JGR Atmospheres

RESEARCH ARTICLE
10.1029/2020JD033165

Identification of Sea Breeze Recirculation and Its Effects on Ozone in Houston, TX, During DISCOVER-AQ 2013

Wei Li[1], **Yuxuan Wang**[1], **Claudia Bernier**[1], **and Mark Estes**[2,3]

[1]Department of Earth and Atmospheric Sciences, University of Houston, Houston, TX, USA, [2]Texas Commission on Environmental Quality, Austin, TX, USA, [3]Now at School of Natural Sciences Mathematics, Saint Edward's University, Austin, TX, USA

Key Points:
- A new method was developed to automatically identify sea breeze recirculation in Houston, TX, by applying K-Means clustering algorithm
- This method has low data requirement and no *ad hoc* location-specific adjustments and thus can be potentially applied to other coastal cities
- Resulting wind clusters in Houston can be linked with specific ozone levels and spatial distributions

Supporting Information:
- Supporting Information S1

Correspondence to:
Y. Wang,
ywang246@central.uh.edu

Citation:
Li, W., Wang, Y., Bernier, C., & Estes, M. (2020). Identification of sea breeze recirculation and its effects on ozone in Houston, TX, during DISCOVER-AQ 2013. *Journal of Geophysical Research: Atmospheres*, *125*, e2020JD033165. https://doi.org/10.1029/2020JD033165

Received 26 MAY 2020
Accepted 22 OCT 2020
Accepted article online 31 OCT 2020

Abstract In coastal environments, sea breeze recirculation has been found to be an important mesoscale meteorological phenomenon that causes high ozone episodes, yet the identification of this small-scale circulation pattern remains difficult. In this study, a new method was developed to automatically identify sea breeze recirculation in Houston, TX, by applying K-Means clustering algorithm to surface winds measurements at near-coast sites during the DISCOVER-AQ (Deriving Information on Surface Conditions from Column and Vertically Resolved Observations Relevant to Air Quality) field campaign period from August to October 2013. The key to the clustering algorithm is seven features derived from site-based surface winds on each day, including zonal (U) and meridional (V) winds in the morning and afternoon, 24-hr transport direction (θ) and the recirculation factor, which is the ratio of net transport distance (L) to wind run distance (S). For comparison, the same clustering was applied to San Antonio, TX, a noncoastal city yet within the synoptic-scale distance from Houston. Four clusters were obtained for each region, including three synoptic patterns common to both regions and one mesoscale pattern that differs by region, classified as Stagnation and Sea Breeze Cluster for San Antonio and Houston, respectively. The clustering outputs were verified by wind profiler data in Houston. By linking the wind clusters with surface and aircraft ozone measurements, we revealed a clear connection between circulation patterns and daily ozone variability showing that maximum daily average 8-hr (MDA8) ozone levels and spatial distributions differed by cluster type (e.g., the highest ozone found in the Stagnation/Sea Breeze Cluster and the lowest ozone in the Southerly Cluster). This automatable method of sea breeze identification we developed can be potentially applied to other coastal cities because it has low data requirement and no *ad hoc* location-specific adjustments.

1. Introduction

The Houston-Galveston-Brazoria (HGB) region in Texas is the fourth largest metropolitan area in the United States and home to more than 6 million residents as well as hundreds of petrochemical factories. Located on the coast of the Gulf of Mexico and adjacent to the Galveston Bay, the HGB region has a warm and humid climate. It is also subject to complex circulation patterns such as the mesoscale land and sea breezes which are resulted from the diurnal heating and cooling differences between land and water (Miller, 2003). All these factors contribute to the HGB's nonattainment status for exceeding the 2015 National Ambient Air Quality Standards (NAAQS) of ozone (TCEQ: Texas Commission on Environmental Quality, 2012).

A significant fraction of summertime ozone exceedances in HGB have been found to associate with sea breeze recirculation (Banta et al., 2005, 2011; Caicedo et al., 2019; Rappenglück et al., 2008). The recirculation typically develops as the land breeze in the morning transports the pollution-laden air mass from the land toward the sea, where ozone loss rate is much lower than over the land, followed by the return of these pollutants in the afternoon by the onshore sea breeze. On days when the synoptic winds in the lower boundary layer are weak, this recirculation can cause high concentrations of surface ozone over a coastal city (Ma & Lyons, 2003; Wu et al., 2010). It is noteworthy that the sea breeze recirculation cases previously identified in HGB were all manually selected, taking advantages of the abundant data sets obtained during past field campaigns, such as the TexAQS II (Texas Air Quality Study II) in summer 2006 and DISCOVER-AQ (Deriving Information on Surface Conditions from Column and Vertically Resolved Observations Relevant to Air Quality) during August–September 2013.

Existing studies that classified wind patterns for a region tend to focus on a single meteorological scale. For the synoptic scale, manual (e.g., by trained experts) or automated (e.g., cluster analysis and regression tree) methods were used and both classification methods typically rely on coarse-resolution (≥0.25°) meteorological fields, such as geopotential height, sea level pressure, and wind fields (Burlando, 2008; Carvalho et al., 2010; Dayan & Levy, 2005; Russo et al., 2014; Saavedra et al., 2012; Toreti et al., 2010). These data are not applicable for identifying the sea breeze because of their coarse resolution. Instead, surface wind measurements at coastal sites were the most commonly used data sets to discern the sea breeze circulation (Caicedo et al., 2019; David & Nair, 2011; Helmis, Asimakopoulos, et al., 1997). While the general features of the sea breeze are known, such as the diurnal change of wind directions from offshore to onshore and wind speed enhancement after the onset of the sea breeze, the quantitative representation of such features into an automatic algorithm is often elusive and a series of *ad hoc* screening criteria had to be added which differ from region to region (Adame et al., 2010; Stauffer & Thompson, 2015). For example, in selecting sea breeze days on the Atlantic coastline of Spain, Adame et al. (2010) restricted the daytime wind speeds to be larger than 2.5 m/s with directions from 120° to 300° (onshore direction for the study location), and nighttime winds from other directions. Stauffer and Thompson (2015) set the bay breeze criteria for a site on the shore of the Chesapeake Bay to contain shifts of daytime wind direction from 160–360° (offshore) to 10–150° (onshore) for two or more consecutive hours.

One difficulty of identifying sea breeze recirculation is to separate it from synoptic-scale influences because the development and penetration of sea breeze can be perturbed by the intensity and direction of synoptic background flows (Abbs & William, 1992; Arritt, 1993; Helmis, Tombrou, et al., 1997). Studies that classified multiscale wind patterns have found that sea breeze can only prevail when the synoptic forces are weak (Ainslie & Steyn, 2007; Flocas et al., 2008; Russo et al., 2016; Zhou et al., 2019). In these studies, classifications of large- and small-scale winds were conducted separately, and both gridded and site-level data sets were needed as inputs so as to separate the two scales. Here, we develop a new automatable method that uses only surface wind measurements at a few coastal sites and exploits the diurnal wind changes from these measurements to classify both synoptic-scale winds and sea breeze recirculation simultaneously for HGB.

For comparison and verification, we apply the same classification method to surface wind measurements at another urban area in Texas, San Antonio (SA), which has the seventh largest population in the United States and was recently designated as marginal ozone nonattainment by the US Environmental Protection Agency (EPA) in 2018. Geographically, SA is about 300 km away from the HGB region and expected to be affected by similar synoptic systems given the typical synoptic scale of 1,000 km or more. The difference between the two regions is that SA is about 200 km inland while HGB is a coastal region uniquely influenced by the land and sea breezes. Thus, these two cities comprise a good contrast for testing the ability of our method to distinguish the sea breeze from synoptic flows. The study period is from August to October 2013, covering the DISCOVER-AQ field campaign which provides wind profiler and aircraft data sets to verify the classification results. During this study period, Texas was not affected by major weather events, such as drought, flood, or tropical storms. Thus, our analysis can be taken to represent the general meteorological conditions in the HGB and SA during August to October. On the contrary, drought years (e.g., 2011) can witness abnormal circulation patterns not representative of other years (Bernier et al., 2019).

2. Data and Methods

2.1. Ozone and Meteorological Data

Hourly ozone mixing ratios and meteorological variables including wind speed and direction, air temperature (T), and relative humidity (RH) in HGB, and SA are available at the continuous ambient monitoring stations (CAMS) maintained by the Texas Commission on Environmental Quality (TCEQ). These data were downloaded from the Texas Air Monitoring Information System (TAMIS) database (https://www17.tceq.texas.gov/tamis/index.cfm). MDA8 ozone calculated from the hourly ozone measurements was obtained from TCEQ's monthly summary reports (https://www.tceq.texas.gov/cgi-bin/compliance/monops/8hr_monthly.pl).

In HGB, we chose 21 sites (red circles in Figure 1) within 20 km from the coastal lines that have wind measurements to capture the sea breeze. Previous studies have demonstrated that the sea breeze can penetrate at least 20 km inland in the HGB region (Banta et al., 2005; Caicedo et al., 2019; Chen et al., 2011). Among these

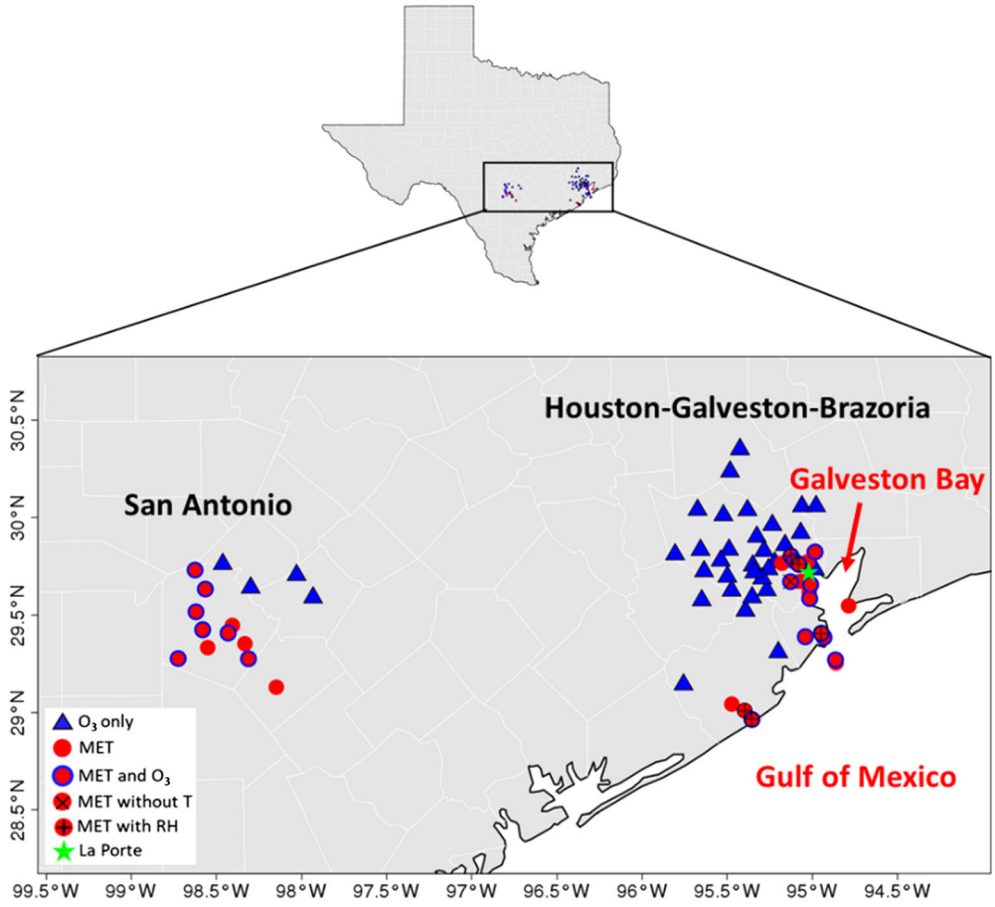

Figure 1. Location of the HGB, SA, and the CAMS sites therein. Filled red circles with red borders are meteorological (MET) sites with wind and temperature measurements; red circles with blue border are those meteorological sites where ozone is also measured; filled blue triangles are ozone sites only; red circles with black plus mark are five meteorological sites that also measure relative humidity; red circles with black cross mark is the only one site that does not have temperature data; green star shows the wind profiler site at La Porte.

sites, five sites have RH measurements (red circles with plus marks), twelve measure surface ozone (red circles with blue border), and only one site does not monitor temperature (red circles with cross mark). Additionally, there are 31 CAMS sites in the HGB region that measure surface ozone only (blue triangles in Figure 1). In SA, there are only eleven sites that measure surface winds, all of which have temperature data and seven measure surface ozone. However, only one site has RH measurements and it is far from the urban center at an elevation of more than 200 m. This site is discarded because it cannot represent the region-mean RH at the surface. Four additional sites in SA monitor ozone. We averaged wind data over all the selected wind sites in each region to obtain hourly zonal (U) and meridional (V) wind component. By averaging, the local-scale between-site differences can be reduced so as to obtain more representative wind conditions in each region. The analysis period is August to October 2013, a total of 92 days in each region.

Wind profiler data during the DISCOVER-AQ field campaign were acquired from the Physical Sciences Division's database of National Oceanic and Atmospheric Administration's Earth System Research Laboratory (NOAA/ESRL). Five profilers were available in Texas during the campaign, but we only show the results of the site in La Porte (Figure 1) because of its coastal location. The time interval of the measurement at this site is 30 min and vertical spacing is 200 m from 150 m above ground level (AGL) up to 3 km aloft.

To examine the ozone spatial distribution, we utilized the ozone and nitrogen dioxide (NO_2) observations from the National Aeronautics and Space Administration (NASA) P-3B and B-200 aircraft during the DISCOVER-AQ field campaign. The data were obtained from NASA Langley Atmospheric Science Data Center (ASDC; https://tad.larc.nasa.gov/index.php). The P-3B aircraft was deployed 9 days in September

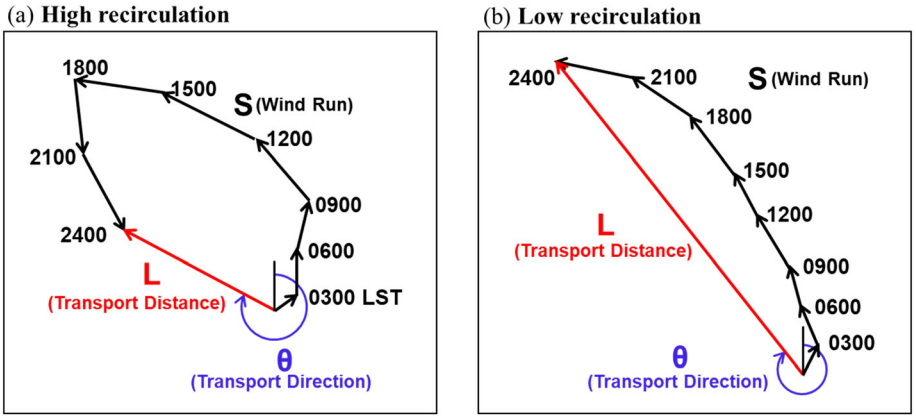

Figure 2. Schematic illustration of wind run S (black arrows), transport distance L (red arrows), and transport direction θ (blue arrows) for high (a) and low (b) recirculation scenario.

and took vertical spiral measurements of ozone at eight sites in the HGB region (Figure 11), including Galveston (Gal), Smith Point (SP), Moody Tower (MT), West Houston (WH), Conroe (Con), Deer Park (DP), Channelview (Cha), and Manvel Croix (MC). The B-200 aircraft equipped with GEOstationary Coastal and Air Pollution Events (GEO-CAPE) Airborne Simulator (GCAS) instrument measured NO_2 column densities on 11 days with four flight loops each day. Details of the data and retrieval algorithm can be found in Nowlan et al. (2018).

As the CAMS and wind profiler data were at point locations, to better visualize and understand different circulation patterns on a regional scale, we used U and V components of winds and geopotential heights at 850 hPa from Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2) with a spatial resolution of 0.5° × 0.625°.

2.2. K-Means Clustering

Clustering analysis has been proved to be an effective technique to discover underlying patterns by dividing objects into a number of groups or clusters based on some features of interest that are similar within each cluster but distinct between any two clusters (Christiansen, 2007; Corte-Real et al., 1998; Huth et al., 2008; MacQueen, 1967). Among all the clustering algorithms, we chose K-means because it is one of the most commonly used algorithms due to its computational and conceptual simplicity. It has been widely applied to classify synoptic wind patterns in the HGB region, which can then be linked to air quality (e.g., Darby, 2005; Davis et al., 1998; Pakalapati et al., 2009). Different from previous studies which relied on large-scale gridded wind or pressure fields for clustering, our clustering algorithm works with surface wind measurements at local sites. As surface wind measurements are readily available for most coastal cities, our algorithm is less contingent on the availability of input data.

Before performing the clustering, features of interest need to be identified. Since our goal is to separate sea breeze recirculation from other circulation patterns, wind diurnal changes are the key features because it is a distinct feature of the sea breeze recirculation whereas synoptic winds are less likely to vary much over the course of a day (except during frontal passages). Based on this proposition, we create seven features to feed into the K-means clustering algorithm. The first four features are zonal (U) and meridional (V) winds averaged across morning (03:00 to 07:00 local time hereafter) and afternoon (14:00 to 18:00) hours, respectively. The reason why we choose these two intervals is that they cover the hours with the lowest and highest wind speeds as illustrated in Figure S1 in the supporting information, which shows the diurnal cycle of wind speeds averaged over all the sites and days. Including both periods allows the algorithm to better capture the synoptic wind signals that would be more persistent over the course of a day. In addition, the sea breeze signal often diminishes over the land after 18:00.

The other features are based on the wind-integral quantities developed by Allwine and Whiteman (1994) (hereafter AW): wind run S, transport distance L, and transport direction θ (Figure 2). Over a given interval (24 hr), S is the scalar summation of transport distance in each time step (an hour) according to the wind speed and direction averaged in that time step. High (low) magnitude of S implies wind ventilation

(stagnation). L is the net displacement of the wind run trajectory, which is the distance from the start point to the end point. The θ is the wind rotation angle always from the north. The mathematical equations of our calculations are as follows:

$$L = \sqrt{U_{24}^2 + V_{24}^2} = \sqrt{\left(\sum_{i=1}^{24} u_i\right)^2 + \left(\sum_{i=1}^{24} v_i\right)^2}$$

$$S = \sum_{i=1}^{24} \sqrt{u_i^2 + v_i^2}$$

$$\theta(\text{radian}) = \begin{cases} \frac{\pi}{2} - \arctan\left(\frac{V_{24}}{U_{24}}\right) & U_{24} > 0 \\ \frac{3\pi}{2} - \arctan\left(\frac{V_{24}}{U_{24}}\right) & U_{24} < 0 \\ 0 & V_{24} > 0,\ U_{24} = 0 \\ \pi & V_{24} < 0,\ U_{24} = 0 \\ \text{undefined} & V_{24} = 0,\ U_{24} = 0 \end{cases}$$

where i ($1 \le i \le 24$) is the time step; u_i and v_i are wind components for each time step; U_{24} and V_{24} are the 24-hr integral wind components.

Based on AW, we define the recirculation factor R as the ratio between L and S to represent the recirculation potential: R = L/S. A low ratio indicates high recirculation because of the wind rotation (Figure 2a), while a high ratio indicates low recirculation due to the persistent wind direction (Figure 2b). The advantage of the AW method is that it only requires hourly wind components at individual stations. The method has been widely applied to other coastal regions due to its low data requirement (e.g., Levy et al., 2008; Nankar et al., 2009; Pérez et al., 2011; Venegas & Mazzeo, 1999). Surkova (2013) calculated R on 208 days with the land-sea breeze occurring from 2002 to 2009 over the coastal zone of the Black Sea and concluded that a low R value is a good indicator of the land-sea breeze circulation. Here we use an hourly time step to calculate L, S, R, and θ on each day based on the site-averaged winds using the equations above.

The fifth and sixth features are $\cos(\theta)$ and $\sin(\theta)$, respectively, because directly using θ for clustering is not practical. For example, both 2° and 358° imply northward transport but mathematically the difference between these two values is large. Therefore, we use the combination of $\cos(\theta)$ and $\sin(\theta)$ to avoid this problem. The last feature is R, which can capture recirculation potential as stated above and is the key feature to distinguish synoptic and local circulations.

The calculations of the features yielded a matrix of size 92×7 for each region, containing 92 days with seven features on each day. Before applying this matrix to K-Means clustering, each feature is normalized to have a zero mean and unit variance. Since the K-Means clustering algorithm calculates Euclidean distance to the centroid, normalization gives each feature the same weight to the clustering. Also, the number of clusters, "k" value, has to be predetermined based on the data. There are several metrics designed for the determination of 'k' and we applied a function in the NbClust R package, which encapsulates 30 indices for selecting "k." As a result, 9 out of 30 metrics selected four as the optimal number of clusters for the HGB region (Figure S2a). For SA, most metrics indicate three as the best value for k, but five metrics suggest four instead (Figure S2b). To be consistent with the HGB region, we decided to use four clusters for both regions. In summary, a matrix containing 92 days with seven normalized features is implemented into the K-Means clustering with k equal to 4 in each region.

3. Clustering Results

3.1. Diurnal Change of Wind Vectors

The diurnal behavior of surface winds from the four patterns identified by the clustering are shown in Figures 3 and 4 for HGB and SA, respectively. The plots are averaged hourly wind speed and directions over all the days in each cluster. Out of the four patterns, three are shared by the two regions, including Oscillation, Northerly and Southerly clusters (a-c in Figures 3 and 4). For Oscillation Cluster, winds

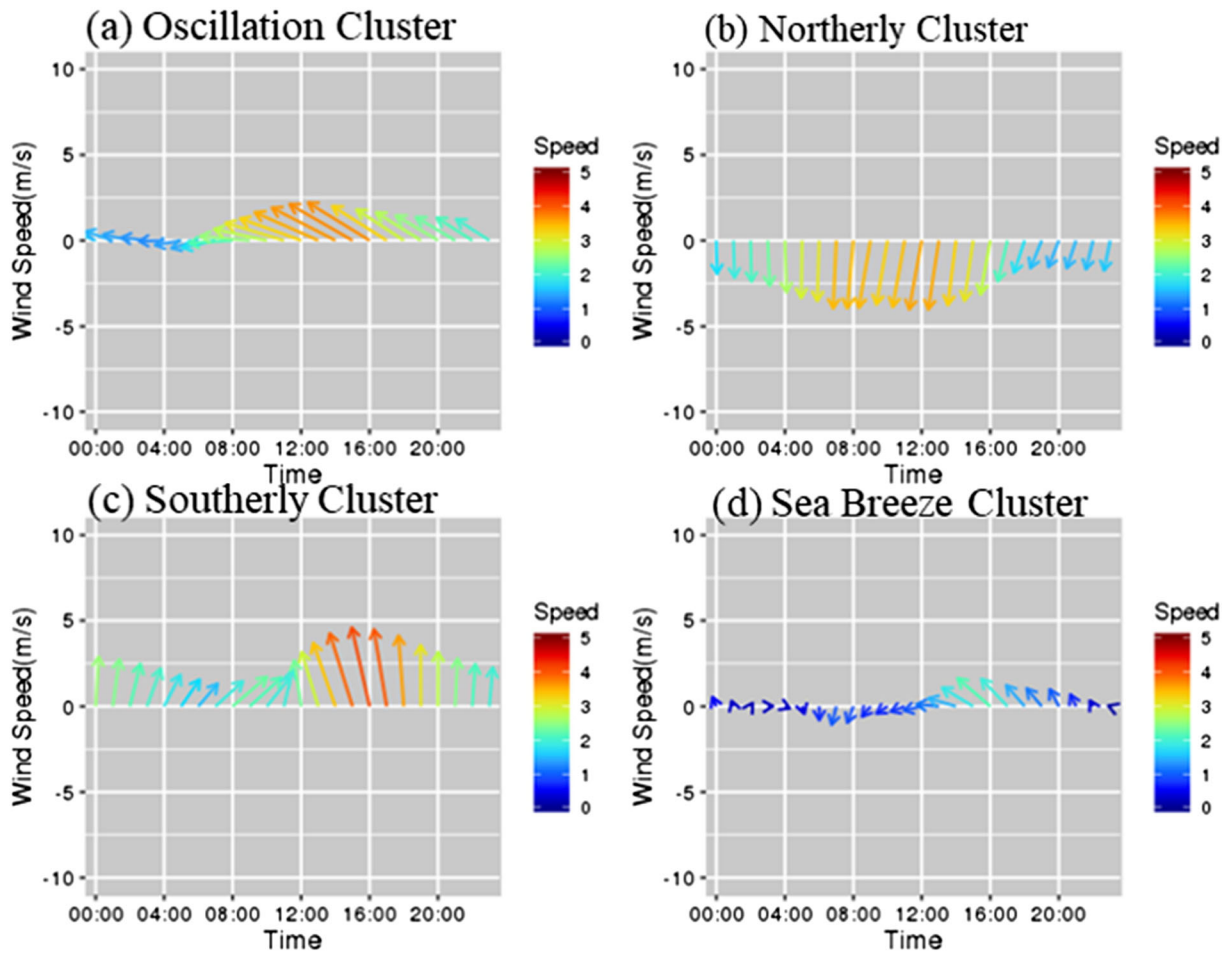

Figure 3. Hourly mean wind direction and speed from August to October in 2013 over HGB in (a) Oscillation, (b) Northerly, (c) Southerly, and (d) Sea Breeze Cluster. Arrows show wind directions; the length and color of the arrows indicate wind speed.

slowly rotate clockwise from east or northeast in the morning to southeast in the afternoon, which is consistent with the characteristics of pure inertial oscillation waves indicated by Nielson-Gammon (2016) under strong large-scale wind conditions. For Southerly and Northerly cluster, winds are consistently from the north and the south, respectively, although there is a slight change in wind direction from southwest to southeast in Southerly Cluster over the HGB region. The consistency between the two regions indicates that those three clusters likely represent synoptic-scale winds that are imposed on both regions as we expected. Despite being implemented separately for each region, the clustering algorithm captures this similarity, suggesting the algorithm is capable of classifying synoptic scale winds.

In addition to the three common clusters, each region has one local cluster (d in Figures 3 and 4) which has lower wind speeds compared to the other three clusters. The local cluster also behaves differently between the two regions. For HGB, winds in the local cluster are light in the morning and then rotate clockwise from southwest to southeast in the afternoon while strengthening in speed. This behavior matches well to the typical sea breeze recirculation in the region reported by a number of earlier studies (Banta et al., 2005; Darby, 2005; Nielsen-Gammon et al., 2005; Rappenglück et al., 2008; Tucker et al., 2010). Thus, we name this cluster Sea Breeze Cluster. For SA, winds in the local cluster are less than 1 m/s on average. Thus, we name it Stagnation Cluster. Although winds in Stagnation Cluster appear to rotate counterclockwise from southwest in the morning to northeast in the afternoon, with this cluster's low wind speed, the rotation is dismissible.

We verified that all seven features used for clustering are necessary to identify the four clusters (Figures S3–S6) and each of the features defines a particular aspect of the circulation. R determines the extent of recirculation; $\sin(\theta)$ and $\cos(\theta)$ represent the general transport direction; morning and afternoon

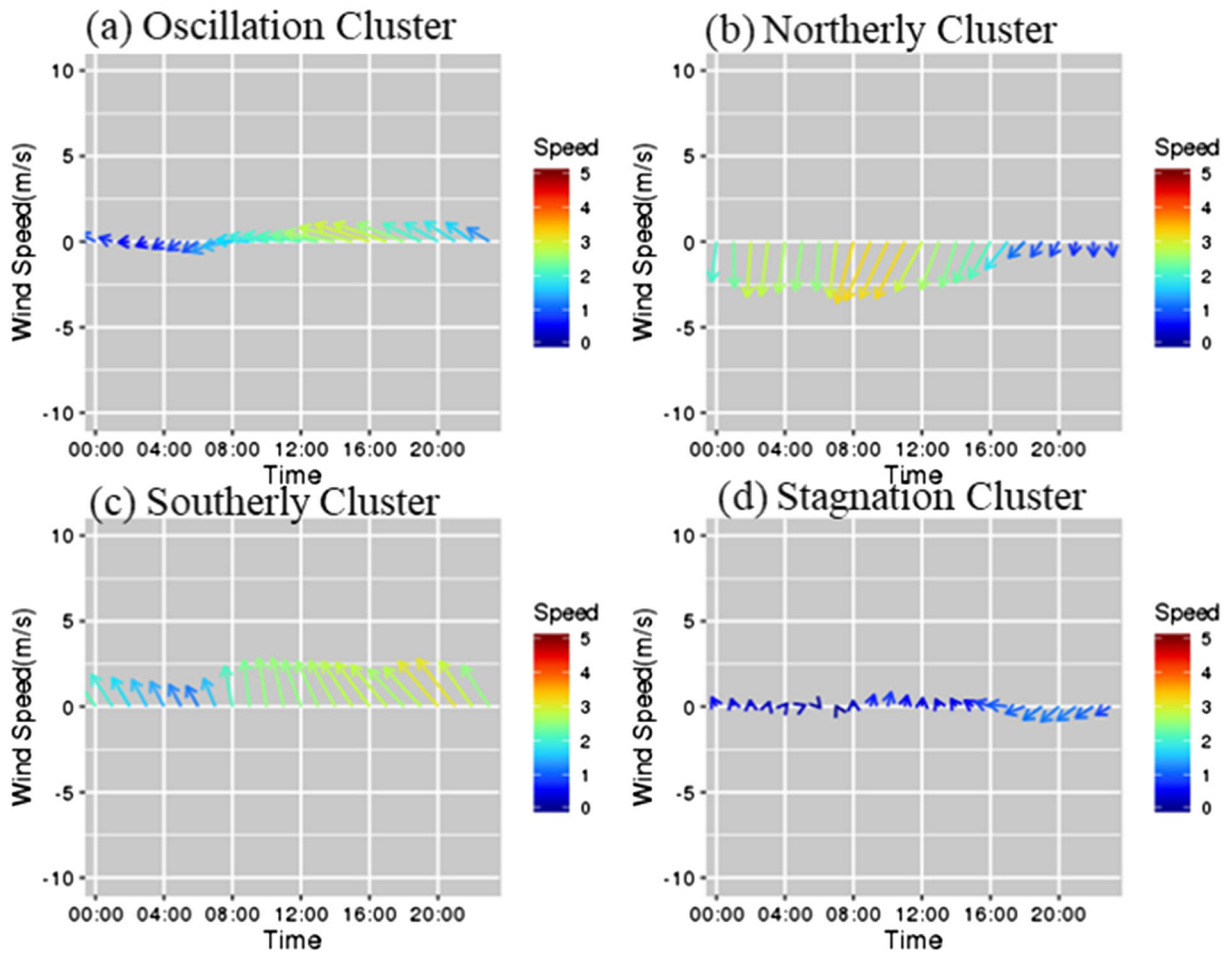

Figure 4. The same with Figure 3 but for SA with (d) showing Stagnation Cluster.

wind U and V describe the actual movement along the general direction. It is the combination of the seven features that identifies a cluster. For example, Oscillation and Sea Breeze/Stagnation Cluster have similar transport direction but they differ in transport pathways, which can be shown by different R. Sometimes, Southerly and Oscillation Clusters have similar winds in the morning and afternoon, but they exhibit different transport directions. Therefore, all the seven features are necessary to obtain meaningful clustering results.

To the best of our knowledge, no prior study investigated wind patterns in SA. Previous studies that classified wind patterns in the Houston area (e.g., Ngan & Byun, 2011; Souri et al., 2016) almost all relied on reanalysis meteorological data with horizontal resolutions no finer than 50 km, therefore not suitable to discern the mesoscale sea breeze recirculation. By comparison, our algorithm uses site-measured surface winds as the only input and such data is readily available from CAMS sites or meteorological stations. In addition, the classification scheme is based on seven mathematically defined features with straightforward physical meanings and no *ad hoc* manual adjustments are needed, which is another advantage compared to prior studies of sea breeze identification. In terms of results, our approach identifies one mesoscale pattern (sea breeze recirculation) and three synoptic patterns (Northerly, Southerly, and Oscillation) for the 3-month period (August to October 2013). The three synoptic wind patterns were also identified by other studies to be the most representative ones for HGB in late summer and early fall (Darby, 2005; Davis et al., 1998; Ngan & Byun, 2011; Souri et al., 2016). These earlier studies used multiyears' meteorological data of all the months and thus revealed more synoptic wind patterns than those three.

The 24-hr progression of winds in each cluster can be more clearly seen in the wind run trajectories (Figure 5) that are calculated using the same cluster-mean hourly wind vectors from Figure 3 (for HGB)

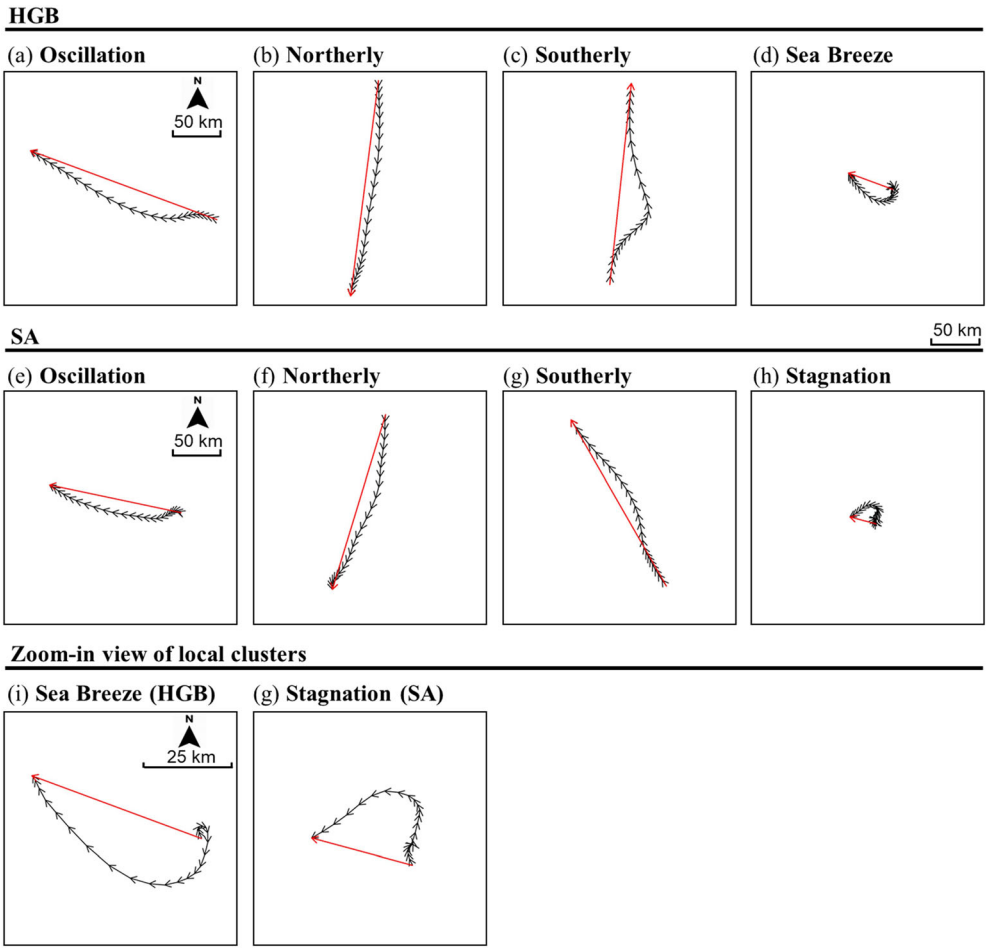

Figure 5. Illustration of wind run (S) and transport distance (L) for Oscillation (a and e), Northerly (b and f), Southerly (c and g), Sea Breeze (d)/Stagnation (h) Cluster in HGB and SA under 50 km scale. (i and g) The enlarged figures for sea breeze and stagnation cluster under 25 km scale.

and 4 (for SA). The cluster-mean values of L, S, θ, and R (cf. section 2.2) and the number of days in each cluster are summarized in Table 1. L and S values are high in three synoptic clusters, ranging from 146.77 to 232.41 km and 165.87 to 245.23 km, respectively. The R values are also higher for those clusters, ranging from 0.86 to 0.94. All these parameters indicate strong wind ventilation and low recirculation, which are evident in Figure 5. The major difference between the three clusters is the transport direction θ. The transport direction is approximately westward (~285°), southward (~195°) and northward (~350°) in Oscillation, Northerly and Southerly Cluster, respectively. Contrary to the synoptic clusters, the local clusters have much smaller L, S, and R values ranging from 67.07 to 72.40 km for L, from 118.46 to 140.55 km for S, and from 0.49 to 0.56 for R. The smaller values of L and S are indications of stagnation, while lower R signifies high recirculation. The wind run distance of the local clusters is a factor of 2–3 smaller than that of the synoptic clusters, as shown in the top and middle panels of the figure on the 50-km scale. Zooming in on the 25-km scale (bottom panel of Figure 5), the sea breeze cluster in HGB clearly shows a clockwise wind rotation over the course of a day, changing from light northerly (onshore) wind in the morning to southeasterly wind in the afternoon with higher speeds, matching well with the typical features of the sea breeze recirculation cases identified previously (Banta et al., 2005; Tucker et al., 2010). Stagnation Cluster in SA has even smaller wind speeds than the sea breeze cluster in Houston, so the anticlockwise rotation can be dismissed because of the small wind speed. The value of L from the cluster-mean wind run trajectory is 27.56 km in SA and 48.12 km in the HGB (i.e., red arrow in Figures 5i and 5g, respectively). These values indicate the upper limit of what spatial resolution an air quality model would need to have in order to capture these mesoscale circulations and simulate their effects on air quality.

Table 1
Mean (Standard Deviation) Meteorological and Ozone Characteristics in Each Cluster

		Oscillation		Northerly		Southerly		Sea breeze/stagnation	
		HGB	SA	HGB	SA	HGB	SA	HGB	SA
Number of Days	Aug	10	14	2	0	13	13	6	4
	Sep	18	14	3	4	5	6	4	6
	Oct	17	3	6	5	3	13	5	10
Total days		45	31	11	9	21	32	15	20
Days ratio (%)		48.91	33.69	11.95	9.78	22.82	34.78	16.30	21.73
Days of ozone exceedance		7	3	1	1	3	2	6	6
L (km)		223.73	146.77	232.41	190.44	210.99	196.15	72.40	67.07
		(66.56)	(34.26)	(57.34)	(57.34)	(67.49)	(55.01)	(35.19)	(28.22)
S (km)		239.88	165.87	245.23	202.48	241.51	206.32	140.55	118.46
		(62.21)	(33.71)	(77.63)	(50.77)	(56.98)	(52.24)	(44.27)	(45.72)
L/S		0.93	0.88	0.94	0.92	0.86	0.94	0.49	0.56
		(0.05)	(0.06)	(0.07)	(0.05)	(0.09)	(0.04)	(0.14)	(0.13)
θ (degree)		288.19	283.42	188.17	199.35	5.62	330.78	281.87	280.22
		(25.36)	(18.82)	(20.91)	(15.90)	(15.61)	(9.35)	(59.30)	(69.29)
Wind speed (m/s)		2.58	1.69	2.73	2.26	2.45	2.26	0.84	0.76
		(0.76)	(0.40)	(0.97)	(0.68)	(0.79)	(0.68)	(0.41)	(0.33)
Temperature (°F)		79.34	81.03	72.39	67.91	83.48	82.45	78.38	75.57
		(5.35)	(4.76)	(6.48)	(6.67)	(3.69)	(5.03)	(6.66)	(7.84)
Relative humidity (%)		76.67	—	72.79	—	75.50	—	73.15	—
		(5.28)		(11.43)		(6.53)		(6.42)	
MDA8 ozone (ppbv)		37.60	42.93	35.40	41.01	34.17	38.14	48.41	51.83
		(10.68)	(11.54)	(10.79)	(11.74)	(13.12)	(10.76)	(13.91)	(13.93)

With regard to the frequency of occurrence, Oscillation and Southerly clusters are the two most frequent patterns in both regions during the study period. In combination, they account for nearly 70% of the study period. The proportion of days in Oscillation Cluster (48.91%) is approximately twice of that in Southerly Cluster (22.82%) in HGB, while the number of days in these two clusters is nearly the same in SA. Northerly Cluster is the least frequent one contributing only about 10% of the study period in both regions. The contribution of days from Sea Breeze/Stagnation Cluster is nearly twice that of Northerly Cluster in each region. Table 1 also lists the number of days in each cluster by month to show the seasonal change of wind clusters. In the HGB region, Oscillation Cluster is more frequent in September and October than August, while Southerly Cluster has the opposite distribution. This should be related to the seasonal displacement of synoptic regimes, which will be discussed in detail in the next section.

3.2. Wind Profiles

To further verify the classification results, we display the wind profiles measured at the La Porte site on a representative day for each cluster (Figure 6). Considering that the K-Means method assigns each day to the nearest cluster centroid based on the Euclidean distance, we select the representative day for a cluster as the one closest to the centroid of that cluster among all the days with available wind profiler data. For the three synoptic clusters, the wind profiler data show that upper-level winds are consistent with those at the lower levels in each cluster, with easterly, northerly, and southerly winds in Oscillation, Northerly and Southerly Cluster, respectively. As winds at the higher levels can better represent the synoptic wind conditions, the consistent winds at all levels in each of the synoptic clusters further support our classification approach. Different from the synoptic clusters, the Sea Breeze Cluster sees a change in wind patterns between the low and high levels. Higher-level winds are offshore from the northeast all day while surface winds are offshore from the northwest in the morning. After the arrival of sea breeze around 12:00, winds start blowing onshore from the surface to about 800 m aloft. Light to stagnant wind conditions can occur during the transition period between offshore to onshore flows.

In summary, the wind profile data on the sea breeze day shows that the winds typically evolve from offshore in the morning to onshore in the afternoon with light to stagnant conditions developing in between. While the offshore flow can occupy a deep layer, the onshore flow usually occurs below 1 km. These features are in

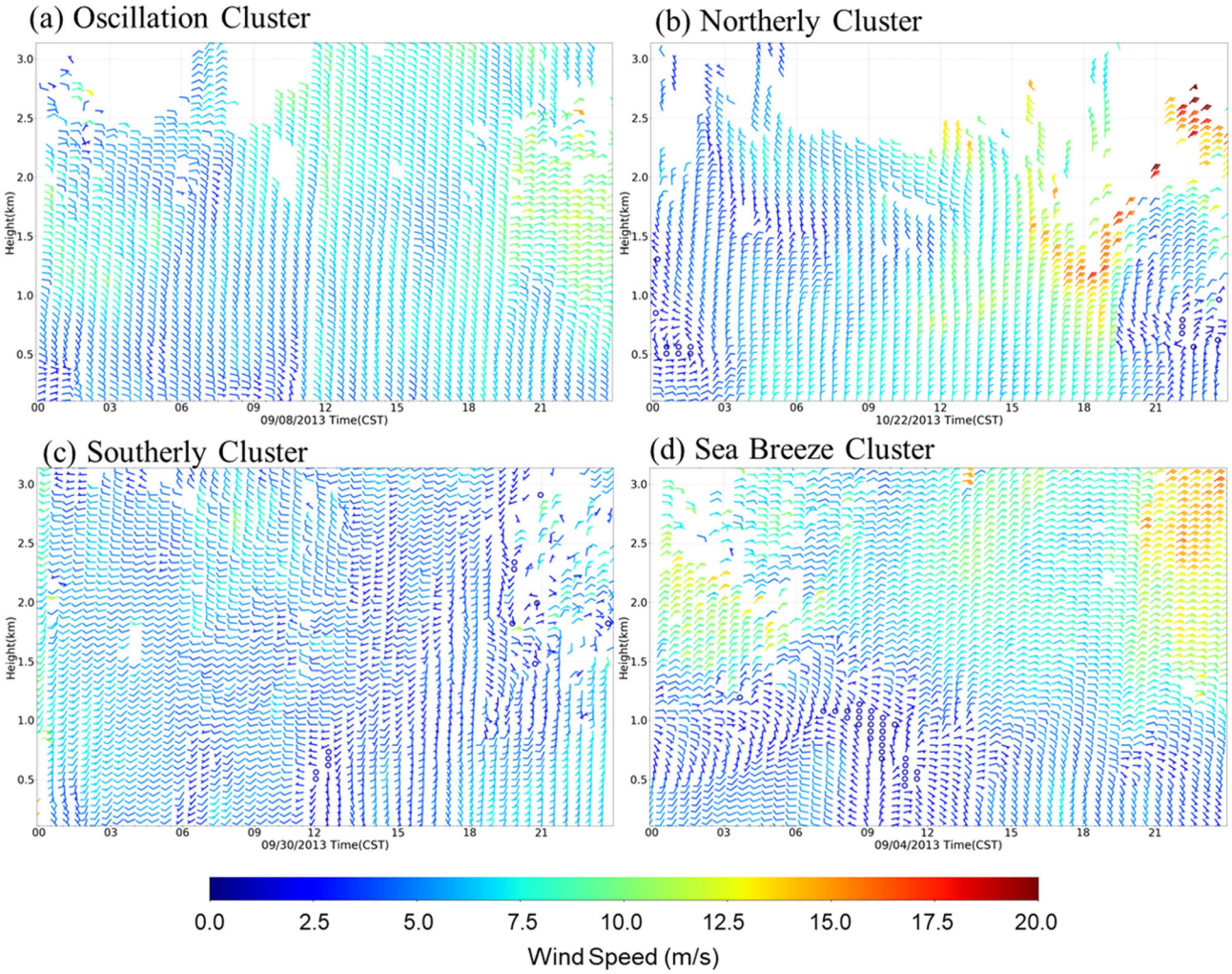

Figure 6. Wind profiles on (a) 8 September, (b) 22 October, (c) 30 September, and (d) 4 September 2013 at La Porte for Oscillation, Northerly, Southerly, and Sea Breeze Cluster, respectively. Colors of the arrows show the magnitude of wind speed. Small circles indicate stagnant wind with a speed less than 2.5 m/s.

good agreement with previously published case studies of sea breeze days over the HGB area (Banta et al., 2005; Caicedo et al., 2019; Chen et al., 2011).

3.3. Meteorological Parameters

Temperature, wind speed, and relative humidity are important meteorological parameters that affect surface ozone concentrations. Table 1 lists the average values and standard deviation of these variables in each cluster. Relative humidity data is not available for SA as stated above. Cluster-mean daily wind speed ranges from 1.69 to 2.73 m/s in the synoptic wind clusters, while it is less than 1 m/s in the local clusters. This wind speed disparity is in agreement with other studies that suggested the sea breeze circulation dominates only under light large-scale winds (Nielson-Gammon, 2016; Tucker et al., 2010). The lower wind speed of the local clusters thus can facilitate the accumulation of ozone and other pollutants due to poor dispersion. Northerly Cluster has the lowest temperature and relative humidity, as air masses from the north are typically colder and drier. Southerly cluster has the highest temperature, attributable to the warm air mass originating from the Gulf of Mexico. For HGB, Oscillation Cluster has the highest RH. This is because we only adopted data from near-coast sites. The easterly and southeasterly winds in Oscillation Cluster can pick up additional moisture from the Galveston Bay which is to the southeast of Houston (Figure 1). Although winds in Sea Breeze Cluster are also from the south in the afternoon, they mainly start after 14:00 from the average wind direction in Figure 3 and the strength of wind is also lower compared

with Southerly Cluster. That makes temperature and humidity levels from the Sea Breeze Cluster fall in the middle among all the clusters.

4. Synoptic Conditions

Figure 7 illustrates the cluster-mean geopotential height and wind fields at 850 hPa over HGB and SA for each established cluster. The similarity in the flow patterns between the two regions indicate that they are affected by similar synoptic processes for each cluster. Oscillation Cluster is identified with a high-pressure system centered over northeast Texas putting HGB and SA under moderate easterly or southeasterly winds. This corresponds to the migratory anticyclone occurring mainly in spring and fall (Davis et al., 1998). Northerly Cluster is associated with a high-pressure sitting to the west of Texas and a low pressure sitting to the farther north inland, resulting in an air mass from the northwest moving to the south. This is a typical flow pattern after the passage of a cold front when southeast Texas is controlled by cool and dry weather brought by the northerly winds (Ngan & Byun, 2011). Southerly Cluster is identified with a high-pressure system located near the coast of the Gulf of Mexico placing HGB and SA under southwesterly and southeasterly flow, respectively. This pattern is associated with the Bermuda High (BH) anticyclone centered over North Atlantic Ocean, which shifts westward from late spring to summer and retreat back eastward in the fall (Shen et al., 2015; Wang et al., 2016; Zhu & Liang, 2013). Stagnation/Sea Breeze Cluster features a small high that situates just over the middle of Texas leading to a very weak synoptic background favorable for mesoscale winds to prevail.

The similarity in synoptic conditions between the two regions is because they have many days in common within the same type of clusters. Figure 8 illustrates the time sequence of the clustering results. Table 2 shows the cross tabulation for the number of days in each cluster and each region. The numbers along the diagonal of the table are the number of days these two regions share for each pair of the same type of synoptic or local clusters (hereafter common-cluster days). For example, if 1 August 2013 was classified as Northerly Cluster in both HGB and SA, it is counted as one common-cluster day for Northerly Cluster. The off-diagonal numbers show the number of days that do not match between HGB and SA (hereafter cross-cluster days). In total, there are 56 (61% out of 92) common-cluster days during the study period, indicating again the common synoptic forcing between the two regions and the ability of our cluster algorithm to capture this commonality despite being implemented independently for each region. Most of these common-cluster days are distributed in Oscillation and Southerly Cluster, the two most frequent clusters in each region. Interestingly, the proportion of common-cluster days is also high (45–60%) for the local-sea breeze cluster pair of the two regions, further indicating the consistency between the two regions as these mesoscale clusters arise under weak synoptic winds.

Despite of the abovementioned similarity, small differences of synoptic systems between the two regions can still be discerned. For example, the average positions of the high-pressure center of Oscillation and Southerly clusters in SA are slightly northward compared with those in HGB (Figures 6a and 6e, and 6c and 6g). These differences can be attributed to the less-frequent cross-cluster days between the two regions. One notable example is the Southerly Cluster of SA (shadings in Figure 8). A proportion of 50% of the days in this cluster belongs to the Oscillation Cluster in HGB. This is due to the 300 km east-west distance between HGB and SA. During some days the same high-pressure system can place HGB under easterly winds (Oscillation Cluster) and SA under southerly winds (Southerly Cluster). Figure S7 shows one case on 28 October 2013, which belongs to Oscillation Cluster in HGB and Southerly Cluster in SA. The winds in HGB and SA are generally from east and south, respectively (Figures S7b and S7c). Cases like this, when included in the cluster-mean calculation, can result in the between-region differences in each cluster.

The seasonal changes of clusters can be seen in Table 1. The number of days in Southerly Cluster over the HGB region decreases from 13 days in August to 5 days in September and 3 days in October, while the number of days in Oscillation Cluster increases from 10 days in August to 18 days in September and 17 days in October. The seasonal displacement of the abovementioned high and low systems is the major cause of the seasonal change of clusters. Our study period is from late summer (August) to early fall (September and October), during which the BH starts to retreat eastward making way for migratory anticyclones to develop. This explains why the southerly winds in our study are less strong and frequent compared with those in other studies focusing on summer alone or longer periods (Bernier et al., 2019; Davis et al., 1998; Ngan & Byun, 2011; Wang et al., 2016). By contrast, SA has an opposite seasonal change in the frequency

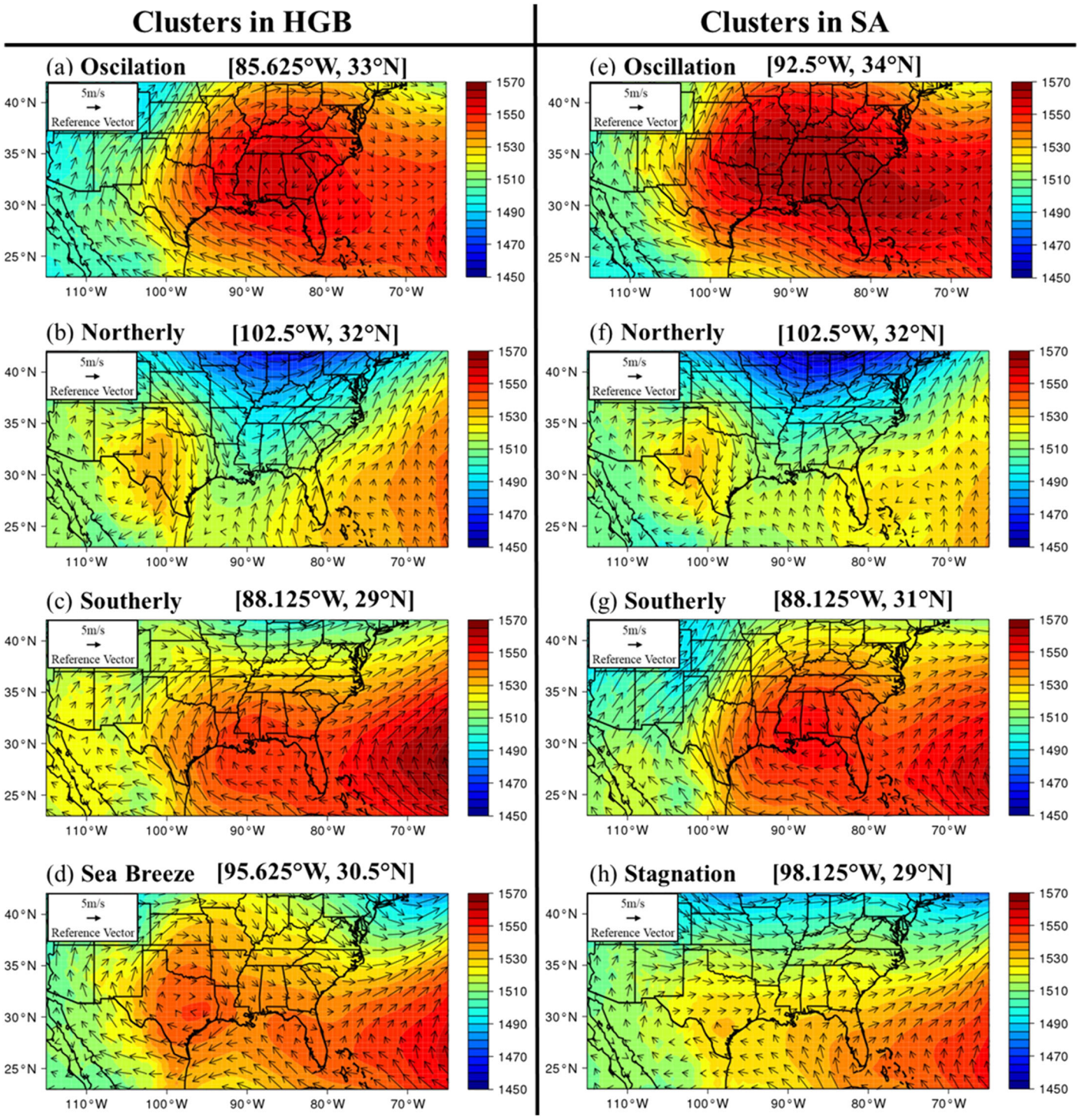

Figure 7. Geopotential height (contours) and winds (black arrows) at 850 hPa averaged over days in each cluster from August to October 2013 in HGB (a–d) and SA (e–h) for Oscillation, Northerly, Southerly, and Sea Breeze/Stagnation Cluster, respectively. The numbers indicate the center of high-pressure systems by locating the grids with the highest geopotential height in each plot.

of these two clusters due to the abnormal number of days in October. This can be explained by the 50% of the cross-cluster days in Southerly Cluster mentioned above, which mainly occur in October (shadings in Figure 8). The frequency of Northerly Cluster increases from August to October in both regions, which is due to the more frequent cold front in fall than summer because of the retreat of the BH anticyclone

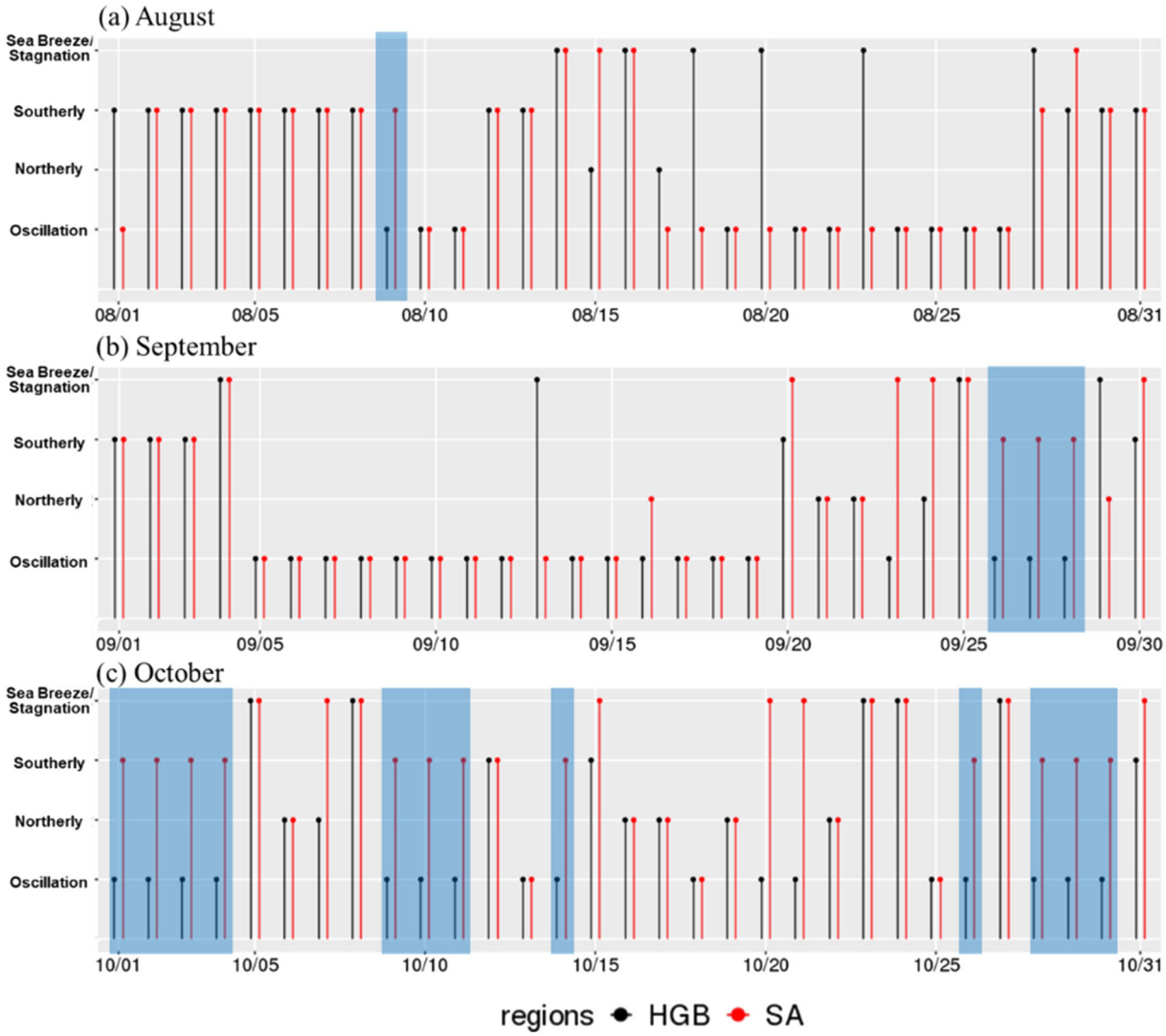

Figure 8. Clustering results on each day from August to October (a–c) 2013 over HGB and SA. The shadings illustrate the days that are classified to Oscillation Cluster in HGB but Southerly Cluster in SA. These days account for 50% of all Southerly Cluster days in SA.

(Souri et al., 2016). From August to October, the number of days in Stagnation Cluster more than doubles in SA, while HGB sees a slight decrease in the days of Sea Breeze Cluster. The latter is consistent with the sea breeze seasonality.

5. Ozone Concentrations and Spatial Distributions

In this section, we compare ozone concentrations and spatial distributions between the different clusters identified above for each region. Here we include all ozone sites in the HGB region, not just the

Table 2
The Number of Days Shared by Each Pair of Clusters Over HGB and SA

	HGB			
SA	Oscillation (45 days)	Northerly (11 days)	Southerly (21 days)	Sea breeze (15 days)
Oscillation (31 days)	25	1	1	4
Northerly (9 days)	1	7	0	1
Southerly (32 days)	16	0	15	1
Stagnation (20 days)	3	3	5	9

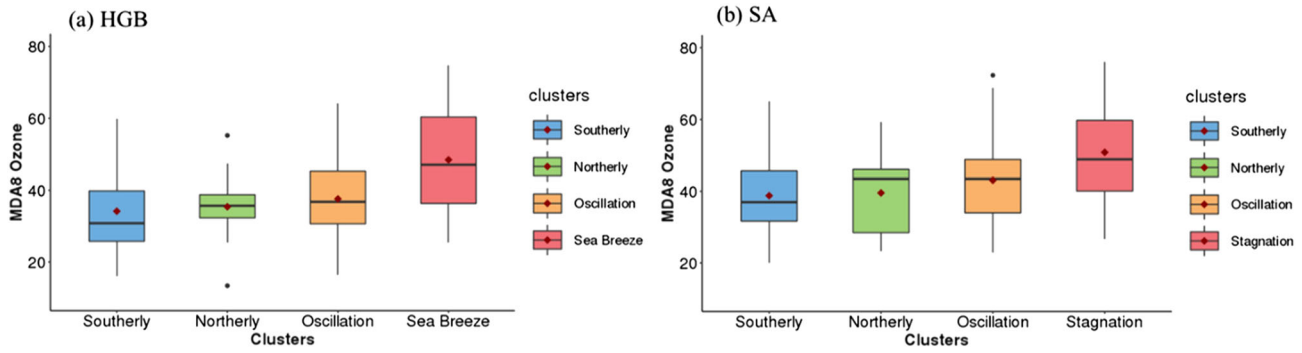

Figure 9. Boxplot of MDA8 ozone in each cluster for HGB (a) and SA (b).

near-coast sites, so as to consider the region-wide influence of the different wind patterns on ozone. The box plots in Figure 9 show the distribution of daily all-site mean MDA8 ozone by cluster in HGB (a) and SA (b). The mean and standard deviation of MDA8 ozone in each cluster are listed in Table 1. Mean ozone levels in both regions follow the same increasing order: Southerly < Northerly < Oscillation < Sea Breeze (Stagnation). The highest MDA8 ozone is in Sea Breeze Cluster for HGB and Stagnation Cluster for SA with the mean value of 48.41 and 51.83 ppbv, respectively. The high ozone is attributed to the pollution recirculation and stagnation (Darby, 2005; Ngan & Byun, 2011). Southerly Cluster experiences the lowest mean MDA8 ozone in HGB (34.17 ppbv) and SA (38.14 ppbv), which is due to the strong southerly wind bringing clean marine air inland to dissipate the pollution. MDA8 ozone level is in the middle for Oscillation and

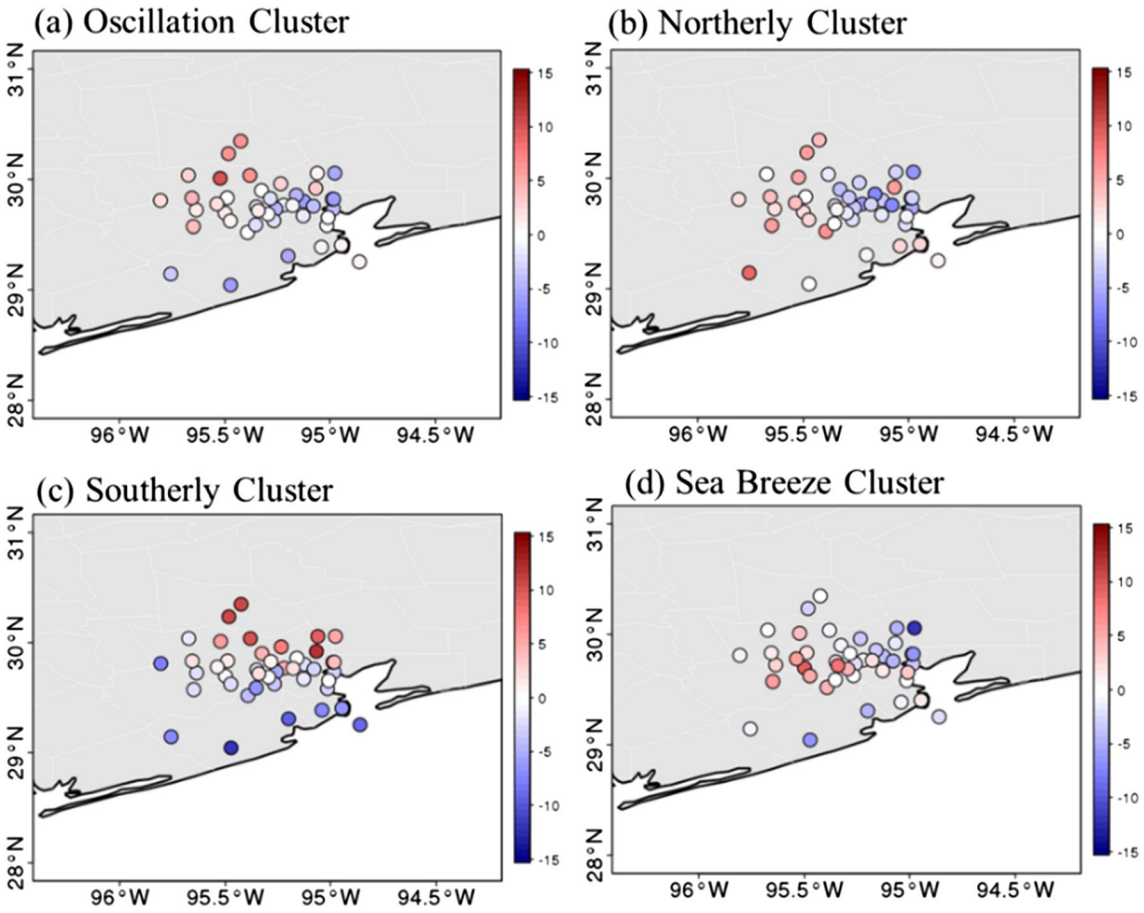

Figure 10. (a–d) MDA8 ozone spatial distribution in each cluster over HGB by subtracting all-site mean MDA8 ozone from each site.

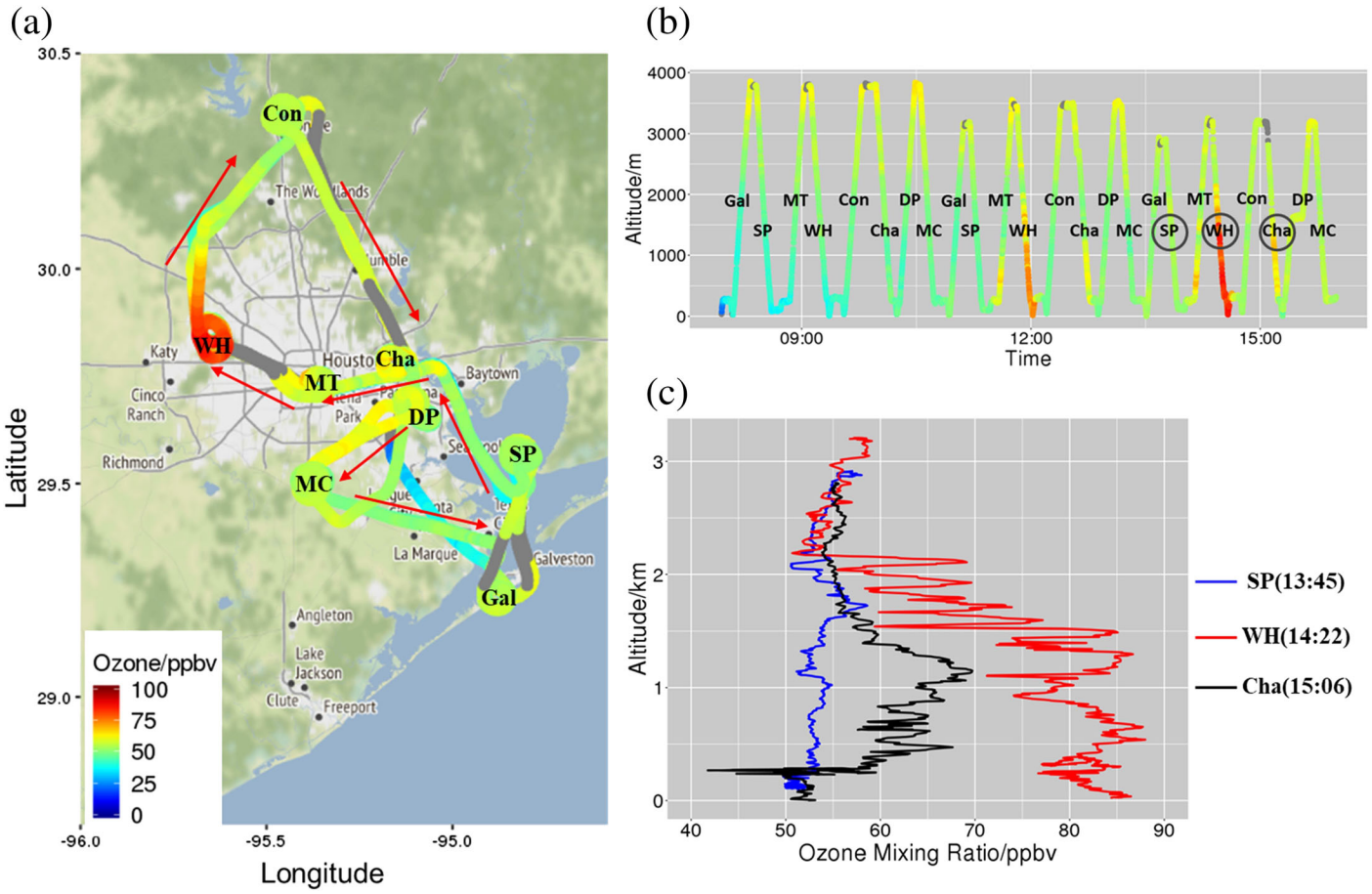

Figure 11. (a) DISCOVER-AQ P-3B flight track on 11 September 2013 (Oscillation Cluster) color coded by ozone mixing ratio. Black letters indicate eight spiral sites including Galveston (Gal), Smith Point (SP), Moody Tower (MT), West Houston (WH), Conroe (Con), Channelview (Cha), Deer Park (DP), and Manel Croix (MC). Red arrows show the flight direction in the order of Gal-SP-MT-WH-Con-Cha-DP-MC. (b) Time-altitude changes of the flight with the same color scales as flight track showing ozone levels. Gray color means missing ozone measurements. Black letters indicate the sites where the upward or downward spiral measurement occurs. (c) Ozone vertical profiles at sites marked by the circles in (b). The time (LST) for each site is the start time of the profile.

Northerly Cluster with the mean values of 37.60 ppbv and 35.40 ppbv in HGB and of 42.93 ppbv and 41.01 ppbv in SA, respectively. Both Northerly and Oscillation Cluster bring polluted continental air masses to the region which can elevate background ozone (Lei et al., 2018; Rappenglück et al., 2008). Oscillation Cluster has slightly higher ozone concentrations than Northerly Cluster because the higher temperature in Oscillation Cluster can result in more production of ozone. In addition, the Ship Channel, which is important emission sources of ozone precursors in HGB, is located to the east of the city. Easterly winds in Oscillation Cluster facilitate the transport of ozone precursors from the Ship Channel to the city (Rappenglück et al., 2008; Souri et al., 2016).

Because the region-mean ozone values could average out the high ozone concentrations at individual sites, we also included ozone exceedance days in Table 1 to better characterize ozone levels in each cluster. The proportion of exceedance days in Sea Breeze and Stagnation Cluster is 40% (6 out of 15) and 30% (6 out of 20), respectively, which is much higher than that in other synoptic clusters ranging from 6% to 15%. This further supports the conclusion drawn above on the different ozone levels by cluster.

We further examine the spatial distribution of surface ozone by cluster. As there are fewer sites in SA, the spatial analysis is conducted for the HGB region only. Figure 10 shows the deviation of MDA8 ozone at each CAMS site from the all-site mean MDA8 ozone in each cluster. By showing the deviation from the mean, Figure 10 removes the between-cluster difference in mean ozone shown above (cf. Figure 9) and highlights the large spatial heterogeneity of ozone (±15 ppbv of the mean) in each cluster. Each cluster appears to exhibit a unique pattern of ozone distribution. For the synoptic clusters, higher ozone concentrations are found

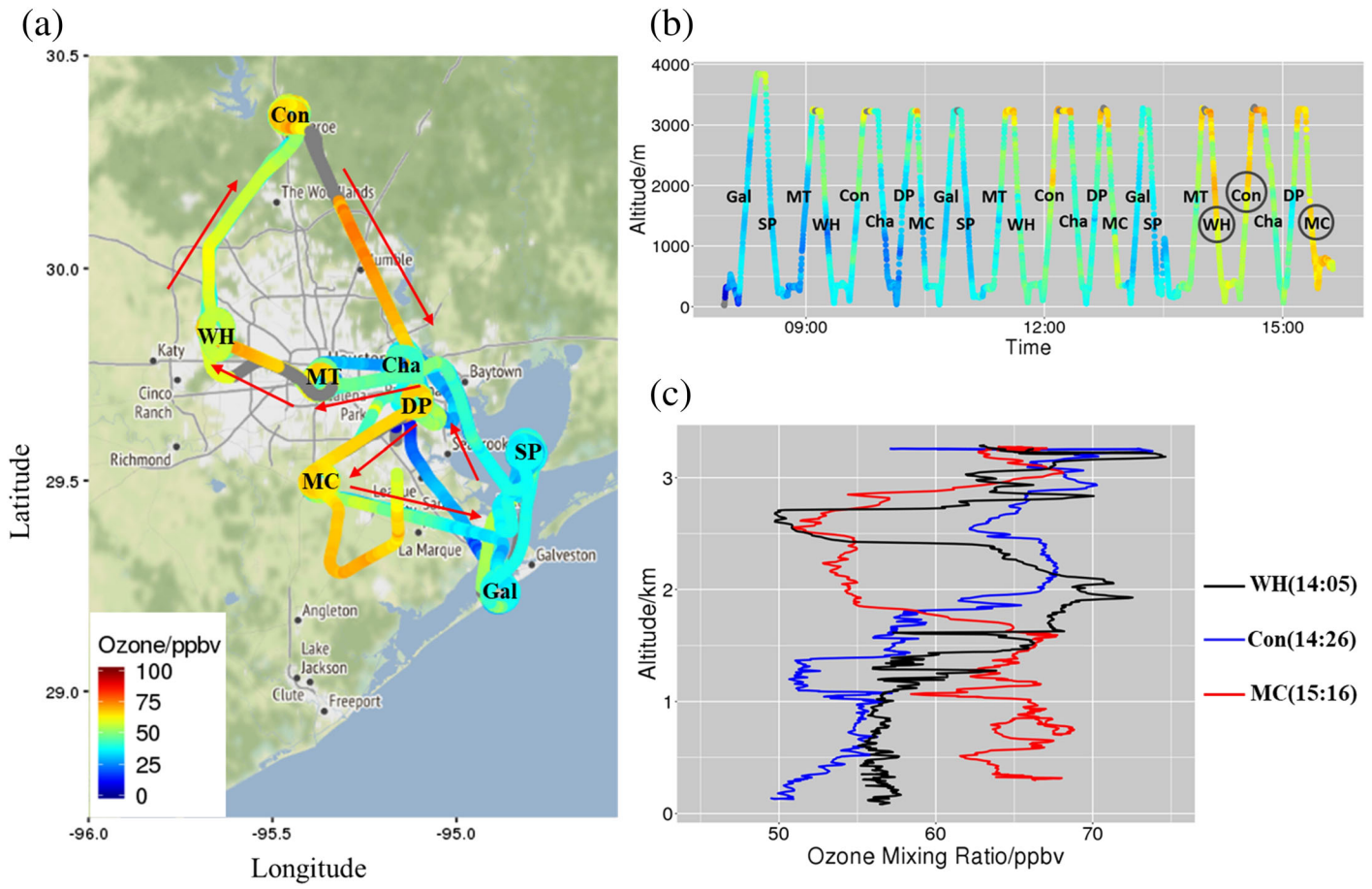

Figure 12. The same with Figure 11 but on 24 September 2013 (Northerly Cluster).

at sites located in the downwind direction of the flow pattern: northwest sites in Oscillation Cluster, southwest sites in Northerly Cluster, and northern sites in Southerly Cluster. For Sea Breeze Cluster, ozone is relatively high at the sites located in the middle of the HGB region where the sea breeze front can penetrate to. The mean transport distance L in this cluster is 72 km with a standard deviation of 35 km (Table 1). As the transport distance in this cluster is highly variable, on some days high ozone sites may not be near emission sources (e.g., Ship Channel). Other factors such as the spatial distribution of emission sources and ozone chemistry also play a role in the spatial heterogeneity of ozone in HGB. For example, more high ozone sites are found downwind of the Ship Channel in all clusters.

The P-3B aircraft observation data during the DISCOVER-AQ field campaign was examined to compare the vertical distribution of ozone for each cluster. Among the 9 days of flights in September, 6 and 2 days belong to Oscillation and Sea Breeze Cluster, respectively. Again, we select a representative day for each cluster to be the one closer to the centroid of the cluster, which is 11 September 2013 for Oscillation Cluster and 4 September 2013 for Sea Breeze Cluster. Only 1 day (24 September) is within Northerly Cluster. No flight occurred on Southerly Cluster days.

The ozone distribution along the flight track on 11 September 2013 (Oscillation Cluster) is shown in Figure 11. In the morning, ozone steadily increased from ~35 ppbv near the surface to ~55 ppbv up to 3 km aloft across the area. In the afternoon, while the ozone at upper levels remained unchanged, surface ozone increased to ~50–90 ppbv due to the ozone photochemical production under high temperature. A positive surface ozone gradient can be seen toward land resulted from the generally easterly winds. Figure 11c depicts this phenomenon by showing the ozone vertical profile at SP, Cha, and WH in the afternoon. Ozone values above 2 km were ~55 ppbv across the three sites. In the mixing layer below 1.5 km ozone increased from ~52 ppbv at SP, to ~62 ppbv at Cha, and ~82 ppbv at WH.

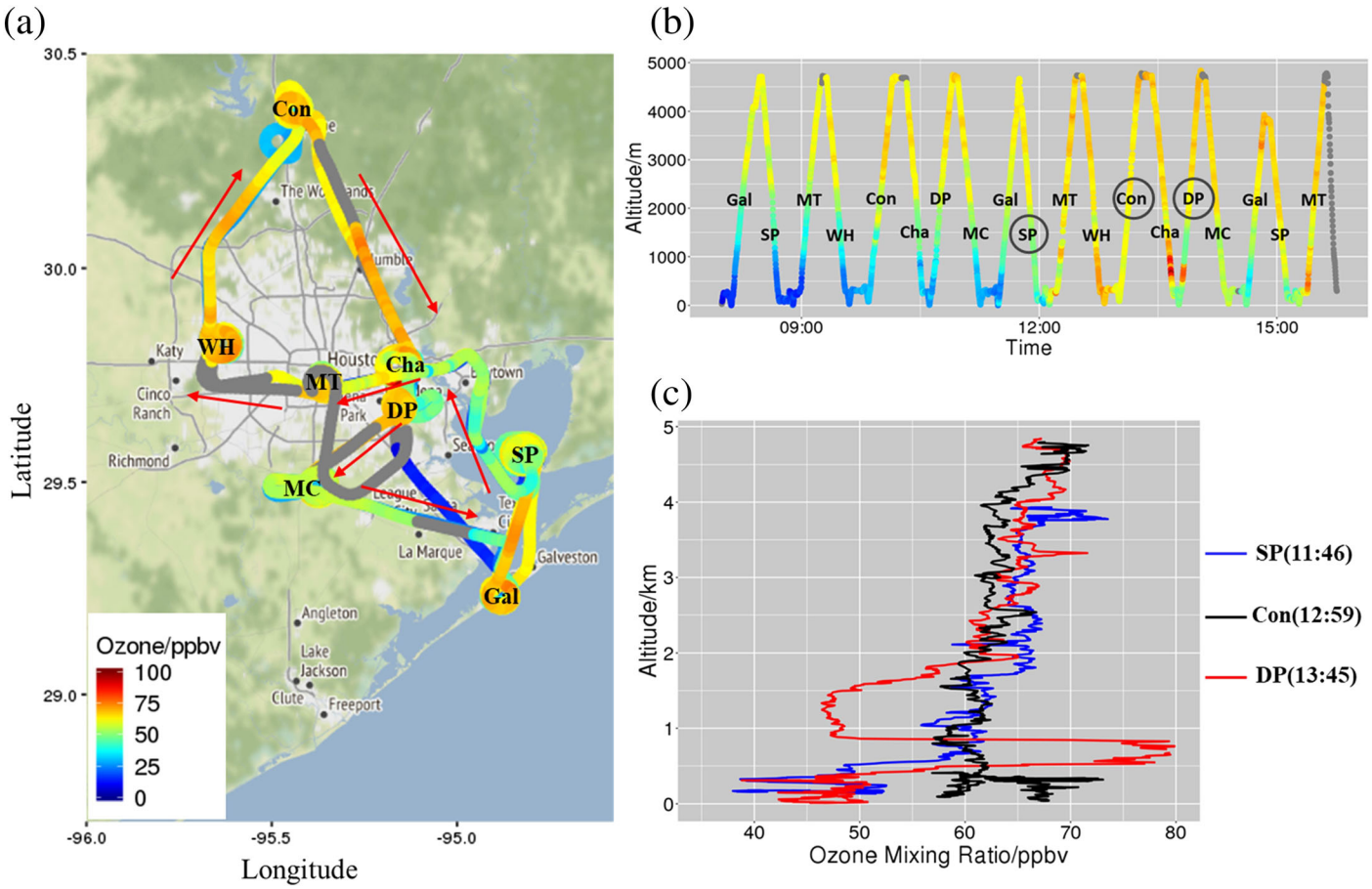

Figure 13. The same with Figure 11 but on 4 September 2013 (Sea Breeze Cluster).

On 24 September 2013 (Northerly Cluster), ozone values at two near-water sites (Gal and SP) were below 50 ppbv over the course of the day at all levels. At inland sites, there was an ~20 ppbv increase of ozone from morning to afternoon at both surface and higher levels. From the afternoon ozone profiles at Con, WH, and MC (Figure 12c), surface ozone increased from ~50 ppbv at Con to ~57 ppbv at WH and ~65 ppbv at MC. However, a southward negative ozone gradient from ~1.5 up to 2.5 km can also be seen, with higher ozone of ~67 ppbv at Con and WH and lower ozone of ~54 ppbv at MC. This gradient above the mixing layer indicates a pollution plume transported from the north. This elevated ozone layer would further travel to the south and stay until the next morning before mixed down as the mixing layer grows as reported by a case study on the next day, 25 September 2019 (Caicedo et al., 2019).

Figure 13 shows the ozone distribution of a sea breeze reversal day on 4 September 2013. Ozone values were stable at ~65 ppbv across the day at high levels due to the northeasterly winds (Figure 6d) transporting continental air. In the morning low ozone values less than ~25 ppbv can be found at all sites near the surface, while they increased by 30–50 ppbv in the afternoon. The highest ozone levels were captured at Cha and DP sites around 13:30. This indicates the arrival of sea breeze on these two sites, which can also be derived from the time when the surface wind directions shift to the southeast after 12:00 at LaPorte (Figure 6d). A high ozone layer with a value of ~80 ppbv can be discerned between 500 m to 1 km at DP (Figure 13c). This is due to the convergence of sea breeze with opposing weak northerly flow as depicted by Banta et al. (2005). In comparison, surface ozone value was ~60 ppbv at Con and ~50 ppbv at SP. This can be explained by the purge effects at SP after the passage of sea breeze.

To further demonstrate the recirculation of emissions, we show the NO_2 column along the B-200 aircraft track on 4 September in Figure 14. Black arrows indicate the forward trajectory starting from La Porte wind profiler. In the morning, high values of NO_2 were mostly concentrated at Ship Channel and Texas City

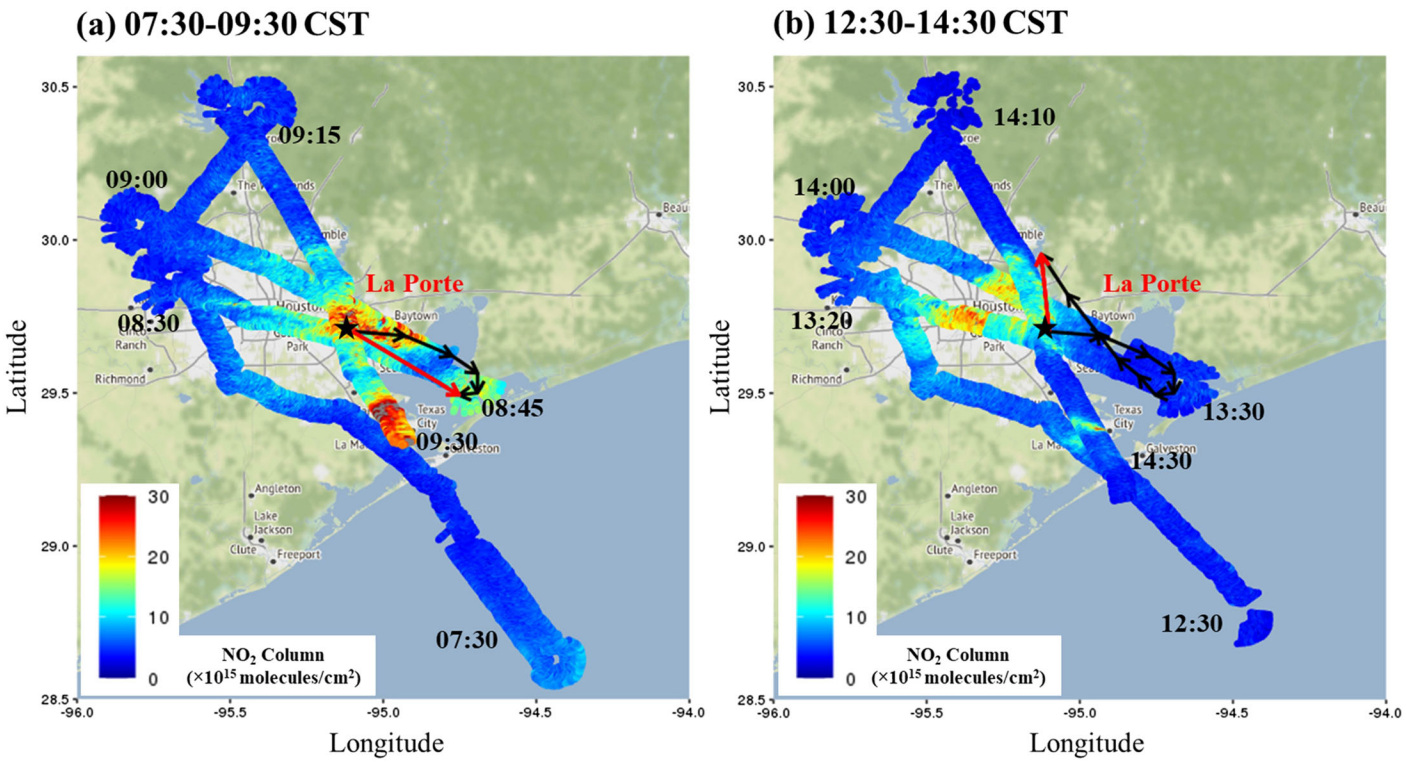

Figure 14. DISCOVER-AQ GCAS measurements of NO_2 columns (color) in the morning (a) and afternoon (b) on 4 September 2013. Overlaid are hourly forward wind trajectory (black arrows) and transport distance (red arrow) from 06:00 to 10:00 (a) and 15:00 (b) at La Porte wind profiler. The trajectory was obtained by averaging hourly winds from 200 to 500 m AGL.

where emissions were high. The trajectory indicates that NO_x emissions were transported from La Porte to Smith Point across the Galveston Bay, resulting in ozone production over the Bay. After the onset of the sea breeze in the afternoon, ozone over the Bay and remaining NO_x were transported onshore to the downtown area at about 14:00, leading to even higher ozone levels there. This corresponds to the high ozone values captured at DP and Cha in Figure 13. This case clearly reveals the effect of sea breeze recirculation on NO_x distribution and how it can cause high ozone.

In general, the aircraft data shows that ozone level above the mixing layer is dependent on wind directions. It is about 10 ppbv higher with northerly winds (~65 ppbv in Northerly and Sea Breeze Cluster) than easterly winds (~55 ppbv in Oscillation Cluster). Although there was no flight within Southerly Cluster, we can expect the lowest ozone level with southerly winds because they usually bring clean marine air inland. Considering upper level ozone is a good representative of background ozone, our results are consistent with the study of Berlin et al. (2013) which concluded that highest background ozone is associated with northeasterly winds over the HGB region. Within the mixing layer, ozone basically follows the spatial patterns established in Figure 10, with high ozone occurring at downwind sites. The most noteworthy feature of ozone vertical profile in Sea Breeze Cluster is an elevated ozone layer between 0.5 and 1 km, which corresponds to the updrafts in the sea breeze convergence zone described by Banta et al. (2005). All these consistencies with previous studies and above-established conclusions further provide support for our clustering results.

6. Conclusions

We developed a new method to automatically identify sea breeze recirculation over coastal areas by applying K-Means clustering algorithm driven by seven well-defined features. To test this methodology, we applied it to the DISCOVER-AQ field campaign period from August to October 2013 and selected the HGB and SA region in Texas as study areas. As a result, four clusters each were obtained for the two regions, including three similar synoptic clusters, referred to as Oscillation, Northerly and Southerly Cluster, and one local cluster, named Sea Breeze Cluster in HGB and Stagnation Cluster in SA. The similar synoptic clusters yet

different local cluster between two study regions indicate that our method is capable of classifying both large-scale and mesoscale wind patterns. Wind direction is persistent and wind speed is high in three synoptic clusters. The major difference among the clusters is the transport direction, which determines the origins of the air mass brought to the study regions. On the contrary, wind direction is variable and speed is low in local wind patterns, which is conducive to the accumulation and recirculation of ozone and other pollutants.

The cluster-mean geopotential height and winds on 850 hPa confirm that the two regions are affected by similar synoptic regimes. The wind pattern in Oscillation Cluster is a result of migratory anticyclone centered over northeast Texas. A high pressure to the west of Texas together with a low pressure to the north inland are the reason for the northerly wind pattern. This phenomenon usually happens after the cold front passage. Southerly wind pattern occurs because of the high-pressure system near the coast of Gulf of Mexico under the influence of the west extension of BH, which starts its retreat back to the east during our study period. The weak wind conditions in local clusters are caused by a relatively weak high pressure sitting over the center of Texas.

We also linked the resulted wind patterns with ozone concentrations. The highest MDA8 ozone exists in Sea Breeze/Stagnation Cluster, in which light and veering winds are conducive for the accumulation and recirculation of pollutants. The southerly wind pattern brings warm, clean marine air inland and helps dissipate the pollutants, therefore resulting in lower ozone concentrations. In comparison, the arrival of polluted continental air from generally northerly and easterly winds is associated with relatively higher ozone levels, although air temperature of this cluster is the lowest among the four clusters. The eastward position of emission sources and higher temperature in Oscillation Cluster make its ozone levels slightly higher than that in Northerly Cluster. For the spatial distribution of surface ozone in the HGB region, it follows the pattern that ozone is higher in the downwind area and the aircraft data further verifies that.

An advantage of our method is its generality because it avoids the *ad hoc* confinements made by other studies to select sea breeze days. In addition, by adapting the transport direction and recirculation parameters from the AW method, we can perform the classification of synoptic and local wind patterns simultaneously. This lowers the data requirement for synoptic wind pattern classification in previous studies which usually involves large coarse-resolution data sets of sea level pressure, geopotential height, or wind fields. We only used in situ measurements of wind speed and direction as inputs for the clustering algorithm and these data are routinely monitored by stations. Thus, it lowers the requirement for computation capacity and improves its applicability at the same time. Since the local recirculation is challenging to reproduce by chemical transport models, the identification of sea breeze recirculation days will allow for a focused evaluation of the model on those days. Although this method can identify sea breeze recirculation, we note that it cannot capture the kind of sea breeze superposed by strong southerly synoptic flow as suggested by Tucker et al. (2010). This type of see breeze, which would be classified as Southerly Cluster instead, does not have recirculation and thus not a concern for air quality.

By classifying wind patterns and linking them to ozone levels, we present a clear picture of the meteorological conditions associated with the daily ozone variability in both the HGB and SA regions. Our method can also be applied to other coastal regions with air quality problems to better comprehend the meteorological effects on the air pollution level and distribution. The present paper mainly focuses on the meteorological factors related to ozone concentration. Further studies in emission and chemistry of ozone precursors are necessary to fully understand what chemical processes are responsible for ozone changes under different meteorological scenarios.

Acknowledgments

This research was supported by Texas Commission on Environmental Quality (TCEQ) (Grant No. 582-19-94456-07). The authors acknowledge TCEQ for providing the hourly wind, temperature, relative humidity and MDA8 ozone data, and Global Modeling and Assimilation Office (GMAO) at NASA Goddard Space Flight Center for the MERRA-2 reanalysis zonal (U) and meridional (V) components of wind, geopotential height data, and NASA Langley Atmospheric Science Data Center for providing the P-3B aircraft data, and researchers involving in the DISCOVER-AQ field campaign for collecting and sharing ozone profiler data. The findings, opinions, and conclusions are the work of the authors and do not necessarily represent findings, opinions, or conclusions of TCEQ. The authors would like to thank the reviewers for their numerous valuable comments that significantly helped improve the article.

Data Availability Statement

The data used for this study are available online (at https://doi.org/10.7910/DVN/IYQCXG).

References

Abbs, D. J., & William, L. (1992). Sea-breeze observations and modelling: A review. *Sea*, *50*, 95s.

Adame, J. A., Serrano, E., Bolívar, J. P., & de la Morena, B. A. (2010). On the tropospheric ozone variations in a coastal area of southwestern Europe under a mesoscale circulation. *Journal of Applied Meteorology and Climatology*, *49*(4), 748–759. https://doi.org/10.1175/2009jamc2097.1

Ainslie, B., & Steyn, D. G. (2007). Spatiotemporal trends in episodic ozone pollution in the lower Fraser Valley, British Columbia, in relation to mesoscale atmospheric circulation patterns and emissions. *Journal of Applied Meteorology and Climatology*, *46*(10), 1631–1644. https://doi.org/10.1175/jam2547.1

Allwine, K. J., & Whiteman, C. D. (1994). Single-station integral measures of atmospheric stagnation, recirculation and ventilation. *Atmospheric Environment*, *28*(4), 713–721.

Arritt, R. W. (1993). Effects of the large-scale flow on characteristic features of the sea breeze. *Journal of Applied Meteorology*, *32*(1), 116–125.

Banta, R. M., Senff, C. J., Alvarez, R. J., Langford, A. O., Parrish, D. D., Trainer, M. K., et al. (2011). Dependence of daily peak O_3 concentrations near Houston, Texas on environmental factors: Wind speed, temperature, and boundary-layer depth. *Atmospheric Environment*, *45*(1), 162–173. https://doi.org/10.1016/j.atmosenv.2010.09.030

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

Berlin, S. R., Langford, A. O., Estes, M., Dong, M., & Parrish, D. D. (2013). Magnitude, decadal changes, and impact of regional background ozone transported into the greater Houston, Texas, area. *Environmental Science & Technology*, *47*(24), 13,985–13,992. https://doi.org/10.1021/es4037644

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*, 13,457–13,474. https://doi.org/10.1029/2019JD031725

Burlando, M. (2008). The synoptic-scale surface wind climate regimes of the Mediterranean Sea according to the cluster analysis of ERA-40 wind fields. *Theoretical and Applied Climatology*, *96*(1–2), 69–83. https://doi.org/10.1007/s00704-008-0033-5

Caicedo, V., Rappenglueck, B., Cuchiara, G., Flynn, J., Ferrare, R., Scarino, A. J., et al. (2019). Bay breeze and sea breeze circulation impacts on the planetary boundary layer and air quality from an observed and modeled DISCOVER-AQ Texas case study. *Journal of Geophysical Research: Atmospheres*, *124*, 7359–7378. https://doi.org/10.1029/2019JD030523

Carvalho, A., Monteiro, A., Ribeiro, I., Tchepel, O., Miranda, A. I., Borrego, C., et al. (2010). High ozone levels in the northeast of Portugal: Analysis and characterization. *Atmospheric Environment*, *44*(8), 1020–1031. https://doi.org/10.1016/j.atmosenv.2009.12.020

Chen, F., Miao, S., Tewari, M., Bao, J.-W., & Kusaka, H. (2011). A numerical study of interactions between surface forcing and sea breeze circulations and their effects on stagnation in the greater Houston area. *Journal of Geophysical Research*, *116*, D12105. https://doi.org/10.1029/2010JD015533

Christiansen, B. (2007). Atmospheric circulation regimes: Can cluster analysis provide the number? *Journal of Climate*, *20*(10), 2229–2250.

Corte-Real, J., Qian, B., & Xu, H. (1998). Regional climate change in Portugal: Precipitation variability associated with large-scale atmospheric circulation. *International Journal of Climatology*, *18*(6), 619–635.

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.

David, L. M., & Nair, P. R. (2011). Diurnal and seasonal variability of surface ozone and NO_x at a tropical coastal site: Association with mesoscale and synoptic meteorological conditions. *Journal of Geophysical Research*, *116*, D10303. https://doi.org/10.1029/2010JD015076

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.

Dayan, U., & Levy, I. (2005). The influence of meteorological conditions and atmospheric circulation types on PM10 and visibility in Tel Aviv. *Journal of Applied Meteorology*, *44*(5), 606–619. https://doi.org/10.1175/JAM2232.1

Flocas, H., Kelessis, A., Helmis, C., Petrakakis, M., Zoumakis, M., & Pappas, K. (2008). Synoptic and local scale atmospheric circulation associated with air pollution episodes in an urban Mediterranean area. *Theoretical and Applied Climatology*, *95*(3–4), 265–277. https://doi.org/10.1007/s00704-008-0005-9

Helmis, C., Asimakopoulos, D., Papadopoulos, K., Kassomenos, P., Kalogiros, J., Papageorgas, P., & Blikas, S. (1997). Air mass exchange between the Athens Basin and the Messogia Plain of Attika, Greece. *Atmospheric Environment*, *31*(22), 3833–3849.

Helmis, C., Tombrou, M., Asimakopoulos, D., Soilemes, A., Güsten, H., Moussiopoulos, N., & Hatzaridou, A. (1997). Thessaloniki'91 field measurement campaign—I. Wind field and atmospheric boundary layer structure over greater Thessaloniki area, under light background flow. *Atmospheric Environment*, *31*(8), 1101–1114.

Huth, R., Beck, C., Philipp, A., Demuzere, M., Ustrnul, Z., Cahynova, M., et al. (2008). Classifications of atmospheric circulation patterns: Recent advances and applications. *Annals of the New York Academy of Sciences*, *1146*(1), 105–152. https://doi.org/10.1196/annals.1446.019

Lei, R., Talbot, R., Wang, Y., Wang, S.-C., & 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*(5). https://doi.org/10.3390/atmos9050159

Levy, I., Dayan, U., & Mahrer, Y. (2008). A five-year study of coastal recirculation and its effect on air pollutants over the East Mediterranean region. *Journal of Geophysical Research*, *113*, D16121. https://doi.org/10.1029/2007JD009529

Ma, Y., & Lyons, T. J. (2003). Recirculation of coastal urban air pollution under a synoptic scale thermal trough in Perth, Western Australia. *Atmospheric Environment*, *37*(4), 443–454.

MacQueen, J. (1967). *Some methods for classification and analysis of multivariate observations*. Paper presented at Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Oakland, CA, USA.

Miller, S. T. K. (2003). Sea breeze: Structure, forecasting, and impacts. *Reviews of Geophysics*, *41*(3), 1011. https://doi.org/10.1029/2003RG000124

Nankar, D., Patra, A., Dole, M., Venkataraman, S., & Hegde, A. (2009). Atmospheric stagnation, recirculation and ventilation characteristics at Kakrapar atomic power station site. *Annals of Nuclear Energy*, *36*(4), 475–480.

Ngan, F., & Byun, D. (2011). Classification of weather patterns and associated trajectories of high-ozone episodes in the Houston–Galveston–Brazoria area during the 2005/06 TexAQS-II. *Journal of Applied Meteorology and Climatology*, *50*(3), 485–499. https://doi.org/10.1175/2010jamc2483.1

Nielsen-Gammon, J., Tobin, J., McNeel, A., & Li, G. (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

Nielson-Gammon, J. W. (2016). Analysis of wind rotation in Houston and surrounding areas as detected by the TEXAQS-II radar wind profiler network. 8-hour ozone SIP coalition.

Nowlan, C. R., Liu, X., Janz, S. J., Kowalewski, M. G., Chance, K., Follette-Cook, M. B., et al. (2018). Nitrogen dioxide and formaldehyde measurements from the GEOstationary coastal and air pollution events (GEO-CAPE) airborne simulator over Houston, Texas. *Atmospheric Measurement Techniques*, *11*(11), 5941–5941, 5964. https://doi.org/10.5194/amt-11-5941-2018

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.

Pérez, I. A., Sánchez, M. L., García, M. Á., & Paredes, V. (2011). Relationship between CO_2 at a rural site and integral measures of atmospheric stagnation, recirculation, and ventilation. *Naturwissenschaften*, *98*(7), 565–574. https://doi.org/10.1007/s00114-011-0800-5

Rappenglück, B., Perna, R., Zhong, S., & Morris, G. A. (2008). An analysis of the vertical structure of the atmosphere and the upper-level meteorology and their impact on surface ozone levels in Houston, Texas. *Journal of Geophysical Research*, *113*, D17315. https://doi.org/10.1029/2007JD009745

Russo, A., Gouveia, C., Levy, I., Dayan, U., Jerez, S., Mendes, M., & Trigo, R. (2016). Coastal recirculation potential affecting air pollutants in Portugal: The role of circulation weather types. *Atmospheric Environment*, *135*, 9–19. https://doi.org/10.1016/j.atmosenv.2016.03.039

Russo, A., Trigo, R. M., Martins, H., & Mendes, M. T. (2014). NO_2, PM_{10} and O_3 urban concentrations and its association with circulation weather types in Portugal. *Atmospheric Environment*, *89*, 768–785. https://doi.org/10.1016/j.atmosenv.2014.02.010

Saavedra, S., Rodriguez, A., Taboada, J. J., Souto, J. A., & Casares, J. J. (2012). Synoptic patterns and air mass transport during ozone episodes in northwestern Iberia. *Science of the Total Environment*, *441*, 97–110. https://doi.org/10.1016/j.scitotenv.2012.09.014

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. M. (2015). Bay breeze climatology at two sites along the Chesapeake bay from 1986–2010: Implications for surface ozone. *Journal of Atmospheric Chemistry*, *72*(3–4), 355–372. https://doi.org/10.1007/s10874-013-9260-y

Surkova, G. (2013). Air recirculation and ventilation in the coastal regions of the Black Sea. *Open Geosciences*, *5*(2), 196–207. https://doi.org/10.2478/s13533-012-0126-7

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: 19 Apr 2020)

Toreti, A., Xoplaki, E., Maraun, D., Kuglitsch, F. G., Wanner, H., & Luterbacher, J. (2010). Characterisation of extreme winter precipitation in Mediterranean coastal sites and associated anomalous atmospheric circulation patterns. *Natural Hazards and Earth System Sciences*, *10*(5), 1037–1050. https://doi.org/10.5194/nhess-10-1037-2010

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

Venegas, L., & Mazzeo, N. (1999). Atmospheric stagnation, recirculation and ventilation potential of several sites in Argentina. *Atmospheric Research*, *52*(1–2), 43–57.

Wang, Y., Jia, B., Wang, S.-C., 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-16-15265-2016

Wu, Y.-L., Lin, C.-H., Lai, C.-H., Lai, H.-C., & Young, C.-Y. (2010). Effects of local circulations, turbulent internal boundary layers, and elevated industrial plumes on coastal ozone pollution in the downwind Kaohsiung urban-industrial complex. *Terrestrial, Atmospheric and Oceanic Sciences*, *21*(2). https://doi.org/10.3319/tao.2009.04.14.01(a)

Zhou, C., Wei, G., Zheng, H., Russo, A., Li, C., Du, H., & Xiang, J. (2019). Effects of potential recirculation on air quality in coastal cities in the Yangtze River Delta. *Science of the Total Environment*, *651*(Pt 1), 12–23. https://doi.org/10.1016/j.scitotenv.2018.08.423

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