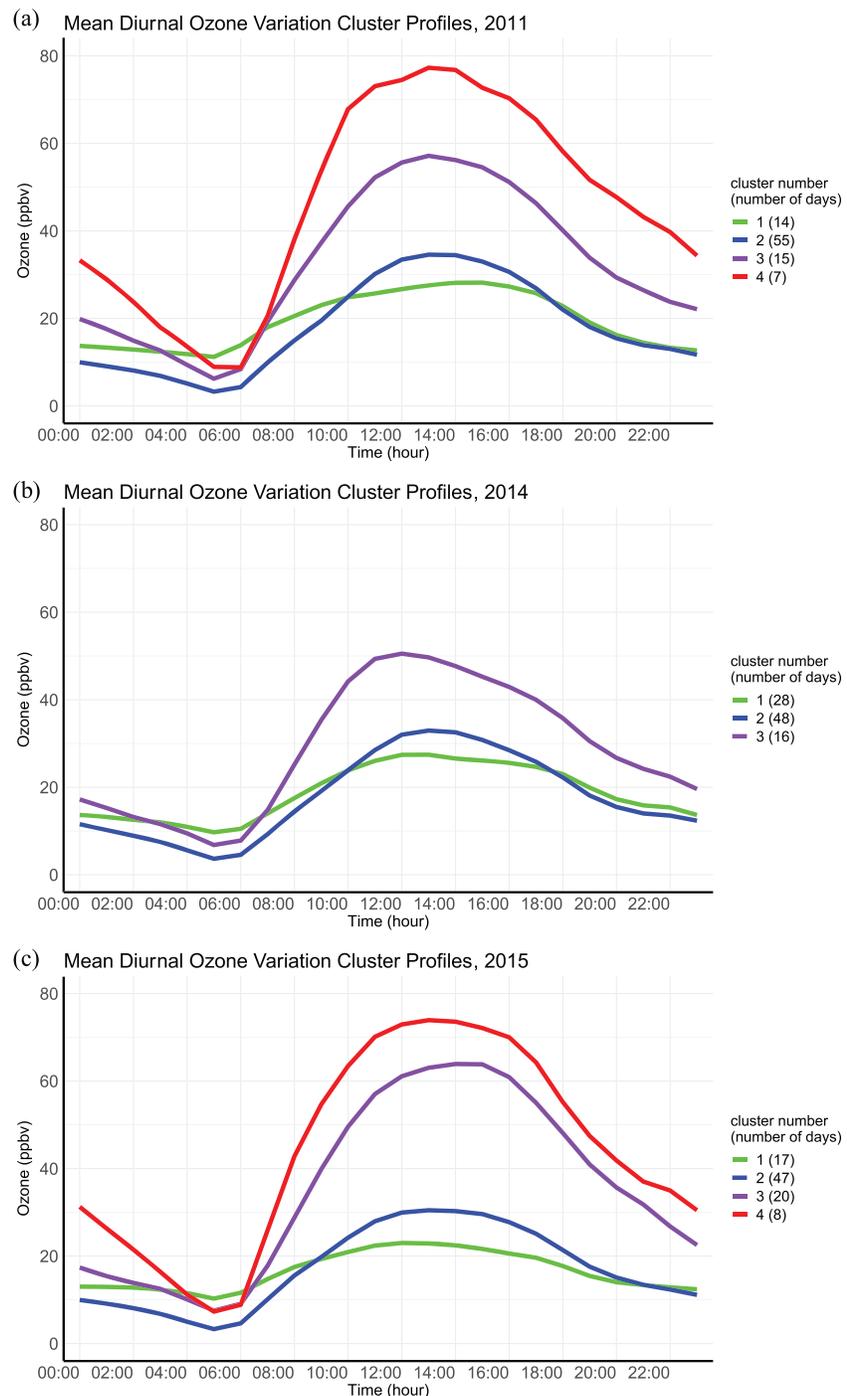


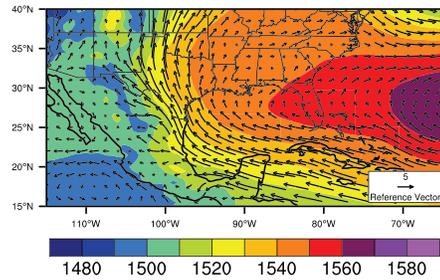
Figure 5. Mean diurnal variations of surface ozone for each cluster and each year (a) 2011, (b) 2014, and (c) 2015.

increase in surface ozone (i.e., Feature 3) through the dispersion of ozone and ozone precursors. For all 3 years, the Bermuda High extension is farther west in Clusters 1 and 2 than in Clusters 3 and 4. This is the reason why Clusters 1 and 2 experience a weak, less variable diurnal surface ozone pattern than Clusters 3 and 4 (Figure 3a).

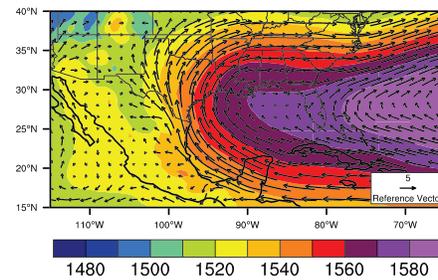
By contrast, Clusters 3 and 4 (Figures 6g–6k) do not have the strong winds that are present in Clusters 1 and 2. It is apparent there is a distinct anticyclonic flow pattern over the HGB region, but this flow appears to be associated with another high-pressure system emerging north or northeast of the HGB region (most

Cluster 1

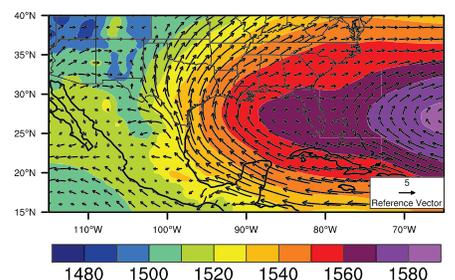
(a) 2011



(b) 2015

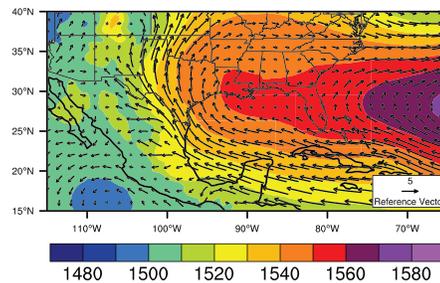


(c) 2014

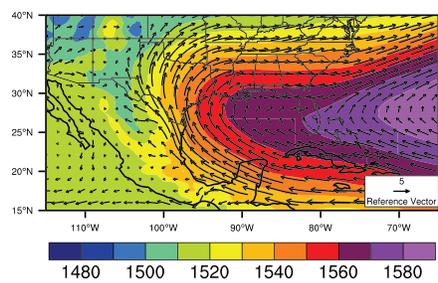


Cluster 2

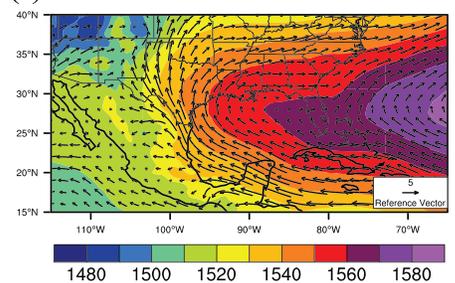
(d) 2011



(e) 2015

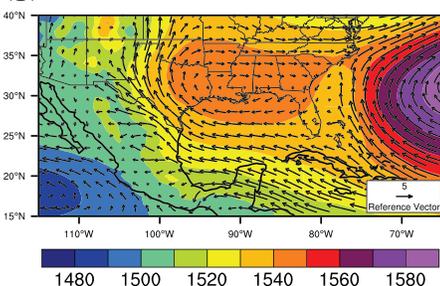


(f) 2014

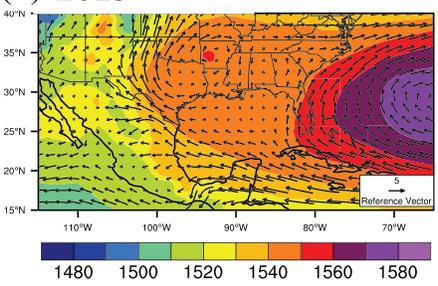


Cluster 3

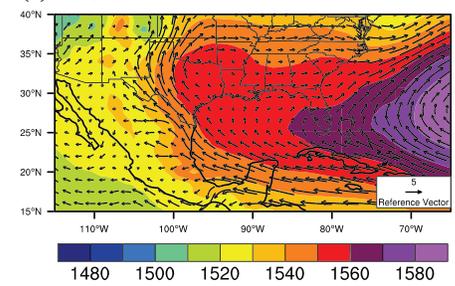
(g) 2011



(h) 2015

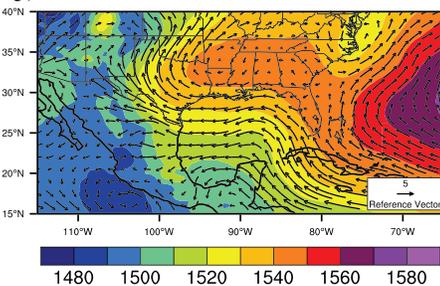


(i) 2014



Cluster 4

(j) 2011



(k) 2015

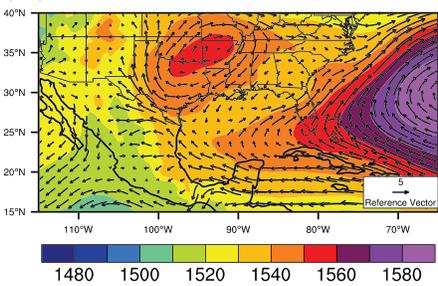


Figure 6. Mean wind fields (black arrows) and geopotential height (contours) at 850 hPa for each established cluster to depict the mean circulation patterns. Cluster 1 (a) 2011, (b) 2015, and (c) 2014. Cluster 2 (d) 2011, (e) 2015, and (f) 2014. Cluster 3 (g) 2011, (h) 2015, and (i) 2014. Cluster 4 (j) 2011 and (k) 2015. Data from ERA-Interim reanalysis.

noticeable in Figures 6g, 6h, 6j, and 6k), unlike the Bermuda High off the east coast of the United States. This second high pressure may be considered a “migratory anticyclone,” as described by Davis et al. (1998). As the Bermuda High begins its retreat eastward (usually occurs in August) or becomes weakened, this sets the stage for the migratory anticyclone to develop or move into the CGS. Indeed, August typically has a greater number of Clusters 3 and Cluster 4 days (Figures 4a–4c). This migratory anticyclone creates stagnant weather conditions over the HGB region, with low winds, high temperatures, clear skies, and subsidence, all of which facilitate ozone production and explains the higher levels of surface ozone concentrations, as well as the more variable diurnal patterns in Clusters 3 and 4. With a lack of strong synoptic-scale air circulations, pollutants are not dispersed efficiently, leading to high ozone production and accumulation, which explains the continuously high surface ozone concentrations in Feature 4 in the diurnal patterns for Clusters 3 and 4.

These circulation plots allow us to investigate the disparities among the years. It is clear that 2011 (Figures 6a, 6d, 6g, and 6j or the first column) is an anomalous year compared to the other years with the Bermuda High on average located farther east in 2011. The absence and lack of westward extension of the Bermuda High in 2011 may be associated with abnormal circulation patterns pertaining to the drought conditions during this summer. For 2011, the 850 hPa heights are overall lower throughout the domain in all four clusters than in the other years. This implies that identifying the Bermuda High extension by a specific height value as adopted in earlier studies (Li et al., 2011; Shen et al., 2015; Wang et al., 2016) may be less effective for abnormal years such as 2011. For 2014 and 2015 (second and third column in Figure 6), the circulation plots show the extension of the Bermuda High is similar for Clusters 1 and 2 for both years. Where the 2 years differ is in Cluster 3 where the westward extension of the Bermuda High is still very present for 2014 but not for 2015. For Cluster 3 in 2015 the Bermuda High is located farther east, and a migratory cyclone is formed over the CGS. Coincidentally, 2015 MDA8 ozone and the diurnal pattern ozone concentrations are drastically higher in contrast to the lower ozone values found in 2014 for Cluster 3 (Figures 4e and 4f and Figures 5b and 5c). The mean diurnal patterns (Figure 5) show that 2015 Cluster 3 ozone is also higher than that in 2011, especially for the last three features. Comparing the circulation plots, we conclude that the reason for this difference is that the migratory anticyclone in Cluster 3, 2015 (Figure 6h) is stronger than the one forming in Cluster 3, 2011 (Figure 6g).

4.2. LLJ

Figure 7 shows the LLJ percentages (LLJ%) for each cluster number. As mentioned before, the LLJ is a pressure gradient flow that dissipates ozone by advection or redistributes ozone through increased vertical wind shear (Hu et al., 2013). The LLJ% clearly decreases from Cluster 1 to Cluster 4 for all three summers. For example, in 2015 Cluster 1 has an average LLJ% of 21.52% and Cluster 2 of 6.53% (Figure 7c). This contributes to lower surface ozone for these cluster's diurnal patterns. In contrast, Clusters 3 and 4 both show zero LLJ occurrences. The lack of LLJ does not facilitate the dispersion of surface ozone and ozone precursor as would occur when LLJ is present at night. This leads to higher surface ozone concentrations the next day. The LLJ% results are similar for 2011 and 2014 (Figures 7a and 7b). This supports the clustering analysis, showing that more frequent LLJ occurrences link to lower ozone concentrations in the HGB region, as perceived in Clusters 1 and 2.

Previous work has demonstrated that 2011 was an anomalous drought year for the HGB region. Pu et al. (2016) suggested that the 2011 drought conditions arose partly because a strong LLJ over the southeast Texas co-occurred with a weak LLJ over the southern Gulf of Mexico, leading to a moisture divergence which transported most of the moisture to the central and northern Great Plains leaving dry conditions in Texas. Indeed, 2011 (Figure 7a) has the greatest LLJ occurrences out of all 3 years which supports the fact that strong cases of LLJ transport a significant amount of moisture from the HGB region to the Great Plains, leaving the HGB region in drier conditions (Pu et al., 2016). Where the 2011 LLJ% differs most is Cluster 3, having an average LLJ% of 7.16%, whereas the other 2 years had a 1.07% and 0% LLJ frequency for that cluster. This can also be associated to the circulation plots which show the Bermuda High is farther east in 2011 than in 2014 and 2015 (section 4.1). Thus, the lack of synoptic influence from the Bermuda High allows for the placement of the strong LLJ over southeast Texas and a weak LLJ over the southern Gulf of Mexico, with the nose of the jet hovering over the HGB region. This specific placement of the LLJ can be one of the factors that led to the drought conditions which allows Cluster 3 in 2011 to have a more

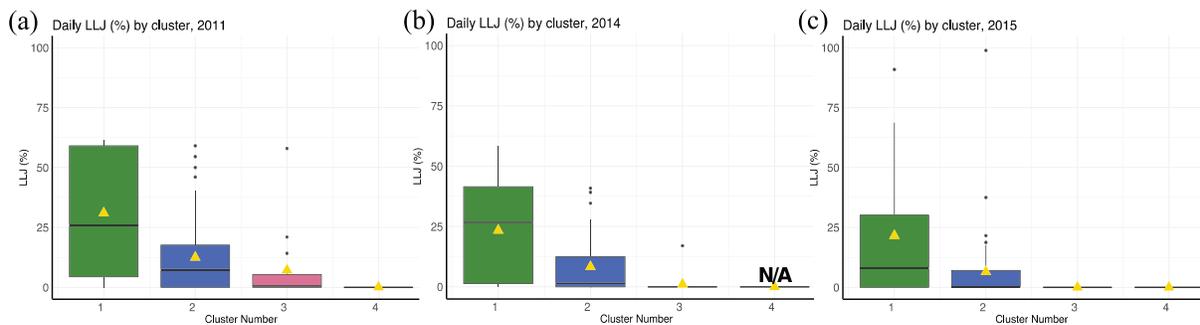


Figure 7. Mean low-level jet percentage by cluster for (a) 2011, (b) 2014, and (c) 2015. Mean depicted by gold triangle.

frequent LLJ% but still have a higher ozone concentration. This is a different case than that of 2014 and 2015 in which the majority of the clusters align with the expected effects of the LLJ. This indicates that the simplistic explanation of the LLJ does not hold up as strongly with an abnormal year such as 2011. The difference in LLJ% between 2014 and 2015 is not substantial, and thus, we cannot draw a conclusion as to what extent the interannual variability in LLJ during nondrought years would contribute to the interannual variability of ozone.

4.3. Bermuda High Analysis

The circulation plots for the clusters in 2014 and 2015 are consistent with our previous analysis of the relationship between the Bermuda High and summertime ozone interannual variability in the HGB region (Wang et al., 2016). Similar to Shen et al. (2015), Wang et al. (2016) identified a daily threshold for the BH-Lon over the HGB region that marks two surface ozone categories which link to distinctive meteorological regimes. The first regime is a clean maritime southeasterly flow to the southeastern Texas which stems from the western edge of the Bermuda High as it moves westward toward the HGB region. This regime leads to lower surface ozone. The other regime is rather the opposite which is a lack of southeasterly flow when the Bermuda High retreats eastward and is associated with a high pressure system over the southeastern United States. Figure 8 shows the bimodal distribution of MDA8 ozone as a function of daily BH-Lon from JJA 2004–2015 which mirrors the threshold from Wang et al. (2016). The lowest ozone concentrations occur when the BH-Lon ranges from 96.2°W to 83.7°W. This is consistent with 2014 in which the extension of the Bermuda High (Figures 6c, 6f, and 6i) is within the “low ozone” range in Figure 8 and consequently hardly any of the three cluster’s MDA8 ozone averages exceed 70 ppbv for that summer (Figures 4b and 4e). This range is also consistent in Clusters 1 and 2 for 2015. Ozone exceedances (MDA8 ozone >70 ppbv) begin to occur from BH-Lon values of 78.7°W and eastward as well as 101.2°W and westward. This is consistent with the circulation plots for Clusters 3 and 4 (Figures 6b, 6e, 6h, and 6k) and the MDA8 ozone (Figure 4c) for 2015. Because of the abnormal circulation pattern in 2011 associated with the extreme drought, the extension of the Bermuda High in the circulation plots for all four clusters in 2011 fall eastward of 78.7°W. Figure 8 demonstrates where ozone air quality in the HGB region benefits the most from the westward extension of the Bermuda High and where the lack of the extension as well as the overextension can permit the buildup of surface ozone.

To analyze BH-Lon and LLJ% among the four clusters on a daily scale, we arbitrarily selected 1 day from each cluster in 2015 (Figure 9). June 13 represents Cluster 1, June 11 represents Cluster 2, June 8 represents Cluster 3, and June 3 represents Cluster 4. June 13 (Cluster 1) has a BH-Lon of 89°W and LLJ% of 7.39 which is a westward extending Bermuda High (Figure 10a) and strong LLJ consistent with what we would expect for a day in Cluster 1 with the least varying diurnal pattern (Figure 9b). June 11 (Cluster 2) has a BH-Lon of 81°W and a 0% LLJ and thus higher ozone than June 13 (Cluster 1). The calculated BH-Lon for June 8 (Cluster 3) is 86°W based on the 1,560 gpm isopleth. Yet looking at Cluster 3 in Figure 10c, the BH-Lon appears to be closer to 72°W based on the usual Bermuda High shape. Although the 1,560 gpm value does indeed reach 86°W, the irregular shape of the Bermuda High (Figure 10c) is different from the mean circulation pattern for that cluster (Figure 6h). The BH-Lon value of 72°W is more indicative of the location of the Bermuda High center and is also more consistent with the Bermuda High in the average circulation plots for

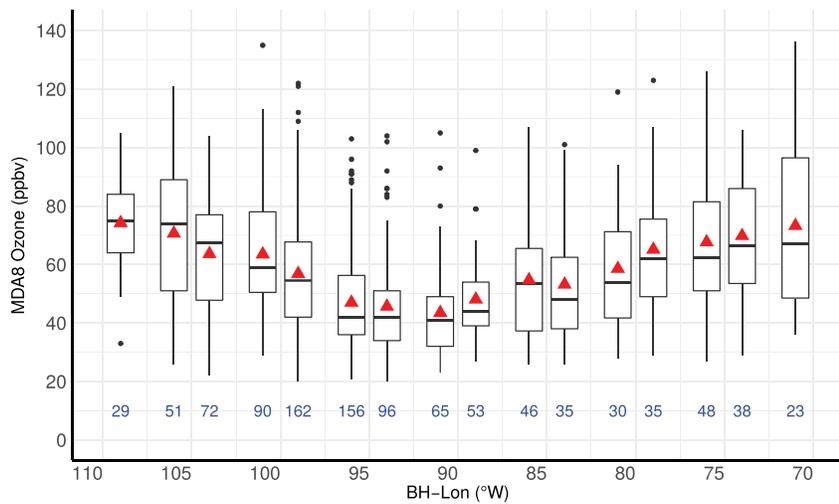


Figure 8. MDA8 ozone observed in HGB area for binned longitudes of the western extent of the Bermuda High (BH-Lon), during JJA 2004–2015. The box shows the 25th and 75th percentiles along with the median (horizontal line) and mean (red triangle), and the whiskers show the 5th and 95th percentiles. The number of observations for each bin is depicted at the 10 ppbv line in blue. Kruskal-Wallis nonparametric 1-way analysis of variance shows significance at $p < 0.0001$, indicating BH-Lon clearly influences MDA8 ozone in the HGB area.

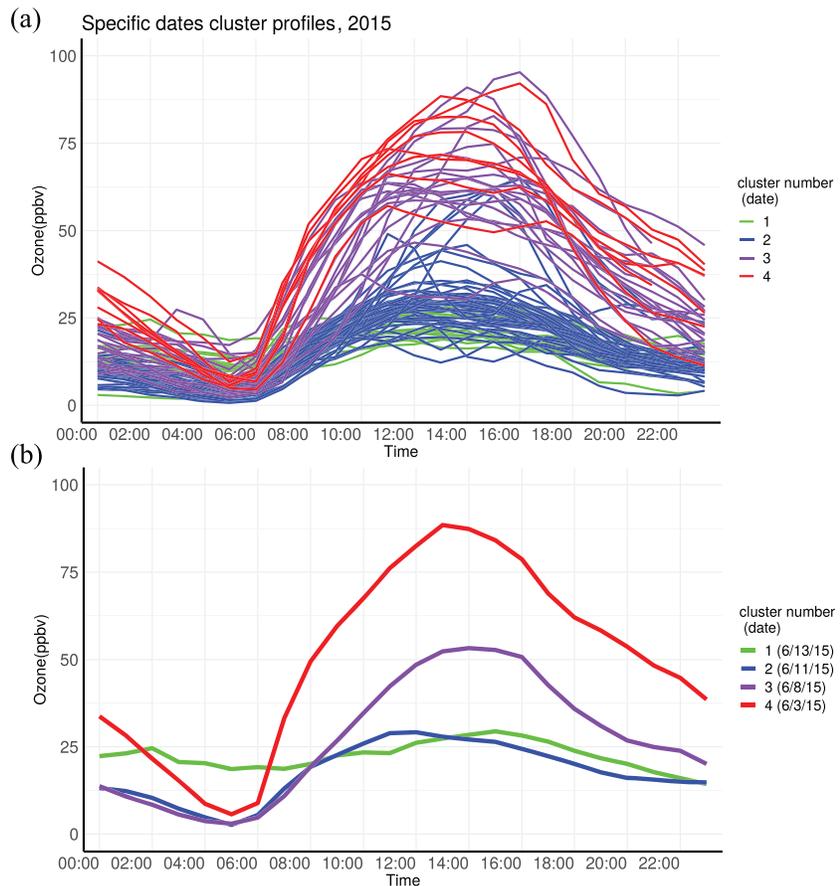


Figure 9. (a) Diurnal surface ozone patterns of all days in JJA for 2015. (b) Specific dates for each cluster: June 13, 2015 represents Cluster 1; June 11, 2015 represents Cluster 2; June 8, 2015 represents Cluster 3; and June 3, 2015 represents Cluster 4.

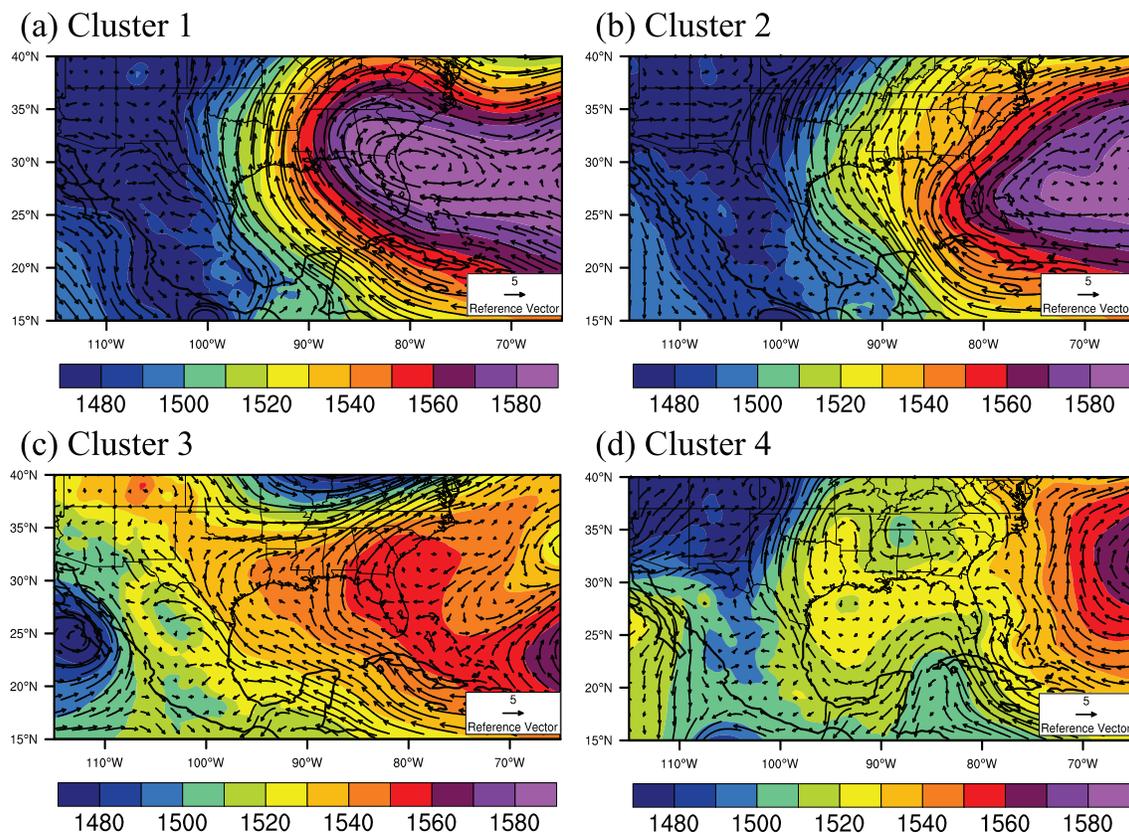


Figure 10. Mean wind fields (black arrows) and geopotential height (contours) at 850 hPa for each 4 days from each cluster to depict the mean circulation patterns. (a) Cluster 1, June 13, 2015; (b) Cluster 2, June 11, 2015; (c) Cluster 3, June 8, 2015; and (d) Cluster 4, June 3, 2015. Data are from ERA-Interim reanalysis.

Cluster 3 in 2015 (Figure 6h). This suggests the need for caution on the criteria and calculation of the BH-Lon on a daily basis in the event of irregular Bermuda High patterns, as we alluded to when discussing the lower 850 hPa heights in 2011. Furthermore, it appears the calculated BH-Lon with a fixed isoline (1,560 gpm) could be confounded with a developing migratory anticyclone that contains the same isoline instead of the true westward extension of the Bermuda High. Determining the BH-Lon at the daily scale thus warrants the need for a future study. The BH-Lon for June 3 (Cluster 4) is 71°W and a 0% LLJ which is expected for a Cluster 4 day. There is a weak migratory anticyclone in Cluster 4 (Figure 10d), which explains the more varying diurnal surface ozone pattern and the faster and stronger buildup of ozone (Feature 3) in Figure 10b. The daily BH-Lon for this day correctly depicts the westward extension of the Bermuda High because the migratory anticyclone developing over the HGB region falls well short of the 1,560 gpm requirement.

5. Conclusion

We used the K-means clustering method to organize diurnal ozone patterns in the HGB region for JJA for the summers 2011, 2014, and 2015 and investigated the effects of synoptic meteorological patterns on hourly surface ozone concentrations. We established four features for the clustering and generated four clusters of diurnal ozone patterns specific to the HGB region during JJA. The four clusters were classified as unique clusters. Cluster 1 is a less common diurnal pattern having generally low ozone with the least varying diurnal pattern, almost appearing as a straight line in all three summer results. Cluster 2 is the most common diurnal pattern (contained the most days out of the four clusters and was closest to the average diurnal pattern for all three years). Clusters 3 and 4 are higher ozone clusters, containing most of the ozone exceedance days. Cluster 3 is categorized as a more common high ozone diurnal pattern expected for a highly polluted region such as HGB. Where Cluster 3 differs from Cluster 4 is in Features 1 (rate of change in ozone from

00:00 AM–05:00 AM), 3 (rate of change in ozone from 06:00 AM–09:00 AM), and 4 (ozone mixing ratio at its maximum averaged from 12:00 PM–04:00 PM). Peak afternoon ozone concentrations in Cluster 3 are about 17 ppbv lower than in Cluster 4. Cluster 4 has the most varying diurnal pattern with the highest surface ozone concentrations and greatest slopes of ozone changes (Features 1 and 3), therefore categorized as the extreme cluster.

We link synoptic meteorological processes specific to the HGB region, the Bermuda High and LLJ, with the ozone clusters to establish their effects on diurnal surface ozone. The cluster averaged circulation patterns showed that for Cluster 1, the westward extension of the Bermuda High was closer to the HGB region, bringing in clean Gulf air and resulting in lower surface ozone, whereas the higher ozone values of Clusters 3 and 4 were linked with a distant easterly Bermuda High. This was consistent with the results for 2014 and 2015 but not as convincing for 2011 which had a farther east Bermuda High for all four clusters. We note that the irregularity of the farther east Bermuda High in 2011 is likely a factor contributing to the drought conditions for this summer. The circulation patterns also show that as the Bermuda High retreats back eastward or becomes weakened, a different high-pressure system (migratory anticyclone) emerges to influence the HGB region which links with the high surface ozone in Clusters 3 and 4. This migratory anticyclone is present in Clusters 3 and 4 for 2011 and 2015. The LLJ also links with the cluster analysis showing strongest LLJ occurrences during Clusters 1 and 2 and nearly zero occurrences for the days in Clusters 3 and 4. The LLJ% for Cluster 1 in all 3 years is significantly higher than the other clusters which links to the extreme “straight line” diurnal pattern established in this cluster. Summer 2011 has overall the highest LLJ% yet also has a farther east Bermuda High in all four clusters. The placement of the Bermuda High (farther east) influenced the location and intensity of the LLJ for the HGB region which probably contributed to the drought conditions for 2011. This analysis proves to be useful in diagnosing the roles of meteorological processes in driving daily and interannual ozone trends.

Compared to the simple approach of splitting the three summers into four quantiles ranging from low to high MDA8 ozone, we found that using the clustering analysis better highlights the higher ozone concentrations found later in the afternoon and also emphasizes the major disparities in the morning hours. This is a direct result of the uneven distribution of the clusters which better exemplifies the unique diurnal ozone patterns than the equally divided quantiles. In using the clustering analysis on only the first two features before sunrise and then again on only the last two features after sunrise, we tested the importance of the entire diurnal ozone pattern when analyzing ozone variability. The results revealed that the before-sunrise hours have greater influence on the clustering process than the after-sunrise hours and a clear linkage with the two different parts of the day.

The present work focuses on the synoptic meteorological processes but acknowledges other processes such as sea breeze and PBL dynamics, which could have a large effect on the diurnal pattern of surface ozone at individual sites. Using the regional-mean ozone, we remove some between-site differences caused by localized effects which are challenging to quantify. Thereby, we focus on the linkage of changes in diurnal ozone patterns with synoptic meteorological processes that affect the region. Future work would be needed to examine the effects of these processes on hourly surface ozone, which would involve a more detailed analysis, for example, by examining the diurnal ozone patterns at the site level. The study also acknowledges the need to re-evaluate the daily-scale BH-Lon calculations, considering the possible confounding effects of the migratory anticyclone. Future work is also needed to examine the impacts of the migratory anticyclone on the HGB region as well as the westward extension of the Bermuda High.

This analysis can be applied to other regions to better comprehend the diurnal surface ozone cycle variability of any region and the effects of region-specific meteorological processes on the diurnal surface ozone cycle. This analysis also helps to identify the annual differences in diurnal ozone concentrations and their associated meteorology. MDA8 (peak) ozone is often heavily studied but less has been said about other hours of the day. The morning hours of a day can possibly tell us how the rest of the day's ozone concentrations will evolve and warrant more attention. The daily and hourly variations of surface ozone concentrations warrant further research to better understand local meteorological effects on surface ozone at different times of the day and should be addressed alongside peak ozone times for the sake of mitigation strategies.

Acknowledgments

This research was partly supported by Texas Commission on Environmental Quality (TCEQ) (582-18-81339). The authors acknowledge the European Centre for Medium-Range Weather Forecasts (ECMWF) for providing the ERA-Interim reanalysis data, Environmental Protection Agency (EPA) Air Quality Station (AQS) AirData for providing the hourly and MDA8 ozone data, and North American Regional Reanalysis (NARR) for the vertical wind profile data. The findings, opinions, and conclusions are the work of the authors and do not necessarily represent findings, opinions, or conclusions of the EPA or TCEQ. The authors would like to thank the reviewers for their numerous valuable comments that significantly helped improve the article. The data used for this study are available online at <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi%3A10.7910%2FDVN%2FBUAIEV&version=DRAFT>.

References

Aksoy, S., & Haralick, R. M. (2001). Feature normalization and likelihood-based similarity measures for image retrieval. *Pattern Recognition Letters*, 22(5), 563–582. [https://doi.org/10.1016/s0167-8655\(00\)00112-4](https://doi.org/10.1016/s0167-8655(00)00112-4)

Balashov, N. V., Thompson, A. M., & Young, G. S. (2017). Probabilistic forecasting of surface ozone with a novel statistical approach. *Journal of Applied Meteorology and Climatology*, 56(2), 297–316. <https://doi.org/10.1175/jamc-d-16-0110.1>

Banta, R. M., Senff, C. J., Nielsen-Gammon, J., Darby, L. S., Ryerson, T. B., Alvarez, R. J., et al. (2005). A bad air day in Houston. *Bulletin of the American Meteorological Society*, 86(5), 657–670. <https://doi.org/10.1175/BAMS-86-5-657>

Baumgardner, R. E., & Edgerton, E. S. (1998). Rural ozone across the eastern United States: Analysis of CASTNet data, 1988–1995. *Journal of the Air & Waste Management Association*, 48(8), 674–688. <https://doi.org/10.1080/10473289.1998.10463722>

Bezdek, J. C. (1981). *Pattern recognition with fuzzy objective function algorithms*. Plenum.

Charrad, M., Ghazzali, N., Boiteau, V., & Niknafs, A. (2014). NbClust: AnRPackage for determining the relevant number of clusters in a data set. *Journal of Statistical Software*, 61(6). <https://doi.org/10.18637/jss.v061.i06>

Darby, L. S. (2005). 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>

Davis, J., Eder, B., Nychka, D., & Yang, Q. (1998). Modeling the effects of meteorology on ozone in Houston using cluster analysis and generalized additive models. *Atmospheric Environment*, 32(14-15), 2505–2520. [https://doi.org/10.1016/S1352-2310\(98\)00008-9](https://doi.org/10.1016/S1352-2310(98)00008-9)

Doubler, D. L., Winkler, J. A., Bian, X., Walters, C. K., & Zhong, S. (2015). An NARR-derived climatology of southerly and northerly low-level jets over North America and coastal environs. *Journal of Applied Meteorology and Climatology*, 54(7), 1596–1619. <https://doi.org/10.1175/JAMC-D-14-0311.1>

Dunn, J. C. (1973). A Fuzzy Relative of the ISODATA process and its use in detecting compact well-separated clusters. *Journal of Cybernetics*, 3(3), 32–57. <https://doi.org/10.1080/01969727308546046>

Haman, C. L., Couzo, E., Flynn, J. H., Vizuete, W., Heffron, B., & Lefer, B. L. (2014). Relationship between boundary layer heights and growth rates with ground-level ozone in Houston, Texas. *Journal of Geophysical Research: Atmospheres*, 119, 6230–6245. <https://doi.org/10.1002/2013JD020473>

Hu, X.-M., Klein, P. M., Xue, M., Zhang, F., Doughty, D. C., Forkel, R., et al. (2013). Impact of the vertical mixing induced by low-level jets on boundary layer ozone concentration. *Atmospheric Environment*, 70, 123–130. <https://doi.org/10.1016/j.atmosenv.2012.12.046>

Jacob, D. J., & Winner, D. A. (2009). Effect of climate change on air quality. *Atmospheric Environment*, 43(1), 51–63. <https://doi.org/10.1016/j.atmosenv.2008.09.051>

Kaufman, L., & Rousseeuw, P. (1990). *Finding groups in data: An introduction to cluster analysis*. Hoboken, NJ: Wiley-Interscience.

Larose, D. T. (2005). *Discovering knowledge in data: An introduction to data mining*. Hoboken, NJ: Wiley-Interscience.

Lei, R., Talbot, R., Wang, Y., Wang, S., & Estes, M. (2018). Influence of cold fronts on variability of daily surface O₃ over the Houston-Galveston-Brazoria Area in Texas USA during 2003–2016. *Atmosphere*, 9, 159. <https://doi.org/10.3390/atmos9050159>

Li, W., Li, L., Fu, R., Deng, Y., & Wang, H. (2011). Changes to the North Atlantic subtropical high and its role in the intensification of summer rainfall variability in the Southeastern United States. *Journal of Climate*, 24(5), 1499–1506. <https://doi.org/10.1175/2010JCLI3829.1>

MacQueen, J. B. (1967). *Some methods for classification and analysis of multivariate observations*. *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability* (pp. 281–297). Berkeley, California: University of California Press.

Markowski, P., & Richardson, Y. (2010). Mesoscale meteorology in midlatitudes. 83–87. <https://doi.org/10.1002/9780470682104>

Mazzuca, G. M., Ren, X., Loughner, C. P., Estes, M., Crawford, J. H., Pickering, K. E., et al. (2016). Ozone production and its sensitivity to NO_x and VOCs: Results from the DISCOVER-AQ field experiment, Houston 2013. *Atmospheric Chemistry and Physics Discussions*, 16(22), 14463–14474. <https://doi.org/10.5194/acp-16-14463-2016>

Nielsen-Gammon, J.W., J. Tobin, A. McNeel, & G. Li (2005). A conceptual model for eight hour ozone exceedances in Houston, Texas Part I: Background ozone levels in eastern Texas. Houston Advanced Research Consortium (HARC), Project H-12

Pakalapati, S., Beaver, S., Romagnoli, J. A., & Palazoglu, A. (2009). Sequencing diurnal air flow patterns for ozone exposure assessment around Houston, Texas. *Atmospheric Environment*, 43(3), 715–723.

Pu, B., Dickinson, R. E., & Fu, R. (2016). Dynamical connection between Great Plains low-level winds and variability of central Gulf States precipitation. *Journal of Geophysical Research: Atmospheres*, 121, 3421–3434. <https://doi.org/10.1002/2015JD024045>

Sáenz, F., & Durán-Quesada, A. M. (2015). A climatology of low level wind regimes over Central America using a weather type classification approach. *Frontiers in Earth Science*, 3, 15. <https://doi.org/10.3389/feart.2015.00015>

Seguel, R. J., Mancilla, C. A., Rondanelli, R., Leiva, M. A., & Morales, R. G. E. (2013). Ozone distribution in the lower troposphere over complex terrain in Central Chile. *Journal of Geophysical Research: Atmospheres*, 118, 2966–2980. <https://doi.org/10.1002/jgrd.50293>

Shen, L., Mickley, L. J., & Tai, A. P. K. (2015). Influence of synoptic patterns on surface ozone variability over the eastern United States from 1980 to 2012. *Atmospheric Chemistry and Physics*, 15(19), 10,925–10,938. <https://doi.org/10.5194/acp-15-10925-2015>

Souri, A. H., Choi, Y., Li, X., Kotsakis, A., & Jiang, X. (2016). A 15-year climatology of wind pattern impacts on surface ozone in Houston, Texas. *Atmospheric Research*, 174-175, 124–134. <https://doi.org/10.1016/j.atmosres.2016.02.007>

Stauffer, R. M., Thompson, A. N., Oltmans, S. J., & Johnson, B. J. (2017). Tropospheric ozonesonde profiles at long-term U.S. monitoring sites: 2. Links between Trinidad Head, CA, profile clusters and inland surface ozone measurements. *Journal of Geophysical Research: Atmospheres*, 122, 1261–1280. <https://doi.org/10.1002/2016JD025254>

TCEQ: Texas Commission on Environmental Quality (2012), Houston-Galveston-Brazoria: Current attainment status, available at: <http://www.tceq.texas.gov/airquality/sip/hgb/hgb-status> (last access: 10 February 2018)

Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, 63(2), 411–423. <https://doi.org/10.1111/1467-9868.0029>

Tucker, S. C., Banta, R. M., Langford, A. O., Senff, C. J., Brewer, W. A., Williams, E. J., et al. (2010). Relationships of coastal nocturnal boundary layer winds and turbulence to Houston ozone concentrations during TexAQs 2006. *Journal of Geophysical Research*, 115, D10304. <https://doi.org/10.1029/2009JD013169>

Wang, Y., Jia, B., Wang, S., Estes, M., Shen, L., & Xie, Y. (2016). Influence of the Bermuda High on interannual variability of summertime ozone in the Houston-Galveston-Brazoria region. *Atmospheric Chemistry and Physics*, 16(23), 15,265–15,276. <https://doi.org/10.5194/acp-2016-596>

Wang, Y., Xie, Y., Chai, L., Dong, W., Zhang, Q., & Zhang, L. (2015). Impact of the 2011 southern US drought on ground-level fine aerosol concentration in summertime. *Journal of the Atmospheric Sciences*, 72(3), 1075–1093. <https://doi.org/10.1175/JAS-D-14-0197.1>

- Wang, Y., Xie, Y., Dong, W., Ming, Y., Wang, J., & Shen, L. (2017). Adverse effects of increasing drought on air quality via natural processes. *Atmospheric Chemistry and Physics*, *17*(20), 12,827–12,843. <https://doi.org/10.5194/acp-17-12827-2017>
- Weaver, S. J., & Nigam, S. (2008). Variability of the Great Plains low-level jet: Large-scale circulation context and hydroclimate impacts. *Journal of Climate*, *21*(7), 1532–1551. <https://doi.org/10.1175/2007JCLI1586.1>
- Zhong, S., In, H., & Clements, C. (2007). Impact of turbulence, land surface, and radiation parameterizations on simulated boundary layer properties in a coastal environment. *Journal of Geophysical Research*, *112*, D13110. <https://doi.org/10.1029/2006JD008274>
- Zhu, J., & Liang, X.-Z. (2013). Impacts of the Bermuda High on regional climate and ozone over the United States. *Journal of Climate*, *26*(3), 1018–1032. <https://doi.org/10.1175/JCLI-D-12-00168.1>