7.3 IMPACT OF SURFACE WIND PATTERNS ON SURFACE OZONE IN HOUSTON, TEXAS AS REVEALED BY CLUSTER ANALYSIS

Lisa S. Darby
NOAA/Environmental Technology Laboratory, Boulder, Colorado

1. Introduction

Cluster analysis is a convenient method for grouping large amounts of data into categories that are easy to work with. The basic principle is that the differences among the data points within a cluster are minimized, while the differences among the clusters are maximized. It is an objective method used in many areas of research (Everitt et al. 2001), including meteorology (Weber and Kaufmann 1995, Kaufmann and Whiteman 1999).

When analyzing meteorological data over the course of a season, there can be an overwhelming amount of data to peruse, particularly when dealing with measurements from a network of instruments. In order to ascertain how the summertime surface winds in Houston affected ozone behavior across the metropolitan area, it was decided to use cluster analysis to organize the measurements from an extensive array of surface meteorological instruments. Hourly averages of wind speed and wind direction from 22 stations in and around Houston (Fig. 1) were clustered into 16 categories using the cluster analysis code included in the Interactive Data Language (IDL) software package. As each hour was assigned to a cluster, the maximum ozone in the network for that same hour was stored with the cluster analysis results.

Examples of three types of results will be shown in this paper: 1) the type of cluster, i.e., wind pattern, most likely to occur coincident with ozone values that exceeded the one-hour ozone concentration standard of 120 ppbv; 2) a sequence of clusters likely to result in an ozone exceedance; and 3) an example of a cluster that was tied to the diurnal cycle, as determined by a frequency-of-occurrence analysis.

2. Method

Data were analyzed for 27 days from mid-August to mid-September 2000, during the time of the Texas Air Quality Study 2000 (TexAQS 2000) (Brock et al. 2003, Roberts et al. 2003). As seen in Fig. 1, most of the stations included in the analysis were within the urbanized area (denoted by the dashed line on the map) adjacent to Galveston Bay. A few stations outside the urbanized area brought additional information regarding the Gulf of Mexico coastal environment and areas to the north and east of Houston.

The input for cluster analysis cannot tolerate any missing data, so data from each station were carefully inspected for missing data. If four or fewer consecutive hours of data were missing, and it was determined that the surrounding good data could be interpolated to fill in the missing data, then a linear data interpolation was performed. If more than 4 consecutive hours of data were missing, or the data were such that a reasonable interpolation could not be made, the station was dropped from the analysis. After the data perusal and interpolations were completed, 22 stations were available for analysis. The u- and v-components for each station, for each hour of the 27 days, were calculated. This gave 648 hourly samples for the cluster analysis, with each sample comprised of the u- and v-components of all 22 stations for that hour.

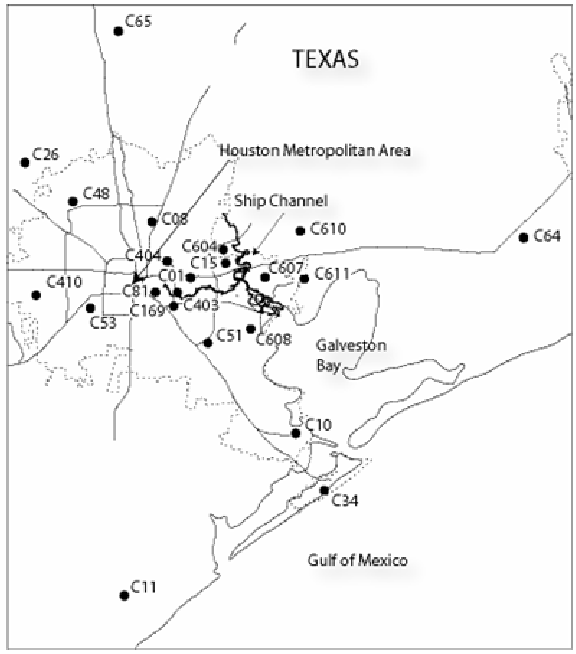


Figure 1*: Map of Houston metropolitan area, including major highways. The dashed line encloses the most populated areas. Black circles indicate locations of surface stations used in the cluster analysis. Information about the stations can be found at http://www.tnrcc.state.tx.us/cgi-bin/monops/site_info.*

* *Corresponding author address*: Lisa S. Darby, NOAA/ETL, R/ET2, 325 Broadway, Boulder, CO, 80305; email: Lisa.Darby@noaa.gov

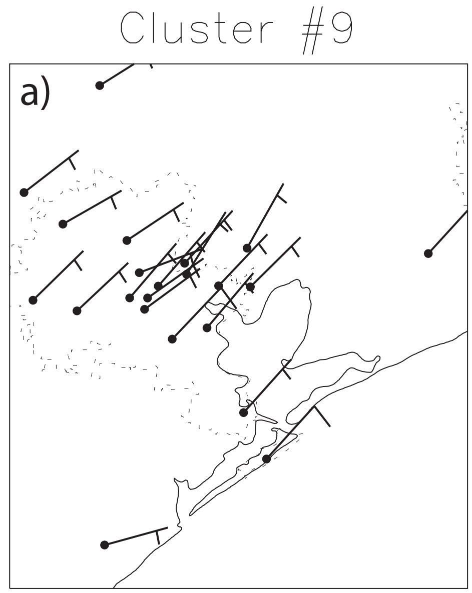
To determine how many clusters were appropriate for the analysis, the cluster analysis was performed first for 8 clusters, and then 12 more times, each time incrementing the number of clusters by one. When more than 16 clusters were requested, the amount of information gained by having additional clusters, for this analysis, declined as the number of clusters increased. Thus, the number 16 was chosen as the optimum number of clusters. Although cluster analysis is objective, determining the number of clusters, as required by the software scheme provided in IDL and other cluster analysis methods, is somewhat subjective.

3. Results

Figure 2 shows winds representing Cluster #9 (Fig. 2a) and the winds from an hour assigned to Cluster #9 (Fig. 2b). To create the plots that represent each cluster, as shown in Fig. 2a, the winds for all hours assigned to a cluster were vector-averaged. In the case of Cluster #9, 55 hours were assigned to it. Thus, Fig. 2a is an average of those 55 hours. Strong similarities exist between the Cluster #9 plot and the plot of hourly-averaged winds for 0900 LST 6 September. All other clusters (not shown) also exhibited a high resemblance to individual hourly plots, with those clusters with a higher number of hours assigned to them having fewer similarities, due to more extensive averaging to create the representative plot. The strong similarities found between the cluster plots and actual hourly plots supports the use of cluster analysis as a means of assessing Houston summertime winds.

During the summer, on clear afternoons with light winds, a thermally forced flow, the Gulf Breeze, forms because of the temperature contrast between the relatively cooler waters of Galveston Bay and the Gulf of Mexico, and the hot land surface. The Gulf breeze brings a new air mass into the Houston area, affecting the city's air quality. Use of cluster analysis helps to demonstrate the impact the Gulf breeze has on Houston's ozone pollution problem.

Figure 3 shows the plot representing Cluster #10 (Fig. 3a), along with the frequency-of-occurrence analysis for this cluster, by hour (Fig. 3b). This cluster only occurred between 1300 and 1900 LST, a clear indication that it was tied to the diurnal cycle. Since the winds in this cluster were perpendicular to the shore and only occurred in the afternoon, when thermally-forced onshore flow is expected to occur, this cluster is a Gulf Breeze cluster. Three other clusters had similar characteristics. The differences among them were wind speed and peak time of occurrence, so they thus represent different phases of the Gulf breeze. It was Cluster #10, however, that was most likely to be coincident with ozone exceedances, i.e., a one-hour average of ozone concentration ≥ 120 ppbv. Cluster #10 occurred for a total of 13 hours when the maximum ozone in the network was ≥ 120 ppbv, accounting for 20% of the exceedances. Cluster #8 (seen in Fig. 4d), is a weaker version of the Gulf Breeze that is likely to occur earlier in the day than Cluster #10. It accounted for the next highest number of exceedances (10). Overall, the four Gulf breeze clusters coincided with 46.8% of all exceedances that occurred during the 27 days of the analysis.



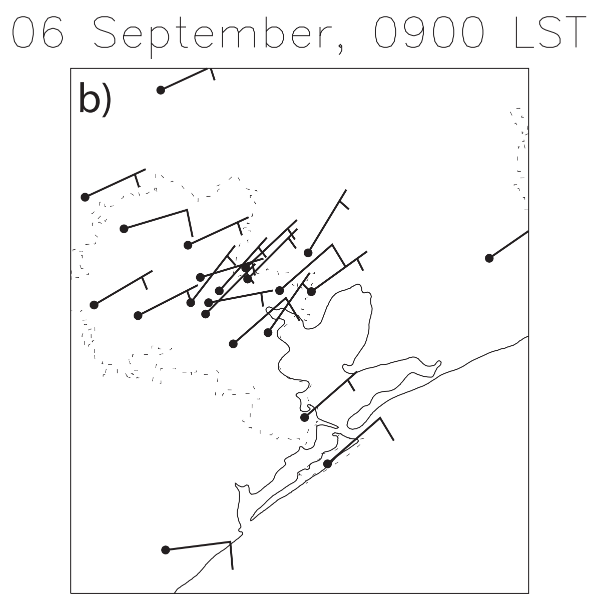


***Figure 2**: Comparison between a representative cluster plot and one of the hours assigned to that cluster. a) Vector-average of all hours assigned to Cluster #9. Wind barbs indicate direction and speed in m s^{-1} (half-barb represents 5 m s^{-1}, full barb 10 m s^{-1}). b) Plot of hourly-averaged winds for 0900 LST, 06 September 2000.*

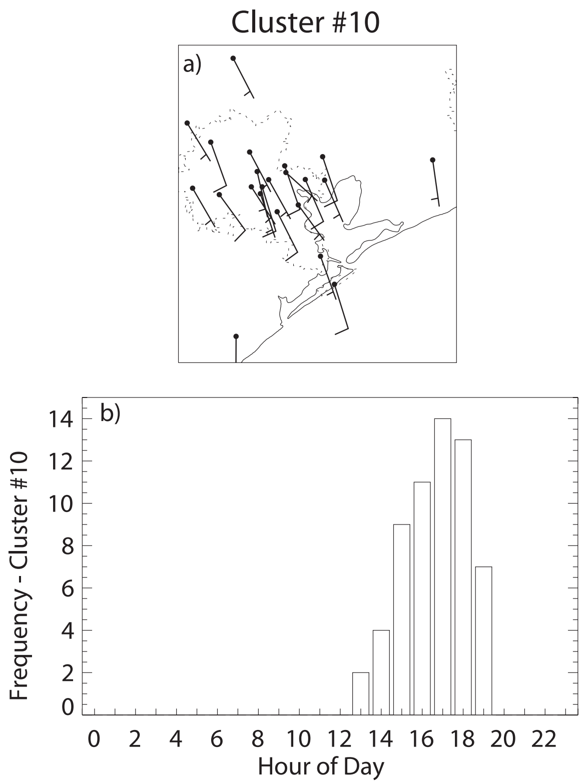


***Figure 3**: Representation of all hours assigned to Cluster #10 and the frequency of occurrence analysis for Cluster #10. a) As in Fig. 2, except for Cluster #10. b) Hour of day (LST) vs. number of occurrences for each hour.*

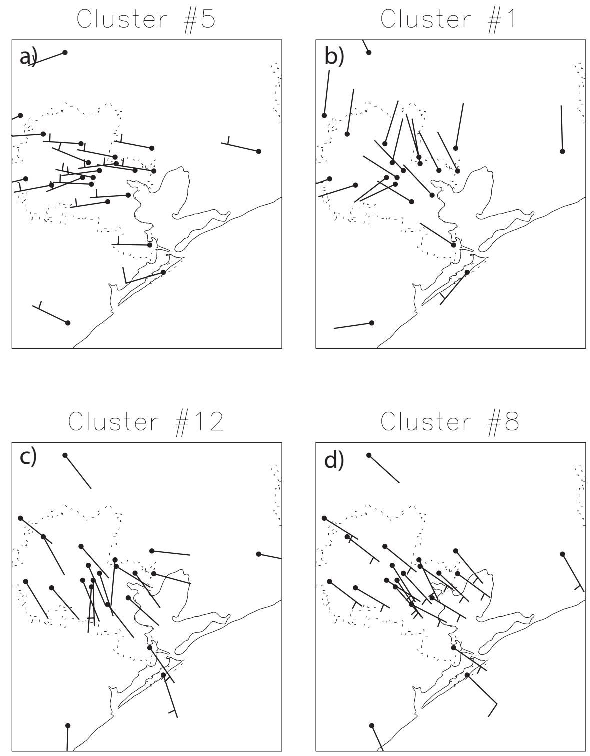


***Figure 4**: Sequence of clusters representing the typical winds associated with summertime high-ozone days in Houston. a) Cluster #5, morning westerly winds. b) Cluster #1, a 'stagnant' cluster. c) Cluster #12, weak Gulf breeze flow. d) Cluster #8, slightly stronger Gulf breeze flow.*

The sequences of clusters that occurred in the 5 hours preceding the maximum ozone of the day, plus the cluster coincident with the maximum ozone of the day, regardless of where or when the maximum occurred, were analyzed for each day. It was found that days with the highest ozone exceedances, particularly days with exceedances ≥ 140 ppbv, tended to have a sequence much like that shown in Fig. 4. This sequence starts with winds with a westerly component (Fig. 4a, Cluster #5). Cluster #5 was most likely to occur between 0500 and 0900 LST, as indicated by the frequency-of-occurrence analysis (not shown). These westerly winds will transport ozone and its precursors from downtown Houston to Galveston Bay. Cluster #1, the most frequently occurring cluster, represents stagnant winds (Fig. 4b). This cluster was most likely to occur at night, but when it occurred during the late morning hours its affect on Houston air quality was significant because it allowed ozone and its precursors to build up over the metropolitan area, as described by Banta et al. (2000). The rest of the sequence is filled out by another light wind cluster that has a weak onshore component (Fig. 4c, Cluster #12), followed by a well-formed Gulf breeze cluster (Fig. 4d, Cluster #8). Winds represented by Clusters #12 and #8 bring the polluted air previously advected toward Galveston Bay back into Houston. Under different circumstances, e.g., mornings without westerly component flow, the Gulf breeze *may* bring cleaner air into Houston. Thus, it is important to consider the complete sequence of clusters throughout the day when assessing their impact on Houston air quality.

4. Discussion and future work

The complex coastline near Houston, TX, including Galveston Bay and the Gulf of Mexico, lead to complex mesoscale wind flows. These winds, along with the large number of oil refineries near the waterways of the area, enhance the ozone problem in Houston because of the low-level transport associated with them. This analysis is presented with the understanding that meteorology alone does not fully explain the high ozone incidents in Houston; the complex chemistry of the region needs to be assessed also (Wert et al. 2003, Ryerson et al. 2003). However, the cluster analysis results indicate that the transition from offshore winds to stagnant winds to onshore winds serve to enhance the ozone concentrations that occur in the city as a result of the chemistry associated with the petrochemical industry. Early morning offshore flow transports ozone and its precursors toward Galveston Bay, the stagnant winds allow newly produced pollutants to accumulate over the city, and the onshore flow recirculates the older pollutants previously advected eastward back into the city to mix with the less-aged pollutants.

Future anticipated work includes using cluster analysis to analyze Houston surface wind data by season. Summertime is not the only time ozone exceedances occur in Houston. Early autumn also has its share of high-ozone days, but the meteorology is different. Cluster analysis will help determine the differences in the meteorology that lead to the summer versus the fall ozone exceedances. Another area of future research includes testing cluster analysis on a network of radar wind profilers deployed to New England to ascertain if it will be a useful tool for understanding how winds above the surface may influence surface ozone concentrations.

5. Acknowledgments

The author wishes to thank the Texas Commission on Environmental Quality for maintaining the data base of surface measurements on their web site for public use. The analysis was funded by NOAA's Health of the Atmosphere program.

6. References

Banta, R.M., C. J. Senff, L. S. Darby, T. B. Ryerson, M. Trainer, and R. J. Alvarez, 2002: 3-D distribution of ozone during the major pollution event of 30 August 2000 during TEXAQS 2000. *Fourth onference on Atmospheric Chemistry: Urban, Regional, and Global Scale Impacts of Air Pollutants*, 13-18 January, 2002, Orlando, FL. 267-269.

Brock, C.A., M. Trainer, T.B. Ryerson, J.A. Neuman, D.D. Parrish, J.S. Holloway, D.K. Nicks Jr., G.J. Frost, G. Hübler, F.C. Fehsenfeld, J.C. Wilson, J.M. Reeves, B.G. Lafleur, H. Hilbert, E.L. Atlas, S.G. Donnelly, S.M. Schauffler, V.R. Stroud, and C. Wiedinmyer, 2003: Particle growth in urban and industrial plumes in Texas. *J. Geophys. Res.*, **108**(d3), 4111, doi:10.1029/2002JD002746.

Everitt, B., S. Landau, and M. Leese, 2001: *Cluster Analysis*, 4th ed. Edward Arnold Publishers, 237 pp.

Kaufmann, P., and C.D. Whiteman, 1999: Cluster-analysis classification of wintertime wind patterns in Grand Canyon region. *J. Appl. Meteor.,* **38**, 1131-1147.

Roberts, J.M., B.T. Jobson, W. Kuster, P. Goldan, P. Murphy, E. Williams, G. Frost, D. Riemer, E. Apel, C. Stroud, C. Wiedinmyer, and F. Fehsenfeld, 2003: An examination of the chemistry of peroxycarboxylic nitric anhydrides and related volatile organic ompounds during Texas Air Quality Study 2000 using ground-based measurements. *J. Geophys. Res.*, **108**(d16), 4495, doi:10.1029/2003JD003383.

Ryerson, T.B., M. Trainer, W.M. Angevine, C.A. Brock, R.W. Dissly, F.C. Fehsenfeld, G.J. Frost, P.D. Goldan, J.S. Holloway, G. Hübler, R.O. Jakoubek, W.C. Kuster, J.A. Neuman, D.K. Nicks Jr., D.D. Parrish, J.M. Roberts, and D.T. Sueper, E.L. Atlas, S.G. Donnelly, F. Flocke, A. Fried, W.T. Potter, S. Schauffler, V. Stroud, A.J. Weinheimer, B.P. Wert, C. Wiedinmyer, R.J. Alvarez, R.M. Banta, L.S. Darby, and C.J. Senff, 2003: Effect of petrochemical industrial emissions of reactive alkenes and NO_x on tropospheric ozone formation in Houston, Texas, 2003: *J. Geophys. Res.*, **108**(d8), 4249, doi:10.1029/2002JD003070.

Weber, R.O., and P. Kaufmann, 1995: Automated classification scheme for wind fields. *J. Appl. Meteor.,* **34**, 1133-1141.

Wert, B.P., M. Trainer, A. Fried, T.B. Ryerson, B. Henry, W. Potter, W.M. Angevine, E. Atlas, S.G. Donnelly, F.C. Fehsenfeld, G.J. Frost, P.D. Goldan, A. Hansel, J.S. Holloway, G. Hübler, W.C. Kuster, D.K. Nicks Jr., J.A. Newman, D.D. Parrish, S. Schauffler, J. Stutz, D.T. Sueper, C. Wiedinmyer, and A. Wisthaler, 2003: Signature of terminal alkene oxidation in airborne formaldehyde measurements during TexAQS 2000. *J. Geophys. Res.*, **108**(d3), 4104, doi:10.1029/2002JD002502.