Use of Cluster Analysis to Define Periods of Similar Meteorology and Precipitation Chemistry in Eastern North America. Part II: Precipitation Patterns and Pollutant Deposition*

MARK E. FERNAU AND PERRY J. SAMSON

Dept. of Atmospheric, Oceanic and Space Sciences, Space Physics Research Laboratory, The University of Michigan, Ann Arbor, Michigan

(Manuscript received 22 August 1989, in final form 7 February 1990)

ABSTRACT

The precipitation chemistry associated with the flow patterns of transport-derived clusters has been examined. Cluster analysis was applied to transport vectors, derived from three years of daily trajectories arriving at monitoring sites, in order to define a synoptic climatology of representative three-day periods of air mass movement. The resulting clusters were successful in defining wet, dry, polluted and nonpolluted clusters, as shown by the spatial patterns of median deposition and by statistical testing. The highest pollutant depositions over the widest areas resulted from mean transport patterns with large areas of slow air mass movement over the regions of high sulfur emissions and which were frequently persistent over several periods or followed persistent clusters. There was a large amount of overlap among the chemistry distributions and large variation within most of the clusters. Seasonal differences exist within each cluster, with sulfur deposition within a given cluster generally being higher in the warmer months. Cluster analysis was shown to be useful in the computer-assisted classification of spatial patterns of weather and pollution data.

1. Introduction

In this paper the spatial precipitation patterns and wet deposition chemistry associated with transport-derived clusters are examined to determine whether representative meteorological events conducive to high acid deposition can be identified. These representative events, if found, could be used in aggregation schemes to extend the results of the episode-based Regional Acid Deposition Model (RADM) (Chang et al. 1987) to determine seasonal or annual deposition loads in eastern North America. The cost of directly simulating seasonal or annual deposition patterns through repeated application of RADM would be high due to the complexity of the model and its demand on computer resources. The process of creation of categories of atmospheric circulation type and assessment of weather and pollution elements in relation to those categories has been called synoptic climatology (Barry and Perry 1973). Cluster analysis has been little used in synoptic climatological and air pollution applications to date.

Crutcher et al. (1986) discussed the application of cluster analysis to aerometric data and its use in conjunction with other multivariate methods, applying three different clustering programs to several ambient pollution datasets and comparing the results. Cluster analysis by variable has been used to group precipitation chemistry concentration variables (Gorham et al. 1984), particulate matter (Gaarenstroom et al. 1977), and aerosol data (Saucy et al. 1987). Slanina et al. (1983) clustered chemical components by sampling period and examined the relative importance of each cluster to acid deposition. Sanchez Gomez and Ramos Martin (1987) clustered urban particulate data and isolated clusters representing high and low amounts of pollutants as well as different source regions.

Kalkstein and Corrigan (1986), using a methodology combining principal component analysis and cluster analysis of weather variables to characterize air masses at a given locale, "were able to identify air mass categories exhibiting particularly high sulfur dioxide concentrations and synoptic scenarios of long duration contributing to severe concentration levels." Kalkstein et al. (1987) refined the method and demonstrated it for two other sites. Moody and Samson (1989) used cluster analysis of trajectory data to stratify meteorologically similar events at a given site and examined the precipitation chemistry associated with each cluster. The concentrations of sulfate, nitrate, hydrogen and ammonium in precipitation were, for the most part, significantly different among clusters at all three sites

* This work was not funded through Argonne National Laboratory.

Corresponding author address: Dr. Mark E. Fernau, Argonne National Laboratory, Environmental Assessment and Information Sciences Division, 9700 South Cass Avenue, Argonne, IL 60439.

studied. There was still a large amount of variability within clusters and it was difficult to isolate clusters leading to high pollutant concentrations. This study extends the earlier work to a larger spatial domain and examines the extent to which synoptic wind flow categories can explain variations in precipitation chemistry.

2. Data and methods

a. Precipitation chemistry

The chemistry data used in this work were restricted to samples from two networks, both of which sample precipitation on a daily or event basis and subsequently analyze the major chemical constituents contained therein. The chemistry measurements come mainly from the Utility Acid Precipitation Study Program (UAPSP), using a subset covering the period from January 1979 to December 1983. The UAPSP dataset is of very high quality and reliability due to rigorous quality assurance programs at the sites and analysis laboratory and efforts to assure consistency among sites (Mueller et al. 1988). The locations and identifications of the UAPSP sites are given in Fig. 1 and Table 1. Data elements used in this study include start and end time of precipitation, precipitation amount, field and laboratory pHs, and the ionic concentrations of sulfate, nitrate, and ammonium in precipitation. Samples with errors in the dates and times of precipitation begin/end and sample removal date/time were corrected or deleted. Midpoints and durations were calculated from the precipitation begin/end information. Hydrogen ion concentration was calculated using the lab pH or, if missing, the field pH according to the formula H^+ (μmoles L^{-1}) = $10^{(6-\text{pH})}$. The sample elements were tested and processed for quality in a manner similar to that described by Endlich et al. (1988).

TABLE 1. Precipitation chemistry sampling sites used in this study.

Site	Location
1	Montague/Turners Falls, MA
2	Scranton/Tunkhannock, PA
3	Indian River, DE
4	Zanesville, OH
5	Rockport, IN
6	Giles County, TN
7	Fort Wayne, IN
8	Raleigh, NC
9	Lewisburg, WV
10	Gaylord, MI
11	Clearfield, KY
12	Alamo, TN
13	Winterport, ME
14	Uvalda, GA
15	Selma, AL
16	Clinton, MS
17	Marshall, TX
18	Lancaster, KS
19	Brookings, SD
20	Underhill, VT
21	Big Moose Lake, NY
22	McArthur, OH
27	Dorset, ONT
28	Longwoods, ONT

The UAPSP data are limited to the United States. Since southwest Ontario and the Sudbury area also have large annual sulfur wet deposition (Finkelstein and Seilkop 1986), two Canadian sites were added to the database. The sites, Dorset and Longwoods, come from the Acid Precipitation in Ontario Study (APIOS) network and are also daily/event samples. The vari-

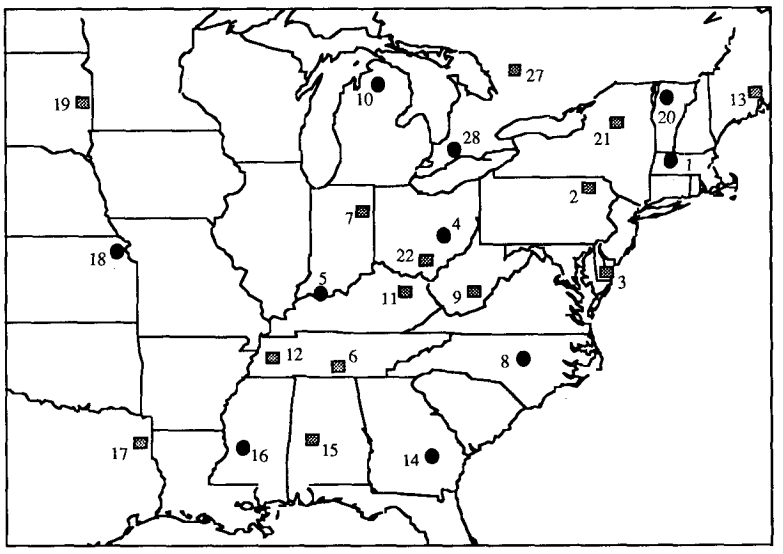

FIG. 1. Locations of UAPSP and APIOS precipitation chemistry sampling sites. Sites with circles are the subset used in the clustering described in the text.

ables and analytes retained are identical to those of the UAPSP data.

Each sample was assigned to the day on which the midpoint of the precipitation duration fell. If the hydrogen ion concentration calculated from laboratory pH was missing the field hydrogen ion concentration was used. Events whose midpoints could not be determined were not included in the dataset. In order to preserve the daily nature of the events, samples with durations longer than 48 hours were not used. The ion concentrations and precipitation amounts for these events were set to missing. The above two criteria generally led to exclusion of less than five events per site, respectively, and more than half of the sites never had a missing midpoint. If multiple events were assigned to the same day they were combined into one event by summing the precipitation deposition and taking either a weighted or arithmetic average of nonmissing ion concentrations, depending on the availability of precipitation amount. For each site, days for which no event (midpoint) was recorded were given precipitation and ion values of zero. Days for which a site was not in operation were assigned missing values for precipitation amount and ion concentrations. At the conclusion of this process all 24 sites had data records for the period from January 1979 to December 1983.

Finally, an event was redefined as consisting of three days rather than one. Storms, on average, traverse the study area in around three days and RADM results are also often presented in terms of three-day cases. Moving three-day totals for precipitation, moving three-day weighted (by precipitation amount) averages for ion concentration, and moving three-day total analyte depositions (precipitation times concentration) were calculated at each site. The use of moving periods, rather than consecutive periods, was felt to be less arbitrary. The events were tagged by the middle day of the three days. If any of the three days constituting an event had a missing value for a variable, that variable was set to missing for the three-day event in order to avoid creation of misleading data, i.e. a wet event appearing dry or a high precipitation event appearing to be lower. The chemical analytes used in this paper will be described in terms of medians and percentiles and statistical testing will utilize nonparametric tests to account for the nonnormal distributions of the precipitation chemistry data. As explained below, a three-year subset of the chemistry data was used in the analyses presented here. Table 2 lists the sample numbers available for sulfate deposition (combined dry periods and actual UAPSP samples) over those three years for each of the clusters described below. The numbers for precipitation will generally be slightly larger. When looking at the analyses presented herein it may be useful to refer to this table to judge the robustness of the percentiles for any given site.

TABLE 2. Number of three-day samples (including dry periods) by cluster at step 7 for sulfate deposition, by site. Site numbers refer to Table 1. *M*: missing data cluster.

	Cluster							
Site	1	2	3	4	5	6	7	*M*
1	54	84	195	87	101	73	173	283
2	32	45	125	48	69	34	103	235
3	14	21	50	19	17	12	37	165
4	47	71	183	85	104	68	167	266
5	55	84	197	91	104	71	174	287
6	15	20	48	18	13	12	33	156
7	42	72	159	79	96	62	160	268
8	52	75	184	81	95	67	158	255
9	13	18	54	19	17	13	41	160
10	17	24	69	26	48	27	65	77
11	21	30	80	35	53	28	74	79
12	22	28	77	32	58	31	69	89
13	22	26	80	37	55	29	71	85
14	24	32	83	38	55	33	73	86
15	23	32	85	38	53	30	75	87
16	21	29	85	37	54	31	74	90
17	22	31	83	38	55	31	75	84
18	20	26	81	33	49	29	72	82
19	22	28	82	33	54	30	68	78
20	20	27	83	39	61	31	69	89
21	14	25	68	35	53	29	64	82
22	23	24	82	34	60	33	71	88
27	31	50	123	57	80	34	110	89
28	34	42	127	56	84	37	110	96

b. Derivation of clusters

The groups described in this work were obtained by clustering transport vectors, derived from daily backwards trajectories arriving at the chemistry monitoring sites. The trajectories represent the mean flow in the mixed layer as interpolated from rawinsonde stations. The latitudes and longitudes of the trajectory intervals were converted to distance vectors from the receptor origin to each interval end point, i.e. the x and y components of the straight line extending from the origin to the end point of the interval were calculated, ignoring any curvature of the trajectory occurring enroute. The distance vectors for a specific upwind interval for each site were merged into one file, arranged in chronological order. The data were arranged in moving three-day period formats, six variables per record (y and x components for days $n-1$, n, and $n+1$), for subsequent cluster analysis studies. The three-day periods allow compatibility with the duration customarily used for RADM simulations.

For the clusters described in this paper a subset of ten of the sampling sites, chosen so as to maintain as well as possible the original spatial domain (Fig. 1), was selected and three years of data were used. 1979, 1981 and 1983 (a total of 1093 three-day moving periods) were selected because most existing RADM simulations are contained in these years. A case or record consisted of six variables for each site: the x and

y components of the distance vector originating twelve hours upwind and arriving at the site at 1800 UTC for each of the three days comprising an event—a total of 60 variables per event. Twenty-seven percent of the three-day records contained missing data and were excluded from the cluster analysis. The clustering was done using Ward's method and Euclidean distance. The details of the back trajectory calculations, data manipulations, cluster analysis and a full description of the meteorological characteristics of the resulting clusters can be found in Fernau and Samson (1990).

3. Results

When discussing clusters, step numbers are defined by the number of clusters existing at that point. The number labels assigned to the clusters are arbitrary. Using the total within-group error sum of squares at each step number, likely breakpoints for terminating the clustering operation were selected. Based on the two factors of sharp changes in slope and retention of a workable number of clusters, steps 7 and 18 were chosen for further examination.

a. Analysis of step 7

1) SUMMARY OF TRANSPORT PATTERNS AT STEP 7

Each of the step 7 clusters can be characterized by plots of the mean transport vectors at all sites for each of the three days, showing the evolution of the trajectory pattern with time. All sites were plotted although the clusters were determined only by ten sites. The resulting plots are presented in Fernau and Samson (1990). Presented here is a summary of those plots and the characteristics of the clusters.

Clusters 1 and 2 both depict the movement of a trough, extending from Canada to the southern United States, eastward across the eastern United States. The trough is followed by northwest flow and the movement of an anticyclone into the southeastern states. Cluster 3 begins with an anticyclone in the southeast, which then slowly drifts off the eastern coast and is replaced by southwest flow over much of the eastern United States. There is a hint of a trough moving into the Minnesota area. Cluster 4 depicts the eastward movement of the system described for cluster 3, with the high pressure now off the coast and the southwest flow being replaced by the eastward moving trough. Cluster 5 is distinct from the others and shows an anticyclone drifting eastward across the Great Lakes. Flow in the vicinity of the high pressure is weak. Cluster 6 contains strong northwest flow over most of the eastern United States behind a trough which has moved out to sea. The flow pattern for cluster 7 remains very similar over the three-day period with westerly winds diminishing and becoming stagnant and ill-defined by the third day. There is a weak ridge present in the Gulf states. The pattern representing the flow at the nonmissing sites when one or more sites had missing data does not change from day 1 to day 3 and seems to represent the average westerly flow found during the course of a year.

The two most common clusters are 3 and 7. Cluster 5 is next most common, with the cyclonic clusters occurring less frequently. Clusters 1 and 2 rarely occur in the warmer months. Clusters 4 and 6 also are less likely to occur in the warmer months. Cluster 6 is most commonly found from January through April. These results reflect the rareness of organized cyclonic storms in the warm months. Clusters 3 and 7 are common all year but have maximum occurrences in the warm months. Cluster 5 is more common in the warm months. In terms of transitions, the high pressure of clusters 3 and 5 and the stagnant, unorganized flow of cluster 7 tend to persist with all three clusters being most likely to follow themselves at least once. Clusters 3 and 4 are prime candidates to be polluted air masses because stagnant clusters 5 and 7 tend to change to cluster 3, cluster 3 is quite persistent itself, and it is then likely to change to cluster 4, which will be shown to be associated with large precipitation amounts.

2) WATER DEPOSITION, STEP 7

Median and 85th percentile three-day precipitation in cm at each sampling site were calculated for each cluster at step 7. The resulting spatial patterns were examined and it was found that the clusters have differing precipitation characteristics, both in space and magnitude. Full descriptions and figures of this and all other analyses discussed in this paper are contained in Fernau (1988). An illustrative subsample is herein shown. The "region" or "map area" referred to in the following discussion is that pictured in the transport figures, encompassing the chemistry sampling sites. The location and amount of precipitation generally reflect the respective mean transport patterns of each cluster. Cluster 5 has the largest area of no precipitation at the median, as one might expect given its anticyclonic nature, with only small quantities of precipitation found to the south and west of the high. At the 85th percentile, it has 1 to 3 cm over most of the sampling region except for the north. This is quite a different pattern from the dry median and suggests the presence of convective precipitation on the south and west sides of the high pressure area. Cluster 7 also has very low precipitation amounts over most of the sampling region at the median, reflecting its stagnant nature. However, it is quite wet at all sites at the 85th percentile with amounts ranging from 1 to almost 4 cm. Given the lack of organized cyclonic flow and the influence of maritime tropical (mT) air, this precipitation is very likely convective in nature. This cluster is pronouncedly skewed, being dry a large fraction of the time but having a "tail" stretching to very wet values. The only cluster showing dry or very low precipitation amounts over a large area at the 85th percentile is cluster 6 (Fig. 2). This probably

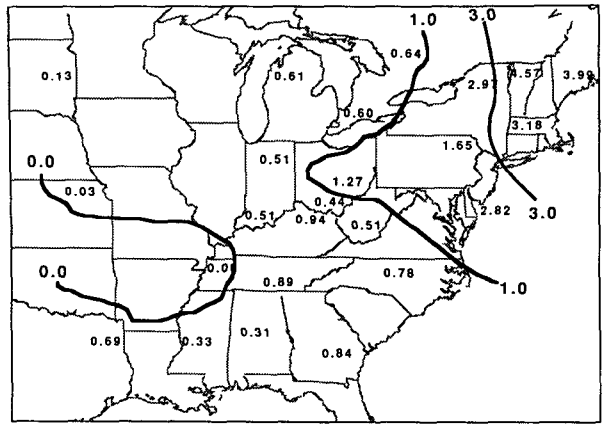

FIG. 2. The 85th percentile water deposition pattern for cluster 6, group A at step 7.

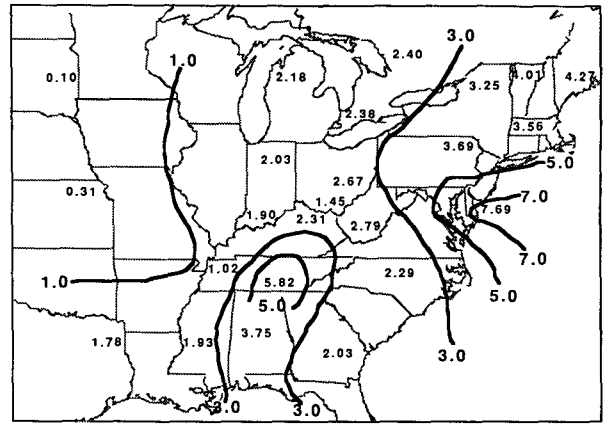

FIG. 4. The 85th percentile water deposition pattern for cluster 2, group A at step 7.

reflects the widespread areas of northwest flow behind the departing trough. This storm's effects are still visible in the northeast with a 5 cm maximum at Underhill, Vermont.

The spatial patterns for clusters 1 through 4 do not vary much from the median to the 85th percentile except that dry areas generally change to relatively light rain. Overall, clusters 2 and 4 are the wettest clusters at step 7 with widespread areas of high precipitation. Clusters 1, 2 and 4 show the effects of troughs moving to the east. Cluster 4 reflects the trough moving in behind the high pressure and also perhaps convective activity. Heavy precipitation is widespread with maximums in New York, Pennsylvania and Tennessee (Fig. 3). Cluster 2 has precipitation over all but the western part of the region, increasing to the east to a median maximum of three cm in Maine and a wet eight cm in Delaware at the 85th percentile (Fig. 4). In cluster 1 the dryer area has spread to cover most of the region and heavier precipitation is confined to the north and east as the trough moves off the coast. The pattern for cluster 3 reflects the influence of the eastern high pressure. It is dryer along the coast and the more western sites with heavier precipitation amounts in the area of the southwest flow.

The 85th percentile precipitation amounts are high at all sites for the "cluster" made up of the nonclassified periods, ranging from 1.5 cm to more than 4 cm. The wettest areas are in the south and in New England, hinting at a possible coastal Atlantic low contributing to this group.

The percent occurrence of the clusters can be compared to the percent of precipitation attributed to each cluster. Table 3 gives the numbers for all sites combined. The wet nature of clusters 2 and 4 can be seen from this table. Clusters 3 and 7 are not disproportionately wet but together contribute over thirty percent of the total precipitation. Most sites receive 20%–28% of their precipitation in cluster 3 but the eastern sites receive only 5%–10%. Brookings, South Dakota, records 37% of its precipitation in cluster 7, probably indicating the presence of a low in the northwest following the ridge. The drier clusters are 1, 5 and 6. However, the five westernmost sites get over ten percent

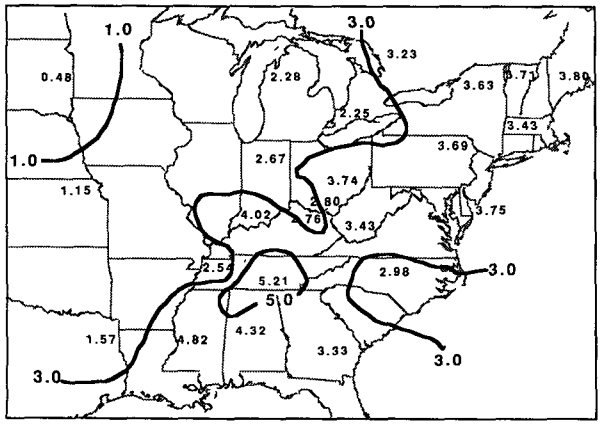

FIG. 3. The 85th percentile water deposition pattern for cluster 4, group A at step 7.

TABLE 3. Comparison of percent occurrence with percent precipitation for step 7, summed over all sites. *M*: missing data cluster, *N*: number of 3-day periods.

Cluster	*N*	Percent occurrence	Percent precipitation	Ratio
1	55	5.0	4.1	0.8
2	85	7.8	9.6	1.2
3	200	18.3	16.9	0.9
4	94	8.6	14.5	1.7
5	106	9.7	6.2	0.6
6	73	6.7	4.4	0.7
7	182	16.7	15.7	0.9
M	298	27.3	28.6	1.0

of their precipitation in otherwise fairly dry cluster 5, with Lancaster, Kansas, getting 23%. The four sites in New York and New England receive 9%–15% of their precipitation in cluster 6, while all other sites are much drier. Indian River, Delaware; Giles County, Tennessee; and Lewisburg, West Virginia, get over 45% of their precipitation in the missing cluster, probably due to a combination of coastal storms and the fact that they were only operative in 1979, the year with the most missing transport data. Additional evidence that this cluster includes coastal storms is the 33%–35% of total precipitation received in it by Uvalda, Georgia; Raleigh, North Carolina; Turners Falls, Massachusetts; and Tunkhannock, Pennsylvania.

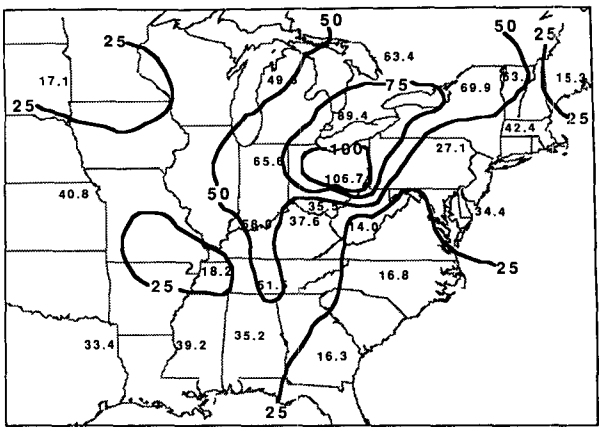

FIG. 5. The 85th percentile sulfate deposition pattern for cluster 3, group A at step 7.

3) PRECIPITATION CHEMISTRY, STEP 7

The chemistry of the precipitation can also be examined by cluster to see if differences exist. This work will concentrate on deposition because that is what ultimately determines the effects of acidic precipitation on the ecosystem. The spatial distribution of sulfate deposition tends to resemble the precipitation fields for most clusters. The median and 85th percentile patterns for three-day sulfur deposition in mg m^{-2} were examined. Cluster 1 has low sulfur deposition, the highest median values being around 18 mg m^{-2} in New York and Ontario and 40 to 50 mg m^{-2} in the same areas at the 85th percentile. Cluster 2 has more deposition than cluster 1. The highest median values are 23 to 27 mg m^{-2} in New York, Massachusetts and Ontario. At the 85th percentile there is a widespread area of greater than 40 mg m^{-2} with peaks of 80 mg m^{-2} in Ontario, Ohio and Tennessee. Ontario and New England receive 10 to 13 percent of their sulfate deposition in this cluster.

Cluster 3 has quite low sulfate depositions at the 50th percentile but at the 85th percentile there are areas of 50 to 107 mg m^{-2} in and downwind of the area of southwest flow (Fig. 5). This situation is sometimes referred to as ducting and has been shown to lead to high ambient and wet pollution levels (Mueller and Hidy 1983). Almost every site in this cluster has a ratio of percent sulfate deposition to percent water deposition greater than 1.0. Turners Falls, Massachusetts; Tunkhannock, Pennsylvania; Winterport, Maine; Big Moose Lake, New York; and Underhill, Vermont, have ratios of 1.91, 1.91, 2.49, 2.15 and 1.81 respectively. The cluster accounts for 21.8% of all sulfate deposition over the three years. At many sites it is even higher: 26% at Zanesville, Ohio, 22% and 24% at Rockport and Fort Wayne, Indiana, 33% at Gaylord, Michigan, 19 to 24% at the four most southern sites, 28% at Lancaster, Kansas, 29% at Brookings, South Dakota, 26% at Underhill, Vermont, 28% at Big Moose Lake, New York, and 37% and 33% at Dorset and Longwoods, Ontario. Remember that this cluster represents 18.3% of all events and 16.9% of all precipitation, although some sites had 5 to 10 percent more precipitation. This is clearly a polluted cluster. This is quite evident at Winterport, Maine, which, in contrast to the other sites, had only 5.2% of its rain yet still received 13% of its sulfate deposition. The high pollution levels make sense because this cluster follows the stagnant clusters 5 and 7 and is quite persistent itself. A relatively large fraction of the pollutants emitted from major sources in the Ohio Valley and perhaps even in the southern states can collect and be transformed to sulfate over a period of days and then be transported to the northeast and deposited.

Cluster 4, which tends to follow cluster 3, also has very high levels of sulfate deposition. The median pattern, with greater than 30 mg m^{-2} in the center, has higher values than does cluster 3. This is presumably due to being more under the influence of the trough and convective storms and less under the influence of the southeast high. The area of high sulfate deposition at the 85th percentile (Fig. 6) is even more widespread than that of cluster 3. A large portion of the center of the map is at 70 to 90 mg m^{-2}. Cluster 4, occurring only 8.6% of the time, is responsible for 13.3% of the total sulfate deposition over all sites. About half of the sites receive 1 to 4 percent more than that. It is a wet cluster, receiving 14.5% of the precipitation. Sites with a greater than expected ratio of sulfate to precipitation deposition are Alamo, Tennessee (1.33), Winterport, Maine (1.04), Marshall, Texas (1.14) and Underhill, Vermont (1.13).

The effects of subsidence associated with the cluster 5 anticyclonic pattern are apparent in the deposition pattern. Fully half the events at all but three sites are dry. At the 85th percentile there are a few sites on the southwest side of the high with depositions of 40 to 60 mg m^{-2}. The rest are still low, with Longwoods, Ontario still showing no deposition. The cluster's 9.7% fraction of events contribute 6.3% of the total deposition. All but six sites have a ratio of percent sulfate

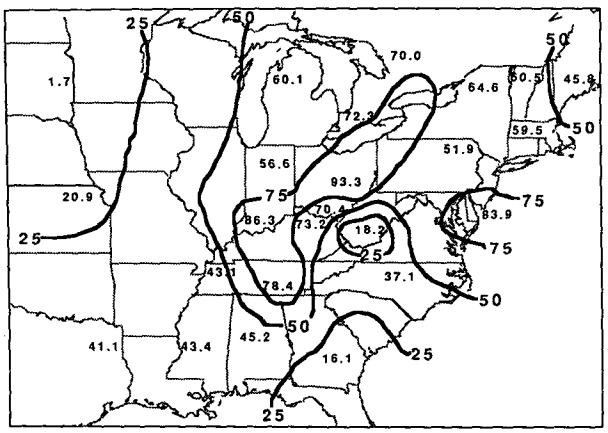

FIG. 6. The 85th percentile sulfate deposition pattern for cluster 4, group A at step 7.

deposition to percent water deposition ratio greater than one with Brookings, South Dakota, the highest with 1.84. Ten sites in the western and southern portion of the region receive more than 6.3% of their sulfate deposition in this cluster. Brookings and Lancaster, Kansas, receive 20% and 23%, respectively.

The median and 85th percentile plots for cluster 6 sulfate deposition are mostly dry except for the northeast. At the 85th percentile Turners Falls, Massachusetts, is at 60 and Indian River, Delaware, is at 80 mg m^{-2}. Of the sulfate deposition 3.3% is found in this cluster, which has 6.7% of the events. Sites with enhanced sulfate deposition relative to precipitation are Rockport, Indiana (1.17), Giles County, Tennessee (1.81), Raleigh, North Carolina (1.92), Lewisburg, West Virginia (1.93), Uvalda, Georgia (2.28) and Selma, Alabama (1.75). The largest percentages of deposition occur at Turners Falls, Massachusetts (6.8%), Winterport, Maine (9.6%) and Underhill, Vermont (7.1%), showing the effects of the eastern low pressure center.

The median pattern for cluster 7 has fairly low sulfate deposition except for Indian River, Delaware (22.5 mg m^{-2}) and Clearfield, Kentucky (15.6 mg m^{-2}). The 85th percentile plot (Fig. 7) shows the effects of stagnation and pollution buildup with widespread areas greater than 40 mg m^{-2}. The highest values are at Indian River, Delaware (88.3 mg m^{-2}) and the center of the map. Cluster 7, which comprises 16.7% of the total events and 15.7% of the total precipitation, contributes 20.3% of the total sulfate deposition. This cluster is a major contributor to deposition at many sites: 36% at Brookings, South Dakota, 33% at Clearfield, Kentucky, 29% at Clinton, Mississippi, 27% at Raleigh, North Carolina, 25% at Gaylord, Michigan, and Longwoods, Ontario, 24% at Uvalda, Georgia, Selma, Alabama, and Big Moose Lake, New York, 23% at Rockport, Indiana, and McArthur, Ohio, and 21% at Tunkhannock, Pennsylvania, Indian River, Delaware and Marshall, Texas. All but five sites located at the extreme east and west of the map are enhanced in sulfate deposition relative to precipitation. The highest ratios are Uvalda, Georgia with 2.14, Big Moose Lake, New York with 1.73, Gaylord, Michigan with 1.67 and Tunkhannock, Pennsylvania with 1.61.

The median and 85th percentile values for three-day sulfate deposition for the missing events omitted from the cluster analysis were examined. Median values mostly range from 5 to 10 mg m^{-2} with no real pattern evident. The 85th percentile values are fairly high at most places. Of total sulfate deposition 24% is accounted for by this cluster. This ranges from 10 percent or less at Gaylord, Michigan; Clearfield, Kentucky; Marshall, Texas; Brookings, South Dakota; Longwoods and Dorset, Ontario to around 40 percent at Tunkhannock, Pennsylvania, and 50 percent at the three sites inoperative after 1980 (Indian River, Delaware; Giles County, Tennessee, and Lewisburg, West Virginia).

The patterns and behavior of the other analytes are much the same as precipitation and sulfate deposition. Table 4 gives the overall percent contributions of each cluster to the various analytes and the ratio of the percent analyte deposition to percent water deposition. Overall, nitrate deposition is greater than sulfate deposition for clusters 1, 2, 4 and 6, which are stormy and tend to occur in the colder months. Ammonium neutralization effects appear to occur in clusters 3, 5 and 7, which are more common in the warmer months. This is especially evident at many sites in cluster 3 whose H^+ depositions, although still high, are several percent lower than the corresponding sulfate depositions.

It was difficult to perform statistical tests on all the chemistry data because of the volume of that data. Because of the nonnormality of the chemistry data, the nonparametric Kruskal-Wallis test (Davis 1986) was

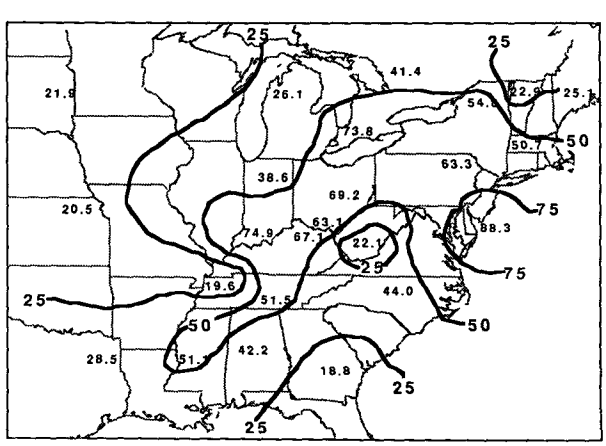

FIG. 7. The 85th percentile sulfate deposition pattern for cluster 7, group A at step 7.

TABLE 4. Percent contributions of each cluster to the various analyte depositions and ratios of analyte to water deposition for step 7. *M*: missing events.

	$SO_4^=$		NO_3^-		NH_4^+		H^+	
Cluster	%	Ratio	%	Ratio	%	Ratio	%	Ratio
1	3.3	0.80	3.8	0.93	3.8	0.91	3.3	0.78
2	7.6	0.79	8.4	0.88	7.6	0.79	8.1	0.84
3	21.8	1.30	20.3	1.21	23.4	1.39	19.7	1.17
4	13.3	0.92	13.5	0.93	13.0	0.90	13.3	0.91
5	6.3	1.02	6.1	0.99	6.8	1.09	6.2	1.00
6	3.3	0.75	4.2	0.95	3.0	0.68	3.9	0.87
7	20.3	1.30	18.7	1.19	20.7	1.32	19.3	1.23
M	24.0	0.84	24.9	0.87	21.8	0.76	26.3	0.92

applied to the step 7 chemistry data for Turners Falls, Massachusetts. The results showed that for all nine chemistry variables (water deposition and the concentrations and depositions of the four analytes) the distributions in the seven clusters are not all the same. Sulfate deposition was looked at in more detail using the nonparametric Mann-Whitney procedure (Davis 1986) to test the null hypothesis that the distributions in two clusters are the same. Table 5 gives the alpha significance level at which the null hypothesis can be rejected for each pair of clusters. Cluster 5 is clearly different from all the other clusters. Clusters 2 and 4 are different from all others except for each other. Among clusters 1, 3, 6 and 7 the null hypothesis cannot be rejected, except between clusters 3 and 6. Testing at several other sites also revealed statistical differences between some of the cluster pairs.

b. Analysis of step 18

1) PRECIPITATION AND CHEMISTRY OF STEP 18 CLUSTERS

The patterns and branching of the step 18 clusters are described in Fernau and Samson (1990) and the patterns are illustrated in Fernau (1988). When the transport, precipitation and wet chemistry at step 18 are compared to step 7 some interesting differences in flow and chemistry are revealed. For example, cluster 3 at step 7 is a mixture of five step 18 clusters (5–9), all of which have different appearances. The feature they hold in common is a high pressure center located in the southeastern United States. However, the exact position and subsequent movement of the high differs in each cluster. Of the five, clusters 5 and 7 are enhanced in precipitation. Cluster 7 contributes almost three times more precipitation than its proportion to the total membership. In cluster 7 the trailing trough is much deeper and has much stronger winds than in the other four clusters. Going from step 7 to step 18 shows that cluster 3 of 7 which was overall deficient in precipitation (Table 3) has within it a wet subset.

In absolute terms, clusters 10, 18 and 5 contribute the most precipitation over the whole region and clusters 5, 9, 10 and 18 deposit the most pollutants. They have the common thread of either being stagnant and persistent or following such conditions. Clusters 5, 6, 8, 9, 16 and 18 are enhanced in analyte deposition relative to precipitation, although less so for hydrogen because of elevated ammonium deposition. Of these, clusters 9 and 16 are deficient in deposition relative to their frequencies of occurrence, reflecting drier natures. Cluster 13 has relatively less ammonium deposition compared to sulfate and nitrate, leading to enhanced H^+ deposition. Cluster 7 is enhanced in analyte deposition relative to its frequency of occurrence but this appears to be mostly due to greatly enhanced precipitation.

In place of showing spatial plots of the median and percentile values of deposition at step 18, the values at a receptor sensitive to acid deposition are examined and the different clusters compared. Figure 8 gives the percentile values for sulfate deposition for each cluster at step 18 for the Big Moose Lake, New York, site, which possesses chemistry data for 15 months of the three-year period. If the median symbol for any one cluster is obstructed it usually means that it is zero. It can be seen that the median and 85th percentile values for the clusters differ and that some clusters clearly cause higher three-day depositions than others. However, it is also clear that there is considerable within-

TABLE 5. Alpha level at which the null hypothesis of identical distributions can be rejected, determined using the Mann-Whitney test. Test performed on sulfate deposition events at Turners Falls, Massachusetts.

	Cluster						
Cluster	1	2	3	4	5	6	7
1	—	.00	.22	.00	.00	.67	.72
2		—	.00	.79	.00	.00	.00
3			—	.00	.00	.02	.17
4				—	.00	.00	.00
5					—	.00	.00
6						—	.27
7							—

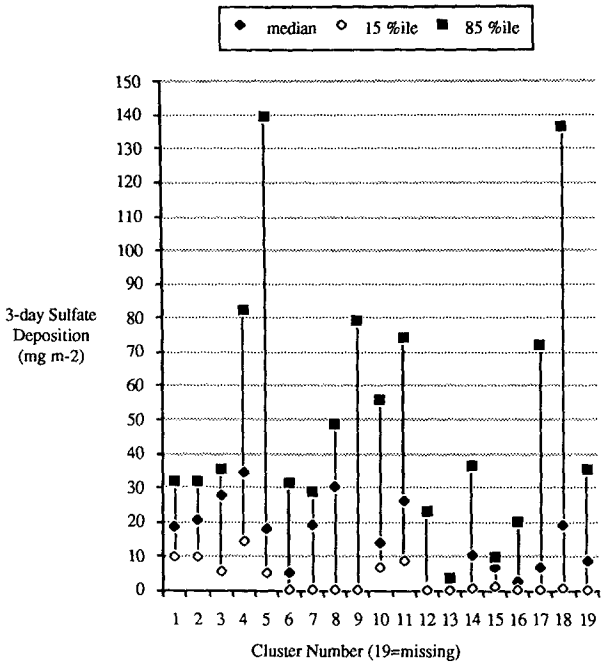

FIG. 8. Percentile statistics by cluster for 3-day sulfate deposition at Big Moose Lake, New York, for group A at step 18.

cluster variation in deposition in some clusters and there is considerable overlap among the clusters, even the ones that are in different clusters at step 7. Several clusters have uniformly low sulfate depositions but no clusters with uniformly high depositions exist. The overlap and variation are partly due to the many dry days present in the clusters. Wet clusters 1 and 2 show a narrower range and fewer dry days. Nonparametric tests show that many of the clusters have significantly different distributions from each other, despite appearances. There is more separation evident among the concentration distributions (not shown) than the deposition distributions and, in general, they appear to have less within-cluster variation as well. The spatial pattern of median sulfate deposition over all sites (not shown) does change from cluster to cluster. One possible reason for overlaps is that changes in precipitation chemistry due to frontal passages may be lost in the three-day running averages, reducing the inter-cluster differences.

Clusters 5 and 18 are good examples of skewed clusters that were often dry but when wet can produce large quantities of precipitation and sulfate deposition. Examination of individual cumulative frequency diagrams showed that, to a greater or lesser extent, all clusters are positively skewed with many values of zero and then a long tail extending in the positive direction. A few of the clusters show bimodal behavior; for example, clusters 7 and 8. The sulfate deposition distributions show that clusters which are combined by step 7 can have differing distributions at step 18.

2) CASE STUDY OF A CLUSTER CONTAINING HIGH POLLUTION EVENTS

An example where observed weather patterns correspond to the mean cluster pattern was described in Fernau and Samson (1990). Nine consecutive moving three-day periods extending from 9 to 17 June 1983 were assigned to cluster 9 at step 18, which is enhanced in sulfate deposition relative to precipitation. As shown by weather maps, the eastern United States during those eleven days is characterized by very light winds aloft, a huge ridge over the area and the polar front well to the North. At the surface, winds are light, precipitation is nonexistent, skies are clear at many stations and high pressure dominates the entire time. The center of the high is located over the Mid-Atlantic states and shows little or no movement for most of the period. Haze and fog begin to be reported on 10 June and quickly spread eastward and northward over wide areas of the eastern United States and Canada, eventually being swept east ahead of a cold front. These conditions all appear to be very conducive to or indicative of a regional buildup of sulfate.

The wet chemistry at every site during this time span was checked. Many sites and days were dry. However, five sites experienced three-day sulfate depositions over 100 mg m^{-2} and weighted sulfate concentrations in the 70–90 μmol L^{-1} range in connection with the passing of the cold front: Turners Falls, Massachusetts; Tunkhannock, Pennsylvania; Zanesville, Ohio; Clearfield, Kentucky and Big Moose Lake, New York. It is likely that sulfur dry deposition and ambient sulfate concentrations were elevated at most sites.

c. Seasonal differences in chemistry

To see if the chemistry in a given cluster varied by season, sulfate deposition at Big Moose Lake for step 7 was examined. Warm season is defined to be from April through September and cold season from October to March. Except for cluster 1, all clusters had higher median and 85th percentile depositions in the warm season. Of those clusters, all but cluster 7 had warm season means at or above the cold season 85th percentile. However, the distributions overlap, perhaps due to the dry periods present all year. One would expect higher sulfate concentration in the warmer months; the behavior of cluster 1 is probably due to greatly increased precipitation in the cold months. The differences are also present in H^+ and nitrate deposition at Big Moose Lake, although the differences are not as pronounced in the nitrate data. Also cluster 2 and cluster 1 both had higher nitrate deposition in the cold season than in the warm season. Apparently, some of the within-cluster variance in the clusters is due to seasonal variation. Whether the clustering should be done separately for each season or how this result should be accounted for in using clusters for aggregation are matters for further research.

d. Other clustering methods

Using the 1979, 1981 and 1983 12-hour upwind distance vectors at the same ten sites, clustering was done using some other clustering methods. Used were centroid, median and average link, all with Euclidean distance, and average link with the correlation coefficient as the distance measure (Fox and Guire 1976; Anderberg 1973). These methods gave the result at step 7 of one extremely large cluster and six smaller clusters representing outlier events. The small clusters generally do not have enough members to allow statistical or spatial examination of their associated chemical deposition. One cluster from the average link with correlation coefficient as distance measure method was found to have a slow-moving high pressure system resembling Ward's method cluster 9. The sites in this cluster are enhanced in sulfate relative to precipitation, with the highest ratios at Tunkhannock, Pennsylvania (2.42), Underhill, Vermont (2.46), Big Moose Lake, New York (2.96) and Dorset, Ontario (2.15). These sites receive little precipitation, but what precipitation occurs is very polluted. This result is in line with the earlier finding that the ducting and stagnation associated with this pattern allow pollutants to accumulate in large quantities.

4. Conclusions

Plots of the spatial patterns of median and 85th percentile three-day water and chemical deposition for clusters defined using transport vectors show intercluster differences in the occurrence, amount and location of precipitation and pollutant deposition. Some clusters are relatively dry over much of eastern North America while others are associated with high precipitation. Some of the clusters have high pollutant deposition over much of the region and/or have a ratio of percent pollutant deposition to percent water deposition greater than 1.0. The highest pollutant depositions over the widest areas resulted from mean transport patterns with large areas of slow air mass movement over regions of high sulfur emissions and which were frequently persistent over several periods or followed persistent clusters. Some of the cluster deposition distributions are significantly different from each other at various sites, although the spatial patterns were not statistically compared. The spatial patterns of the pollutants are a plausible result of the transport, transition, and persistence characteristics of the clusters. However, within-cluster variability was large and there was overlap among clusters. Some of this may be due to changes in deposition due to frontal passages, something not discriminated by the three-day deposition period. It was not possible to define clusters that were almost always heavily polluted. This is especially true given that some clusters with very polluted precipitation also included many dry periods. Within clusters, sulfur deposition was generally higher in the warmer months. Overall, cluster analysis of air-mass transport vectors has been able to point out those weather patterns which are likely to either lead to large deposition events or contribute large amounts of deposition over the course of a year. Additional meteorological variables may need to be included in future analyses in order to reduce the within-cluster variance.

Acknowledgments. This work was funded as part of the National Acid Precipitation Assessment Program by the United States Environmental Protection Agency. The results have not been subject to the agency's peer and policy review and therefore do not necessarily reflect the views of the agency, and no official endorsement should be inferred. Chemistry data were provided by the Electric Power Research Institute and the Ontario Ministry of the Environment. Dr. Jennie L. Moody provided the initial idea for this work and invaluable advice along the way. The EPA project reviewers pointed out many useful revisions, corrections, and areas for further research.

REFERENCES

Anderberg, M. R., 1973: *Cluster Analysis for Applications.* Academic Press, 359 pp.

Barry, R. G., and A. H. Perry, 1973: *Synoptic Climatology—Methods and Applications.* Methuen & Co., 555 pp.

Chang, J. S., R. A. Brost, I. S. A. Isaksen, S. Madronich, P. Middleton, W. R. Stockwell and C. J. Walcek, 1987: A three-dimensional Eulerian acid deposition model: Physical concepts and formulation. *J. Geophys. Res.*, **92,** 14 681–14 700.

Crutcher, H. L., R. C. Rhodes, M. E. Graves, B. Fairbairn and A. C. Nelson, 1986: Application of cluster analysis to aerometric data. *J. Air Pollut. Control Assoc.*, **36,** 1116–1122.

Davis, J. C., 1986: *Statistics and Data Analysis in Geology.* Wiley & Sons, 646 pp.

Endlich, R. M., B. P. Eynon, R. J. Ferek, A. D. Valdes and C. Maxwell, 1988: Statistical analysis of precipitation chemistry measurements over the eastern United States. Part I: Seasonal and regional patterns and correlations. *J. Appl. Meteor.*, **27,** 1322–1333.

Fernau, M. E., 1988: Use of cluster analysis to define periods of similar meteorology and precipitation chemistry in eastern North America. Ph.D. dissertation, University of Michigan, 331 pp. [Available from University Microfilms International, Ann Arbor, MI.]

——, and P. J. Samson, 1990: Use of cluster analysis to define periods of similar meteorology and precipitation chemistry in eastern North America, part I: Transport patterns. *J. Appl. Meteor.*, **29,** 735–750.

Fox, D. J., and K. E. Guire, 1976: *Documentation For MIDAS.* Statistical Research Laboratory, The University of Michigan, 203 pp.

Gaarenstroom, P. D., S. P. Perone and J. L. Moyers, 1977: Application of pattern recognition and factor analysis for characterization of atmospheric particulate composition in southwest desert atmosphere. *Environ. Sci. Technol.*, **11,** 795–800.

Gorham, E., F. B. Martin and J. T. Litzau, 1984: Acid rain: Ionic

correlations in the eastern United States, 1980–1981. *Science,* **225,** 407–409.

Kalkstein, L. S., and P. Corrigan, 1986: A synoptic climatological approach for geographical analysis: Assessment of sulfur dioxide concentrations. *Ann. Assoc. Amer. Geogr.,* **76,** 381–395.

——, G. Tan and J. A. Skindlov, 1987: An evaluation of three clustering procedures for use in synoptic climatological classification. *J. Climate Appl. Meteor.,* **26,** 717–730.

Moody, J. L., and P. J. Samson, 1989: The influence of atmospheric transport on precipitation chemistry at two sites in the midwestern United States. *Atmos. Environ.,* **23,** 2117–2132.

Mueller, P. K., and G. M. Hidy, 1983: The Sulfate Regional Experiment: Report of findings, Vol. 3. EA-1901, Vol. 3, Res. Proj. 862-1, 80-2, prepared by ERT, West Lake Village, CA for Electric Power Research Instit.

——, M. A. Allan and J. J. Jansen, 1988: Utility Acid Precipitation Study Program third summary report. UAPSP 114, contract U101-90. Summary Rep. for Utility Acid Precip. Study Program, Washington, DC and Electric Power Research Instit., [Available from EPRI Research Reports Center, P.O. Box 50490, Palo Alto, CA 94303.]

Sanchez Gomez, M. L., and M. C. Ramos Martin, 1987: Application of cluster analysis to identify sources of airborne particles. *Atmos. Environ.,* **21,** 1521–1527.

Saucy, D. A., J. R. Anderson and P. R. Buseck, 1987: Cluster analysis applied to atmospheric aerosol samples from the Norwegian Arctic. *Atmos. Environ.,* **21,** 1649–1657.

Slanina, J., J. H. Baard, W. L. Zijp and W. A. H. Asman, 1983: Tracing the sources of the chemical composition of precipitation by cluster analysis. *Water, Air and Soil Pollut.,* **20,** 41–45.