Appendix 8B

Determination of Representativeness

of 2002 Ozone Season for

Ozone Transport Region SIP Modeling

FINAL REPORT

DETERMINATION OF REPRESENTATIVENESS OF 2002 OZONE SEASON FOR OZONE TRANSPORT REGION SIP MODELING

Prepared for

Tom Frankiewicz
Ozone Transport Commission
444 North Capitol St., NW, Suite 638
Washington, DC 20001

Prepared by

Till Stoeckenius
Sue Kemball-Cook
ENVIRON International Corporation
101 Rowland Way, Suite 220
Novato, CA 94945

June 3, 2005

TABLE OF CONTENTS

APPENDICES

TABLES

FIGURES

1. INTRODUCTION

The Ozone Transport Commission is coordinating a photochemical modeling study of the Ozone Transport Region (OTR) in support of State Implementation Plan development for certain areas recently designated by the United States Environmental Protection Agency (U.S. EPA) as being in nonattainment of the 8-hour ozone National Ambient Air Quality Standard (NAAQS). The OTR is comprised of 12 states (DC, CT, DE, ME, MD, MA, NH, NJ, NY, PA, RI, VT) and that portion of Virginia contained within the Washington DC Consolidated Metropolitan Statistical Area (see Figure 1-1). Areas within the OTR designated nonattainment for the 8-hour ozone NAAQS are shown in Figure 1-2; detailed attainment demonstrations are required for the nonattainment areas within the OTR classified as "moderate".

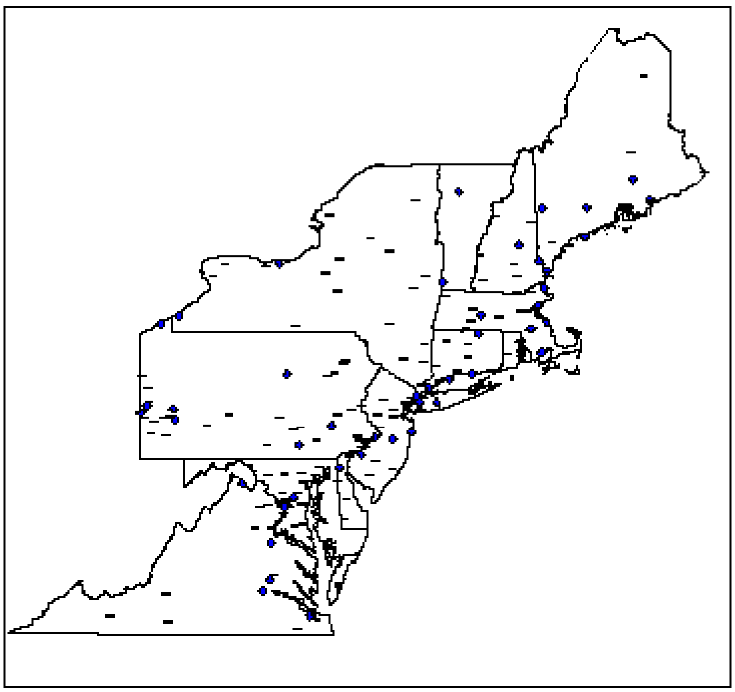


Figure 1-1. Ozone monitoring sites in the Ozone Transport Region which is comprised of DC, CT, DE, ME, MD, MA, NH, NJ, NY, PA, RI, VT, and that portion of Virginia contained within the DC Consolidated Metropolitan Statistical Area.

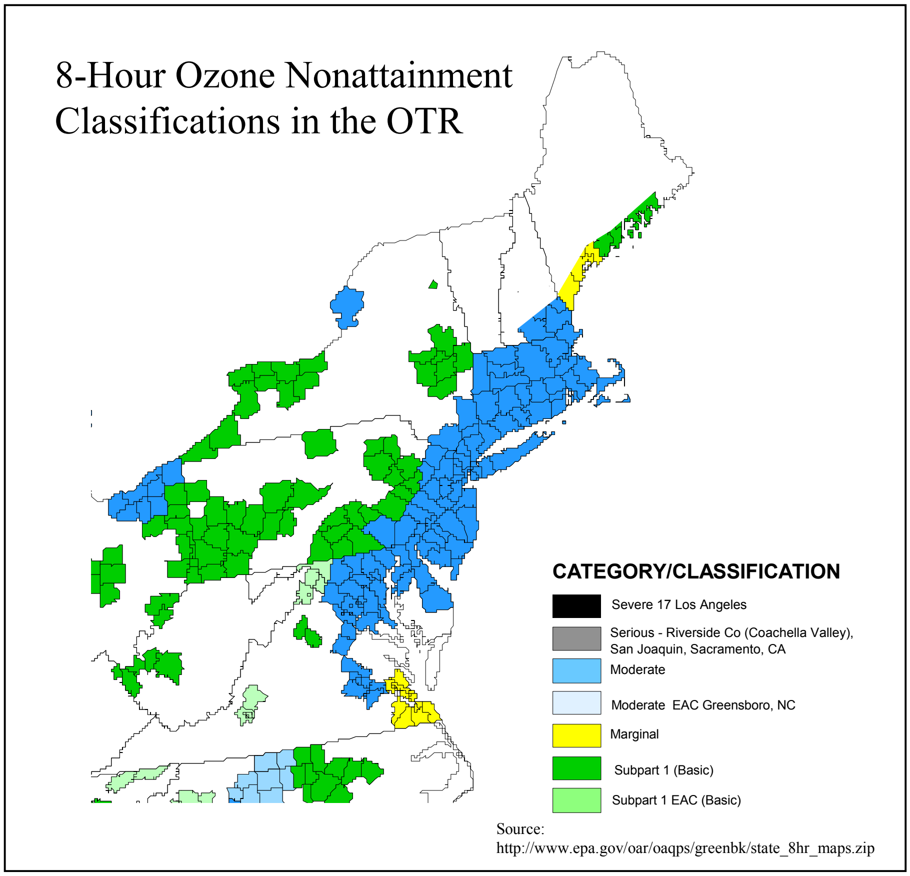


Figure 1-2. 8-hour ozone nonattainment classifications in the OTR and adjacent areas.

Development of effective 8-hour ozone attainment strategies requires application of photochemical models to a set of episodes that adequately represent the range of meteorological conditions associated with violations of the ambient standard. EPA's 8-hour ozone NAAQS modeling guidance (EPA, 1999) lists four criteria for episode selection:

1. Select episodes that both represent a variety of meteorological conditions and frequently correspond to exceedances of the 8-hour ozone standard.
2. Select episodes during which the daily maximum 8-hour ozone averages are close to the 8-hour ozone design value, i.e., the average annual 4th highest daily maximum 8-hour ozone average.
3. Select episodes for which extensive meteorological and air quality data sets are available.
4. Select a sufficient number of episode days for modeling so that the modeled attainment test specified in EPA's guidance is based on several days.

In practice, it is difficult, if not impossible, to meet all of these criteria simultaneously. In general, it is important to include episodes that represent as completely as possible the full range of meteorological conditions associated with exceedances of the ozone standard. Differences among episode types are important in so far as they influence the predicted effectiveness of alternative emission control strategies.

Because the OTR is a large region that experiences a wide variety of weather patterns associated with 8-hour ozone NAAQS exceedances, the OTC has decided to perform ozone SIP modeling of the full 2002 ozone season, May 15 - September 15, to incorporate a fairly large number of episode days in different portions of the OTR. Thus, there should be a good chance that all of the important episode types are covered within this period. However, the 2002 season includes some of the most prolonged and severe ozone episodes in recent years, raising the possibility that one or more episode types of interest are not adequately represented within the 2002 season. The goal of this study, therefore, is to assess the representativeness of conditions during the 2002 season with respect to exceedance events that have occurred in other years and determine if there are any types of episodes that are not adequately represented within the 2002 season.

EPA's 1999 draft guidance recommends joint use of subjective and statistical methods for characterizing and classifying 8-hr ozone episodes. Subjective methods include "typing" of episode meteorological conditions in which episodes are classified via inspection on the basis of similarities in meso- and synoptic-scale weather patterns. In contrast, statistical methods can produce objective classifications either by use of tree models[1] or and various forms of cluster analysis (often in conjunction with a principal components analysis). A predictive classification procedure such as a classification tree model (which can be viewed as a non-parametric form of least-squares regression) does not actually classify episodes, although it can be used to identify potential episodes with common meteorological features. This information can then be used to inform the episode selection process. A cluster analysis, on the other hand, is designed to identify natural groupings of conditions within the set of candidate episodes. In either case, considerable expert judgment is required in variable selection, selection of different modeling methods, and interpretation of results so even the statistical methods are not wholly objective. Nevertheless, these approaches are well suited to the development of valid, defensible episode classification schemes that are sufficiently robust to explain the major characteristics of ozone episode types.

In this study, we apply a combination of exploratory statistical techniques, cluster analyses, and classification tree building algorithms to ozone and meteorological data from the OTR to assess the representativeness of 8-hour ozone episodes occurring during the 2002 season. Data sources and preliminary analyses are described in Section 2. Procedures and results used to identify the major Northeastern U.S. ozone episode types and their key characteristics are presented in Section 3 along with a comparison of the frequency of occurrence and features of each episode type in 2002 versus those in other recent years. Our conclusions regarding the representativeness of the 2002 season are detailed in Section 4.

[1] A commonly used tree modeling approach is based on the CART methodology (Breiman et al., 1984).

2. DATA GATHERING AND INITIAL ANALYSIS

DATA

Daily ozone and meteorological data required for the episode representativeness analysis were obtained from a variety of sources. To capture the full range of OTR episode characteristics and insure statistical significance, a seven year period (1997 – 2003) was chosen for analysis. Data prior to 1997 were not used to avoid any confounding influences of long-term air quality trends. For purposes of this study, data from the warm season months (May – September) were used to capture most if not all high ozone events during the year.

Ozone and meteorological data were separated into two groups: data from 1997 – 2001 and 2003 were treated as the "historical" period and were used to define the types of ozone episode conditions occurring in the OTR. Data from 2002 were treated as an independent data set with data in this year to be compared against the types of conditions found in the historical period.

Hourly ozone concentrations at monitoring sites throughout the OTR for the period 1997-2003 were provided by the New York State Department of Environmental Conservation. Stations missing more than one year of data were excluded from the study, leaving a total of 158 stations with nearly complete data. Daily maximum 8-hour averages were calculated from the hourly data using the data handling conventions specified in 40 CFR 50, Appendix I. Because the spatial pattern analysis procedure requires a complete data set, missing daily maxima were set to the station mean daily maximum (this conforms to the procedure used by Cox, 1997).

Hourly surface meteorological data (winds, temperature, etc.) from airports and other locations in the OTR were obtained from the National Center for Atmospheric Research (NCAR) as dataset ds472. Upper air data were extracted from the ETA Data Assimilation System (EDAS) files available from the National Climatic Data Center. EDAS contains 3-hourly objective analysis initialization and forecast fields from the National Center for Environmental Prediction's (NCEP) ETA model at 40 km resolution. By using the EDAS data, we were able to obtain a consistent set of surface and upper air variables covering the entire eastern half of the U.S. at high temporal resolution.

IDENTIFICATION OF OZONE MONITORING SUB-REGIONS

Monitoring sub-regions were defined within the OTR to emphasize the spatial ozone patterns associated with different types of ozone episodes and to reduce the number of variables required to describe the spatial ozone distribution under different episode patterns. Sub-regions were defined by combining results of a station clustering analysis with information on typical ozone concentration patterns provided by air quality analysts from several OTR states. A variable clustering procedure (VARCLUS) based on principal components analysis was used to group the OTR ozone monitoring sites into disjoint geographic clusters (Sarle, 1990, Harrell 1999). This procedure essentially divides the monitoring stations into groups of highly correlated sites. Station clusters are selected to explain most of the day-to-day variation in ozone levels over the OTR using a small number of station groups. VARCLUS works by performing a principal components analysis on the ozone values in each candidate cluster and seeks to find the set of clusters that maximize the total (across clusters) of the variance explained by the first principal components.

Required input for the VARCLUS procedure is the number of clusters to be formed. As with any clustering procedure, this introduces an element of subjectivity that can be minimized by repeating the analysis several times, each time varying the number of clusters to be formed and examining the robustness of the cluster memberships as the number of requested clusters (k) changes.

Application of the clustering algorithm for various values of k showed that, for a given value of k, the VARCLUS procedure produced several spatially coherent clusters as well as other clusters which were not spatially coherent. Clusters which were not spatially coherent were always made up of just 5 or fewer member stations. For example, setting k=5 produced 2 coherent clusters (clusters 1 and 2) and 3 smaller clusters (clusters 3-5) whose members tended to be widely separated in space (see Figure 2-1). The version of VARCLUS used for our analysis assigns the lowest cluster identification numbers to the "tightest" (i.e., most easily identifiable and robust) clusters. As the results in Figure 2-1 show, these lowest numbered clusters (in this case Clusters 1 and 2) turned out to also be the most spatially coherent (note that the clustering is based on ozone correlations only – the locations of each monitoring site are not an input to the clustering algorithm). This is consistent with our expectation that sites located close to one another will be highly correlated. Clusters 1 and 2 are similar to the two northeast clusters found by Cox (1997), who used a similar analysis technique applied over the entire eastern U.S.. Successive increases in k over the range 6-10 produced additional coherent clusters which subdivided the two large clusters seen in Figure 2-1. The smaller, non-contiguous clusters remained largely unchanged for all values of k.

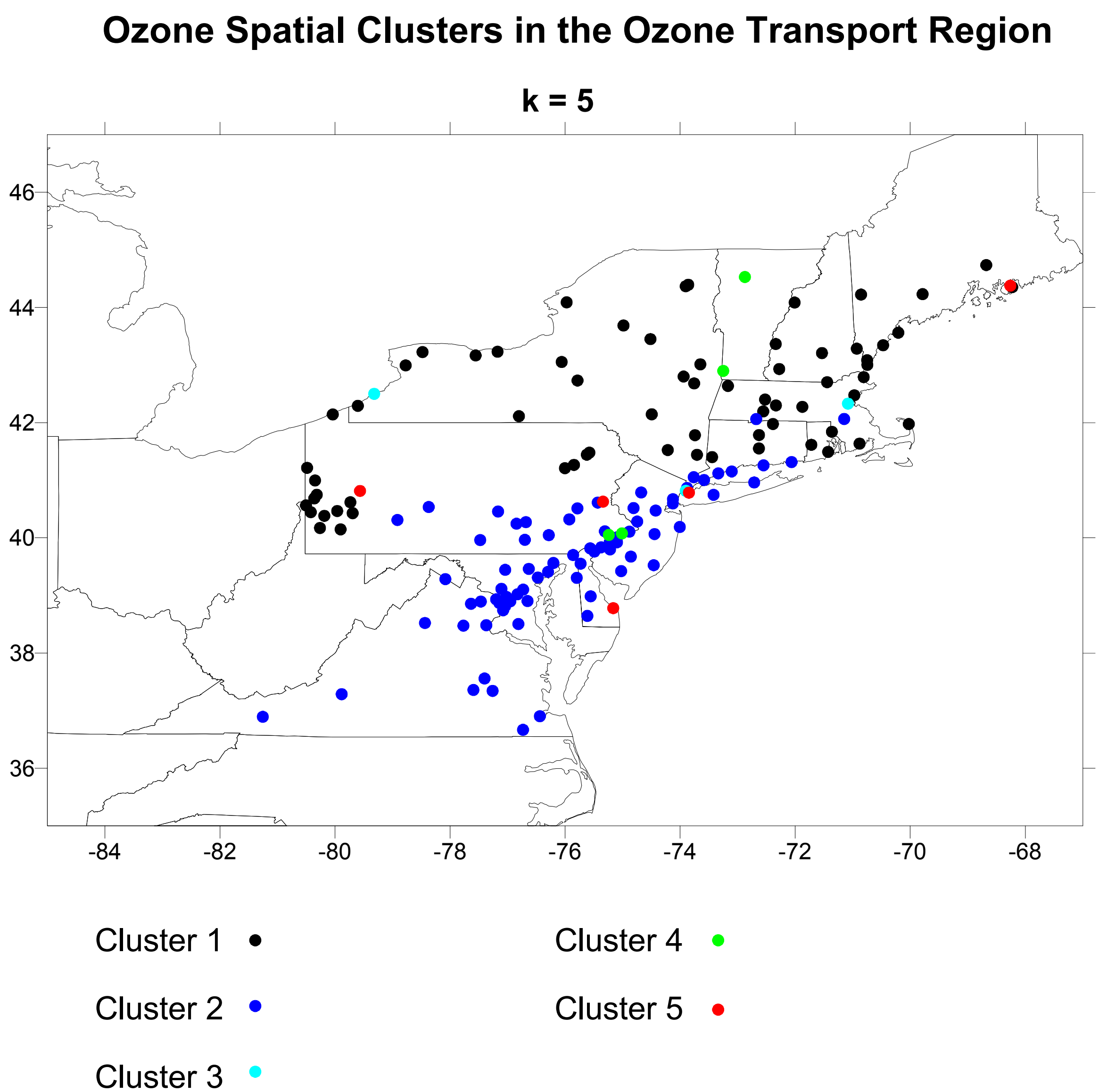


Figure 2-1. Ozone monitoring station cluster assignments for k = 5 clusters.

With k = 10, VARCLUS produced 6 spatially coherent clusters, and 4 smaller, non-coherent clusters (Figure 2-2)[2]. As in the k=5 case, the spatially coherent clusters are the lowest numbered clusters, 1-5, and the non-contiguous clusters are 6-8, and 10. The k=10 case is unusual because cluster 9 (located on the Rhode Island/Massachusetts coast) turned out to be spatially coherent, even though the lower numbered clusters 6-8 were not. We investigated the possibility that cluster 9 should be treated as a separate sub-region. After examining the way exceedances in cluster 9 vary with those in surrounding clusters, however, we concluded that this area could be adequately treated by including it in with cluster 4 (along the Washington – New York City corridor). In order to use only the clusters which seemed robust under variations in k, we therefore based the final ozone monitoring sub-regions largely on the first five clusters obtained under the k=9 scenario (which were slightly more coherent than those under the k=10 case).

[2] Ask was increased beyond 10, the coherent clusters produced were judged to be too small in spatial dimension to be useful in classifying ozone exceedance regimes.

However, the RI – MA coastal sites were associated with the New York City metropolitan area sites rather than the other MA sites based both on the k=10 result described above and input from several state air quality analysts. In addition, all stations on the ME coast were assigned to the southern New England group (Cluster 1) based on input from state air quality analysts. Stations from the other higher numbered, non-contiguous clusters were integrated into the surrounding clusters; there were no such stations for which the appropriate cluster assignment was ambiguous.

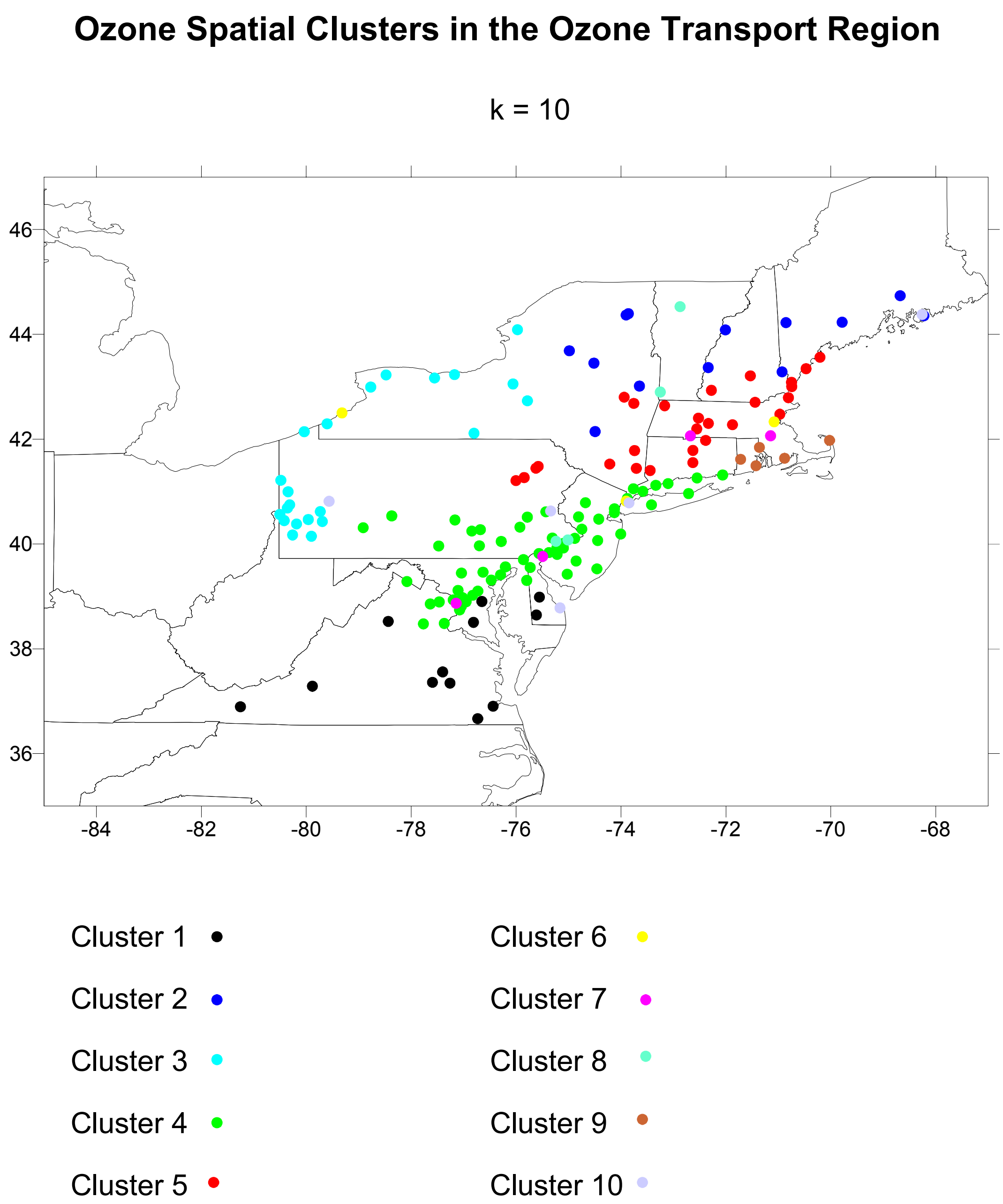


Figure 2-2. Ozone monitoring station cluster assignments for k = 10 clusters.

Another adjustment to the VARCLUS results was made along the Philadelphia – New York corridor. Figure 2-3 shows the number of 8-hour exceedances at each monitoring site during the period analyzed. Exceedance events in the Washington – Philadelphia corridor are more frequent than within and downwind (northeast) of the New York City metropolitan area. Furthermore, based on our discussions with state ozone forecasters in the OTR, we expect transport of ozone and ozone precursors along the I-95 corridor to play an important role in exceedance events. This suggests that leaving the entire Washington to New York City cluster intact might cause our final episode classification scheme to overlook events in which transport northeast from Washington-Baltimore-Philadelphia to New York is an important feature. We therefore decided to split this cluster into two parts as shown in the final ozone monitoring sub-region assignments presented in Figure 2-4. Cluster 5 extends from the Washington area through Trenton and a new cluster 6 covers the New York City-Long Island-Southern Connecticut region. A list of the monitoring sites assigned to each cluster is provided in Appendix D.

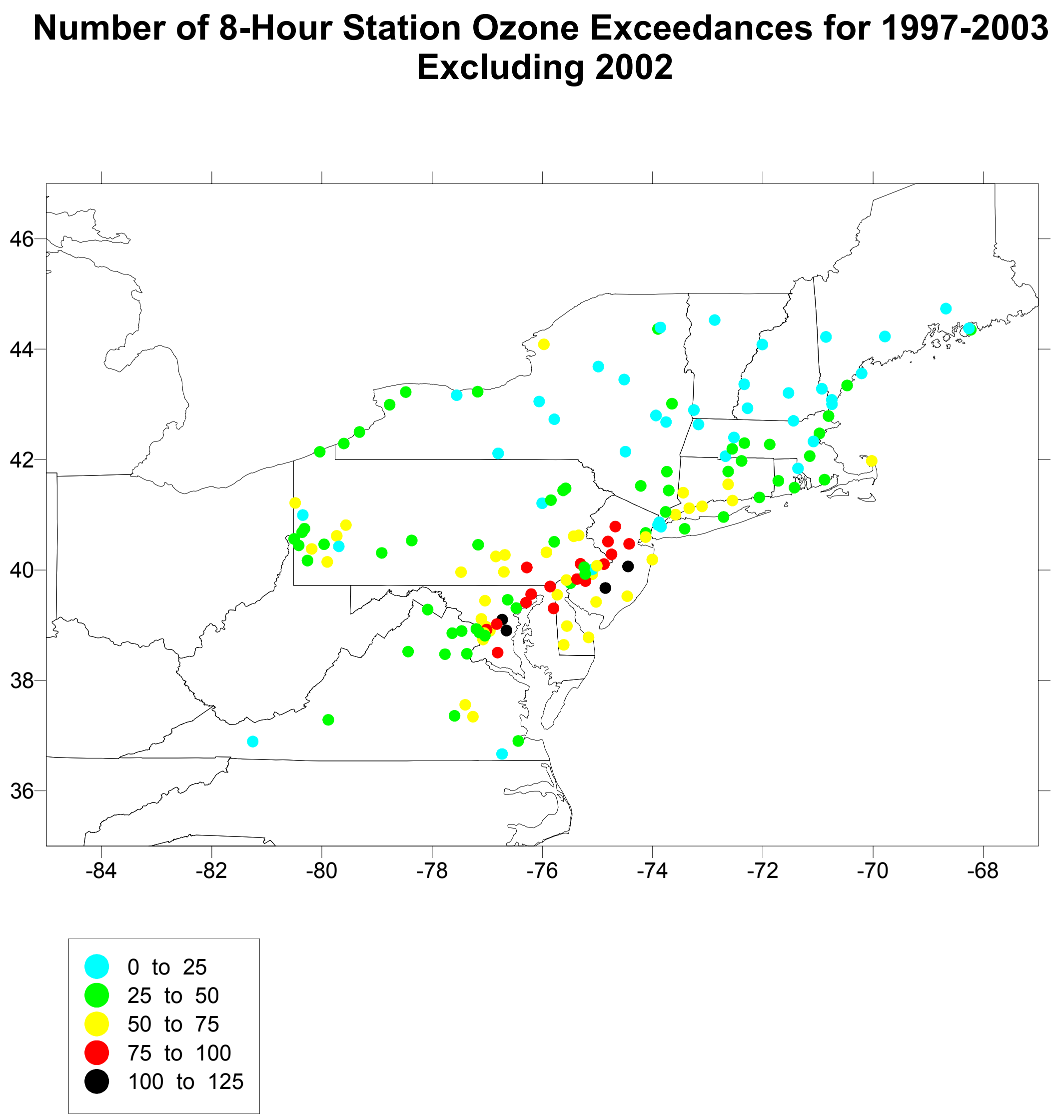


Figure 2-3. Number of 8-hour ozone NAAQS exceedance days at each monitoring site during the study period (1997 – 2001 and 2003).

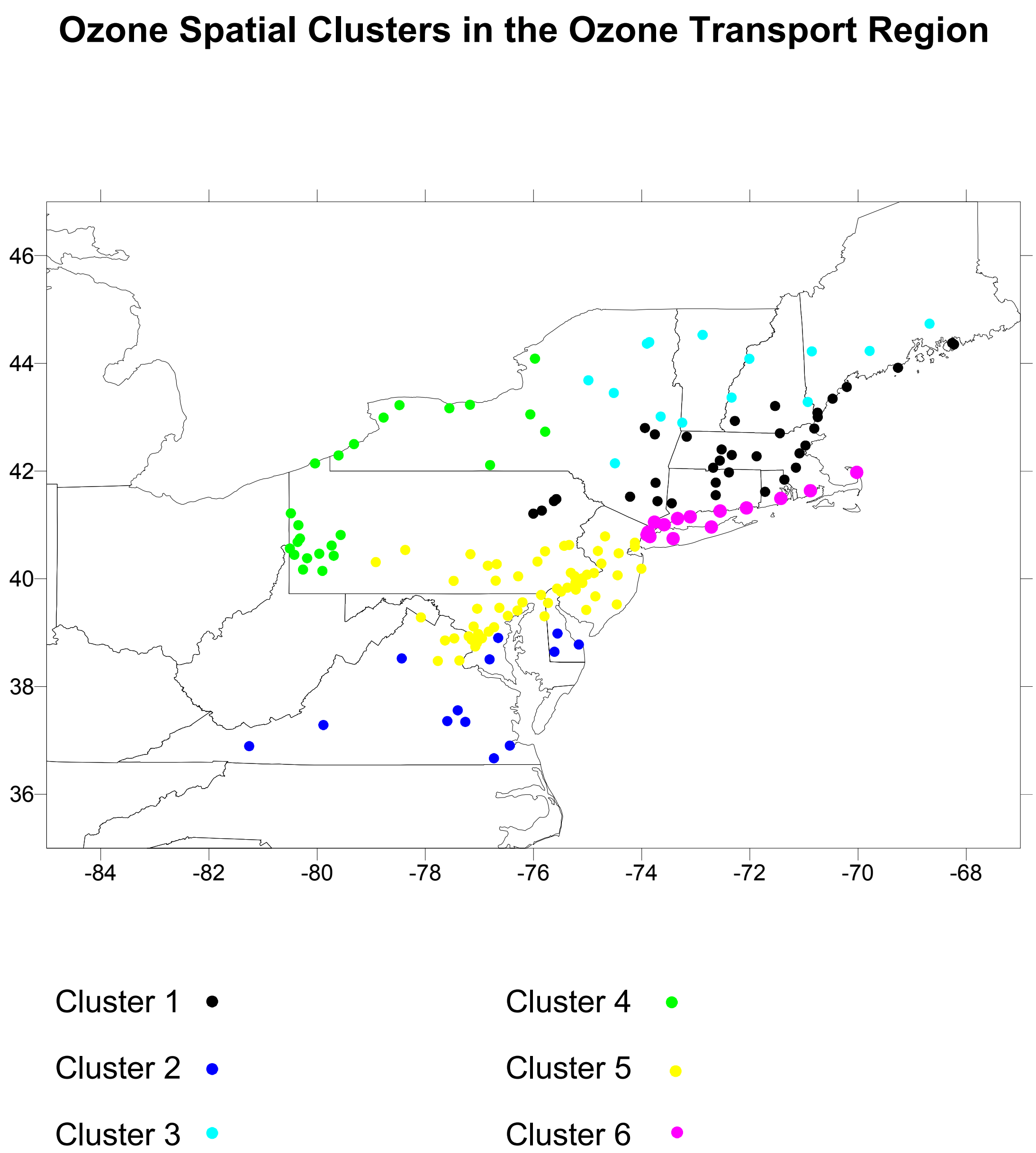


Figure 2-4. Ozone monitoring sub-regions in the OTR.

SPATIAL OZONE PATTERN ANALYSIS

An initial analysis of episode patterns was performed based on 8-hour ozone concentrations within the sub-regions (spatial clusters) described above. For each day, a cluster was determined to be in exceedance if any one monitoring site in the cluster recorded an exceedance. We then counted the number of joint exceedance events between each pair of clusters and examined exceedance patterns across all six clusters. Detailed results from this analysis were provided in a technical memorandum to the OTC (Stoeckenius and Kemball-Cook, 2004) but are not repeated here because this approach was eventually discarded in favor of an integrated analysis approach in which the daily ozone levels in each sub-region were combined with daily meteorological data to determine the key characteristics of the major types of ozone episodes occurring in the OTR.

METEOROLOGICAL VARIABLE SELECTION

The extensive amount of meteorological data collected for this study was reduced to allow processing of days into groups with similar conditions as described in Section 3. Selection of key meteorological variables that best represent conditions across the OTR on exceedance days was based on a review of previous studies (Deuel and Douglas, 1996; McHenry et al., 2004) and on discussions with state and local agency air quality personnel involved in ozone forecast programs within the OTR. Key variables focused on both surface conditions (maximum temperature, morning and afternoon average wind direction and speed, pressure) and conditions aloft (500 and 850 mb heights, temperatures, and winds). The final selected set of key daily meteorological parameters are:

> Surface resultant wind speed and direction computed for both morning (05:00 – 10:00 EST) and afternoon (12:00 – 17:00 EST) hours at New York City (LaGuardia), NY; Philadelphia, PA; Boston, MA; Buffalo, NY; Albany, NY; Washington, DC; Portland, ME, Atlantic City, NJ; Islip (Long Island), NY; Hyannis (Cape Cod), MA; Worcester, MA; and Hartford, CT.[3]
>
> Surface daily maximum temperatures at New York City (LaGuardia), NY; Philadelphia, PA; Boston, MA; Buffalo, NY; Albany, NY; Washington, DC; Portland, ME, Atlantic City, NJ; Islip (Long Island), NY; Hyannis (Cape Cod), MA; Worcester, MA; and Hartford, CT.[3]
>
> Temperatures, heights, and winds at 850 mb pressure surface at Washington, DC; New York, NY; Boston, MA; Pittsburgh, PA; Buffalo, NY; and Portland, ME.
>
> Surface pressure gradients across the OTR computed as pressure differences between:
> - Washington, DC and New York City, NY;
> - Washington, DC and Boston, MA;
> - Washington, DC and Pittsburgh, PA;
> - Pittsburgh, PA and Buffalo, NY;
> - Buffalo, NY and Boston, MA.

[3]Surface wind and temperature data from Concord, NH and New Haven, CT were also examined but these sites had a high frequency of missing data which prevented their use in this analysis.

3. EPISODE CLASSIFICATION ANALYSIS

In this section we describe a series of clustering and exploratory analyses performed on the ozone and meteorological data discussed in Section 2. Clustering was performed with data from the historical (1997 – 2001 and 2003) period to identify the major types of ozone episodes in the OTR and their key characteristics. Once the key episode types were identified, we developed a decision rule for classifying any given day into one of the identified episode types based on ozone levels and meteorological conditions. This decision rule was then used to classify days during the 2002 ozone season by episode type. The resulting distribution of episode types and the ozone and meteorological conditions occurring within each type in 2002 were subsequently compared with results from the historical period to determine the representativeness of the 2002 with respect to conditions during the historical period.

CLUSTERING ANALYSIS

Clustering was performed with data for the 329 days in the 1997-2001/2003 historical period on which an 8-hour ozone exceedance was recorded at one or more of the monitoring sites shown in Figure 2-4. As the clustering algorithms require numerical variables, wind directions were decomposed into u (east-west) and v (north-south) components. Meteorological data were prepared for clustering by first filling in missing values with exceedance day means. This step was necessary as the clustering procedures cannot process any days that have missing values for one or more variables. While the fraction of data that are missing for any individual variable is fairly small, roughly two-thirds of the 329 8-hour ozone exceedance days in our historical dataset had at least one missing value, so it was important to impute the missing values in some fashion even though the clustering results are not likely to be too sensitive to the exact method of imputation. All of the data were then standardized by computing z-scores (i.e., subtracting the mean and dividing by the standard deviation) prior to clustering so that variables with different scales of measure are given equal weight.

Ozone data were also prepared for use in the clustering analysis. Two daily ozone summary statistics, AvgEx08 and AvgEx00 were computed for each monitoring sub-region shown in Figure 2-4. AvgEx08 was defined as the average, over all sites in a given sub-region, of the amount by which the daily maximum 8-hour average exceeded 0.08 ppm (with values for sites below 0.08 ppm set equal to zero). AvgEx00 is identical to AvgEx08 but with the exceedance threshold set to 0 ppm. As with the meteorological data, z-scores were computed for the daily AvgEx08 in each sub-region for use in the clustering analysis. Preliminary clustering analyses were performed using the methods described below with first the AvgEx08 measure and then the AvgEx00 measure. Of the two, cluster results based on the AvgEx00 measure were chosen, as they were more robust and physically meaningful then results based on the AvgEx08 measure.

Initially, clustering was applied to the meteorological variables only. Both agglomerative and divisive hierarchical clustering techniques were used. Classifications of days under the resulting meteorological clusters were compared with the classification of days by ozone exceedance pattern, which had previously been reported (Stoeckenius and Kemball-Cook, 2004). These comparisons showed that, while some pairs of exceedance and meteorological patterns showed a dominant one-to-one relationship, others did not. In other words, some of the exceedance patterns were typically associated with more than one meteorological pattern and some meteorological patterns were typically associated with more than one exceedance pattern. This result was found to be robust in the sense that it occurred under a variety of clustering approaches. We interpreted this to mean that at least some of the ozone exceedance patterns described by Stoeckenius and Kemball-Cook were not sufficiently unique from a meteorological perspective to serve as adequate archetypes of different types of ozone episodes. Given this result, we decided to examine clustering approaches based on using both the meteorological and ozone (AvgEx00) data simultaneously.

Before proceeding further, we performed a principal components analysis (PCA) on the combined ozone and meteorological data set prior to clustering to determine if it would be possible to reduce the number of variables required for the analysis. Preliminary results showed, however, that the first four components only explained 14% of the total variance. As a result, we did not pursue the PCA any further but simply retained all of the key variables in the clustering analysis.

Several different clustering procedures were applied to the data. Application of single and complete linkage hierarchical agglomerative clustering methods (Venables and Ripley, 1994) to the combined ozone and meteorological data resulted in the formation of one large cluster containing most of the days in the dataset and a large number of additional clusters containing at most a few days each. Use of Ward's method (Ward, 1963) produced a more even distribution of cluster membership at each stage of the agglomeration but with fairly evenly spaced reductions in deviance (see resulting dendrogram in Figure 3-1). In other words, these results did not provide much guidance as to what would constitute a reasonable number of clusters to use in describing the data.

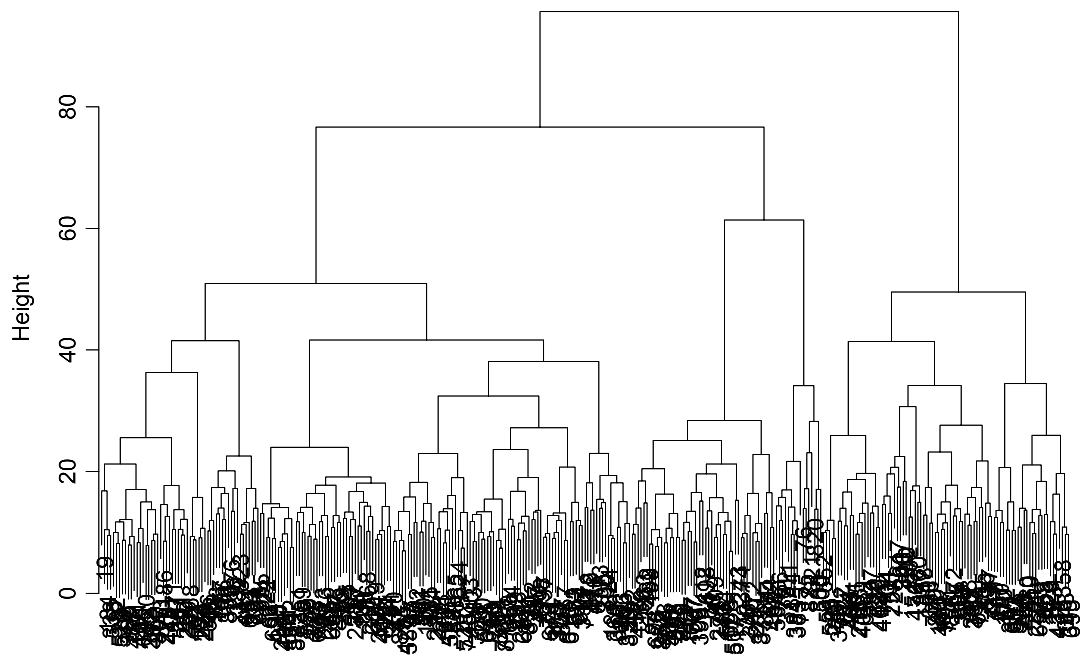


Figure 3-1. Dendrogram from application of Ward's hierarchical agglomerative clustering to the combined daily ozone and meteorological data. Each leaf at the bottom of the figure represents one day; the vertical height at which pairs of leaves (or pairs of clusters of leaves) are joined represents a measure of the distance between the leaves (or cluster centroids) in the multivariate data space.

Based on the agglomerative clustering results, we decided to apply Hartigan's k-means clustering algorithm (Hartigan, 1979) several times, specifying a different value for the number of clusters to form in each application. Under the k-means algorithm, data are arranged into a pre-specified number of clusters so as to minimize the total within-cluster sum of squares. Initial cluster centroids are determined via agglomerative hierarchical clustering. After this initial step, each day is assigned to the nearest cluster centroid where "nearest" is in this case defined as the minimum least squares distance computed over all of the standardized variables. After this initial assignment phase, the algorithm iteratively reassigns days to different clusters until the sum of the within-cluster sums of squares is minimized.[4]

Due to the large number of variables used in the clustering procedure, it is difficult to obtain a complete picture of the meteorological and air quality conditions associated with days falling in each cluster, especially when looking at several alternative cluster configurations. As one of the most important features of each cluster is the spatial ozone distribution, we tabulated the mean

[4] As finding the global minimum of this objective function is not computationally feasible, Hartigan's algorithm actually finds a local minimum such that switching any single observation from one cluster to another does not reduce the objective. As a result, the final cluster assignments may be sensitive to the selection of initial cluster centroids.

values of the ozone measure described above (AvgExc00) for each sub-region within each cluster identified by the k-means algorithm when the data are divided into between 4 and 7 clusters (see Table 3-1). We also examined similar sets of results for each key meteorological variable. Inspection of these results revealed the presence of five distinct sets of ozone and meteorological conditions that are robust in the sense that they show up consistently whether the data are divided into 4, 5, 6 or 7 clusters.

Table 3-1. Mean z-scores for the AvgEx00 ozone summary statistic within each monitoring sub-region under four different candidate sets of cluster designations. The episode pattern ID in the far right-hand column is keyed to the episode patterns described in the text.

a) 4 clusters							**Episode Pattern**
Cluster #	**Sub-Region 1**	**Sub-Region 2**	**Sub-Region 3**	**Sub-Region 4**	**Sub-Region 5**	**Sub-Region 5a**	**ID**
1	0.51	0.08	0.40	-0.08	0.26	0.56	C
2	-0.86	0.49	-0.86	-0.40	-0.15	-0.37	B
3	0.27	0.05	0.22	0.30	0.30	0.20	A
4	-0.72	-0.57	-0.45	-0.31	-0.93	-0.96	E
b) 5 clusters							
Cluster #	**Sub-Region 1**	**Sub-Region 2**	**Sub-Region 3**	**Sub-Region 4**	**Sub-Region 5**	**Sub-Region 5a**	
1	-0.74	0.07	-0.40	-0.58	-0.34	-0.78	E
2	-0.70	0.45	-0.73	-0.37	-0.06	-0.17	B
3	0.45	0.16	0.32	0.35	0.45	0.46	A
4	-0.49	-0.97	-0.35	0.22	-1.07	-0.91	D
5	0.54	-0.04	0.47	-0.10	0.17	0.42	C
c) 6 clusters							
Cluster #	**Sub-Region 1**	**Sub-Region 2**	**Sub-Region 3**	**Sub-Region 4**	**Sub-Region 5**	**Sub-Region 5a**	
1	-0.83	0.47	-0.88	-0.47	-0.17	-0.27	B
2	0.38	0.21	0.23	0.44	0.49	0.46	A
3	0.05	-1.62	0.17	0.58	-1.32	-0.86	D
4	-1.13	-0.17	-0.91	-0.52	-0.89	-1.10	E1
5	0.49	0.03	0.43	-0.10	0.21	0.48	C
6	-0.19	-0.14	0.11	-0.39	-0.17	-0.48	E2
d) 7 clusters							
Cluster #	**Sub-Region 1**	**Sub-Region 2**	**Sub-Region 3**	**Sub-Region 4**	**Sub-Region 5**	**Sub-Region 5a**	
1	0.59	-0.30	0.56	-0.39	-0.02	-0.01	C
2	0.08	0.02	0.02	0.32	0.32	0.10	D1
3	0.08	-1.77	0.21	0.65	-1.52	-1.00	D2
4	-1.17	-0.15	-0.97	-0.54	-0.88	-1.11	E1
5	0.60	0.44	0.39	0.45	0.60	1.03	A
6	-0.38	-0.12	-0.03	-0.45	-0.27	-0.58	E2
7	-0.89	0.47	-0.89	-0.47	-0.19	-0.39	B

We prepared summaries of the meteorological characteristics of each of these five episode types as follows:

1) Composite maps of surface and upper air (850 mb) meteorological variables for each cluster,

2) Side-by-side box plots comparing the distributions of selected key meteorological variables within each cluster, and

3) Tables of morning and afternoon resultant wind direction frequencies within each cluster.

Full results of items 1 – 3 above are presented in Appendix A, B, and C, respectively. By way of example, we show the 850 mb height and wind fields, 850 mb temperature, surface pressure, and surface daily maximum temperature and 10 m wind fields composited for each episode type in Figures 3-2 to 3-5, respectively. Comparing these composite fields for different episode types reveals that each episode type is characterized by a distinct meteorological pattern and these patterns are consistent with the ozone patterns noted in Table 3-1. Key characteristics of the five episode types are presented in Table 3-2. In the description of each episode type, “average” refers to averages over all OTR exceedance days used in the cluster analysis.

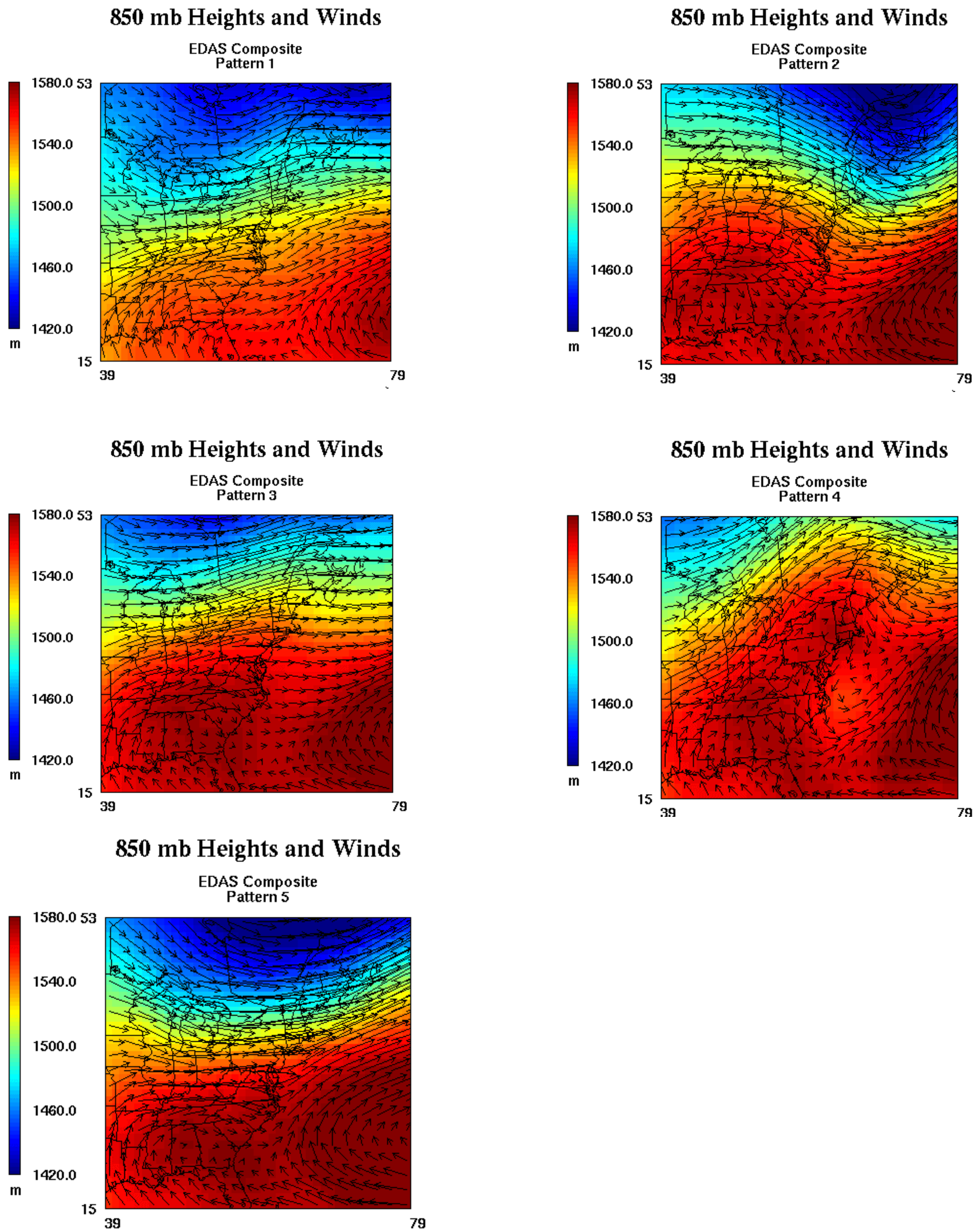


Figure 3-2. Average 850 mb height and wind fields for each episode (pattern) type (pattern numbers refer to the episode types listed in Table 3-2): Pattern 1 = Episode Type E, 2 = B, 3 = A, 4 = D, 5 = C).

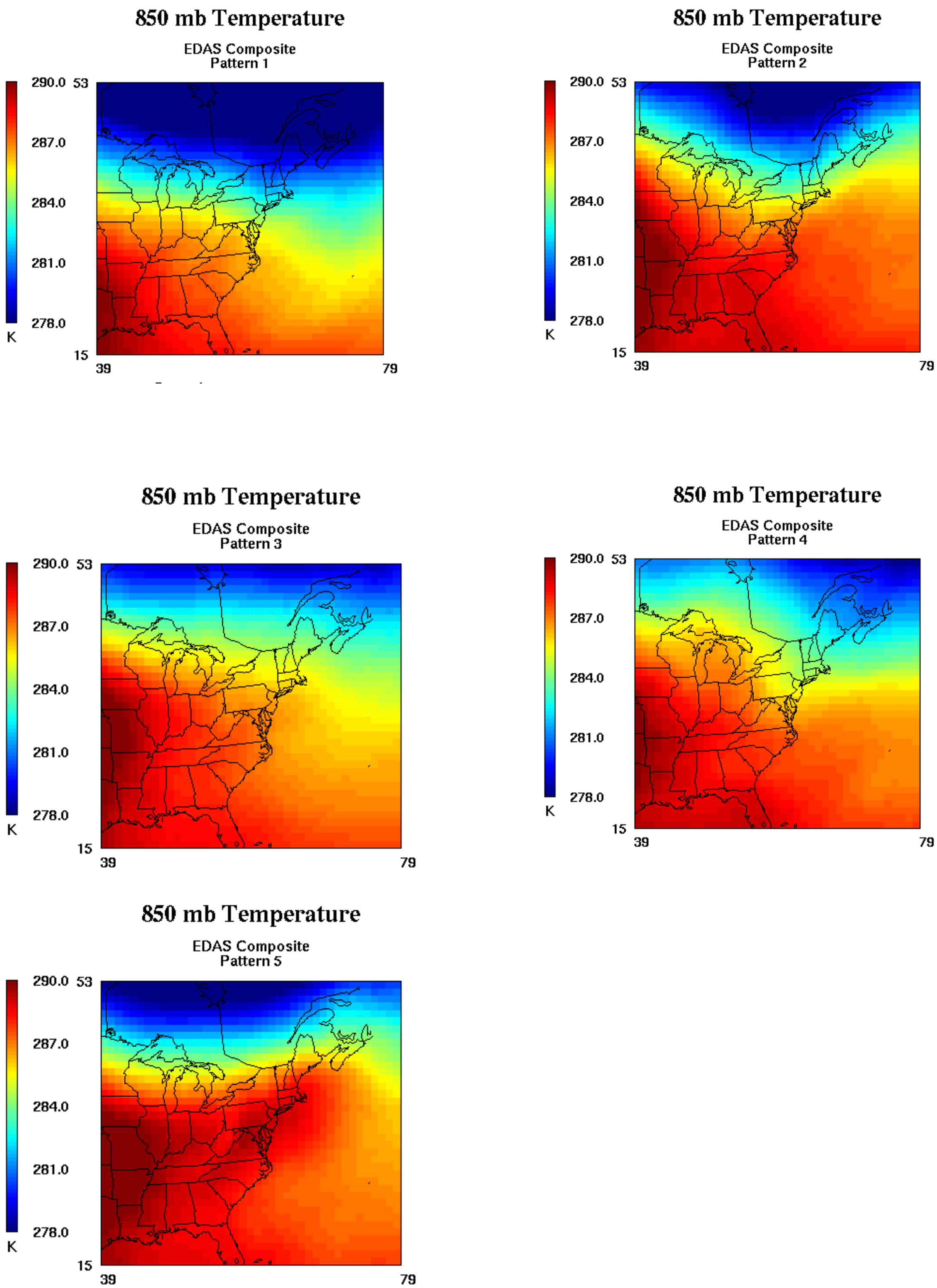


Figure 3-3. Average 850 mb temperature fields for each episode (pattern) type (pattern numbers refer to the episode types listed in Table 3-2): Pattern 1 = Episode Type E, 2 = B, 3 = A, 4 = D, 5 = C).

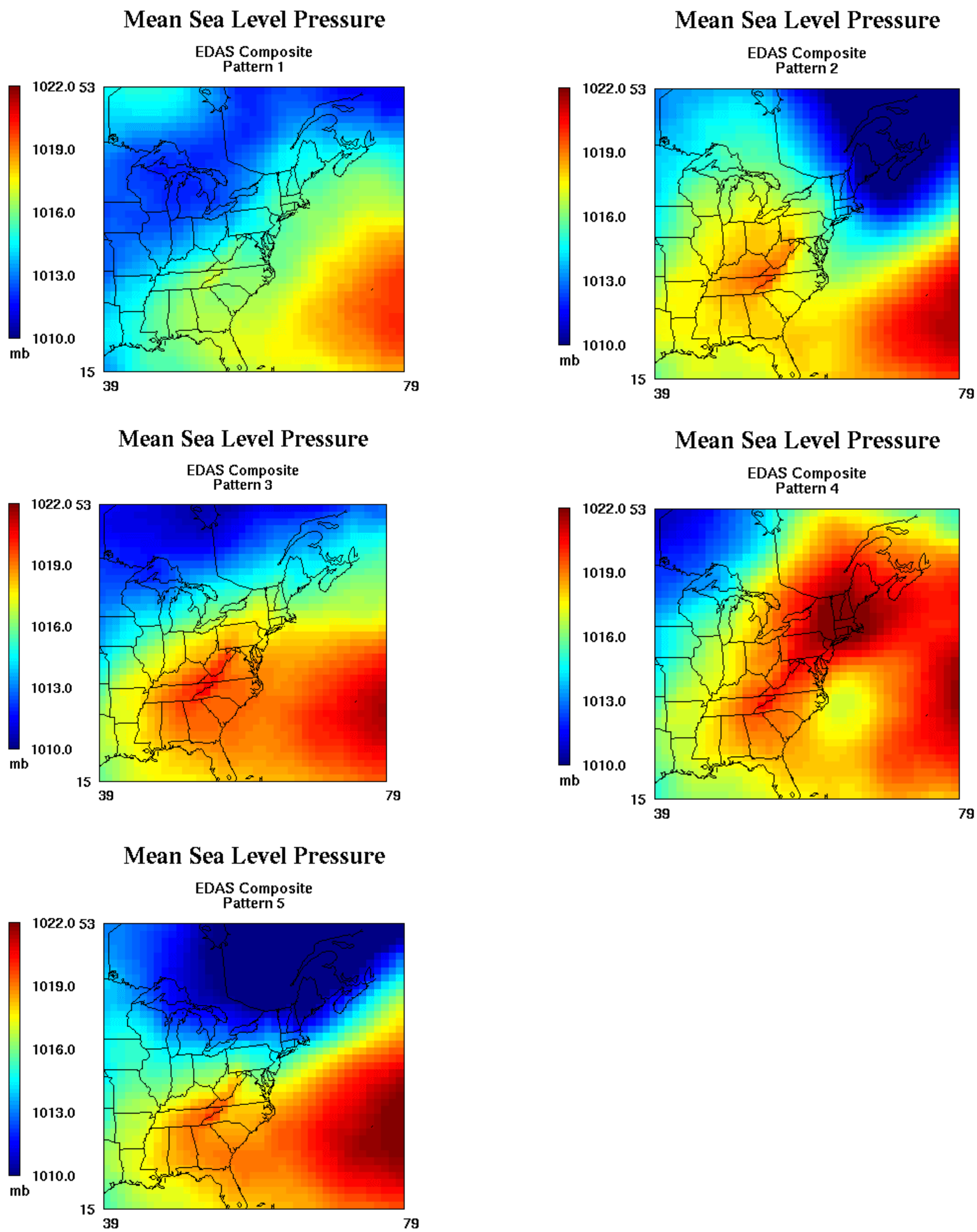


Figure 3-4. Average surface sea level pressure for each episode (pattern) type (pattern numbers refer to the episode types listed in Table 3-2). Pattern 1 = Episode Type E, 2 = B, 3 = A, 4 = D, 5 = C).

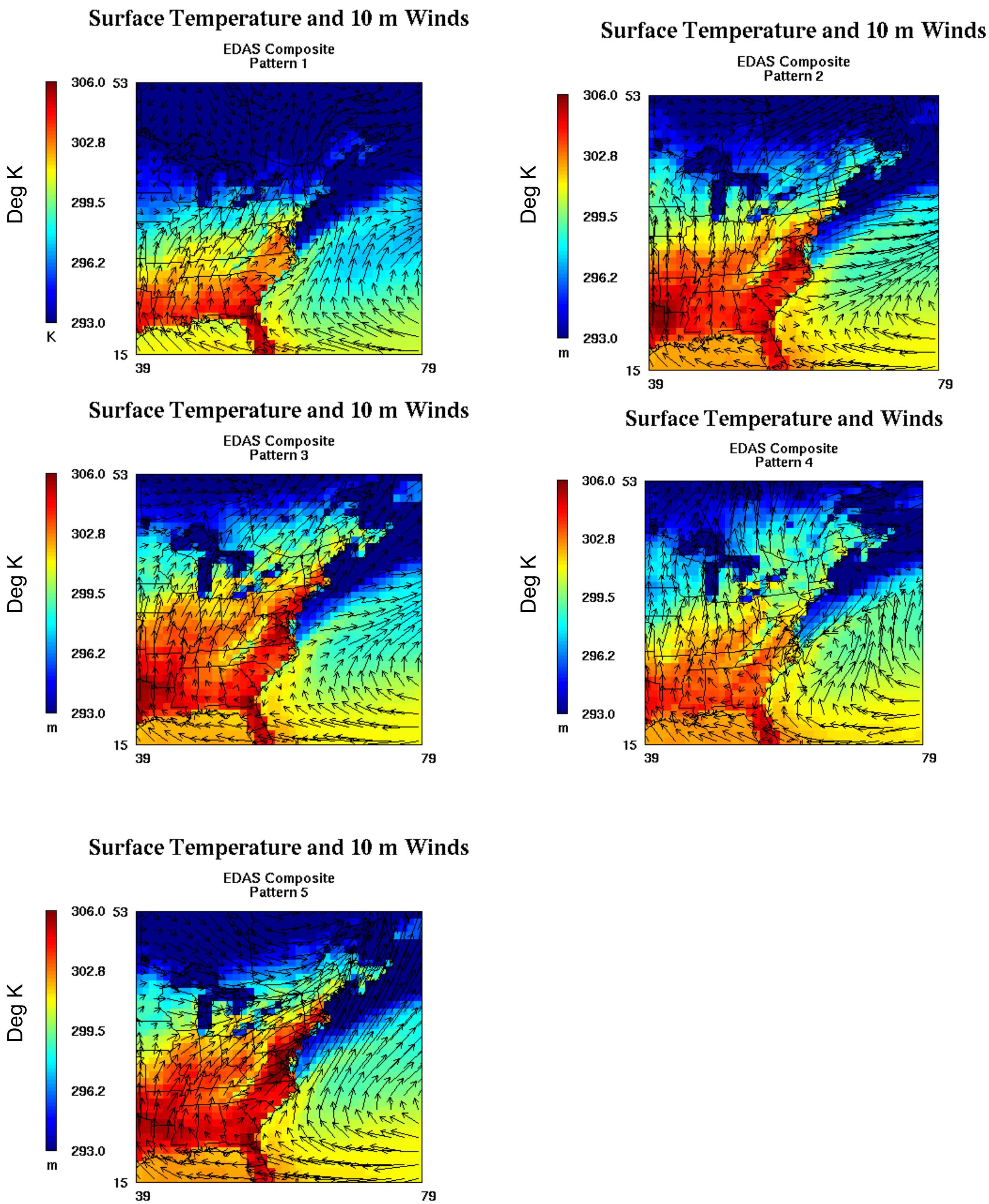


Figure 3-5. Average surface temperature and 10 m wind fields for each episode (pattern) type (pattern numbers refer to the episode types listed in Table 3-2). Pattern 1 = Episode Type E, 2 = B, 3 = A, 4 = D, 5 = C).

Table 3-2. Key characteristics of each OTR episode type.

Episode Type	Pattern No.	Description
A	3	*High ozone throughout the OTR.* This pattern is characterized by strong high pressure over the southeastern states extending from the surface to 500 mb with high temperatures extending into New England and southwest surface winds throughout the OTR. 850 mb temperatures and heights, and surface temperatures are above average at all locations except Washington DC; winds are SW to W throughout the OTR except more variable at LaGuardia and magnitudes of resultant wind vectors are higher than average (indicative of a fairly steady, well defined flow regime), E-W surface pressure gradients are near neutral but SW-NE gradients both along the I-95 corridor and in the west (Pittsburgh to Buffalo) are positive which is consistent with the SW flow. Ozone formation under these conditions is promoted throughout the OTR by the stable air mass and high temperatures.
B	2	*High ozone confined to the extreme southeastern OTR.* This pattern is characterized by an upper-level trough offshore of the OTR and a surface high centered over Kentucky. This results in cooler air advection over nearly all of the OTR with northwest flow aloft and a more westerly flow at the surface. 850 mb heights are lower than average (especially in New England) and surface winds are more frequently from NW along the I-95 corridor than under Type A. Temperatures at 850 mb along the I-95 corridor are only slightly cooler than under Type A but inland temperatures, especially in the north, are much cooler (e.g., at Buffalo); similarly, surface temperatures along the I-95 corridor are about the same as under Type A but temperatures are cooler in Buffalo and Albany. Type B events have the strongest positive W – E surface pressure gradients of any category, consistent with the NW winds but gradients from Washington to New York and Boston are positive. The cooler air over the western OTR and westerly to northwesterly flow result in the higher ozone levels being confined to just the extreme southern portion of the OTR under this pattern.
C	5	*High ozone along I-95 corridor and northern New England.* This pattern is characterized by an extension of the semi-permanent Bermuda high into the southeastern U.S. and an area of high surface and 850 mb temperatures extending from Maryland to Maine; the 500 mb pattern is nearly zonal (east – west flow) while flow at the surface is generally from the SW. 850 mb heights intermediate between Type A and Type B but 850 mb temperatures are very high along the I-95 corridor and slightly cooler further inland. Winds are more consistently S - SW at all sites than under other episode types and almost no NW-N-NE winds are seen at LaGuardia in contrast to other types. Resultant wind vector magnitudes are much higher than average, consistent with the steady SW flow. SW – NE pressure gradients along I-95 corridor and from Pittsburgh to Buffalo are positive, consistent with the SW flow. Average E-W pressure gradients are near zero. These conditions result in above average ozone levels all along the I-95 corridor with advection north into coastal and interior New England. Ozone levels are slightly below average in the extreme southeastern and western OTR (subregions 2 and 4).

Episode Type	Pattern No.	Description
D	4	*High ozone in the western OTR*. This pattern is characterized by an area of mean upper level divergence with associated cut-off low at 850 mb off the Outer Banks of North Carolina. A relatively vigorous mean low pressure center can be seen at the surface. An east-west temperature gradient across the OTR is evident at 850 mb. Surface temperatures along the I-95 corridor and in Albany are below average but surface temperature is above average at Buffalo. 850 mb heights are the highest of any episode type due to a strong ridge over New England. Surface winds are mostly E - NE along I-95 corridor from DC to NY but more variable further north. In contrast to episode types A, B, or C, SW – NE pressure gradients along the I-95 corridor are negative, consistent with the NE surface winds. W – E pressure gradients are flat. These conditions result in below average ozone in the eastern OTR (sub-regions 1, 2, 3, 5, and 6) due to the on-shore flow in the north and cyclonic conditions in the south but above average ozone levels in the western OTR (sub-region 4) due to stable, warm conditions with light winds.
E	1	*Generally low ozone throughout OTR*. This category includes days with moderately low to lowest average ozone readings of all OTR exceedance days included in the cluster analysis. The Bermuda high is shifted east relative to the other types and flow over the southeastern U.S. is only weakly anti-cyclonic with a nearly zonal flow pattern at the 850 and 500 mb levels over the OTR. Temperatures at the surface and aloft are the coolest of any episode type. While winds aloft are nearly westerly, surface winds are generally S – SE over most of the OTR. SW – NE pressure gradients are negative along the I-95 corridor and E-W gradients are positive, consistent with the SE flow. These conditions result in below average ozone throughout the OTR due to the relatively low temperatures and southeasterly onshore flow at coastal locations.

The five episode types described in Table 3-2 exhibit characteristics, which are largely consistent across the different cluster allocations noted in Table 1 (4, 5, 6 or 7 clusters). When four clusters are specified, the Type D events are subsumed into the remaining four episode types. Finer division of days into six clusters results in a split of the Type E events into two groups (denoted as E1 and E2 in Table 1) with generally very similar meteorological conditions but distinguished in part by E-W pressure gradient anomalies that are slightly greater under type E2. Further division into seven clusters appears to preserve the Type A, B, and, to a lesser extent, Type C events along with the Type E1 and E2 events found in the seven cluster result while the Type D events are split into two new categories (denoted D1 and D2 in Table 1). Both D1 and D2 events are associated with high ozone in the west (sub-region 4) under S – SW flow as is typical of Type D but differ in the surface wind pattern, and hence ozone anomalies, along the I-95 corridor.

It is important to keep in mind that there is no *a priori* expectation that all ozone exceedance events in the OTR fall into one of a finite number of distinct patterns: daily conditions differ from one another to varying degrees and some days will always have characteristics that cross over any predetermined classification boundaries. This means that an episode classification system will always have a certain degree of arbitrariness to it and division of days into bins will always result in some days that do not fit particularly well into any single bin. Nevertheless, for purposes of this study, we seek a reasonable classification system based on a handful of pattern types each of which is uniquely identifiable by a set of characteristics related to ozone formation across the *entire* OTR.[5] Based on the clustering results described above, it appears that the episode Types A – E meet these requirements reasonably well. Frequencies of occurrence for these five types are shown in Table 3-3.

Table 3-3. Frequencies of occurrence of OTR episode types.

	Type A	Type B	Type C	Type D	Type E
No. Days	123	50	66	44	46
Pct.	37%	15%	20%	13%	14%

DEVELOPMENT OF AN OBJECTIVE CLASSIFICATION PROCEDURE

In order to complete our analysis, we needed to develop a final episode classification rule based on results of the above analysis of the 1997 – 2001 and 2003 data which can then be applied to the 2002 data to determine the classification of episodes in 2002 to the five ozone event types described above. A classification tree model was created for this purpose using the ozone and meteorological data from 1997-2001 and 2003 as predictors and the episode pattern type as the response variable. In the classification tree model, data from all exceedance days start out together in the root node of the tree and are then split into two daughter nodes based on the value of one of the predictor variables. For example, a split might consist of sending all days with resultant afternoon wind speed at Hartford, CT less than 4.8 m/s to one node and all remaining days to the other. The variable and value of that variable used to perform a split is determined by examining all possible splits and finding the one which results in the greatest reduction in deviance in the response variable (deviance is a measure of the degree of heterogeneity of the response variable in a node). The splitting process is then repeated for each resulting daughter node and so on until a stopping criterion is reached. The daughter nodes resulting from the last split along each branch of the tree are referred to as *terminal* nodes. The resulting classification tree, grown using the 1997-2001/2003 data as the *learning* dataset, can then be applied to the 2002 data for which the episode classifications are unknown by running the 2002 daily data down the tree, separating days at each node according to the previously determined splitting criteria. Each day from the 2002 data will fall into one of the terminal nodes of the tree, and the probability of that day belonging to the ith episode type is estimated from the fraction of days from the learning dataset in the terminal node belonging to the ith episode type. The *predicted* episode type for days in 2002 falling in the terminal node is taken to be the episode type with the highest probability of occurrence.

Initially, the classification tree was grown by making successive splits until only a small number of days (in this case five) ends up in each terminal node. This results in a relatively large tree with many terminal nodes, each of which will typically be very homogeneous: most of the days in any one terminal node will belong to the same episode type. This large tree represents an over fit to the data in the learning dataset. In other words, if the tree were to be validated against an independent set of days for which the episode types are known (i.e., a test dataset) the frequency of misclassification will generally be higher than the low misclassification frequency determined

[5] It is worth reiterating here that we are seeking a general classification system applicable to the whole of the OTR. More precise classification systems could be developed for individual sub-regions within the OTR but the resulting two dimensional system (consisting of a unique set of episode types for each of several sub-regions) would not only be very time-consuming to develop but would lead to results from which it would most likely be very difficult to draw any conclusions regarding the representativeness of a single season with respect to conditions over the whole of the OTR.

by applying the full tree against the learning dataset. Thus, a smaller tree (one with fewer splits and therefore fewer terminal nodes), is likely to perform at least as well against a test dataset as the initial, large tree. We therefore applied a recursive tree-pruning algorithm known as *cost-complexity pruning* to the large tree (Venables and Ripley, 1994). This results in a sequence of trees, each of which can be characterized by the number of terminal nodes and the cost-complexity parameter, which is a measure of the trade off between growth in tree size and reduction in deviance. The resulting tree sequence is shown in Figure 3-6. As this figure shows, there is a diminishing return in deviance reduction as the size of the tree increases beyond about 5 terminal nodes.

To further evaluate the relative value of different size trees, we performed a ten-fold cross-validation using the learning dataset. The ten-fold cross-validation consists of setting aside 1/10th of the days in the learning sample as a test sample, building a tree using the remaining 90% of days, and evaluating the deviance reduction using the reserved days. This process is repeated 10 times with a different set days set aside in each case. Results from the cross-validation (Figure 3-7) suggests that the residual deviance is minimized at a tree size of about five or six terminal nodes. These results, together with an examination of the misclassification rates from the learning dataset for the pruned tree sequence shows that the 6 terminal node tree is about the optimal size.

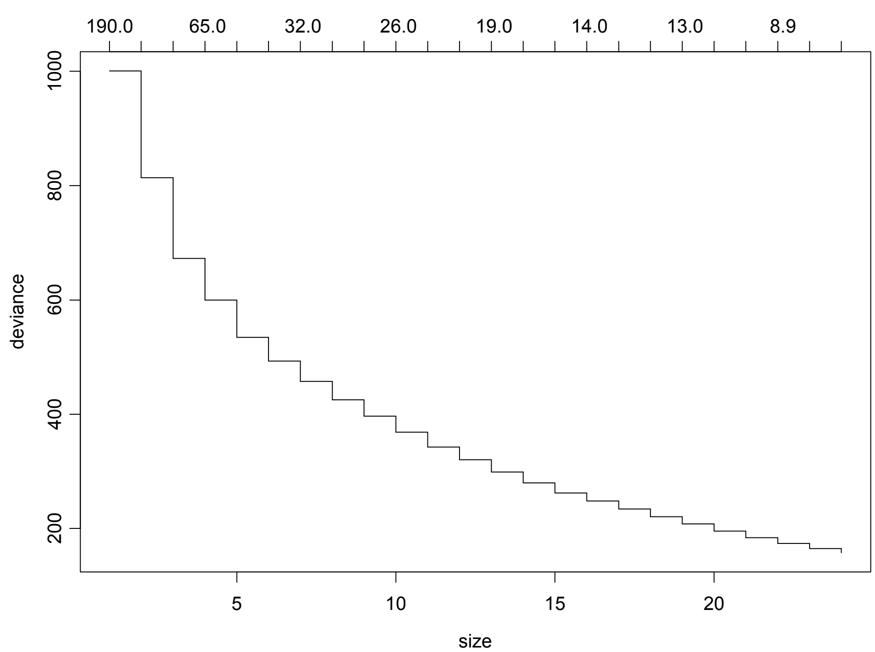


Figure 3-6. Deviance as a function of tree size (number of terminal nodes) for sequence of trees generated by the pruning algorithm.

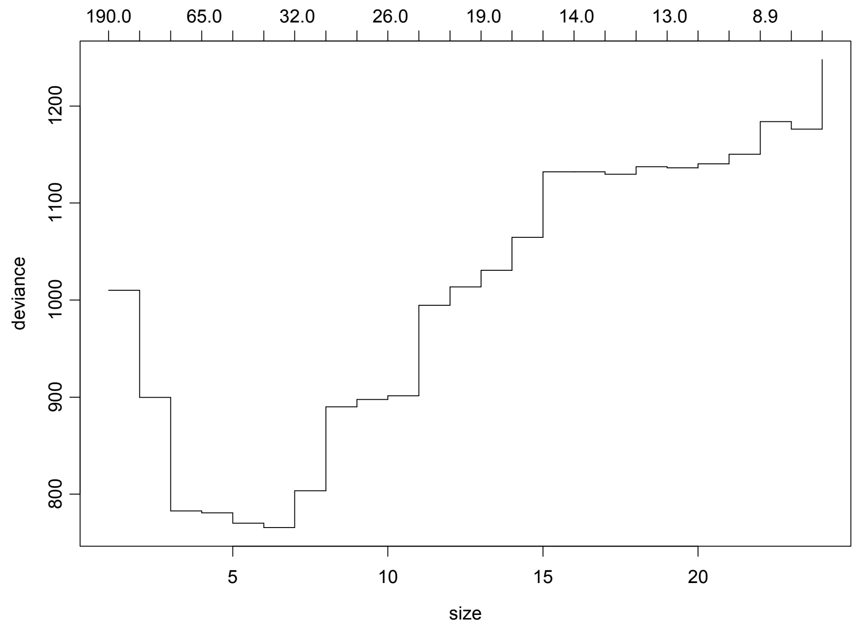


Figure 3-7. Deviance from 10-fold cross-validation as a function of tree size (number of terminal nodes) for sequence of trees generated by pruning algorithm.

The selected classification tree is shown in Figure 3-8; Table 3-4 summarizes the distribution of days by episode type in each terminal node. Two nodes are made up of predominantly Type E days, each of the rest are most representative of one of the four other episode types. Each terminal node has a dominant episode type accounting for between 64 and 81% of days assigned to the node. To use the classification tree for assigning an episode type to a previously unclassified day, we define the *predicted* episode type for all days reaching a given terminal node as the dominant episode type for the node as shown by the shaded boxes in Table 3-4. When this rule is applied to the 329 episode days during the historical period, a comparison of the predicted episode types with the episode types assigned by the cluster analysis shows an overall misclassification rate of 23%.

Table 3-4. Distribution of episode types during the 1997-2001/2003 historical period (as determined via the clustering analysis) for days in each terminal node of the classification tree shown in Figure 3-8.

	Episode Type					
Node No.	**A**	**B**	**C**	**D**	**E**	**Total**
4	6	3	0	0	16	25
5	100	16	9	1	0	126
7	5	3	53	0	5	66
8	1	25	3	0	2	31
10	5	0	0	4	19	28
11	6	3	1	39	4	53
Total	123	50	66	44	46	329

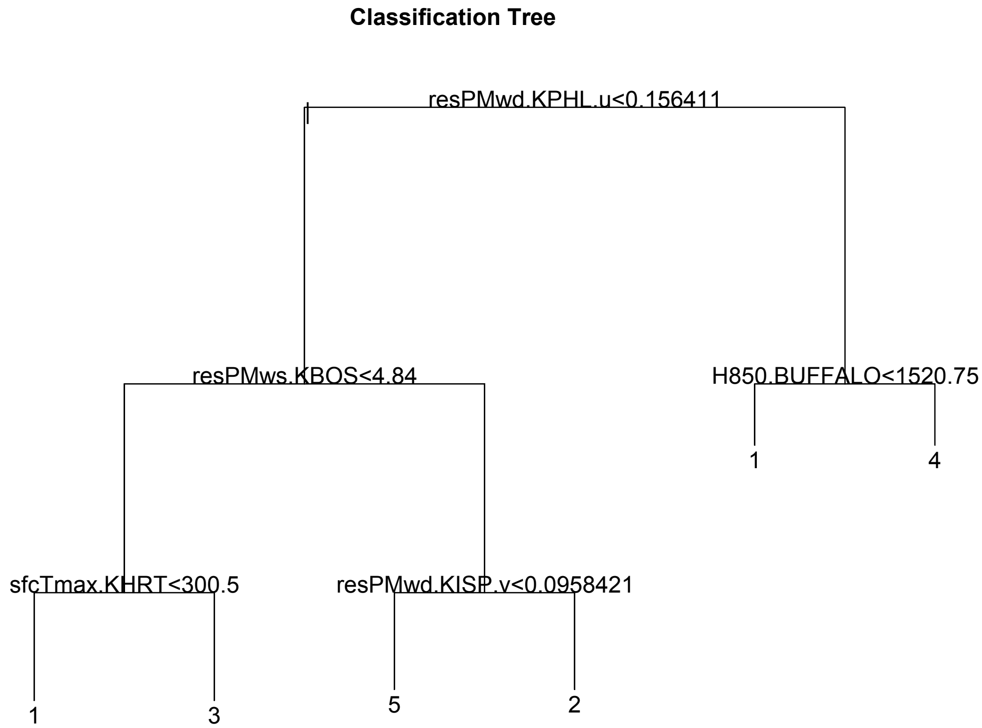


Figure 3-8. Classification tree used to group days by episode type. Variable names and values used to divide data at each splitting node are shown: days meeting the specified criterion are moved down the left branch in each case (resPMwd.KPHL.u = easterly component of the resultant afternoon wind direction at Philadelphia [m/s]; resPMws.KBOS = resultant afternoon wind speed at Boston [m/s]; H850.BUFFALO = 850 mb pressure height at Buffalo [m]; sfcTmax.KHRT = daily maximum surface temperature at Hartford, CT [K]; resPMwd.KISP.v = northerly component of afternoon wind direction at Islip, NY [m/s]). Terminal nodes are numbered 1 – 5 and are keyed to the summary in Table 3-4.

CLASSIFICATION OF 2002 OZONE EPISODES

Data from the 2002 ozone season were analyzed using the classification tree described above to yield a division of the ozone exceedance days into the five episode types. The resulting frequency distribution of episode types in 2002 was then compared with the historical episode type frequency distribution shown in Table 3-4, thereby providing an indication of the degree to which conditions during 2002 are representative of conditions observed in other years. We also compared ozone concentration distributions and composite meteorological fields by episode type in 2002 with those during the historical period as a way of further evaluating the representativeness of conditions during the 2002 ozone season.

Episode Type Classification

We applied the 6-node tree shown in Figure 3-8 to all 8-hour ozone exceedance days in 2002. Of the 71 exceedance days, 69 could be assigned to terminal nodes on the tree; missing data prevented classification of two of the days. Examination of the classification results showed that if surrogate splits[6] were used to assign these two days to one of the terminal nodes, the number of days falling into the node would change by no more than 3 percentage points, so the two days with missing data were simply ignored. The predicted episode type for each exceedance day in 2002 was taken to be the predominant episode type in the terminal node to which it was assigned (as indicated by the shaded boxes in Table 3-4). Appendix E lists the resulting episode type associated with each exceedance day in 2002. The resulting distributions of days by episode type for the 2002 season and the 1997-2001/2003 historical period are shown in Figure 3-9.[7] For the historical days, both the episode type assignments based on the classification tree and the episode types as originally assigned in the clustering analysis are shown. The overall pattern of episode type occurrence frequencies for the historical period is similar between the classification tree and the clustering analysis, as we would expect. Frequencies of occurrence of the episode types are within two percentage points of each other except for Type D events (slightly more Type D days assigned by the classification tree) and Type B events (about a third fewer Type B days determined by the classification tree).

Comparison of the occurrence frequencies over the historical period with the 2002 data also suggest a generally similar pattern of episode types. Note that the error bars in Figure 3-9 show the 10th and 90th percentile range in the frequencies of occurrence of each episode type observed within individual years during the historical period: an individual year would be expected to fall within this range with 80% probability. The 2002 type frequencies generally fall within these error bars except for a somewhat higher frequency of Type C events and a lower frequency of Type E events. As Type E events are characterized by below average ozone (relative to all exceedance days) throughout all but the southernmost OTR, this difference reflects the higher frequency of exceedance days in 2002 relative to the historical period as noted above. If we ignore the Type E events and renormalize (see Figure 3-10), the occurrence frequencies in 2002 of the remaining episode types are found to be similar to those in the historical period and fall within the 10th to 90th percentile range in each case. Thus, each of the event types A – D appear to be well represented within the 2002 season.

[6]Surrogate splitting uses the best alternative splits (based on the non-missing variable that produces nearly the same split as the primary splitting variable).

[7]The bars in this figure are scaled to the fraction of OTR exceedance days assigned to each episode type. Thus, these comparisons are not effected by the above average number of exceedance days in 2002 noted earlier.

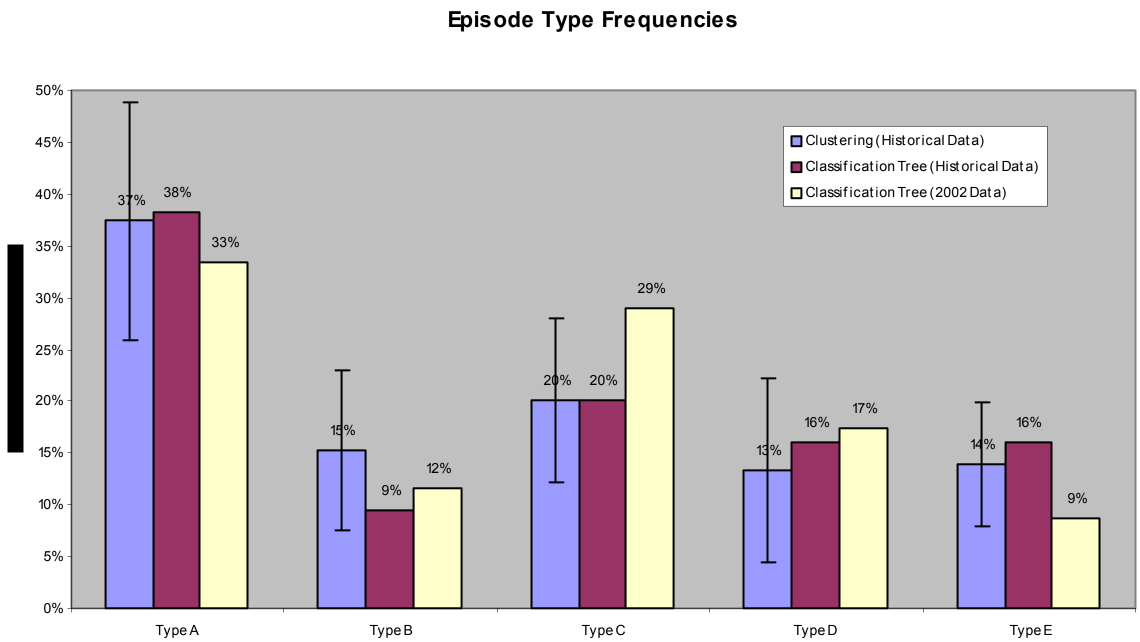


Figure 3-9. Percent of episode days by type in the 1997 – 2001/2003 historical period (as determined by the cluster analysis and by the classification tree) and in 2002; error bars show 10th and 90th percentiles of annual frequencies of occurrence during 1997-2001&2003.

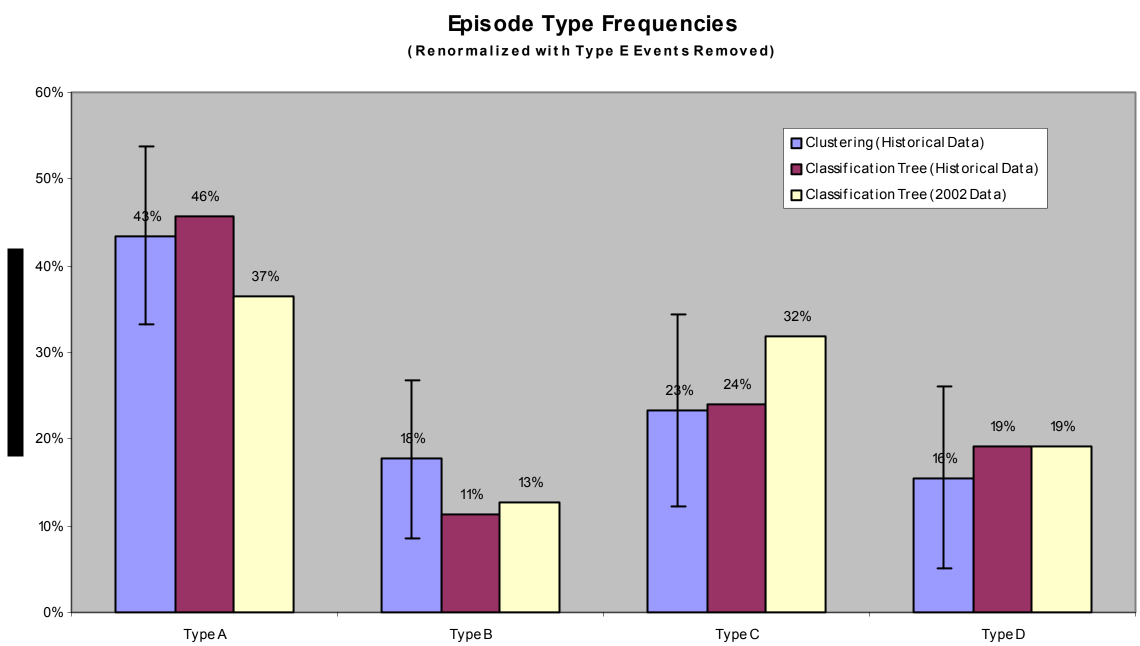


Figure 3-10. Percent of episode days by type in the 1997 – 2001/2003 historical period (as determined by the cluster analysis and by the classification tree) and in 2002 with Type E events removed and frequencies re-normalized; error bars show 10th and 90th percentiles of annual frequencies of occurrence during 1997-2001&2003.

Ozone Concentration levels

An exceedance of the 8-hour ozone standard occurred at one or more sites in the study region on 71 days during 2002, representing 46% of the 153 days during the May – September season analyzed in this study. For the 1997 – 2001/2003 historical period, the corresponding percentage was 36% so exceedances were more frequent during 2002. The greater frequency of ozone exceedance events was distributed throughout the OTR as shown by the comparison by monitoring sub-region in Table 3-5. Exceedances occurred with 20 – 50% greater frequency in 2002 in all sub-regions (100% greater in sub-region 3). This difference in the frequency of exceedances in 2002 as compared to the historical period does not necessarily mean, however, that the exceedance events themselves have characteristics that significantly differ from those seen during the historical period.

Table 3-5. Number of days during May-September with 8-hour daily maximum ozone greater than 0.08 ppm in each monitoring sub-region averaged over the 1997-2001/2003 historical period and in 2002.

	Sub-Region					
	1	**2**	**3**	**4**	**5**	**6**
>0.08 ppm						
1997-2001/2003	22.3	31.0	8.5	27.3	42.0	30.8
2002	34	38	17	39	58	44
Pct. Difference	52%	23%	100%	43%	38%	43%

Distributions of daily maximum 8-hour average ozone concentrations averaged over monitors in each sub-region (the AvgExc00 statistic) in 2002 and the historical period are compared for each event type in Figures 3-12(a-e), a key to the boxplot symbols used to summarize the ozone distribution is shown in Figure 3-11. Overall, the range of ozone under each event type in 2002 is similar to that under the corresponding event type in the historical period. The most notable exceptions are higher ozone levels during Type D events in 2002 along the Washington – New York City corridor (sub-regions 2, 5, and 6). This is consistent with a less pronounced low pressure center off the NC coast in the 2002 Type D events as compared to the historical period (see further discussion below). Aside from this difference, the overall ozone levels during the 2002 exceedance events were generally very consistent with those observed during the historical period, not withstanding the fact that exceedance days were more frequent during 2002.

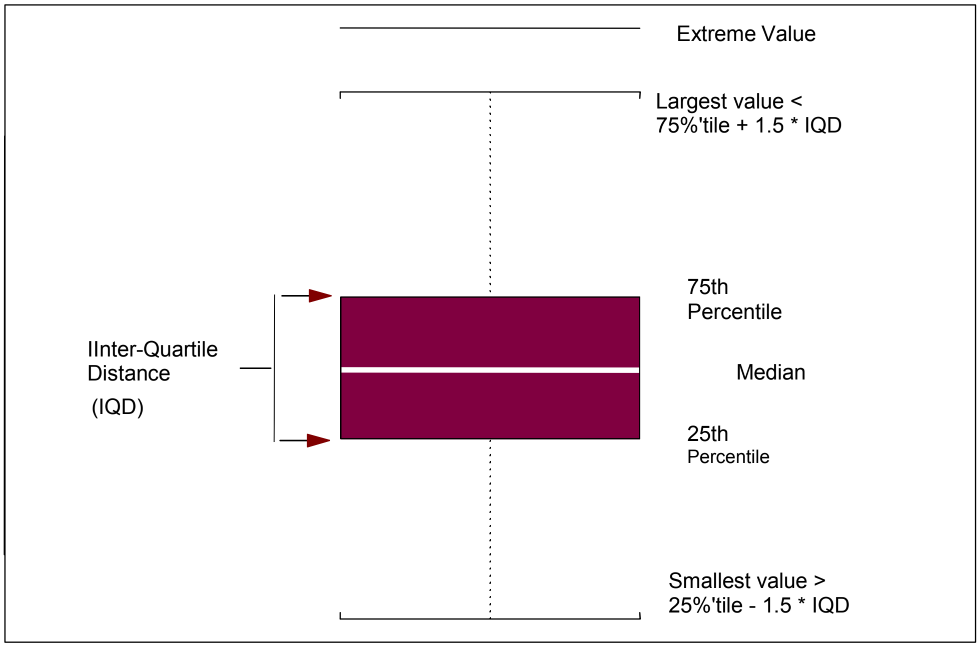


Figure 3-11. Key to boxplot symbols.

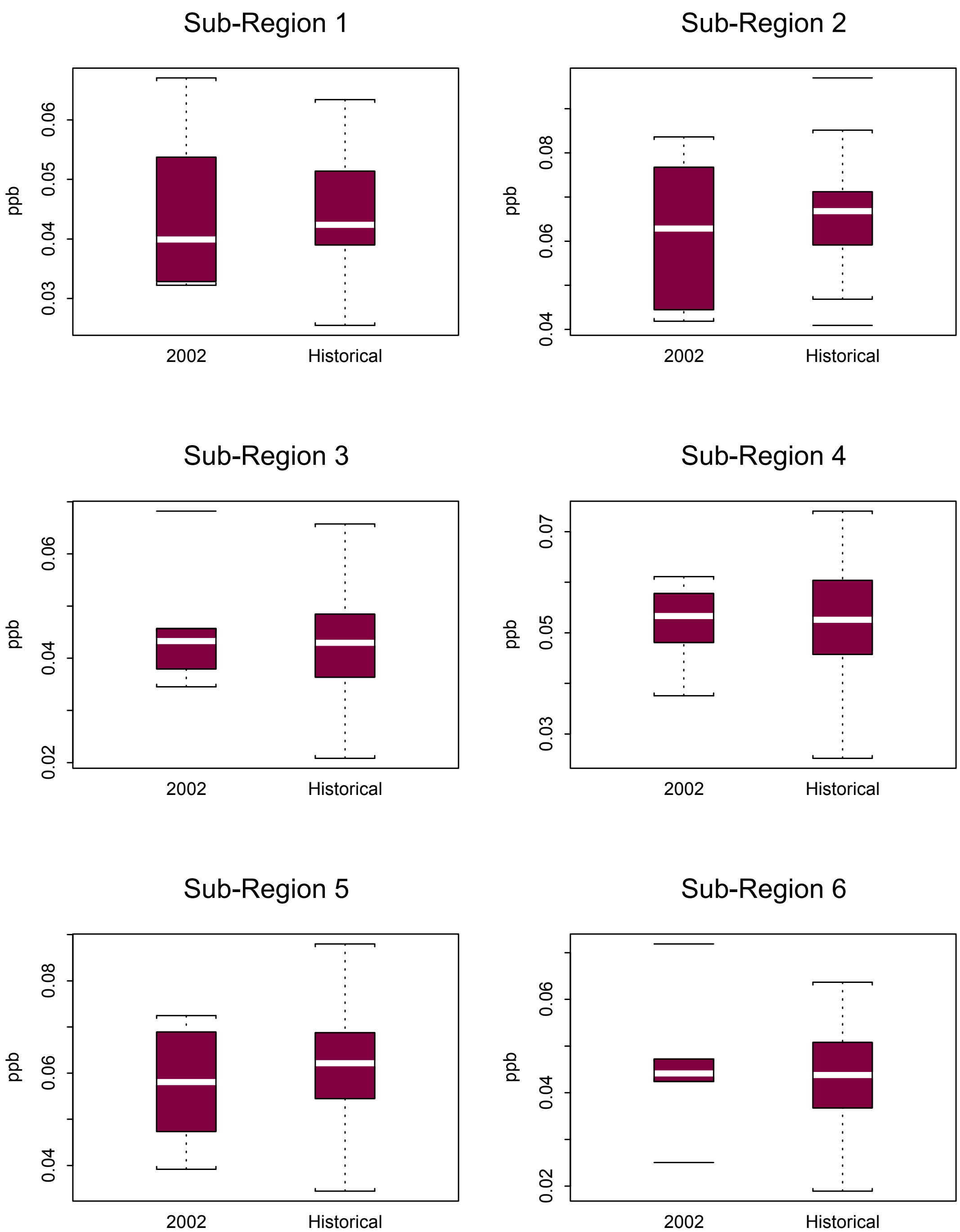


Figure 3-12a. Boxplots of average 8-hour daily maximum ozone (ppb) in each monitoring sub-region during 2002 and during the 1997-2001 and 2003 "historical" period: Type E (Pattern No. 1) events.

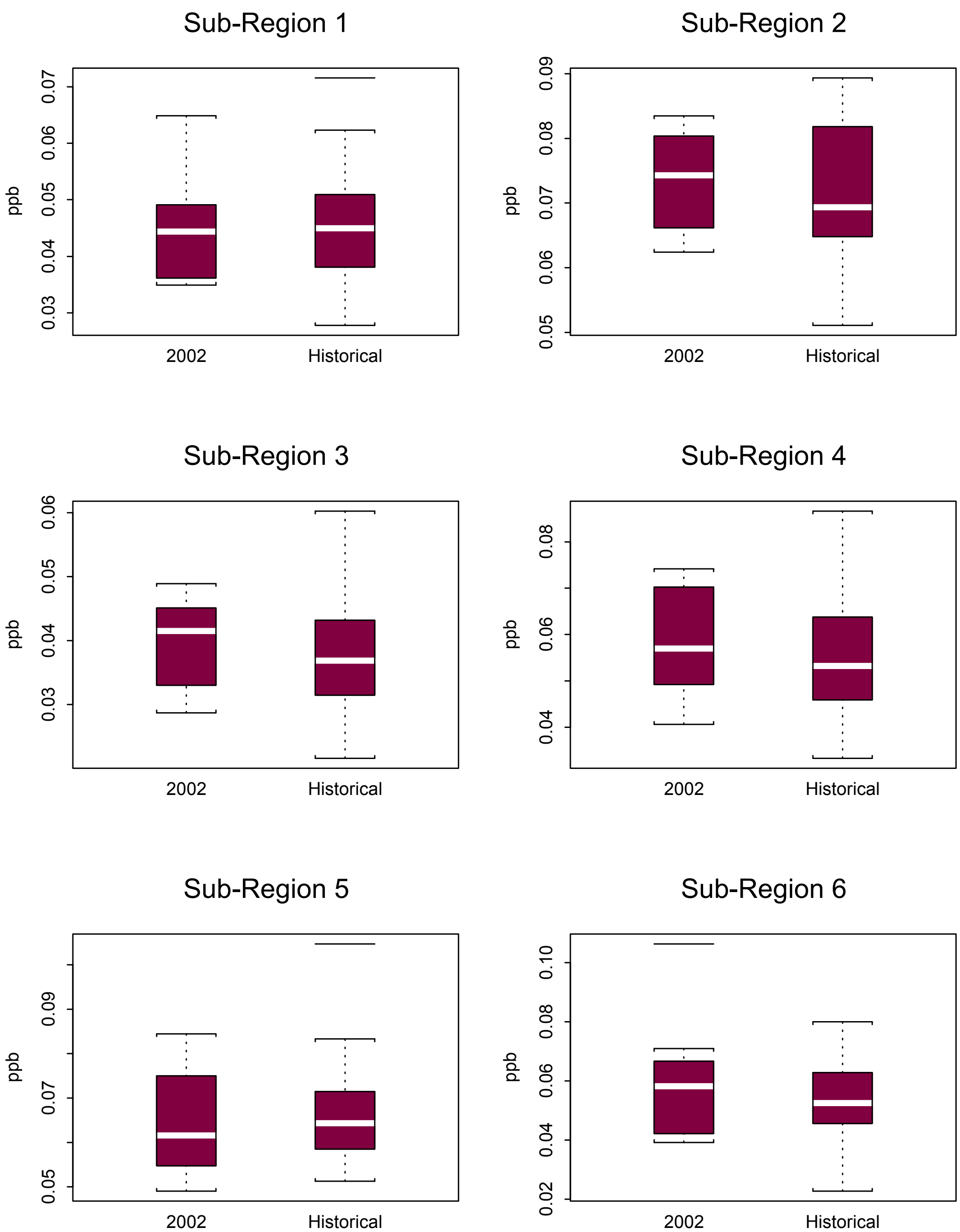


Figure 3-12b. Boxplots of average 8-hour daily maximum ozone (ppb) in each monitoring sub-region during 2002 and during the 1997-2001 and 2003 “historical” period: Type B (Pattern No. 2) events.

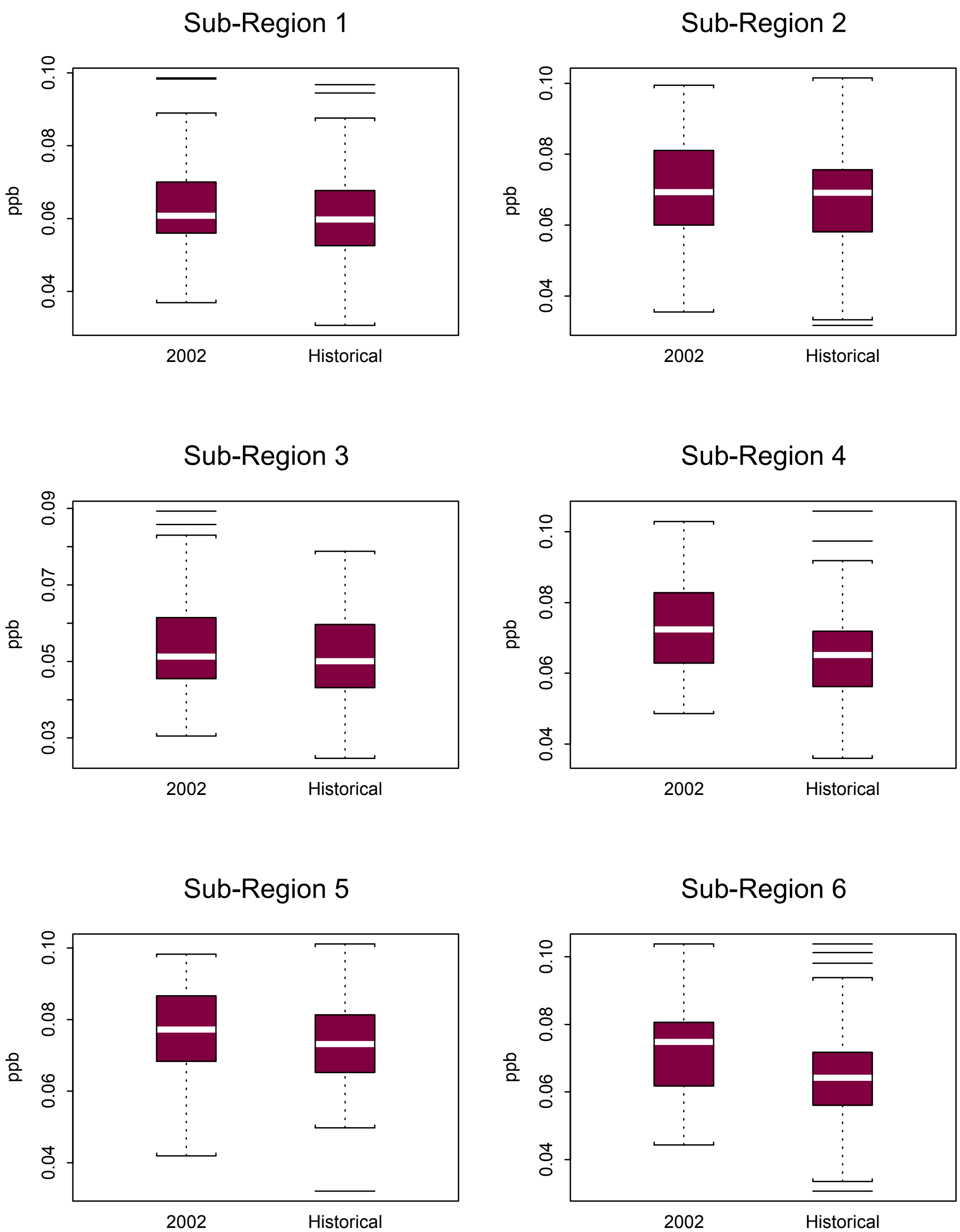


Figure 3-12c. Boxplots of average 8-hour daily maximum ozone (ppb) in each monitoring sub-region during 2002 and during the 1997-2001 and 2003 "historical" period: Type A (Pattern No. 3) events.

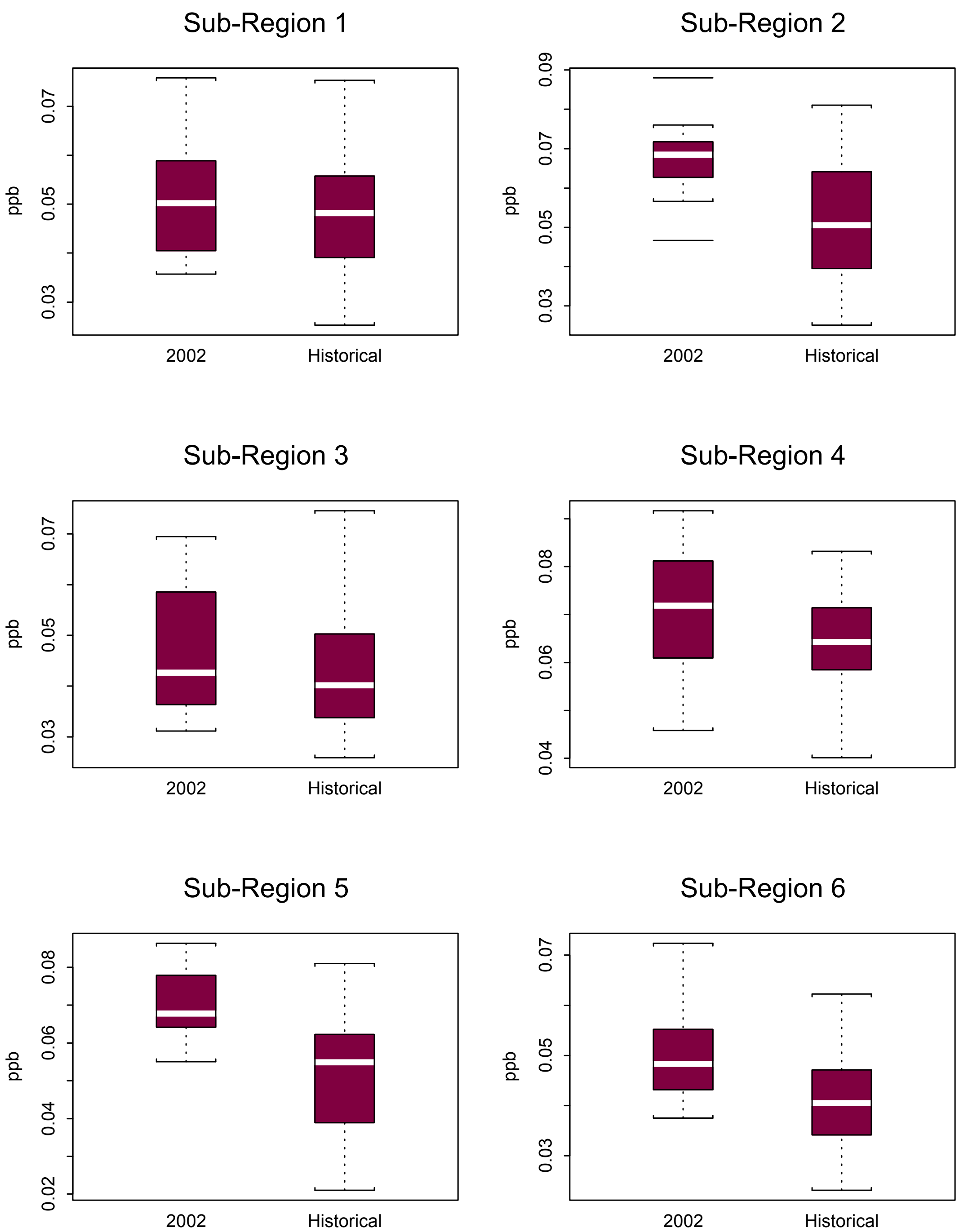


Figure 3-12d. Boxplots of average 8-hour daily maximum ozone (ppb) in each monitoring sub-region during 2002 and during the 1997-2001 and 2003 “historical” period: Type D (Pattern No. 4) events.

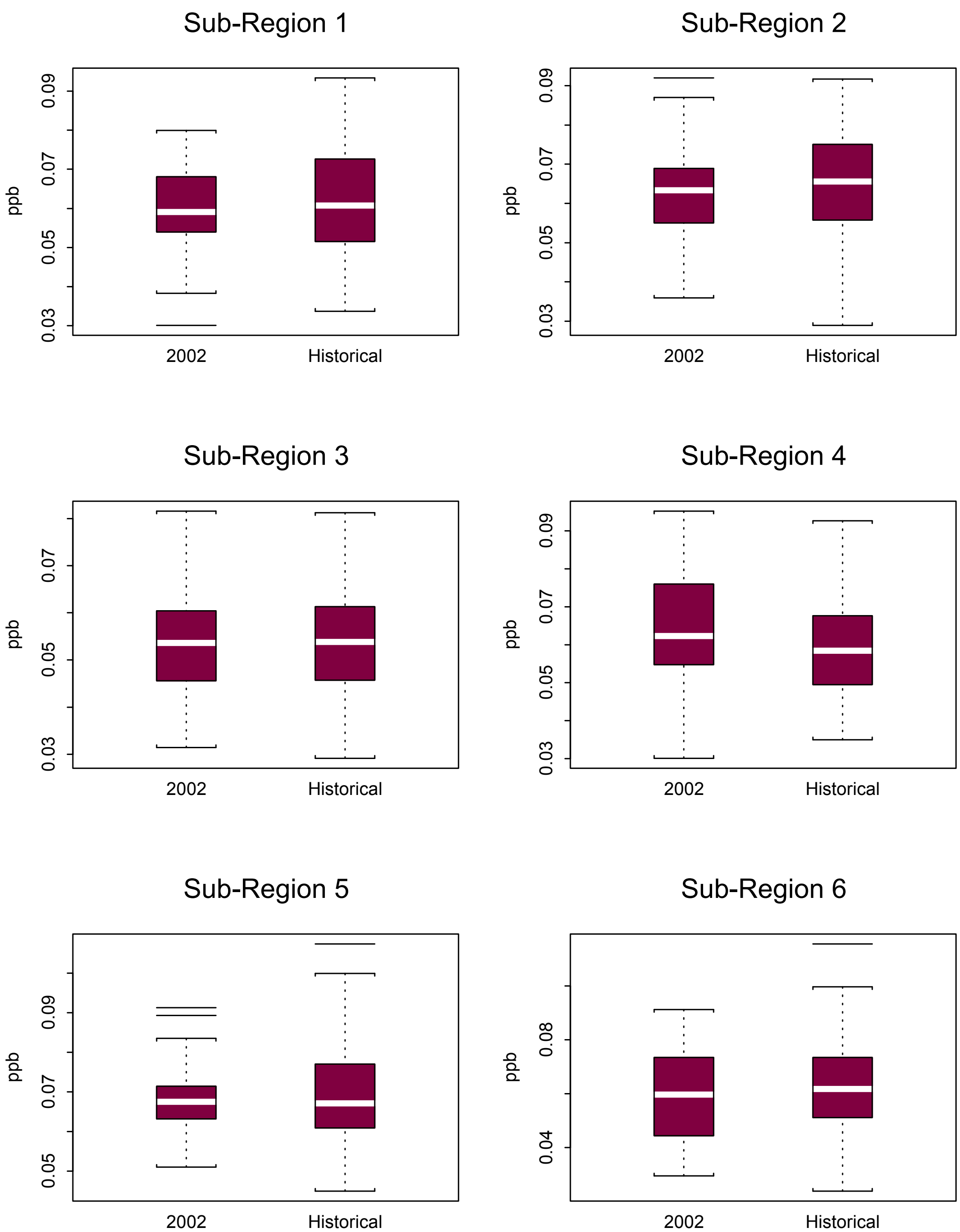


Figure 3-12e. Boxplots of average 8-hour daily maximum ozone (ppb) in each monitoring sub-region during 2002 and during the 1997-2001 and 2003 "historical" period: Type C (Pattern No. 5) events.

Meteorological Conditions

Selected composite meteorological fields for each episode type in 2002 as predicted by application of the classification tree were computed and displayed for comparison with the historical period composite fields. Results are shown in Figure 3-13 through 3-16. Comparing of these results with those for the historical period (Figures 3-2 to 3-5), we see a remarkable degree of similarity:[8] the surface and upper air meteorological patterns for a given episode type in 2002 are very similar to those for the same episode type observed in the historical period. In other words, the key characteristics of each type observed in the historical dataset are reproduced within the 2002 data. Perhaps the most significant difference is the less pronounced low pressure center off the NC coast under Type D events in 2002 which allowed for the formation of higher ozone concentrations along the Washington – New York City corridor for these event types in 2002 as compared to the historical period. Overall, however, the close match in weather patterns associated with each event type in 2002 and the historical period strongly supports the conclusion that the 2002 ozone episodes, although more numerous than in other years, are of substantially similar character.

[8]In making these comparisons, note that different color and wind vector scales had to be used in some plots of the 2002 data.

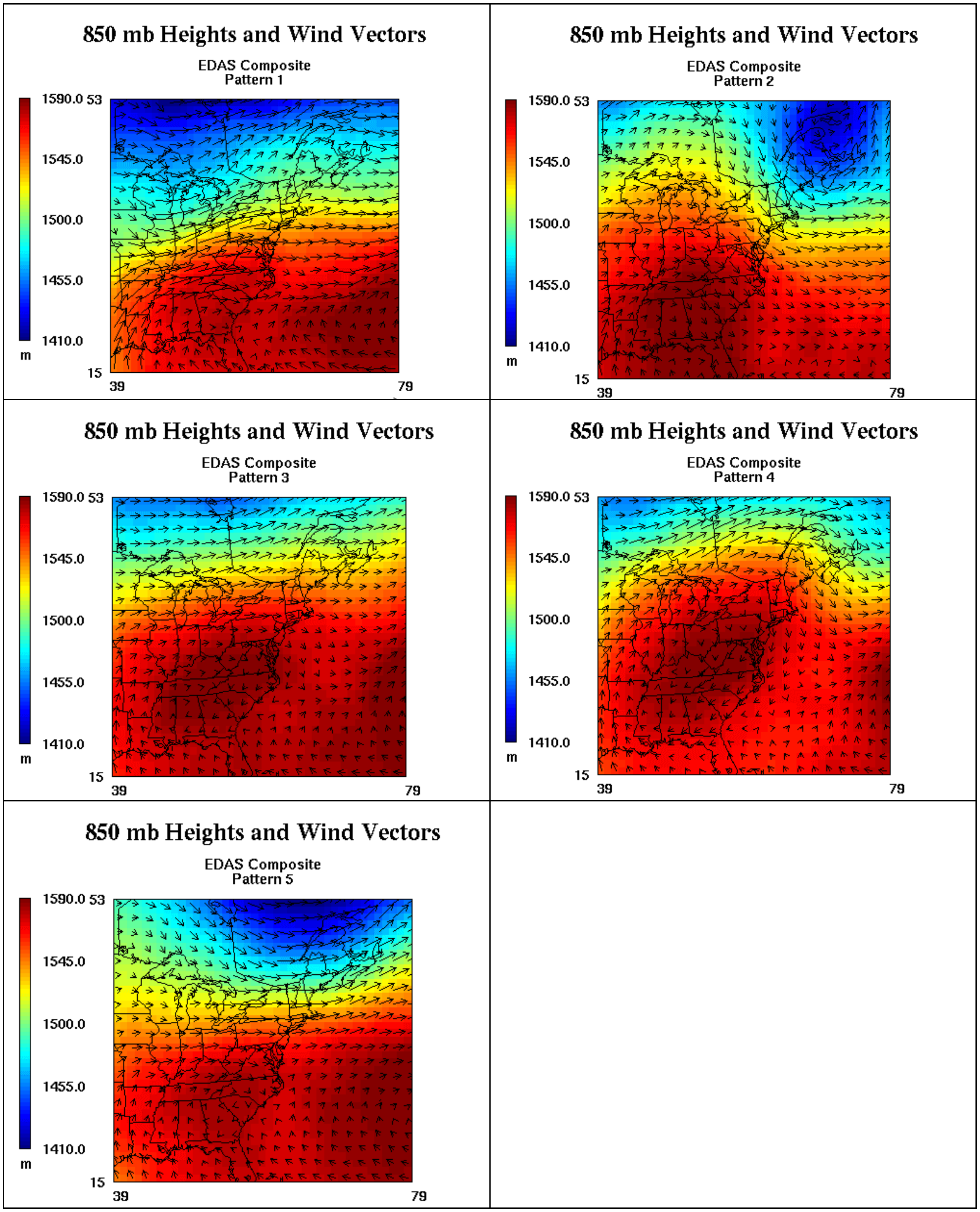


Figure 3-13. Average meteorological fields by episode type (pattern) in 2002: 850 mb heights and winds. Pattern numbers refer to the episode types listed in Table 3-2: Pattern 1 = Episode Type E, 2 = B, 3 = A, 4 = D, 5 = C).

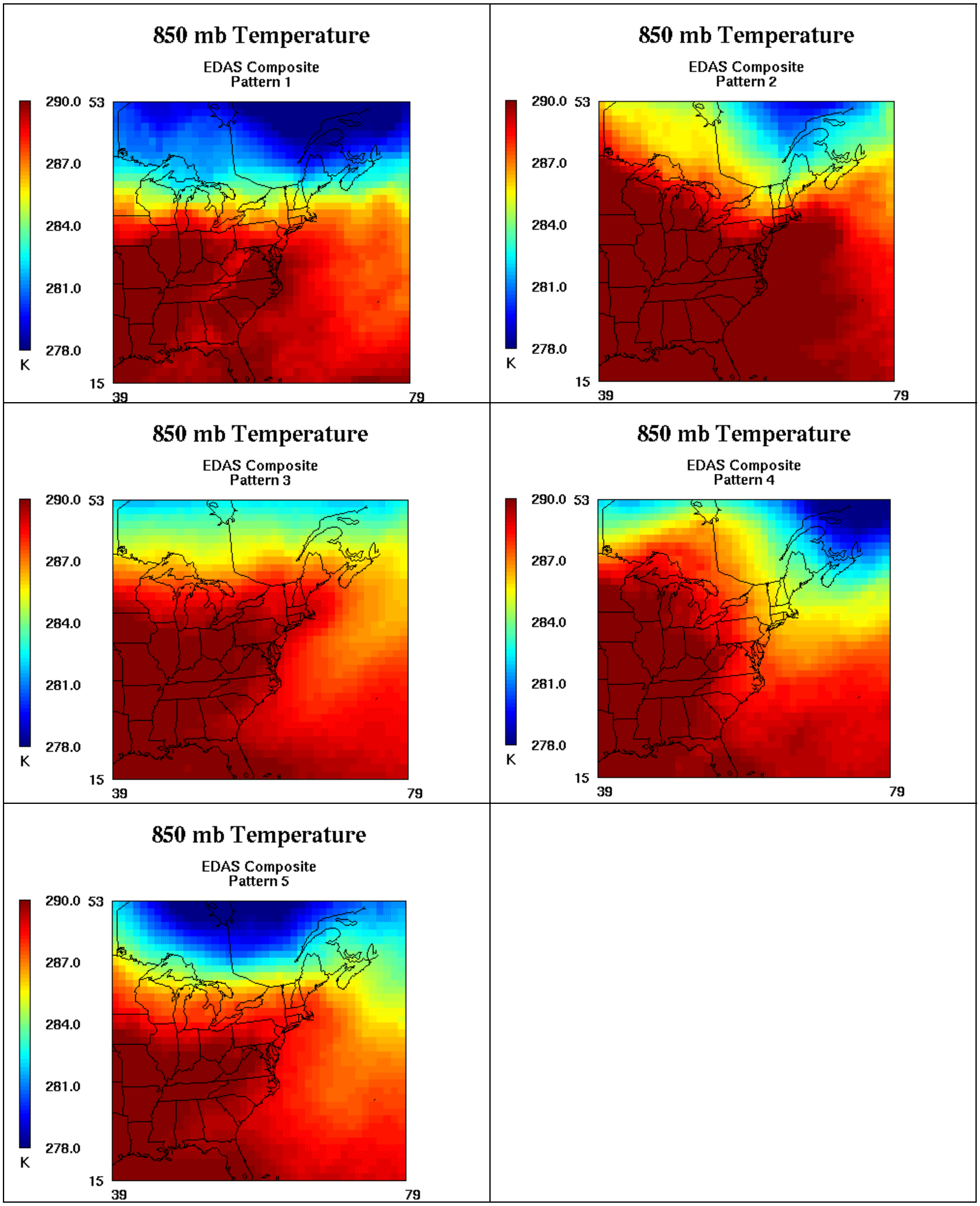


Figure 3-14. Average meteorological fields by episode type (pattern) in 2002: 850 mb temperature. Pattern numbers refer to the episode types listed in Table 3-2: Pattern 1 = Episode Type E, 2 = B, 3 = A, 4 = D, 5 = C).

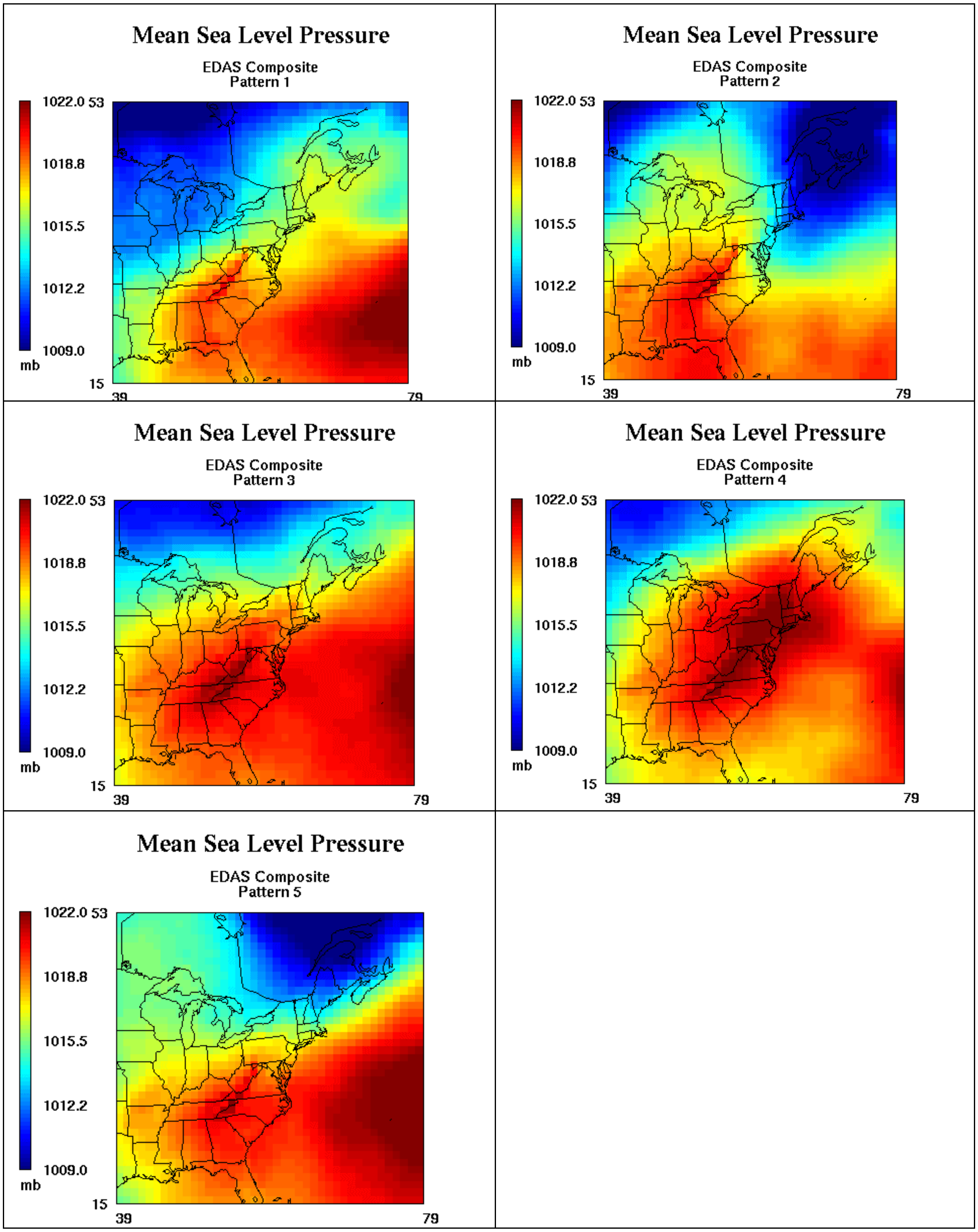


Figure 3-15. Average meteorological fields by episode type (pattern) in 2002: sea level pressure. Pattern numbers refer to the episode types listed in Table 3-2: Pattern 1 = Episode Type E, 2 = B, 3 = A, 4 = D, 5 = C).

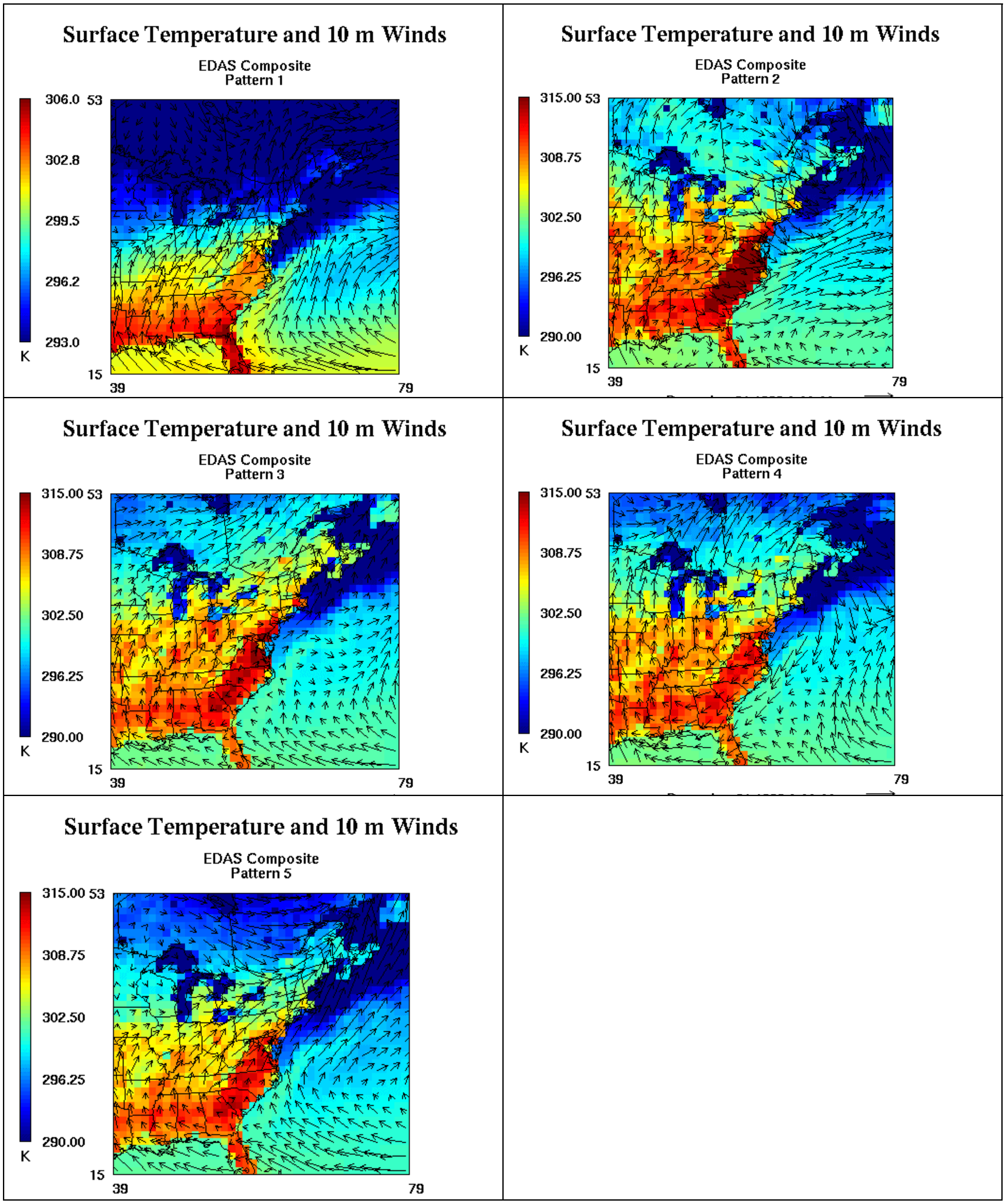


Figure 16. Average meteorological fields by episode type (pattern) in 2002: surface temperature and 10 m winds. Pattern numbers refer to the episode types listed in Table 3-2: Pattern 1 = Episode Type E, 2 = B, 3 = A, 4 = D, 5 = C).

4. CONCLUSIONS AND RECOMMENDATIONS

Results from the application of statistical clustering analyses presented in Section 3 show that regional ozone episode conditions over the OTR can be reasonably well described by a set of five different episode types. Our examination of mean ozone and meteorological conditions shows that each of these episode types is associated with a unique set of distinguishing characteristics. While we would not expect every exceedance day to exhibit all of the characteristics of one type or another, our results provide no clear evidence for the existence of any other additional sufficiently unique types that occur frequently enough to be distinguishable within the six year historical period analyzed.

Data from the 2002 ozone season were analyzed within the framework of the five identified episode types with respect to: a) frequencies of occurrence of each type and b) characteristics of the ozone and meteorological conditions within each type in 2002 as compared to the 1997 – 2001/2003 historical period.

A key feature of the 2002 season is that ozone episodes (defined as an exceedance of the 8-hour ozone NAAQS at one or more monitoring sites within the OTR) occurred more frequently than during the historical period (71 exceedance days during the May – September season in 2002 as compared to an average of 55 days per season during the historical period). Taken by itself, however, this difference does not necessarily mean that region-wide meteorological and ozone concentration patterns *during exceedance days* were significantly different in 2002 as compared to other years: the greater number of exceedance days in 2002 may just reflect a lower than average frequency of days with meteorological conditions not conducive to ozone formation in 2002. The higher than average exceedance rate in 2002 is by itself not an indication of any lack of representativeness of the 2002 exceedance events.

Our examination of conditions during exceedance days in 2002 showed that:

- Except for the Type E events during which ozone exceedances are typically confined to the extreme southeastern corner of the OTR, each of the five episode types identified in the historical period was found to occur on about as many days in 2002 as one would expect based on their rate of occurrence during the historical record. Thus the meteorological conditions on episode days in 2002 exhibit a normal range of variation and each of the five types of episodes are well represented.
- Type E events are under represented in the 2002 season. This is consistent with the higher than average frequency of exceedance days in 2002. The relative lack of Type E events in 2002 should not be of concern from a SIP modeling standpoint, however, as these events are characterized by relatively low ozone levels throughout nearly all of the OTR (except the Washington and Virginia area).
- The distribution of daily maximum 8-hour average ozone levels during each event type in 2002 is generally very similar to that within the same event type during the historical period. The only significant exception is higher ozone along the Washington – New York City corridor under Type D events in 2002 as compared to the historical average.
- Regional-scale meteorological conditions during each event type in 2002 exhibit the same key characteristics as observed for the event types during the historical period. A less

pronounced low pressure center off of the NC coast under the 2002 Type D events appears to be responsible for the higher Washington – New York City ozone levels under this event type noted in the previous bullet.

In summary, while ozone exceedances were more frequent during 2002, conditions during the 2002 exceedance events were for the most part very similar to those found to occur in other years. This leads us to conclude that the 2002 season can be considered to be representative for purposes of photochemical modeling in support of SIP development.

REFERENCES

Breiman, L., J.H. Friedman, R.A. Olshen and C.J. Stone, 1984. Classification and Regression Trees. Wadsworth, Belmont, CA

Cox, W.M., 1997. Spatial Clustering of Ambient Ozone in Eastern U.S., Draft, U.S. Environmental Protection Agency, January (available at http://www.epa.gov/ttn/naaqs/ozone/areas/misc/ozclus2a.htm)

Deuel, H.P. and S.G. Douglas, 1996. "Regional Ozone Patterns in the Eastern U.S. 1985-1995: Statistical Pattern Analysis and Classification and Regression Tree (CART) Analysis in Support of the Ozone Transport and Assessment Group (OTAG) Modeling Effort." Systems Applications Inc. San Rafael, CA October.

EPA, 1999. "Draft Guidance on the Use of Models and Other Analyses in Attainment Demonstrations for the 8-Hour Ozone NAAQS." EPA-54/R-99-004, U.S. Environmental Protection Agency, May.

Harrell, F.E., 1999. Hmisc S-PLUS function library. Programs available from lib.stat.cmu.edu.

Hartigan, J. A. and M.A. Wong, 1979. A k-means clustering algorithm. *Applied Statistics 28*, 100-108.

McHenry, J.N., W.F. Ryan, N.L. Seaman, C.J. Coats, J. Pudykiewicz, S. Arunachalam and J.M. Vukovich, 2004. A real-time Eulerian photochemical model forecast system: Overview and initial ozone forecast performance in the northeastern U.S. corridor. *Bulletin of the American Meteorological Society*, 85(4), April.

Sarle, WS: The VARCLUS Procedure. SAS/STAT User's Guide, 4th Edition, 1990. Cary NC: SAS Institute, Inc.

Stoeckenius, T.E. and S. Kemball-Cook, 2004. Ozone episode classification project: spatial ozone pattern analysis (Task 2a). Memo to T. Frankiewicz, Ozone Transport Comission, 21 September.

Venables, W.N. and B.D. Ripley, 1994. "Modern Applied Statistics with S-Plus." Springer-Verlag, New York.

Ward, J.H., 1963. Hierarchical groupings to optimize an objective function. *J. Amer. Statistical Assoc.*, 58:236-244.

Appendix A

Composites Representing Mean Meteorological Conditions During Each Ozone Episode Pattern

Composites Representing Mean Meteorological Conditions During Each Ozone Episode Pattern

Mean meteorological fields were computed over days falling into each of the five ozone episode patterns in the Ozone Transport Region defined in the text. The five episode patterns and their composite pattern identifiers are:

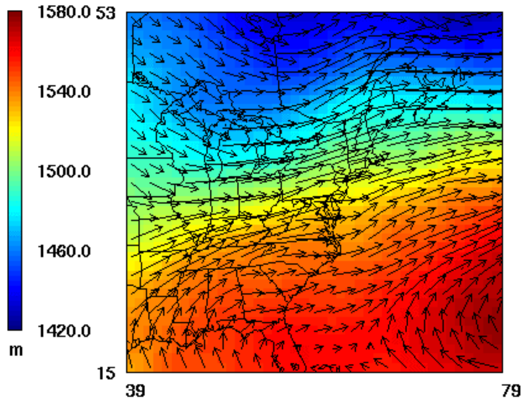
Composite Pattern	Episode Type
3	Type A: High ozone throughout the OTR
2	Type B: High ozone confined to extreme southeastern OTR
5	Type C: High ozone along I-95 corridor and northern New England
4	Type D: High ozone in the western OTR
1	Type E: Generally low ozone throughout the OTR

Mean fields were computed for the following parameters extracted from the EDAS data:

Parameter ID	Description
H850	850 mb height
850 mb Wind	Resultant wind vector at 850 mb
T(850 mb)	850 mb temperature (deg K)
MSLP	Mean sea level pressure (mb)
TSFC	Surface temperature (deg K)
10m Wind	Resultant wind vector at 10 m height
w_500	w (vertical) component of 500 mb wind vector
H500	500 mb heights

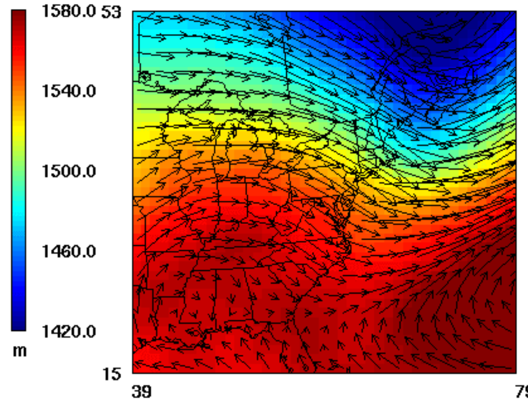
850 mb Heights and Winds

EDAS Composite
Pattern 1



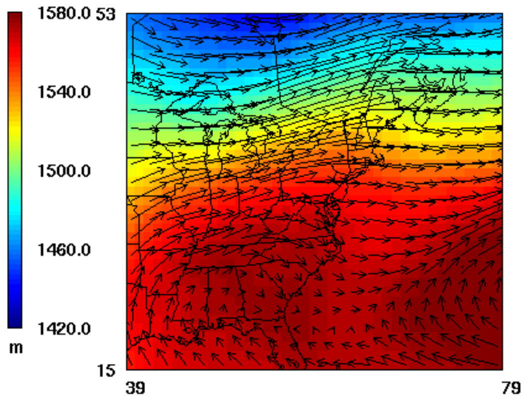
850 mb Heights and Winds

EDAS Composite
Pattern 2



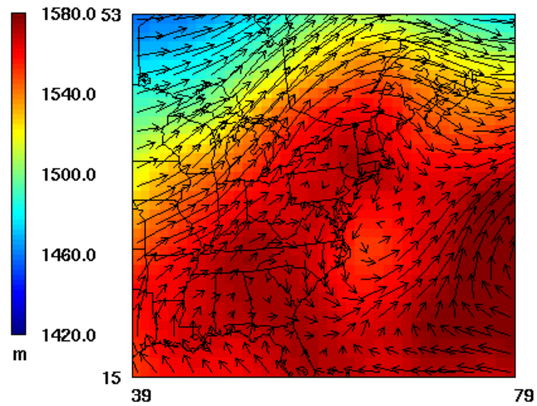
850 mb Heights and Winds

EDAS Composite
Pattern 3



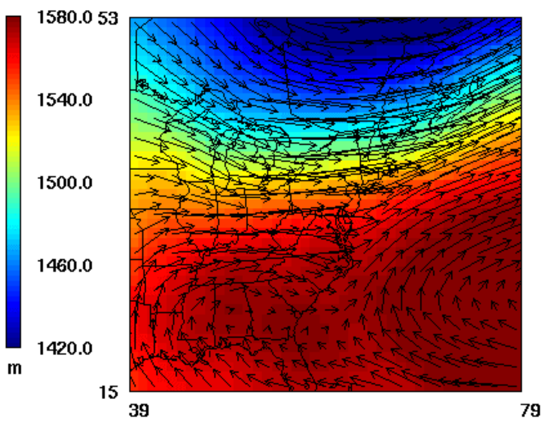
850 mb Heights and Winds

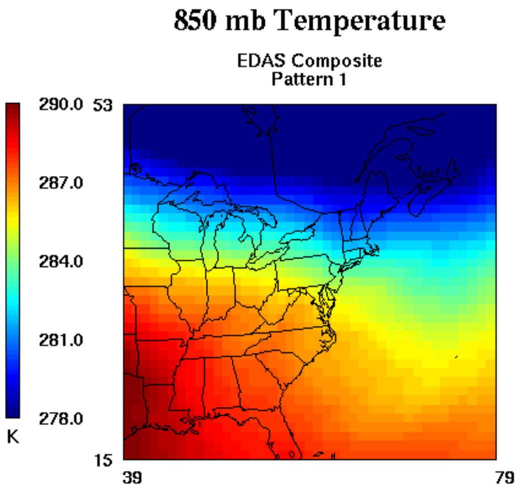
EDAS Composite
Pattern 4

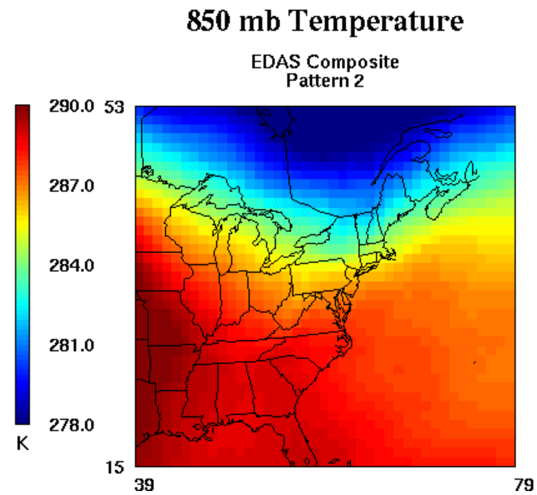


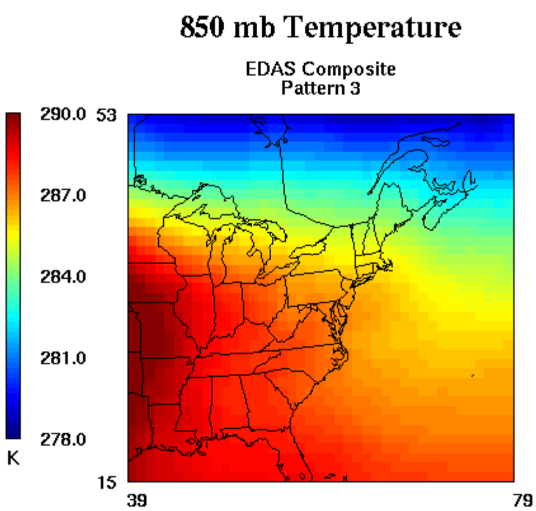
850 mb Heights and Winds

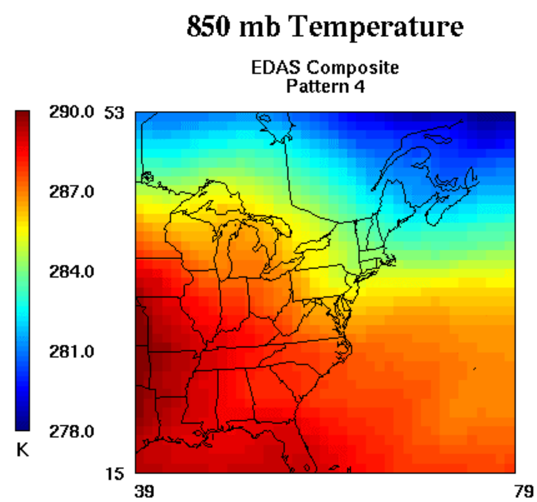
EDAS Composite
Pattern 5

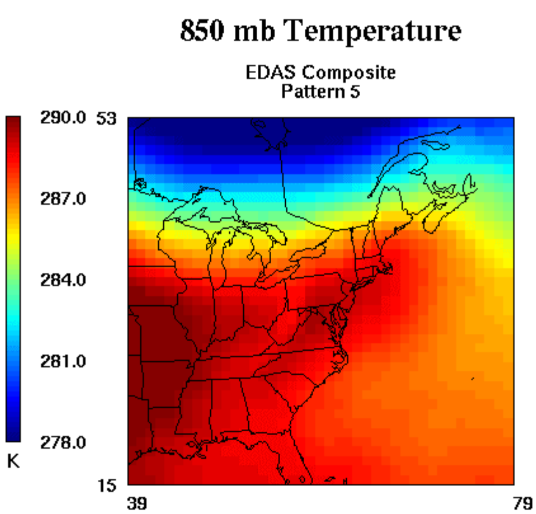


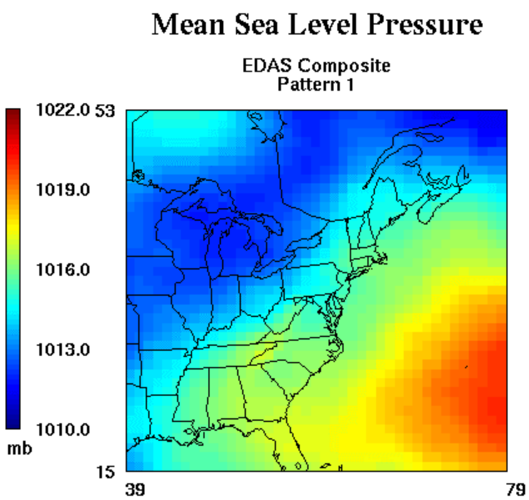

850 mb Temperature
EDAS Composite
Pattern 1
290.0
287.0
284.0
281.0
278.0
K
53
15
39
79

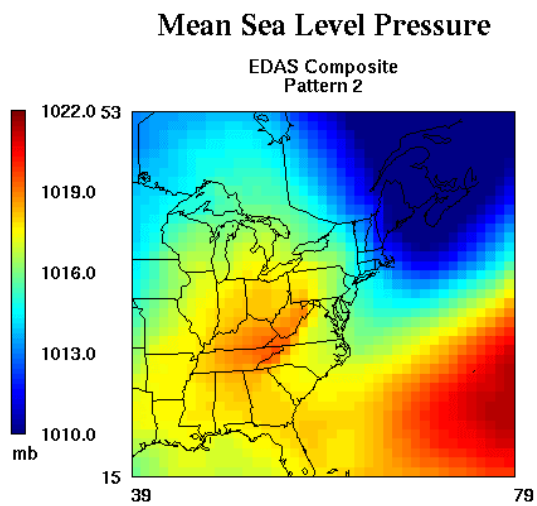

850 mb Temperature
EDAS Composite
Pattern 2
290.0
287.0
284.0
281.0
278.0
K
53
15
39
79

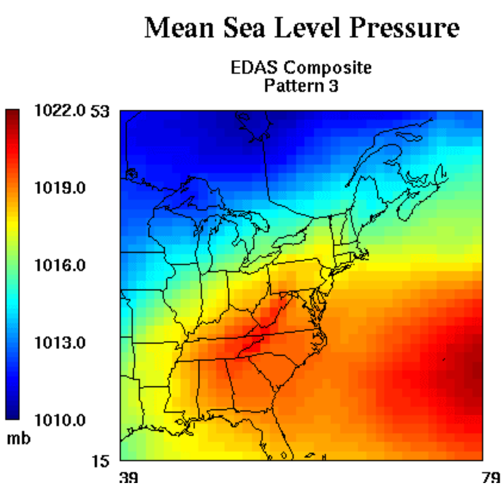

850 mb Temperature
EDAS Composite
Pattern 3
290.0
287.0
284.0
281.0
278.0
K
53
15
39
79

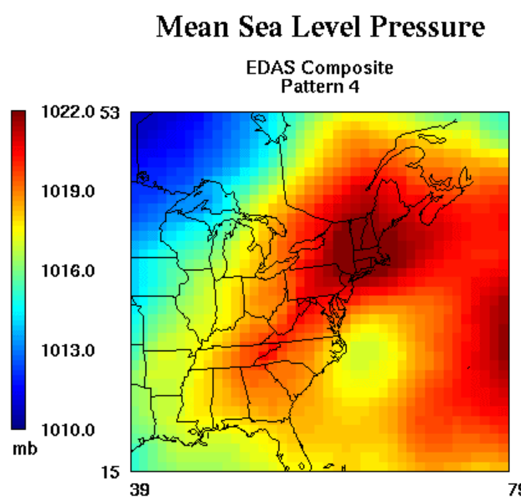

850 mb Temperature
EDAS Composite
Pattern 4
290.0
287.0
284.0
281.0
278.0
K
53
15
39
79

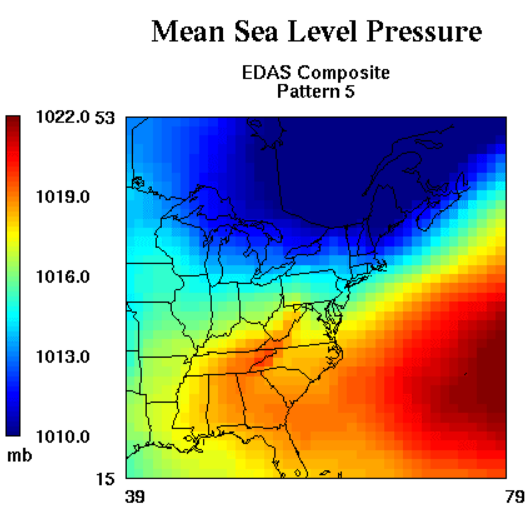

850 mb Temperature
EDAS Composite
Pattern 5
290.0
287.0
284.0
281.0
278.0
K
53
15
39
79

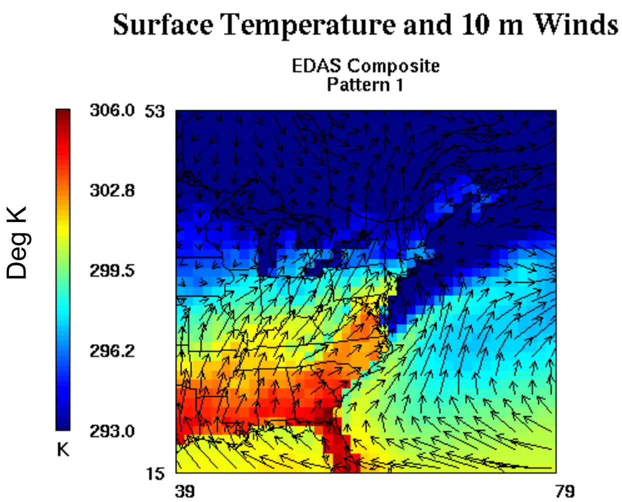

Mean Sea Level Pressure
EDAS Composite
Pattern 1
1022.0
1019.0
1016.0
1013.0
1010.0
mb
53
15
39
79

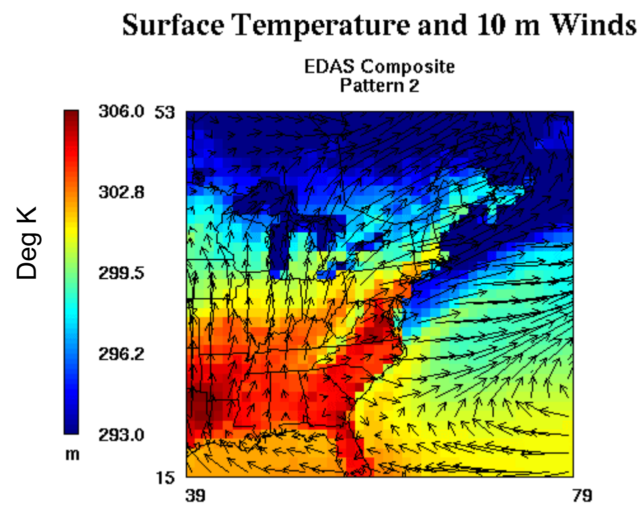

Mean Sea Level Pressure
EDAS Composite
Pattern 2
1022.0
1019.0
1016.0
1013.0
1010.0
mb
53
15
39
79

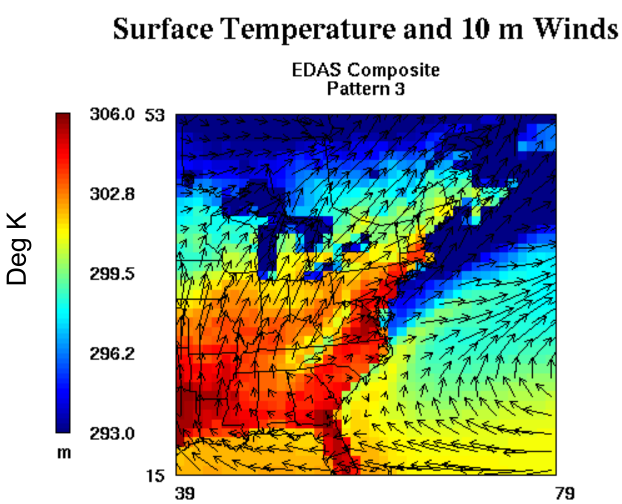

Mean Sea Level Pressure
EDAS Composite
Pattern 3
1022.0
1019.0
1016.0
1013.0
1010.0
mb
53
15
39
79

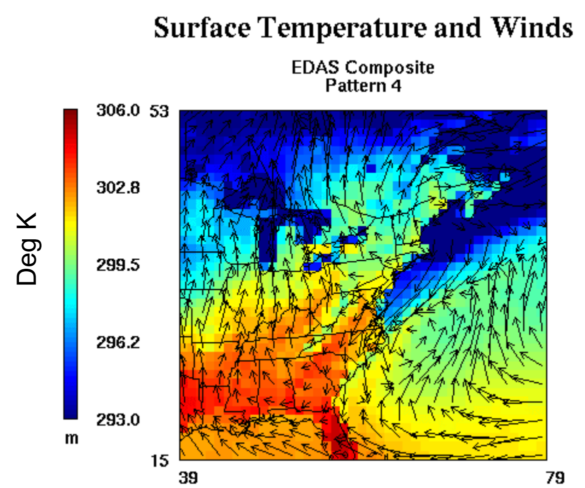

Mean Sea Level Pressure
EDAS Composite
Pattern 4
1022.0
1019.0
1016.0
1013.0
1010.0
mb
53
15
39
79

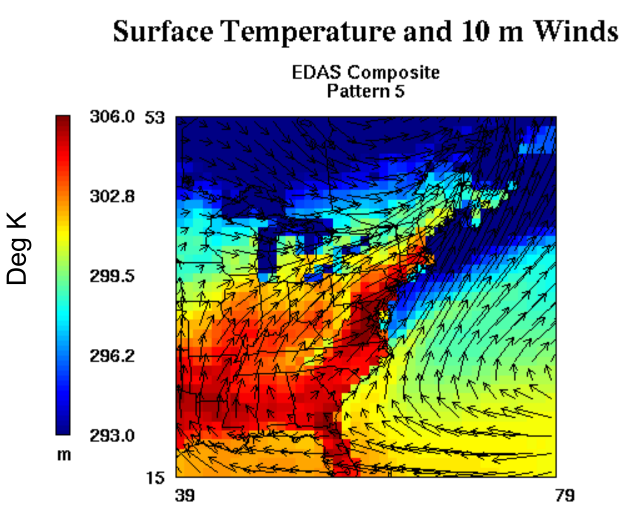

Mean Sea Level Pressure
EDAS Composite
Pattern 5
1022.0
1019.0
1016.0
1013.0
1010.0
mb
53
15
39
79

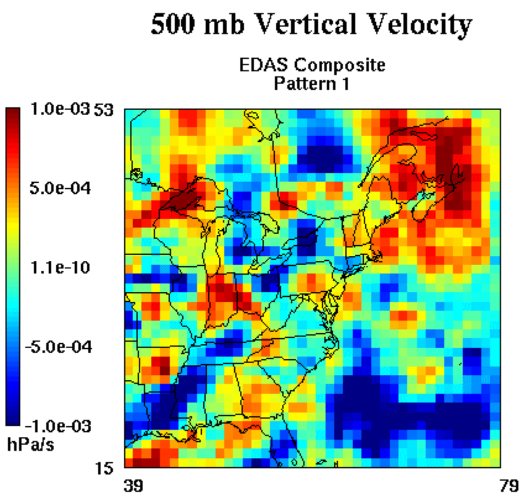

Surface Temperature and 10 m Winds
EDAS Composite
Pattern 1
Deg K
306.0
302.8
299.5
296.2
293.0
K
53
15
39
79

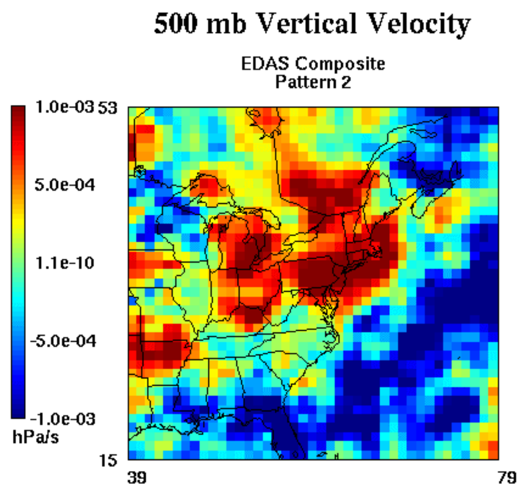

Surface Temperature and 10 m Winds
EDAS Composite
Pattern 2
Deg K
306.0
302.8
299.5
296.2
293.0
m
53
15
39
79

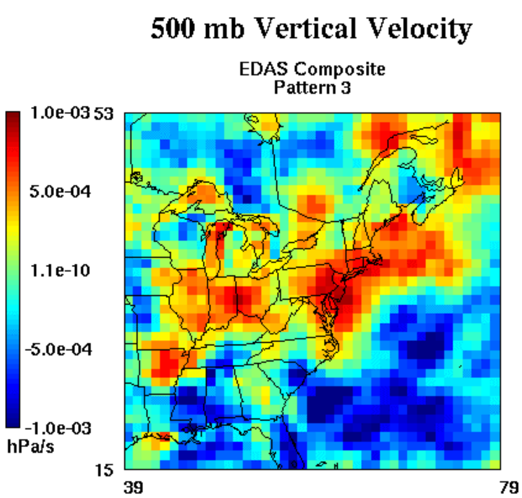

Surface Temperature and 10 m Winds
EDAS Composite
Pattern 3
Deg K
306.0
302.8
299.5
296.2
293.0
m
53
15
39
79

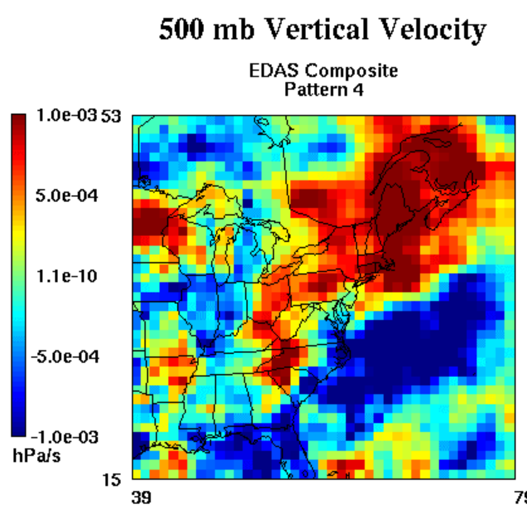

Surface Temperature and Winds
EDAS Composite
Pattern 4
Deg K
306.0
302.8
299.5
296.2
293.0
m
53
15
39
79

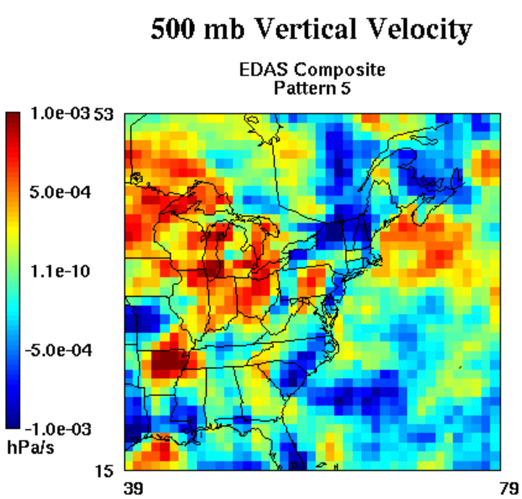

Surface Temperature and 10 m Winds
EDAS Composite
Pattern 5
Deg K
306.0
302.8
299.5
296.2
293.0
m
53
15
39
79

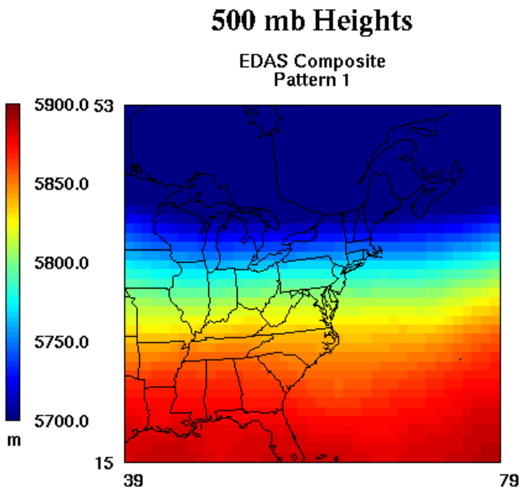

500 mb Vertical Velocity
EDAS Composite
Pattern 1
1.0e-03
5.0e-04
1.1e-10
-5.0e-04
-1.0e-03
hPa/s
53
15
39
79

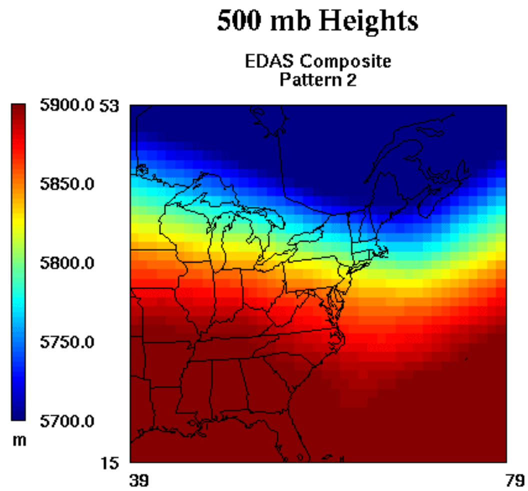

500 mb Vertical Velocity
EDAS Composite
Pattern 2
1.0e-03
5.0e-04
1.1e-10
-5.0e-04
-1.0e-03
hPa/s
53
15
39
79

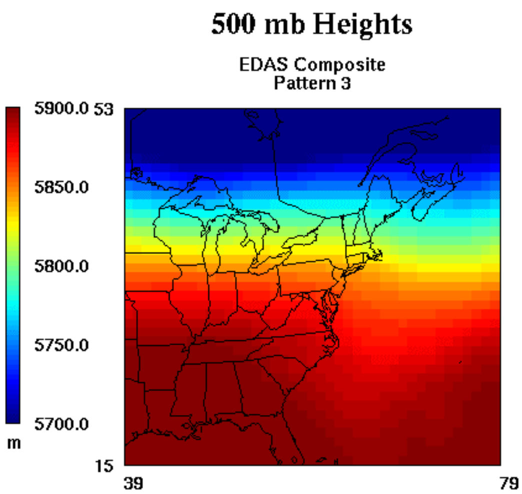

500 mb Vertical Velocity
EDAS Composite
Pattern 3
1.0e-03
5.0e-04
1.1e-10
-5.0e-04
-1.0e-03
hPa/s
53
15
39
79

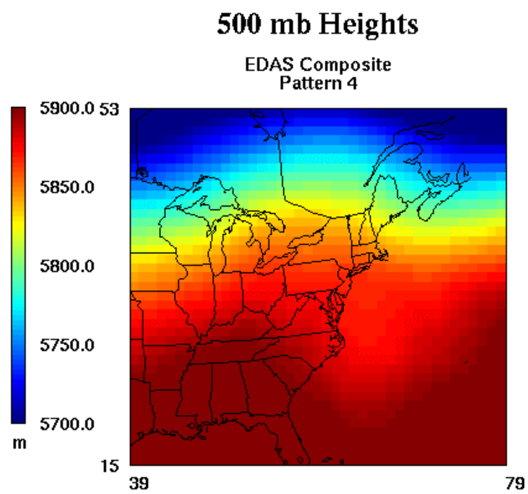

500 mb Vertical Velocity
EDAS Composite
Pattern 4
1.0e-03
5.0e-04
1.1e-10
-5.0e-04
-1.0e-03
hPa/s
53
15
39
79

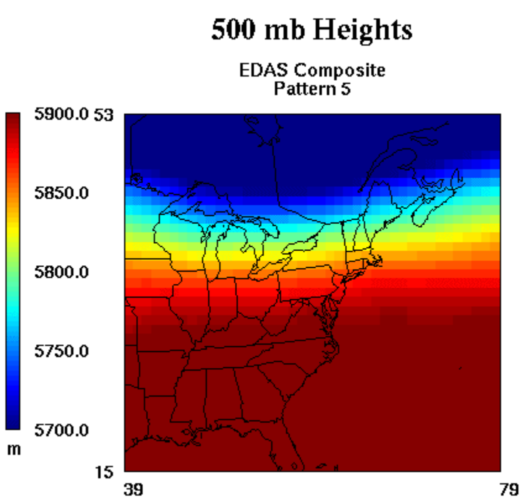

500 mb Vertical Velocity
EDAS Composite
Pattern 5
1.0e-03
5.0e-04
1.1e-10
-5.0e-04
-1.0e-03
hPa/s
53
15
39
79


500 mb Heights
EDAS Composite
Pattern 1
5900.0
5850.0
5800.0
5750.0
5700.0
m
53
15
39
79


500 mb Heights
EDAS Composite
Pattern 2
5900.0
5850.0
5800.0
5750.0
5700.0
m
53
15
39
79


500 mb Heights
EDAS Composite
Pattern 3
5900.0
5850.0
5800.0
5750.0
5700.0
m
53
15
39
79


500 mb Heights
EDAS Composite
Pattern 4
5900.0
5850.0
5800.0
5750.0
5700.0
m
53
15
39
79


500 mb Heights
EDAS Composite
Pattern 5
5900.0
5850.0
5800.0
5750.0
5700.0
m
53
15
39
79

Appendix B

Boxplots of Key Meteorological Variables

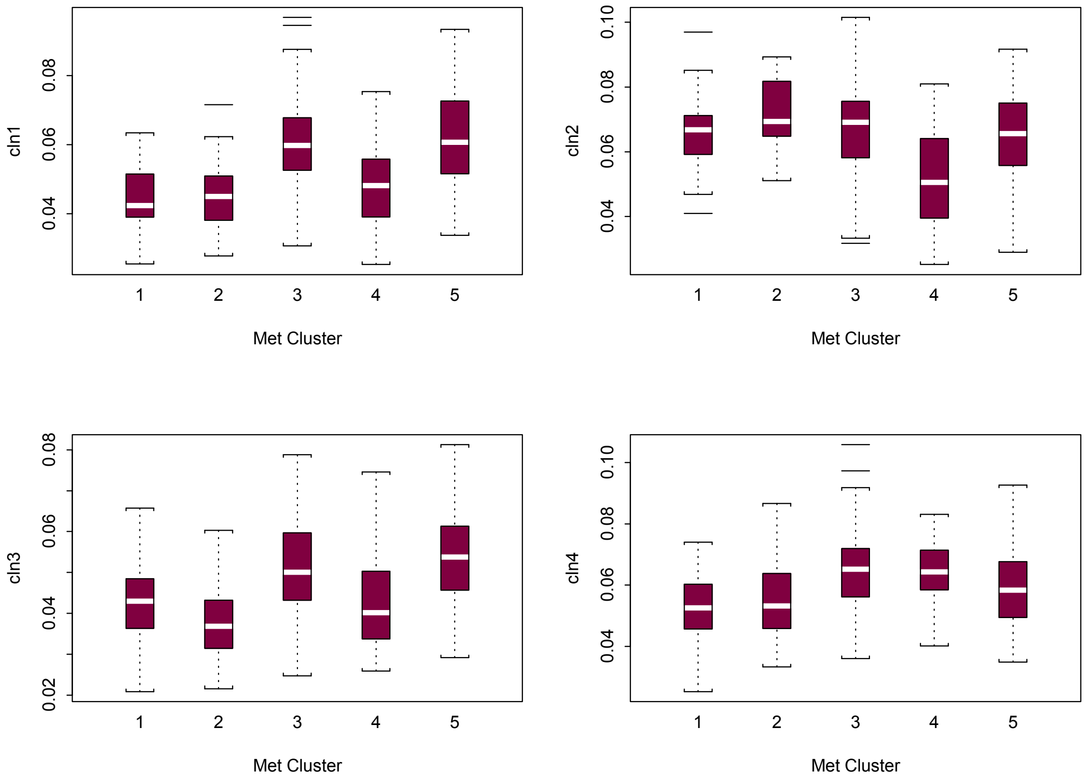
Boxplots Of Key Meteorological Variables

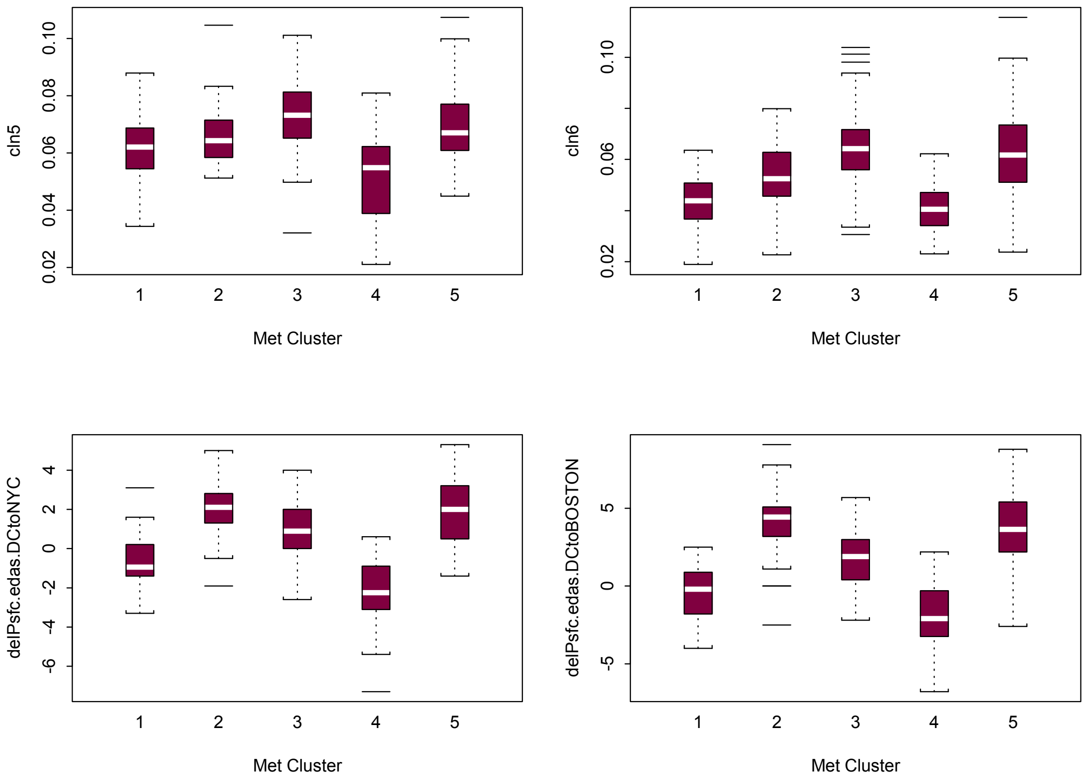
Boxplots in this Appendix summarize distributions of the sub-regional ozone summary statistic, AvgEx00, described in the text along with selected key daily meteorological variables by episode pattern membership for the five cluster case. Pattern membership identifiers ("Met Cluster") used in these plots correspond to the episode types described in the text as follows:

Met Cluster	Episode Type
1	Type E
2	Type B
3	Type A
4	Type D
5	Type C

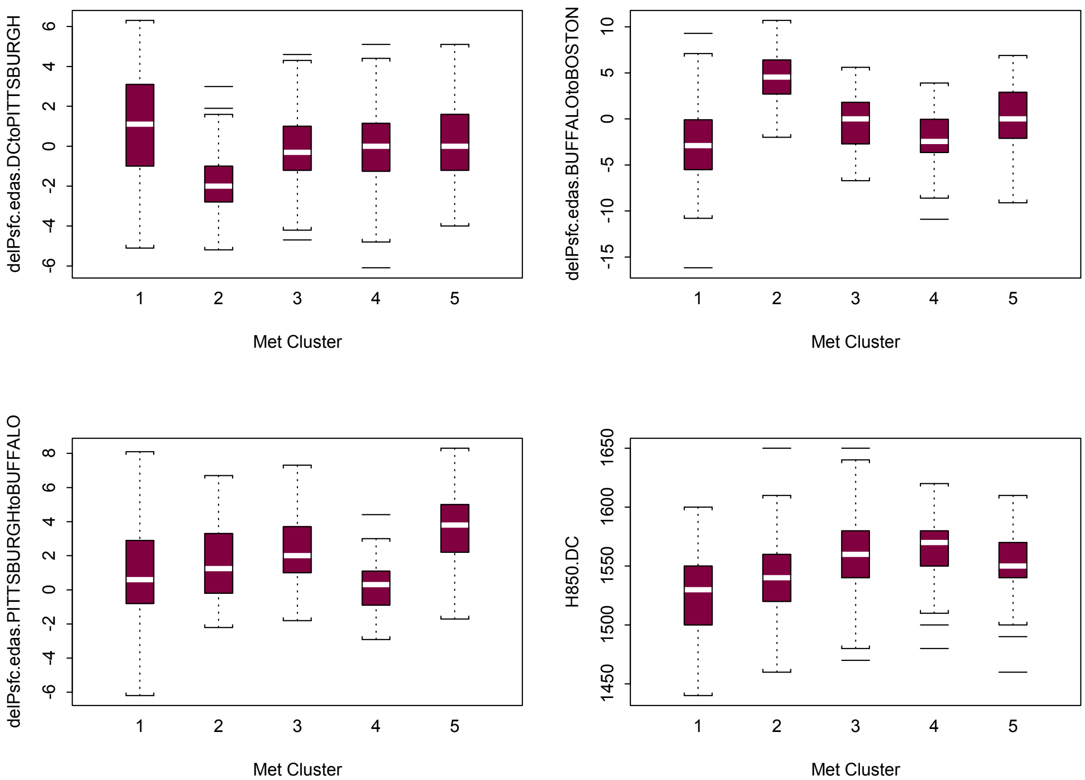
Ozone and meteorological variables are:

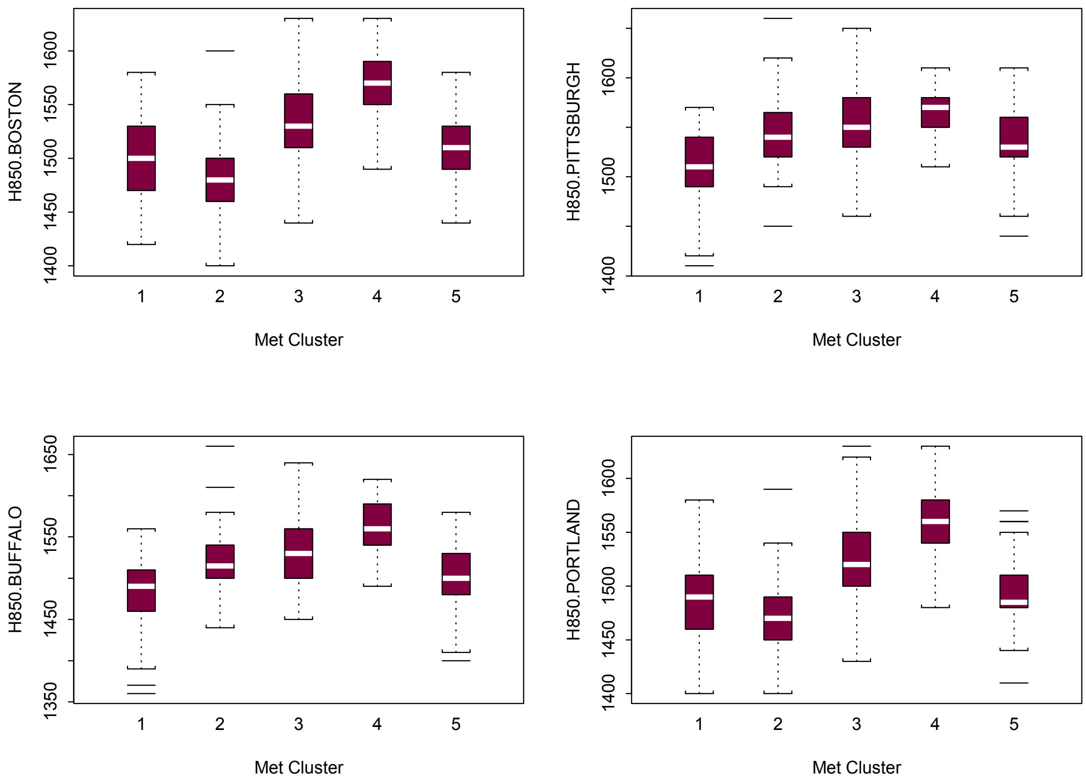
Variable	**Description**
clnx	AvgEx00 ozone summary statistic for ozone monitoring cluster x (x = 1,2…6; see Figure 1 in text)
DelPsfc.edas.DCtoNYC	Surface pressure gradient: Washington DC – New York City
DelPsfc.edas.DCtoBOSTON	Surface pressure gradient: Washington DC to Boston
DelPsfc.edas.DCtoPITTSBURGH	Surface pressure gradient: Washington DC to Pittsburgh
DelPsfc.edas.BUFFALOtoBOSTON	Surface pressure gradient: Buffalo to Boston
DelPsfc.edas.PITTSBURGHtoBUFFALO	Surface pressure gradient: Pittsburgh to Buffalo
H850.DC	850 mb height: Washington DC
H850.BOSTON	850 mb height: Boston
H850.PITTSBURGH	850 mb height: Pittsburgh
H850.BUFFALO	850 mb height: Buffalo
H850.PORTLAND	850 mb height: Portland, ME
H850.NYC	850 mb height: New York City
T850.DC	850 mb temperature: Washington DC
T850.BOSTON	850 mb temperature: Boston
T850.PITTSBURGH	850 mb temperature: Pittsburgh
T850.BUFFALO	850 mb temperature: Buffalo
T850.PORTLAND	850 mb temperature: Portland, ME
T850.NYC	850 mb temperature: New York City
sfcTmax.KLGA	Daily max surface temperature: La Guardia
sfcTmax.KPHL	Daily max surface temperature: Philadelphia
sfcTmax.KBOS	Daily max surface temperature: Boston
sfcTmax.KBUF	Daily max surface temperature: Buffalo
sfcTmax.KALB	Daily max surface temperature: Albany
sfcTmax.KDCA	Daily max surface temperature: Washington DC

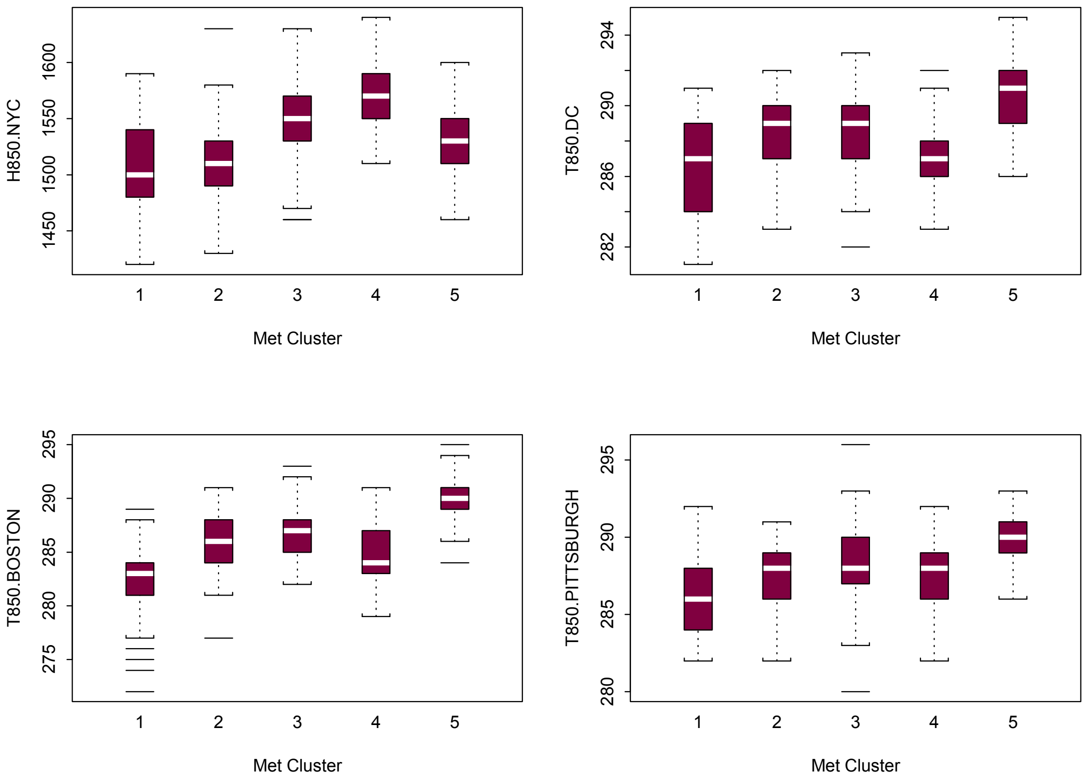

cln1
0.04
0.06
0.08
1
2
3
4
5
Met Cluster
cln2
0.04
0.06
0.08
0.10
1
2
3
4
5
Met Cluster
cln3
0.02
0.04
0.06
0.08
1
2
3
4
5
Met Cluster
cln4
0.04
0.06
0.08
0.10
1
2
3
4
5
Met Cluster



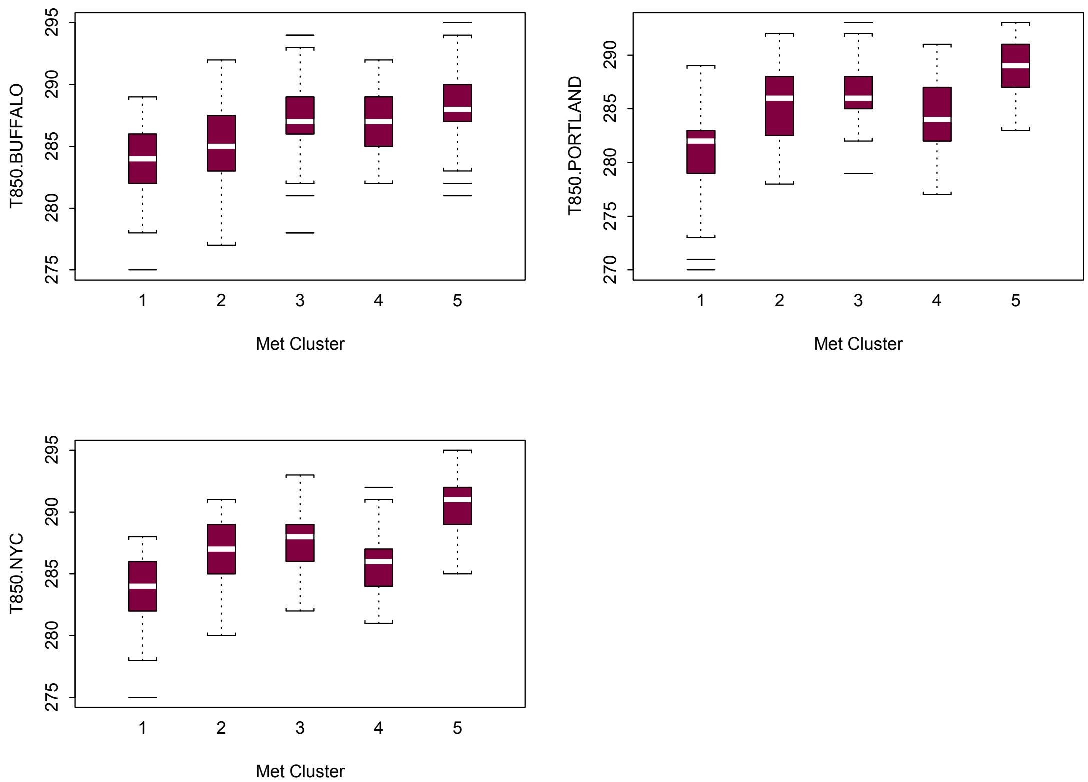
cln5
0.02
0.04
0.06
0.08
0.10
1
2
3
4
5
Met Cluster
cln6
0.02
0.06
0.10
Met Cluster
delPsfc.edas.DCtoNYC
-6
-4
-2
0
2
4
Met Cluster
delPsfc.edas.DCtoBOSTON
-5
0
5
Met Cluster


delPsfc.edas.DCtoPITTSBURGH
Met Cluster
delPsfc.edas.BUFFALOtoBOSTON
Met Cluster
delPsfc.edas.PITTSBURGHtoBUFFALO
Met Cluster
H850.DC
Met Cluster

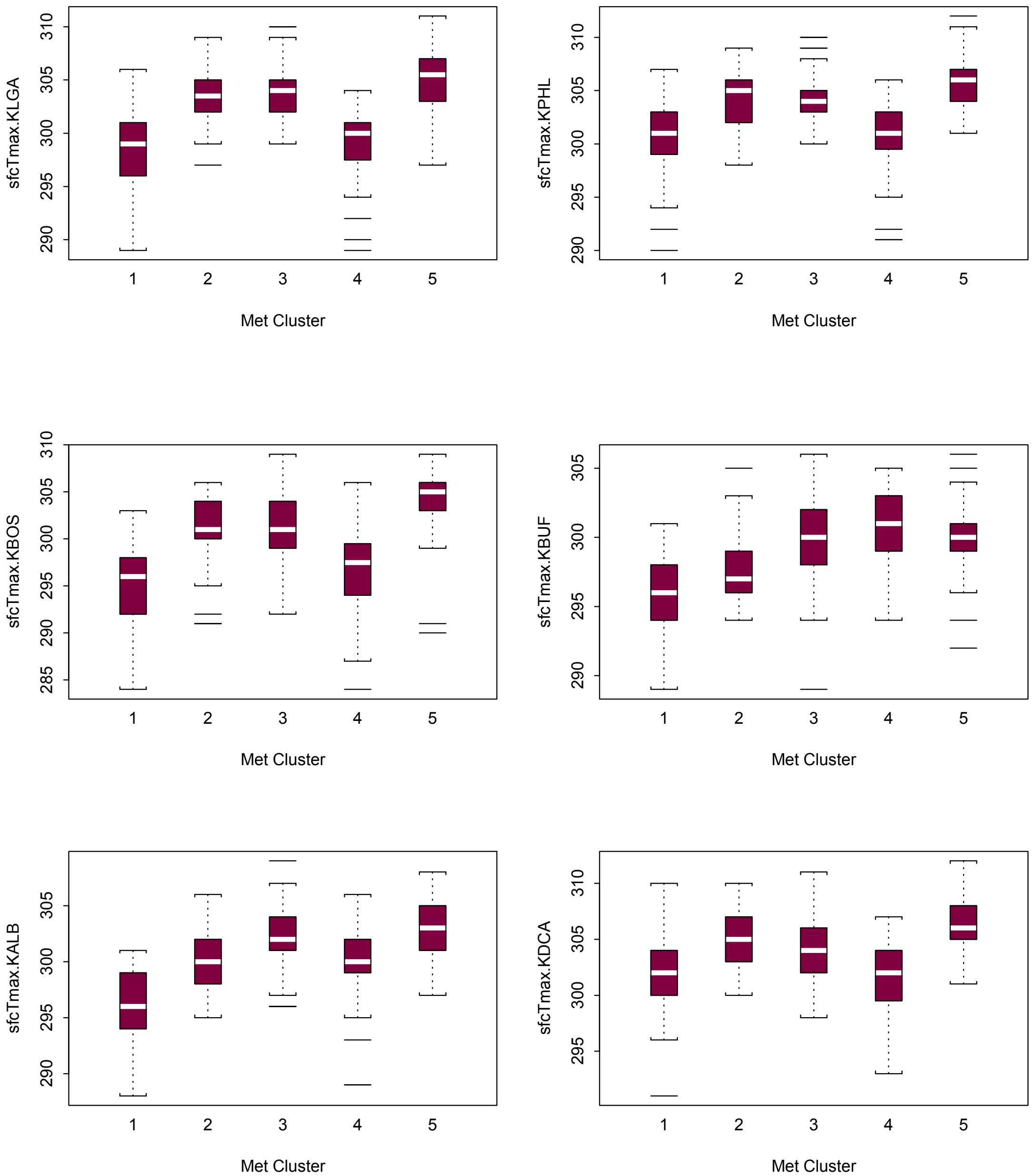

H850.BOSTON
H850.PITTSBURGH
H850.BUFFALO
H850.PORTLAND
Met Cluster
Met Cluster
Met Cluster
Met Cluster

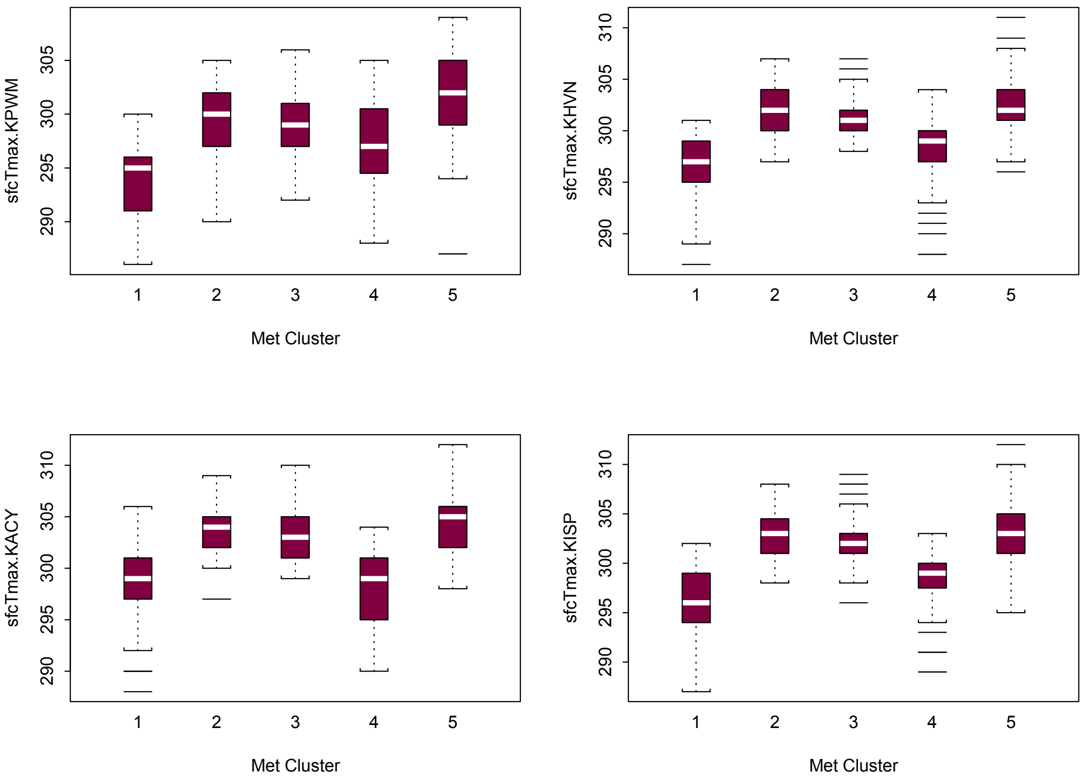


H850.NYC
1450
1500
1550
1600
1
2
3
4
5
Met Cluster
T850.DC
282
286
290
294
1
2
3
4
5
Met Cluster
T850.BOSTON
275
280
285
290
295
1
2
3
4
5
Met Cluster
T850.PITTSBURGH
280
285
290
295
1
2
3
4
5
Met Cluster



T850.BUFFALO
295
290
285
280
275
1
2
3
4
5
Met Cluster
T850.PORTLAND
290
285
280
275
270
1
2
3
4
5
Met Cluster
T850.NYC
295
290
285
280
275
1
2
3
4
5
Met Cluster


sfcTmax.KLGA
290
295
300
305
310
1
2
3
4
5
Met Cluster
sfcTmax.KPHL
290
295
300
305
310
1
2
3
4
5
Met Cluster
sfcTmax.KBOS
285
290
295
300
305
310
1
2
3
4
5
Met Cluster
sfcTmax.KBUF
290
295
300
305
1
2
3
4
5
Met Cluster
sfcTmax.KALB
290
295
300
305
1
2
3
4
5
Met Cluster
sfcTmax.KDCA
295
300
305
310
1
2
3
4
5
Met Cluster



sfcTmax.KPWM
sfcTmax.KHVN
sfcTmax.KACY
sfcTmax.KISP
290 295 300 305 310
1 2 3 4 5
Met Cluster

Appendix C

Wind Direction Frequency Tables

Wind Direction Frequency Tables

Contingency tables showing resultant surface wind direction frequencies were prepared for the five cluster membership cases. These results show relative frequency of days with the indicated wind direction in each cluster, i.e., the values for each cluster (column) sum to 100%. Tabulations are shown for both morning (AM) and afternoon (PM) resultant wind direction. Site location codes referenced in these tables are shown below.

Site Code	Location
KLGA	LaGuardia airport, New York, NY
KPHL	Philadelphia, PA
KBOS	Boston, MA
KBUF	Buffalo, NY
KALB	Albany, NY
KDCA	Washington, DC
KPWM	Portland, ME
KHVN	New Haven, CT
KACY	Atlantic City, NJ
KISP	Islip, Long Island, NY
KHYA	Hayannis, Cape Cod, MA
KWOR	Worcester, MA (KORH)
KHRT	Hartford, CT (KHFD)

Table B-1. Morning and afternoon daily resultant wind direction frequencies (%) by cluster membership for the five cluster case (columns sum to 100%). Header row for each table indicates AM or PM and four letter site ID as described in text (e.g., KLGA = LaGuardia, NY). Cluster identifier (A, B, C, D, E) is shown in first row of each table.

a) Morning wind directions

$resAMwd.KLGA:

	E	B	A	D	C
E	2	2	5	47	3
N	4	4	1	0	0
NE	27	4	7	35	0
NW	7	42	10	0	0
S	33	6	12	5	6
SE	11	0	7	7	2
SW	16	10	34	2	76
W	0	32	23	5	14

$resAMwd.KPHL:

	E	B	A	D	C
E	9	0	2	32	2
N	4	4	1	2	0
NE	7	0	0	34	0
NW	4	28	4	5	0
S	24	4	17	9	29
SE	20	0	4	7	2
SW	26	34	55	9	64
W	7	30	17	2	5

$resAMwd.KBOS:

	E	B	A	D	C
E	2	0	4	9	2
N	2	0	1	0	0
NE	9	4	1	14	0
NW	9	22	13	20	0
S	17	8	11	16	11
SE	15	2	2	7	3
SW	22	26	43	20	77
W	24	38	25	14	8

$resAMwd.KBUF:

	E	B	A	D	C
E	16	2	3	16	0
N	0	0	1	0	0
NE	13	0	0	5	3
NW	0	10	0	0	5
S	31	29	44	16	17
SE	24	0	24	57	6
SW	16	40	25	7	59
W	0	19	2	0	11

$resAMwd.KALB:

	E	B	A	D	C
E	2	0	8	6	0
N	10	7	5	20	2
NE	2	0	1	3	0
NW	10	20	4	3	0
S	34	20	55	49	87
SE	15	0	9	6	3
SW	20	26	12	9	8
W	7	28	5	6	0

$resAMwd.KDCA:

	E	B	A	D	C
E	4	0	0	26	0
N	2	14	2	2	0
NE	7	0	1	28	0
NW	9	24	5	5	0
S	46	24	38	16	68
SE	4	0	5	16	2
SW	22	28	45	2	29
W	7	10	5	5	2

$resAMwd.KPWM:

	E	B	A	D	C
E	10	8	3	2	3
N	2	0	1	7	0
NE	7	2	0	0	2
NW	12	17	14	17	2
S	12	8	11	5	27
SE	2	4	7	2	5
SW	15	8	16	29	50
W	39	52	49	38	12

$resAMwd.KACY:

	E	B	A	D	C
E	7	0	1	26	0
N	2	2	2	10	0
NE	5	0	1	23	0
NW	5	22	7	10	0
S	29	0	17	10	41
SE	12	2	6	8	3
SW	22	29	45	8	52
W	17	45	22	5	5

$resAMwd.KISP:

	E	B	A	D	C
E	9	0	5	14	0
N	9	9	6	19	0
NE	9	0	7	36	0
NW	0	34	10	5	0
S	32	2	8	2	11
SE	14	0	3	14	2
SW	11	21	44	7	80
W	16	34	18	2	8

$resAMwd.KHYA:

	E	B	A	D	C
E	6	0	0	8	0
N	3	13	3	14	0
NE	11	2	1	43	0
NW	6	20	6	0	0
S	14	7	6	11	3
SE	8	4	6	3	0
SW	31	22	41	19	60
W	22	33	38	3	37

$resAMwd.KWOR:

	E	B	A	D	C
E	5	2	0	7	0
N	0	4	3	10	0
NE	10	0	1	15	2
NW	5	23	10	17	0
S	5	2	0	5	2
SE	5	0	2	7	0
SW	33	4	19	12	62
W	38	65	65	27	35

$resAMwd.KHRT:

	E	B	A	D	C
E	5	5	5	19	0
N	5	2	2	6	0
NE	8	2	2	9	0
NW	3	7	0	0	0
S	51	36	71	41	94
SE	5	12	13	16	3
SW	19	19	2	6	3
W	3	17	5	3	0

Table B-1 (concl).
b) Afternoon wind directions

$resPMwd.KLGA:

	E	B	A	D	C
E	9	0	7	20	0
N	0	6	1	0	0
NE	30	0	15	68	0
NW	4	68	26	0	8
S	30	6	12	2	8
SE	13	2	8	7	0
SW	11	6	17	0	48
W	2	12	14	2	36

$resPMwd.KPHL:

	E	B	A	D	C
E	13	2	4	57	0
N	0	8	2	0	0
NE	4	0	0	32	0
NW	2	40	5	0	2
S	22	0	20	2	8
SE	26	4	3	9	2
SW	30	22	50	0	73
W	2	24	16	0	17

$resPMwd.KBOS:

	E	B	A	D	C
E	20	2	13	34	5
NE	9	2	2	9	0
NW	4	30	2	0	0
S	17	4	23	18	8
SE	15	2	20	25	0
SW	26	10	24	7	65
W	9	50	16	7	23

$resPMwd.KBUF:

	E	B	A	D	C
E	7	2	3	19	0
N	0	2	0	0	0
NE	11	4	1	5	5
NW	2	6	6	0	8
S	28	12	20	33	3
SE	11	6	2	16	0
SW	37	51	61	28	68
W	4	16	7	0	17

$resPMwd.KALB:

	E	B	A	D	C
E	2	0	3	8	0
N	2	2	2	11	0
NE	4	2	2	3	0
NW	7	41	6	0	2
S	59	2	45	50	58
SE	15	2	5	18	2
SW	7	20	17	8	24
W	4	31	20	3	15

$resPMwd.KDCA:

	E	B	A	D	C
E	7	4	5	23	0
N	0	4	2	2	0
NE	9	6	2	48	0
NW	2	34	4	0	5
S	57	14	53	9	59
SE	11	6	9	11	3
SW	11	6	17	2	21
W	4	26	8	5	12

$resPMwd.KPWM:

	E	B	A	D	C
E	11	8	17	20	5
N	0	4	2	0	0
NE	13	2	2	7	2
NW	11	42	11	0	2
S	20	6	24	27	29
SE	20	4	16	25	6
SW	16	14	15	14	33
W	9	20	13	7	24

$resPMwd.KACY:

	E	B	A	D	C
E	15	2	3	42	0
N	0	6	1	2	0
NE	7	4	1	47	0
NW	2	48	11	0	0
S	33	4	21	0	23
SE	24	0	10	9	0
SW	13	8	26	0	55
W	7	28	27	0	22

$resPMwd.KISP:

	E	B	A	D	C
E	11	0	6	36	0
N	2	10	5	2	0
NE	11	0	3	36	0
NW	2	52	11	2	3
S	38	0	17	2	14
SE	20	0	8	18	0
SW	11	8	39	2	68
W	4	29	13	0	15

$resPMwd.KHYA:

	E	B	A	D	C
E	16	11	7	9	0
N	0	9	2	7	0
NE	9	4	5	27	0
NW	0	11	6	0	0
S	13	9	12	20	5
SE	18	6	8	20	0
SW	38	21	48	16	75
W	7	30	11	0	20

$resPMwd.KWOR:

	E	B	A	D	C
E	14	4	3	17	2
N	5	4	1	2	0
NE	14	4	4	22	0
NW	9	35	9	2	0
S	9	0	7	22	5
SE	5	0	3	10	0
SW	37	4	28	7	54
W	7	49	44	17	40

$resPMwd.KHRT:

	E	B	A	D	C
E	11	2	16	16	0
N	4	9	3	3	0
NE	0	4	7	22	0
NW	2	22	1	0	0
S	56	13	43	22	52
SE	13	11	18	24	2
SW	11	11	9	11	35
W	2	28	4	3	11

Appendix D

Ozone Monitoring Stations by Sub-Region

Cluster	Site ID	State	City	Location
6	90010017	Connecticut	GREENWICH	GREENWICH POINT PARK
1	90011123	Connecticut	DANBURY	TRAILER, W. CONNECTICUT STATE UNIVERSITY
6	90013007	Connecticut	STRATFORD	USCG LIGHTHOUSE, PROSPECT STREET
6	90019003	Connecticut	WESTPORT	SHERWOOD ISLAND STATE PARK
1	90031003	Connecticut	EAST HARTFORD	MCAULIFFEE PARK
1	90070007	Connecticut	MIDDLETOWN	CONN. VALLEY HOSP., SHEW HALL, EASTERN D
6	90093002	Connecticut	MADISON	HAMMONASSET STATE PARK
6	90110008	Connecticut	GROTON	UNIVERSITY OF CONNECTICUT, AVERY POINT
1	90131001	Connecticut	STAFFORD	ROUTE 190, SHENIPSIT STATE FOREST
2	100010002	Delaware	NOT IN A CITY	STATE ROAD 384
5	100031003	Delaware	NOT IN A CITY	RIVER ROAD PARK, BELLEFONTE
5	100031007	Delaware	NOT IN A CITY	LUMS POND STATE PARK
5	100031010	Delaware	NOT IN A CITY	BRANDYWINE CREEK STATE PARK
2	100051002	Delaware	SEAFORD	350 VIRGINIA AVE SEAFORD
2	100051003	Delaware	LEWES	UNIV. OF DE COLLEGE OF MARINE STUDIES
5	110010025	Washington DC	NOT IN A CITY	TAKOMA SC. PINEY BRANCH RD & DAHLIA ST N
5	110010041	Washington DC	NOT IN A CITY	34TH. AND DIX STREETS, N.E.
5	110010043	Washington DC	NOT IN A CITY	S.E. END MCMILLIAN RESERVOIR, WASH. DC.
1	230052003	Maine	CAPE ELIZABETH	TWO LIGHTS STATE PARK
1	230090102	Maine	BAR HARBOR	TOP OF CADILLAC MOUNTAIN
1	230090103	Maine	BAR HARBOR	MCFARLAND HILL-DISPRO SITE
3	230112005	Maine	GARDINER	PRAY STREET SCHOOL
1	230130004	Maine	NOT IN A CITY	PORT CLYDE, MARSHALL POINT LIGHTHOUSE
3	230173001	Maine	NOT IN A CITY	ROUTE 5, NORTH LOVELL DOT
3	230194008	Maine	NOT IN A CITY	SUMMIT OF RIDER BLUFF (WLBZ TRANSMITTER)
1	230312002	Maine	NOT IN A CITY	OCEAN AVE/PARSONS WAY, KENNEBUNKPORT
1	230313002	Maine	KITTERY	FRISBEE SCHOOL, GOODSOE ROAD
2	240030014	Maryland	NOT IN A CITY	QUEEN ANNE AND WAYSON ROADS
5	240030019	Maryland	FORT MEADE	9001 'Y'STREET, FT. MEADE, ANNE ARUNDEL MD
5	240051007	Maryland	COCKEYSVILLE	GREENSIDE DRIVE, COCKEYSVILLE MD
5	240053001	Maryland	ESSEX	WOODWARD & DORSEY RDS, ESSEX MD
5	240130001	Maryland	NOT IN A CITY	1300 W. OLD LIBERTY ROAD, WINFIELD, MD
5	240150003	Maryland	NOT IN A CITY	RTE.273, FAIR HILL, CEIL CO., MARYLAND
2	240170010	Maryland	NOT IN A CITY	SO MD CORRECTIONAL CAMP, HUGHESVILLE MD
5	240251001	Maryland	EDGEWOOD	EDGEWOOD ARMY CHEM CENTER EDGEWOOD MD
5	240259001	Maryland	NOT IN A CITY	3538 ALDINO ROAD, HARFORD COUNTY MARYLAND
5	240290002	Maryland	NOT IN A CITY	KENT COUNTY; MILLINGTON
5	240313001	Maryland	ROCKVILLE	LOTHROP E SMITH ENV.ED CENTER ROCKVILLE
5	240330002	Maryland	GREENBELT	GODDARD SPACE FLIGHT CENTER
6	250010002	Massachusetts	TRURO	FOX BOTTOM AREA-CAPE COD NAT'L SEASHORE
1	250034002	Massachusetts	ADAMS	MT. GREYLOCK SUMMIT
6	250051002	Massachusetts	FAIRHAVEN	LEROY WOOD SCHOOL
1	250051005	Massachusetts	EASTON	1 BORDERLAND ST.
1	250092006	Massachusetts	LYNN	390 PARKLAND AVE. (LYNN WATER TREATMENT)
1	250094004	Massachusetts	NEWBURY	SUNSET BOULEVARD
1	250130003	Massachusetts	AGAWAM	152 SOUTH WESTFIELD STREET, FEEDING HILL

Cluster	Site ID	State	City	Location
1	250130008	Massachusetts	CHICOPEE	ANDERSON ROAD AIR FORCE BASE
1	250150103	Massachusetts	AMHERST	NORTH PLEASANT ST. U. MASS PATHOLOGY DEPT
1	250154002	Massachusetts	WARE	QUABBIN SUMMIT
1	250250042	Massachusetts	BOSTON	HARRISON AVENUE
1	250270015	Massachusetts	WORCESTER	WORCESTER AIRPORT
1	330050007	New Hampshire	KEENE	RAILROAD STREET
3	330090008	New Hampshire	HAVERHILL	HAVERHILL ARMORY, RT 116, HAVERHILL, NH
1	330111010	New Hampshire	NASHUA	SANDERS ASSOCIATES, PARKING LOT D
1	330130007	New Hampshire	CONCORD	STORRS STREET
1	330150012	New Hampshire	RYE	RYE HARBOR STATE PARK OCEAN BLVD, RTE. 1A
3	330173002	New Hampshire	ROCHESTER	ROCHESTER HILL ROAD, ROCHESTER
3	330190003	New Hampshire	CLAREMONT	SOUTH STREET
5	340010005	New Jersey	NOT IN A CITY	BRIGANTINE WILDLIFE REFUGE, NACOTE CREEK
5	340070003	New Jersey	NOT IN A CITY	COPEWOOD E. DAVIS STS; TRAILER
5	340071001	New Jersey	NOT IN A CITY	ANCORA STATE HOSPITAL, ANCORA
5	340110007	New Jersey	NOT IN A CITY	LINCOLN AVE.&HIGHWAY 55,NE OF MILLVILLE
5	340150002	New Jersey	NOT IN A CITY	CLARKSBORO, SHADY LANE REST HOME
5	340170006	New Jersey	BAYONNE	VETERANS PARK ON NEWARK BAY
5	340190001	New Jersey	FLEMINGTON	RARITAN STP, RTE.613S, THREE BRIDGES
5	340210005	New Jersey	NOT IN A CITY	RIDER COLLEGE; LAWRENCE TOWNSHIP
5	340230011	New Jersey	NOT IN A CITY	R.U. VEG RESEARCH FARM 3,RYDERS LN, NEWB
5	340250005	New Jersey	WEST LONG BRANC	MONMOUTH COLLEGE, WEST LONG BRANCH
5	340273001	New Jersey	NOT IN A CITY	BLDG.#1, BELL LABS, OFF ROUTE 513
5	340290006	New Jersey	NOT IN A CITY	COLLIERS MILLS, JACKSON TOWNSHIP
1	360010012	New York	ALBANY	LOUDONVILLE RESERVOIR
6	360050083	New York	NEW YORK CITY	200TH STREET AND SOUTHERN BLVD
6	360050110	New York	NEW YORK CITY	E 156TH ST BET DAWSON AND KELLY
4	360130006	New York	DUNKIRK	STP LAKESIDE BLD DUNKIRK
4	360130011	New York	NOT IN A CITY	TOWN OF WESTFIELD
4	360150003	New York	ELMIRA	SULLIVAN ST., WATER TR. PL.
1	360270007	New York	NOT IN A CITY	VILLAGE OF MILLBROOK
4	360290002	New York	AMHERST	AUDUBON GOLF COURSE, MAPLE ROAD
3	360310002	New York	NOT IN A CITY	SUMMIT, WHITEFACE MTN, WEATHER STATION
3	360310003	New York	NOT IN A CITY	BASE WHITEFACE MTN, ASRC, SUNY
3	360410005	New York	NOT IN A CITY	PISECO LAKE AIRPORT
3	360430005	New York	NOT IN A CITY	NICKS LAKE CAMPGROUND
4	360450002	New York	NOT IN A CITY	VADAI ROAD, PERCH RIVER, BROWNVILLE
4	360530006	New York	NOT IN A CITY	TOWN OF GEORGETOWN
4	360551004	New York	NOT IN A CITY	TRAILER, WEST END OF FARMINGTON ROAD
4	360631006	New York	NOT IN A CITY	MIDDLEPORT STP, NORTH HARTLAND RD
4	360671015	New York	NOT IN A CITY	5895 ENTERPRISE PARKWAY,
1	360715001	New York	NOT IN A CITY	1175 ROUTE 17K, MONTGOMERY
1	360790005	New York	NOT IN A CITY	NYSDEC FIELD HQTRS GYPSY TRAIL ROAD
6	360810098	New York	NEW YORK CITY	120-07 15TH AVE
5	360850067	New York	NEW YORK CITY	SUSAN WAGNER HS, BRIELLE AVE.& MANOR RD,
3	360910004	New York	NOT IN A CITY	SARATOGA NATIONAL HISTORICAL PARK
1	360930003	New York	SCHENECTADY	MT.PLEASANT HS, NORWOOD AVE.& FOREST RD.

Cluster	Site ID	State	City	Location
6	361030002	New York	BABYLON	EAST FARMINGDALE WATER DIST., GAZZA BLVD.
6	361030004	New York	RIVERHEAD	39 SOUND AVENUE, RIVERHEAD
3	361111005	New York	NOT IN A CITY	BELLEAYRE MOUNTAIN
4	361173001	New York	NOT IN A CITY	WAYNE EDUCATIONAL CENTER, WILLIAMSON
6	361192004	New York	WHITE PLAINS	WHITE PLAINS PUMP STATION, ORCHARD STREET
4	420030008	Pennsylvania	PITTSBURGH	BAPC 301 39TH STREET BLDG #7
4	420030067	Pennsylvania	NOT IN A CITY	OLD OAKDALE ROAD SOUTH FAYETTE
4	420031005	Pennsylvania	NOT IN A CITY	CALIFORNIA & 11TH, HARRISON TWP
4	420050001	Pennsylvania	KITTANNING	GLADE DR. & NOLTE RD. KITTANNING
4	420070002	Pennsylvania	NOT IN A CITY	ROUTE 168 & TOMLINSON ROAD
4	420070005	Pennsylvania	NOT IN A CITY	1015 SEBRING ROAD
4	420070014	Pennsylvania	BEAVER FALLS	EIGHT STREET AND RIVER ALLEY
5	420110001	Pennsylvania	KUTZTOWN	KUTZTOWN UNIVERSITY - GRIM SCIENCE BLDG
5	420110009	Pennsylvania	READING	UGI CO MONGANTOWN RD AND PROSPECT ST
5	420130801	Pennsylvania	ALTOONA	2ND AVE & 7TH ST
5	420170012	Pennsylvania	BRISTOL (BOROUG	ROCKVIEW LANE
5	420210011	Pennsylvania	NOT IN A CITY	MILLER AUTO SHOP 1 MESSENGER ST
5	420430401	Pennsylvania	HARRISBURG	1833 UPS DRIVE HARRISBURG PA
5	420431100	Pennsylvania	HERSHEY	SIPE AVE & MAE STREET
5	420450002	Pennsylvania	CHESTER	FRONT ST & NORRIS ST
4	420490003	Pennsylvania	NOT IN A CITY	10TH AND MARNE STREETS
5	420550001	Pennsylvania	NOT IN A CITY	FOREST ROAD - METHODIST HILL
1	420690101	Pennsylvania	NOT IN A CITY	WILSON FIRE CO. ERIE & PLEASANT
1	420692006	Pennsylvania	SCRANTON	GEORGE ST TROOP AND CITY OF SCRANTON
5	420710007	Pennsylvania	LANCASTER CITY	ABRAHAM LINCOLN JR HIGH GROFFTOWN RD
4	420730015	Pennsylvania	NEW CASTLE	CROTON ST & JEFFERSON ST.
5	420770004	Pennsylvania	ALLENTOWN	STATE HOSPITAL REAR 1600 HANOVER AVE
1	420791100	Pennsylvania	NANTICOKE	255 LOWER BROADWAY (NEXT TO LEON&EDDY'S)
1	420791101	Pennsylvania	WILKES-BARRE	CHILWICK & WASHINGTON STS
4	420850100	Pennsylvania	NOT IN A CITY	PA518 (NEW CASTLE ROAD) & PA418
5	420910013	Pennsylvania	NORRISTOWN	STATE ARMORY - 1046 BELVOIR RD
5	420950025	Pennsylvania	NOT IN A CITY	WASHINGTON & CAMBRIA STS. FREEMANSBURG
5	420990301	Pennsylvania	NOT IN A CITY	ROUTE 34 LITTLE BUFFALO STATE PARK
5	421010004	Pennsylvania	PHILADELPHIA	1501 E LYCOMING AVE AMS LAB
5	421010014	Pennsylvania	PHILADELPHIA	ROXY WATER PUMP STA EVA-DEARNLEY STS
5	421010024	Pennsylvania	PHILADELPHIA	GRANT-ASHTON ROADS PHILA NE AIRPORT
5	421010136	Pennsylvania	PHILADELPHIA	AMTRAK, 5917 ELMWOOD AVENUE
4	421250005	Pennsylvania	CHARLEROI	CHARLER01 WASTE TREATMENT PLANT
4	421250200	Pennsylvania	WASHINGTON	MCCARRELL AND FAYETTE STS
4	421255001	Pennsylvania	NOT IN A CITY	HILLMAN STATE PARK - KINGS CREEK ROAD
4	421290006	Pennsylvania	NOT IN A CITY	OLD WILLIAM PENN HWY & SARDIS AVE
5	421330008	Pennsylvania	YORK	HILL ST.
1	440030002	Rhode Island	NOT IN A CITY	W. ALTON JONES CAMPUS URI PARKERFIELD WE
1	440071010	Rhode Island	EAST PROVIDENCE	FRANCIS SCHOOL, 64 BOURNE AVE
6	440090007	Rhode Island	NARRAGANSETT	TARWELL ROAD, NARRAGANSETT
3	500030004	Vermont	BENNINGTON	AIRPORT RD, BENNINGTON, VERMONT
3	500070007	Vermont	UNDERHILL	PROCTOR MAPLE RESEARCH FARM

Cluster	Site ID	State	City	Location
5	510130020	Virginia	NOT IN A CITY	S 18TH AND HAYES ST
2	510360002	Virginia	NOT IN A CITY	SHIRLEY PLANTATION, ROUTE 5
2	510410004	Virginia	NOT IN A CITY	BEACH, INTERSECTION OF CO.ROADS 655 & 654
5	510590005	Virginia	NOT IN A CITY	CUBRUN LEE RD CHANT, (CUBRUN TREAT PLANT)
5	510590018	Virginia	NOT IN A CITY	MT.VERNON 2675 SHERWOOD HALL LANE
5	510591004	Virginia	SEVEN CORNERS	6100 ARLINGTON BLVD MONTG WARD
5	510595001	Virginia	MC LEAN	LEWINSVILLE 1437 BALLS HILL RD
5	510610002	Virginia	NOT IN A CITY	RT651 C PHELPS WILDLIFE MANAGEMENT AREA
5	510690010	Virginia	NOT IN A CITY	RTE 669, BUTLER MANUF. CO NEAR REST VA
2	510870014	Virginia	NOT IN A CITY	2401 HARTMAN STREET MATH & SCIENCE CTR
2	511130003	Virginia	NOT IN A CITY	SHENANDOAH NP BIG MEADOWS
5	511530009	Virginia	NOT IN A CITY	JAMES S. LONG PARK
2	511611004	Virginia	VINTON	EAST VINTON ELEMENTARY SCHOOL
5	511790001	Virginia	NOT IN A CITY	WIDEWATER ELEM. SCH., DEN RICH ROAD
2	511970002	Virginia	NOT IN A CITY	16-B RURAL RETREAT SEWAGE DISPOSAL
5	515100009	Virginia	ALEXANDRIA	517 N SAINT ASAPH ST, ALEXANDRIA HEALTH
2	518000004	Virginia	SUFFOLK	TIDEWATER COMM. COLLEGE, FREDERIC CAMPUS
2	518000005	Virginia	SUFFOLK	TIDEWATER RESEARCH STATION, HARE ROAD

Appendix E

Episode Types Associated with 8-Hour Ozone Exceedance Days in 2002

Episode Pattern	Year	Month	Day
5	2002	5	16
3	2002	5	24
1	2002	5	25
5	2002	6	1
5	2002	6	5
5	2002	6	6
5	2002	6	9
4	2002	6	10
5	2002	6	11
1	2002	6	12
4	2002	6	20
5	2002	6	21
3	2002	6	22
3	2002	6	23
2	2002	6	24
4	2002	6	25
5	2002	6	26
5	2002	6	27
5	2002	6	29
3	2002	6	30
5	2002	7	1
3	2002	7	2
3	2002	7	3
3	2002	7	4
2	2002	7	5
3	2002	7	7
3	2002	7	8
5	2002	7	9
4	2002	7	12
3	2002	7	13
3	2002	7	14
5	2002	7	15
2	2002	7	16
3	2002	7	17
2	2002	7	18
3	2002	7	19
4	2002	7	20
4	2002	7	21
5	2002	7	22
5	2002	7	23
1	2002	7	27
1	2002	7	28
2	2002	7	29
2	2002	7	30
2	2002	7	31

Episode Pattern	Year	Month	Day
3	2002	8	1
4	2002	8	2
3	2002	8	3
3	2002	8	4
5	2002	8	5
4	2002	8	9
3	2002	8	10
3	2002	8	11
3	2002	8	12
3	2002	8	13
3	2002	8	14
5	2002	8	15
3	2002	8	16
3	2002	8	17
3	2002	8	18
1	2002	8	19
4	2002	8	21
5	2002	8	22
2	2002	8	23
3	2002	9	7
4	2002	9	8
4	2002	9	9
5	2002	9	10
5	2002	9	13
1	2002	9	14
4	2002	9	18