EVALUATOIN OF AGGLOMERATIVE HIERARCHICAL CLUSTERING METHODS

by

Min-Tang Li, Ph.D., Research Associate
Fang Zhao, Ph.D., P.E., Associate Professor
Yifei Wu, Graduate Research Assistant
Albert Gan, Ph.D., Assistant Professor

Lehman Center for Transportation Research
Department of Civil and Environmental Engineering
Florida International University
University Park, EAS 3740, Miami, FL 33199
Tel: 305-348-2577, Fax: 305-348-2802
E-mail: lim@fiu.edu

TRB Paper Number: 03-4147
Words = 4,100
Tables = 3 x 250 = 750
Figures = 7 x 250 = 1,750
Total Word Count = 6,600

Prepared for Consideration for
Presentation and Publication at the
82nd Annual Meeting of the Transportation Research Board
Washington, D.C.
January 12-16, 2003

November 2002

ABSTRACT

This paper describes the findings from evaluating the performance of agglomerative hierarchical cluster methods for determining seasonal factor groups. Seasonal factor groups are usually determined by traditional cluster analysis based on various similarity measures. Agglomerative hierarchical methods merge telemetry traffic monitoring sites (TTMSs) into groups according to their similarities. A wide variety of similarity measures may be used in cluster analysis. This study evaluated a total of eight agglomerative clustering methods: average linkage method, centroid method, EML method, flexible-beta method, McQuitty's similarity analysis method, median method, single linkage method, and Ward's minimum-variance method. Multi-year data collected between 1997 and 2000 from 21 TTMSs in Florida Department of Transportation (FDOT) District 4 were utilized in this study. The Pseudo *F* (PSF) statistic was employed as the criterion for determining the number of clusters. The average linkage, centroid, and single linkage methods were found to be more robust to outliers than the other methods. The study also found that the McQuitty's (MCQ) method performed better than the other methods on grouping TTMSs after outliers were eliminated. When the MCQ method was applied to analyze the historical data collected between 1997 and 1999, TTMSs were not consistently assigned to the same cluster group across years. Roadway functional classes were found to be insignificant in determining seasonal groups, while spatial location was a more significant factor because a TTMS tended to be clustered with those in its proximity.

INTRODUCTION

Traffic data are the foundation of highway transportation planning and are used to assist highway engineers in maintaining and designing safe, efficient, and cost effective facilities *(1)*. It is well known that traffic variations occur at different time scales, e.g., time of day, day of week, and season (month) of the year *(2)*. Consequently, it is important to accurately interpret the temporal variation effects on collected traffic data in order to achieve better design decisions.

Among the known temporal fluctuations of traffic stream, seasonal variation is probably the most important characteristic that must be accounted for in traffic monitoring. Currently, the Florida Department of Transportation (FDOT) districts determine seasonal factor (SF) categories from a group of selected permanent telemetry traffic monitoring sites (TTMSs) and assign them to short-term traffic count sites for the purpose of estimating AADT, assuming that seasonal variability and traffic characteristics of short-term and permanent counts are similar *(1)*. It is reported that, when the true factor group for a site is known, traditional short-counts could provide estimates of mean daily traffic with a PI95 (precision achievable with 95 percent confidence) between 10 and 23 percent *(3)*.

The SFs used in Florida are determined by interpolating between the monthly seasonal factors (MSFs) for two consecutive months. The MSF for a specific month at a particular location is derived by dividing the monthly average daily traffic (MADT) at a given location by its AADT. SFs are also applied in the estimation of peak-season conversion factor (PSCF) and model output conversion factor (MOCF) for modeling purposes. PSCF is used to convert a short-term traffic count (ADT) to peak season weekday average daily traffic (PSWADT). MOCF is used to convert the PSWADT to average annual daily traffic (AADT).

The problem of constructing factor groups from TTMS sites and then determining the number of permanent count stations to estimate monthly factors with a given precision have attracted a fair amount of attention over the years. Among the research efforts devoted to determining seasonal groups in traffic monitoring, cluster analysis is probably the most widely used and preferred approach. For example, Sharma *et al*. applied a hierarchical cluster method to group 45 permanent traffic counters (PTCs) in Alberta, Canada based on their 12 monthly factors *(4, 5, 6)*. The Scheffe's S-method of multiple comparisons of group means was used to determine the optimal number of groups ranging from 6 to 10 obtained from the hierarchical process, each containing more than two counters. Flaherty *(7)* used the hierarchical clustering method and the k-means method to analyze the monthly factor data collected over a 5-year period from 28 PTCs in Arizona. He concluded that the similarity in the patterns of the monthly factors was more a function of geography and topography than functional classification of the associated highways, and that the population of the surrounding area did not appear to be an explanatory factor for the factor groups. Flaherty also found that the mean-square error for the within group was the lowest for the estimates obtained by comparing the MSFs of each ATR with the mean values from the ATR's corresponding factor group. Four clusters were also found to be the most stable over the 5-year period. Aunet *(8)* used cluster analysis to examine the variation in Wisconsin's traffic data and recommended a procedure with the following steps:

- Examining plots of monthly traffic at each permanent automatic traffic counter (ATR);
- Examining tables of coefficient of variations (CVs) for each permanent ATR;
- Examining the results of cluster analysis; and
- Examining geographic mappings of ATRs in preliminary groupings.

The preliminary results from Aunet's procedure applied in Wisconsin revealed that the seasonal patterns remained stable over time. Additionally, although significant variations existed in the monthly seasonal factors at each permanent ATR in the same group, seasonal factor groups could be generally defined according to roadway functional classifications.

Although it is difficult to determine how many "groups" should be formed *(9)* and some research concluded that clusters were not consistent across years *(9, 10)*, cluster analysis has been recommended for identifying roadway sections having similar traffic patterns within a given group *(2)*. In cluster analysis, agglomerative hierarchical methods merge TTMSs into groups according to their similarities. A wide variety of similarity measures have been used. The following agglomerative hierarchical clustering methods are available in SAS (Statistical Analysis System) Version 8 *(11)* for quantifying the distance (or dissimilarity) between two clusters:

1. Average Linkage (AVE)
2. Centroid Method (CEN)
3. EML
4. Flexible-beta Method (FLE)
5. McQuitty's Similarity Analysis (MCQ)
6. Median Method (MED)
7. Single Linkage (SIN)
8. Ward's Minimum-Variance Method (WAR)

Each clustering method utilizes a different formula to estimate the distance between two clusters and tends to create clusters of certain types *(11)*. For example, average linkage tends to join clusters with small variances and is slightly biased toward producing clusters with the same variance. Ward's method tends to join clusters with a small number of observations and is strongly biased toward producing clusters with roughly equal number of observations. The EML method is similar to Ward's minimum-variance method but removes the bias toward equal-sized clusters. Practical experience has indicated that the EML method is somewhat biased toward unequal-sized clusters *(11)*.

In SAS, the penalty option is used to adjust the degree of bias toward unequal-sized clusters for the EML method. The value specified as the penalty should be greater than zero. In this study, four additional penalty values other than its default value (i.e., 2.00) were applied in the cluster analysis to test the parameter's sensitivity to the results. They were 1.00, 1.25, 1.50, and 1.75. Additionally, two values, i.e., -0.25 and -0.50, were specified for the beta option for the flexible-beta method. This gives a total of 13 methods that were implemented in the evaluation of the agglomerative hierarchical clustering methods.

This paper documents the process and findings from cluster analyses using the 13 aforementioned clustering methods on the MSFs from 21 TTMSs in a FDOT district. The Pseudo *F* (PSF) statistic was used to determine the number of clusters in the data. The clusters obtained from each method were then examined to validate their performance. The method with the optimal performance was then employed to investigate the historical clustering patterns.

STUDY DATA

For this study, the traffic data for the permanent TTMSs located within the jurisdiction of FDOT District 4 (covering Broward, Indian River, Martin, Palm Beach, and St. Lucy counties) were investigated. The data source of this study was the 1997-2000 Traffic Count Information CDs published by FDOT. Each CD contains the MSFs derived for 1997, 1998, 1999, and 2000. For 1999 and 2000, detailed hourly traffic count data are also recorded on the CD. For 1997 and 1998, however, only monthly seasonal factors are available. The AADTs for these TTMSs in the four-year period vary from 2,593 to 228,518. Table 1 shows the number of TTMSs in the study area in different years. Over the four-year period, only 19 stations were consistently recorded with MSF information. By including two more TTMSs located on the Florida Turnpike in the region, the MSFs from a total of 21 TTMSs were analyzed. The MSFs from each TTMS for a given year were stored in a 12-element vector, one element for each month. There were, therefore, 21 vectors for the 21 TTMSs in a given year for analysis.

TABLE 1 Number of TTMSs in FDOT District 4 from 1997 through 2000

Year	1997	1998	1999	2000
Number of TTMSs*	29	31	27	29

* Excluding TTMSs located on the Florida Turnpike section (District 8)

RESEARCH PROCEDURE

The research process entails the following steps to evaluate agglomerative hierarchical clustering methods:

1. Verify seasonal factors at each TTMS,
2. Perform preliminary cluster analyses to identify outliers,
3. Perform cluster analyses and evaluate the factor groups using the data without outliers, and
4. Select the optimal clustering method to verify if seasonal groups are temporally stable.

The following sections explain each of these steps in detail.

Data Verification

Before applying cluster analysis to determine seasonal groups, the historical MSFs recorded on the FDOT Traffic Count Information CDs were examined. The purpose was to identify possible outliers in the dataset by verifying the temporal pattern in the data collected in four consecutive years at the same TTMSs. Although the days with missing data were excluded from the calculation of MADT, extremely low daily volumes in a given month, most likely caused by

equipment failures and other unknown reasons, were not eliminated from the data. Consequently, the corresponding MADTs were higher than expected. On the other hand, higher MADTs could also have resulted from excluding days with low traffic volumes, e.g., on weekends.

Figure 1 illustrates the problems related to temporal stability in the multi-year MSFs at one of the TTMSs (Station No. 860214). The figure shows a susceptible high MSF value in September 2000. The CV[1] (coefficient of variation) for the MSFs at the station for September alone was 8.296% while the same statistics for the other months ranged from 0.558% to 2.663%. Figure 2 illustrates the daily volumes at the same location in September of 1999 and 2000. From Figure 2, it can be seen that the MSF in year 2000 is likely to be overestimated due to missing data since the higher volumes that tend to occur near the end of the month were excluded from the calculation of MADT, resulting in a lower than expected MSF.

In order to identify probable data outliers, the monthly CVs were first calculated for the multi-year MSFs at each TTMS. The daily volumes collected at each TTMS in 1999 and 2000 were then examined for those months with CVs greater than or equal to 3%. The 3% threshold value was selected by observing the resulted monthly CVs for the 21 TTMSs. However, since no daily volumes were available in 1997 and 1998, no verification effort was performed to validate the MSFs in these two years. When possible outliers were identified, they were replaced with the median MSF of the respective month obtained from the multi-year data.

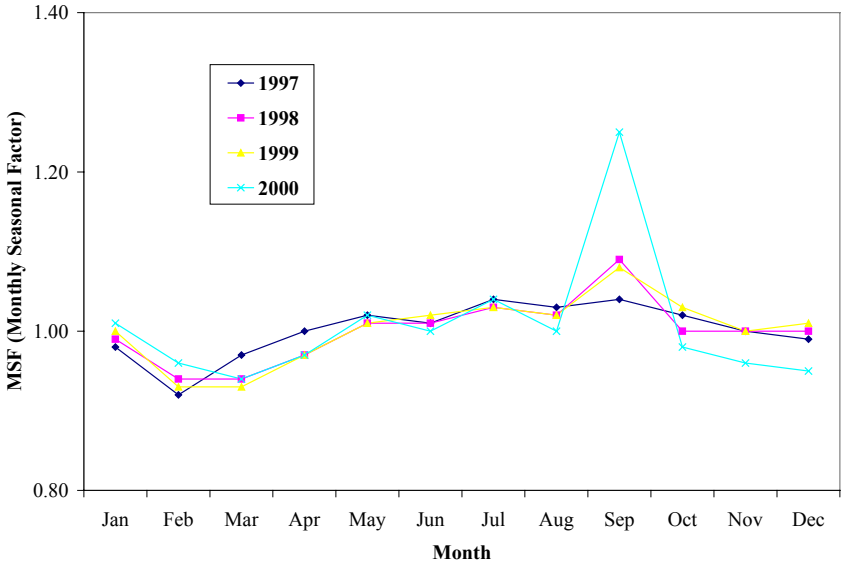


FIGURE 1 MSF versus Month at Station 820614 in 2000.

[1] Defined as the standard deviation divided by the mean MSF, and multiplied by 100 to get a percentage.

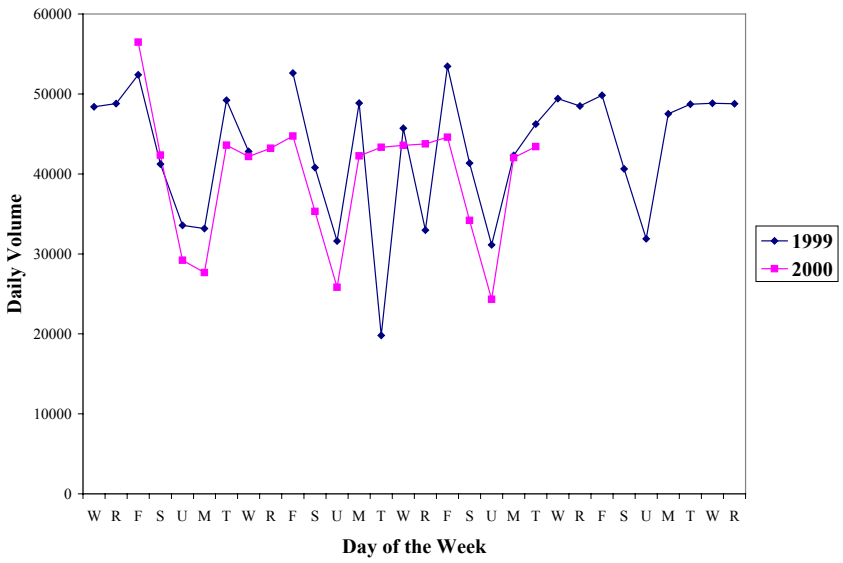


FIGURE 2 Daily Volume versus Day of the Week[2] at Station 820614 in 1999 and 2000.

Preliminary Cluster Analysis

After the MSFs were verified, the year 2000 MSFs were analyzed with the various clustering methods to identify outliers. The Pseudo *F* (PSF) statistic was used in the study as the criterion for determining the number of clusters in the data. The relatively large PSFs, i.e., a local peak in the graph of the PSFs plotted against the number of clusters, indicate a stopping point. The PSF for a given level is calculated as follows *(11)*:

$$PSF = \frac{\frac{T - P_G}{G - 1}}{\frac{P_G}{n - G}}$$

where

T = $\sum_{i=1}^{n} \left\| x_i - \bar{x} \right\|^2$;

P_G = $\sum W_j$, where summation is over the G clusters at the G^{th} level of the hierarchy;

G = number of clusters at a given level of the hierarchy;

n = number of observations (i.e., TTMSs);

x_i = i^{th} obervation;

$\bar{x}$ = sample mean vector;

W_k = $\sum_{i \in C_k} \left\| x_i - \bar{x}_k \right\|^2$; and

$\bar{x}_k$ = mean vector for cluster C_k .

[2] M for Monday, T for Tuesday, W for Wednesday, R for Thursday, F for Friday, S for Saturday, and U for Sunday.

Note that T is the sum of squared Euclidean distances from each observation to the overall mean while P_G is the distance measure from the observations in a given cluster to its cluster mean *(12)*. The resulted clusters were examined after the preliminary cluster analyses were performed. The TTMSs that formed a single member cluster were treated as outliers and excluded from the evaluation process, since their monthly fluctuation patterns were much different from the others.

Evaluation

The 13 clustering methods were applied again to the year 2000 data after the TTMS outliers were eliminated. Denote the resulted clusters by C_i, $i \in \{1, 2, 3, \ldots n\}$, where n is the number of clusters. Let the number of TTMSs within a cluster C_i be m and the SF variance for a particular month k and cluster group C_i be v^{ik}. The cluster groups from each method were evaluated by first calculating the pooled estimate of common variance, denoted by v^i_{pool}, which was simply the arithmetic average of the monthly MSF variances for the TTMSs in the same group, i.e., $v^i_{pool} = 1/12 \sum_k v^{ik}$. The average of the estimates of common variances from all of the resulted cluster groups was then calculated, i.e., $1/n \sum_{i=1}^{n} v^i_{pool}$. The methods with lower average pooled variance indicate smaller dissimilarities within cluster groups and, consequently, are considered to have a better performance than the other methods *(13)*. The spatial locations of the TTMSs clustered in the same seasonal factor groups by the method with the least pooled variance were further examined to verify if they were logical and reasonable. The identified method(s) was then used to cluster the MSFs for the data from other years.

Temporal Stability

Outliers were first identified and excluded by performing a preliminary cluster analysis using the cluster method(s) identified in the evaluation step. The remaining TTMSs were then analyzed again to verify if cluster groups were consistently stable from year to year.

RESULTS AND DISCUSSIONS

All of the clustering methods revealed that Stations 860306, 890259, and 940144 were outliers during the preliminary analysis step. As a result, these stations were excluded in the subsequent analysis. Additionally, methods such as AVE, CEN, and SIN were more robust to outliers than the other hierarchical methods since the above three stations were not merged with the other station groups until the last step when these methods were applied. The results suggest that the AVE, CEN, and SIN methods may be the preferred methods for screening out outliers in practice. The above findings may be illustrated using a tree diagram (also known as the dendrogram or phenogram) shown in Figure 3. The figure gives the results of hierarchical clustering using the SIN method in the SAS CLUSTER procedure. As shown in the figure, the three outliers were the last three single-member clusters that were merged with the others. In other words, the SFs from the three outliers were considered to be much dissimilar to the others and, consequently, were not merged until the end.

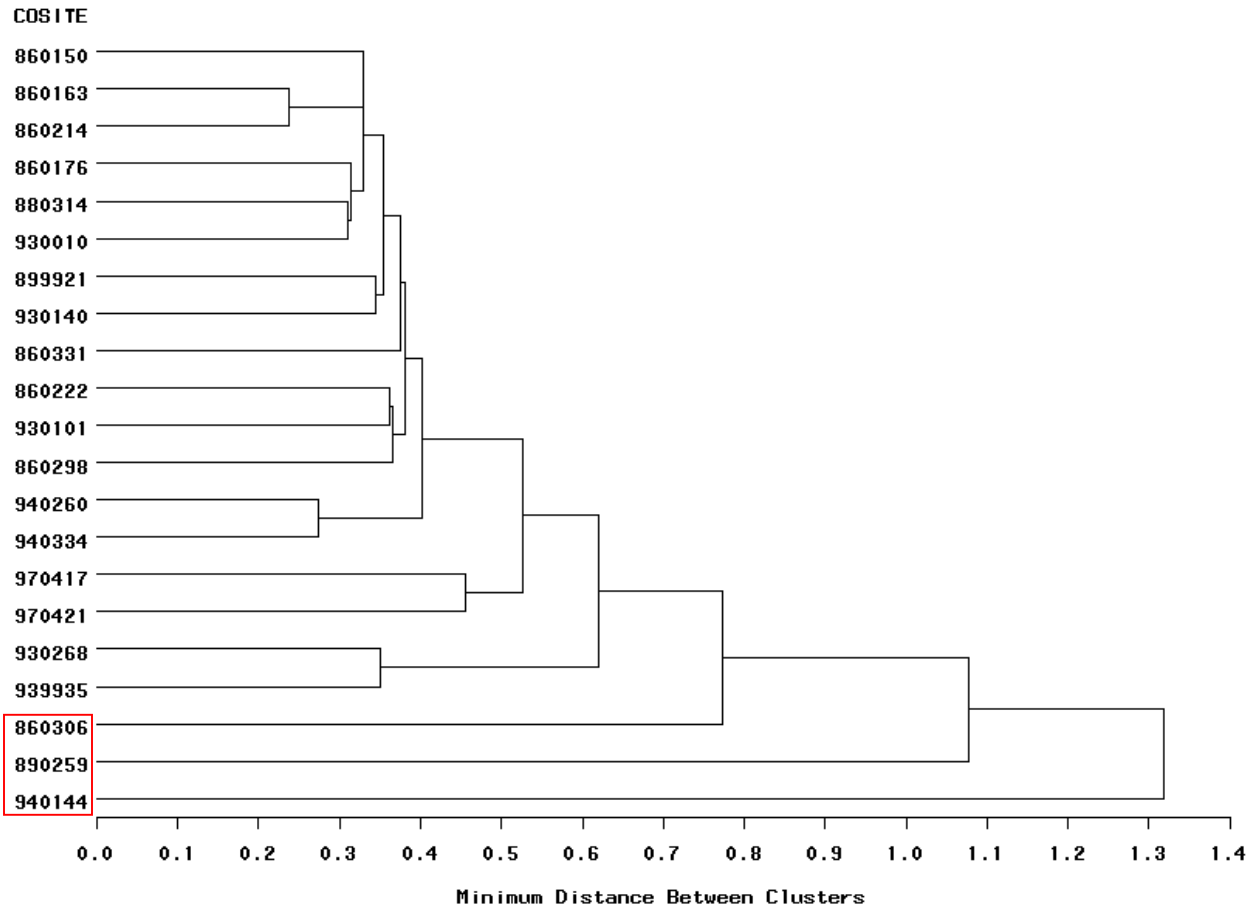


FIGURE 3 Preliminary Hierarchical Clustering Result from the SIN Method.

The 13 clustering methods were then applied again to analyze the data from the remaining 18 TTMSs after the outliers were excluded. Table 2 presents the PSFs obtained. In this table, the cells with larger PSFs than their adjacent cells, which suggest a possible stopping point from merging groups further, are highlighted. Table 2 indicates that different penalty values do not alter the results from the EML method, since PSFs were unchanged. Moreover, regardless of what beta value was specified, the same stopping point was obtained by the FLE method. In order to confirm that the same PSFs represented the same clustering groups, the tree structure diagrams that indicated the disjoint clusters at a specified level from the EML and FLE methods were examined. They were found to be identical. Therefore, the default value in SAS may adequately serve the purpose of constructing seasonal factor groups.

Table 2 also shows that the optimal number of clusters for almost all of the methods is five. Based on practical experience and as suggested by the TMG (i.e., 4 seasonal cluster groups), it is unlikely to have less than two or more than seven seasonal factor categories. Subsequently, if the PSF criterion suggested more than one optimal number of clusters for a method such as the MED method, only the pooled variances for those that fell between 2 and 7 were calculated.

TABLE 2 PSFs at Different Hierarchical Levels for Various Clustering Methods

Number of Cluster	AVE	CEN	EML					FLE		MCQ	MED	SIN	WAR
			1	1.25	1.5	1.75	2	-0.25	-0.5				
17	10.90	10.90	10.90	10.90	10.90	10.90	10.90	10.90	10.90	10.90	10.90	10.90	10.90
16	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00
15	9.20	9.20	9.20	9.20	9.20	9.20	9.20	9.20	9.20	9.20	9.20	9.20	9.20
14	8.60	8.20	8.60	8.60	8.60	8.60	8.60	8.60	8.60	8.60	8.20	8.20	8.60
13	8.30	8.30	8.40	8.40	8.40	8.40	8.40	8.40	8.40	8.30	8.30	7.20	8.40
12	8.50	8.50	8.50	8.50	8.50	8.50	8.50	8.50	8.50	8.50	8.50	4.70	8.50
11	8.80	8.80	8.80	8.80	8.80	8.80	8.80	8.80	8.80	8.80	8.80	5.40	8.80
10	9.00	9.00	9.00	9.00	9.00	9.00	9.00	9.00	9.00	8.70	9.00	6.10	9.00
9	9.20	9.00	9.20	9.20	9.20	9.20	9.20	9.20	9.20	9.20	9.00	4.30	9.20
8	9.60	8.40	9.60	9.60	9.60	9.60	9.60	9.60	9.60	9.60	7.70	5.10	9.60
7	10.30	7.80	10.30	10.30	10.30	10.30	10.30	10.30	10.30	10.30	8.20	5.90	10.30
6	10.50	8.40	10.50	10.50	10.50	10.50	10.50	10.50	10.50	9.40	5.60	6.30	10.50
5	10.70	8.40	10.70	10.70	10.70	10.70	10.70	10.70	10.70	9.70	6.90	3.90	10.70
4	8.90	10.90	9.60	9.60	9.60	9.60	9.60	8.90	9.60	8.90	7.40	4.50	9.60
3	6.60	6.60	9.30	9.30	9.30	9.30	9.30	7.70	9.30	6.40	8.60	6.60	9.30
2	5.40	5.40	10.20	10.20	10.20	10.20	10.20	7.20	10.20	7.20	8.20	5.40	10.20

In addition, for the CEN, MED, and SIN methods, the corresponding pooled variances at the number of clusters of five comparions purpose. For simplicity, the 13 clustering methods were first categorized into five method groups according to their clustering results. They are:

Group 1: AVE, EML (penalty = 1, 1.25, 1.5, 1.75, 2), FLE (beta = -0.25, -0.5), WAR
Group 2: CEN
Group 3: MCQ
Group 4: MED
Group 5: SIN

Table 3 shows the calculated pooled variance for each group at different levels of the hierarchy. The gray cells indicate the pooled variance for the number of clusters recommended by the PSF criterion. The cells shaded with horizontal lines indicate more outliers, i.e., single TTMS in a cluster. These outliers were detected and eliminated from the calculation of pooled variances. The "crossed-out" cells were discarded due to the difference in sample size. Table 3 indicates that the Group 3 method, i.e., the MCQ method, produced the least pooled variance at five clusters and was defined as the optimal clustering method in this study.

TABLE 3 Pooled Variance for Various Clustering Groups

Method Group	Number of Clusters			
	3	4	5	6
1			4.37	
2		4.92	4.78	
3			4.18	
4	7.40		4.22	
5	6.09		5.08	4.02

The spatial patterns of the TTMS clusters from the MCQ method were examined. Figure 4 illustrates the seasonal factor groups determined by the MCQ method for five cluster groups, with each TTMS labeled by its group number between 1 and 5. As shown in Figure 4, the two TTMSs on the Florida Turnpike were assigned to Category 1. Category 2 included the TTMSs that were located on major roads close to Florida Turnpike on its west side. Category 3 included two TTMSs that were located on major roads far west of Palm Beach County. The two TTMSs located on major roads near the Palm Beach and Martin county line were assigned to Category 4. The last category included nearly all the rest of the TTMSs, including four located on Interstate 95 (I-95) and six on major roads that were on the east side of Florida Turnpike and/or I-95, whichever lied further west. The spatially clustering patterns shown in Figure 4 suggest that it is not appropriate to merge the TTMSs on the Florida Turnpike with those located on regular major roads or interstate highways. In addition, roadway functionality does not seem to play an important role in determining seasonal groups. It is the location of a given TTMS that matters since a TTMS tends to be clustered with those in its proximity.

Because the seasonal groups determined by the MCQ method were logical and reasonable, the method was implemented to analyze the MSFs collected from 1997 through 1999. The three TTMSs that were previously excluded from the year 2000 data were revealed again as outliers for the data from 1997 through 1999. The three stations were thus excluded from the three-year dataset. Figures 5 to 7 illustrate the seasonal factor groups determined by the MCQ method for years 1997, 1998, and 1999, respectively, after the outlier stations were eliminated. The numbers of seasonal groups were 7, 6, and 6 for these three years, respectively.

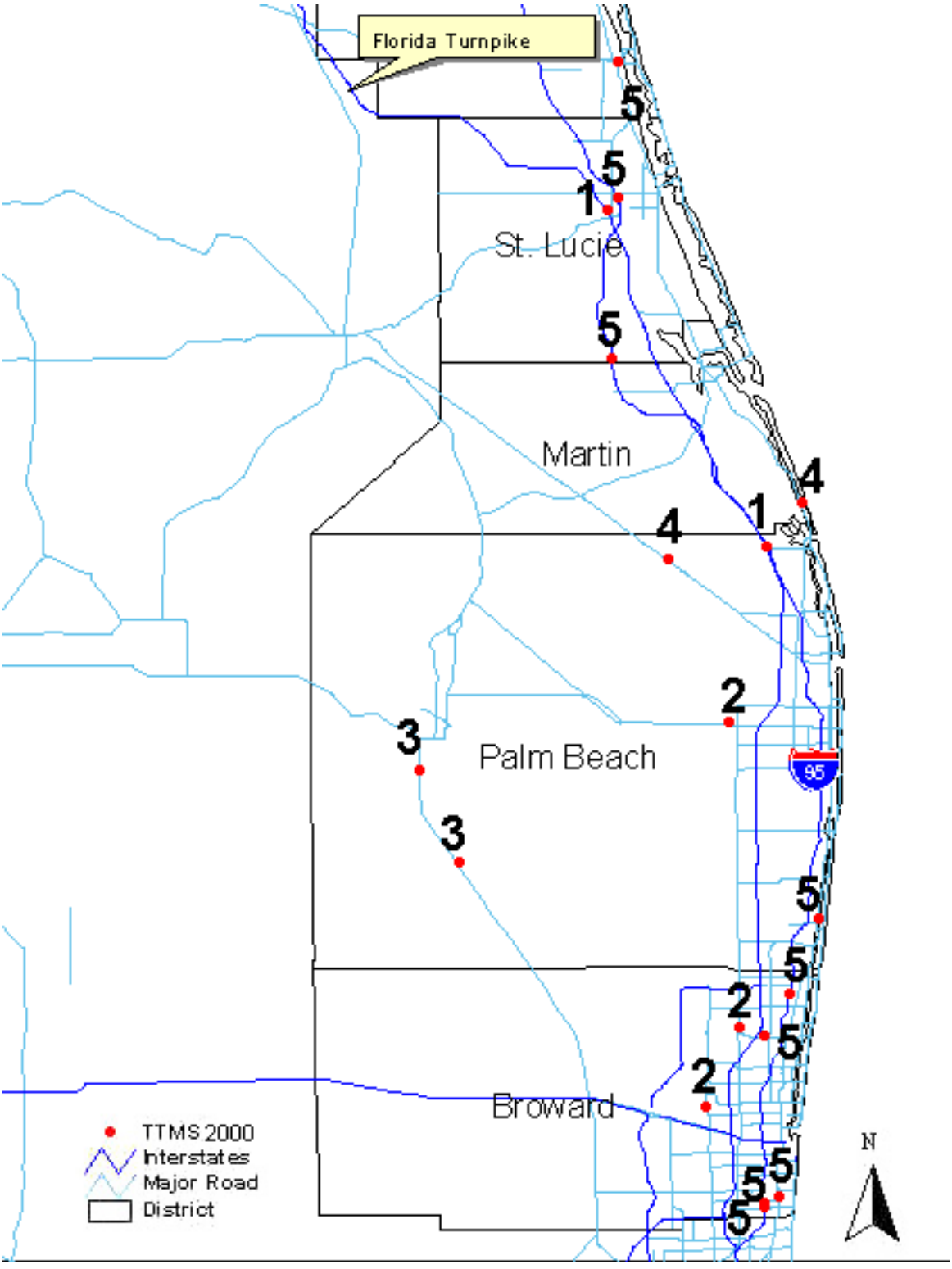


FIGURE 4 Seasonal Cluster Groups Determined by the MCQ Method (Year 2000).

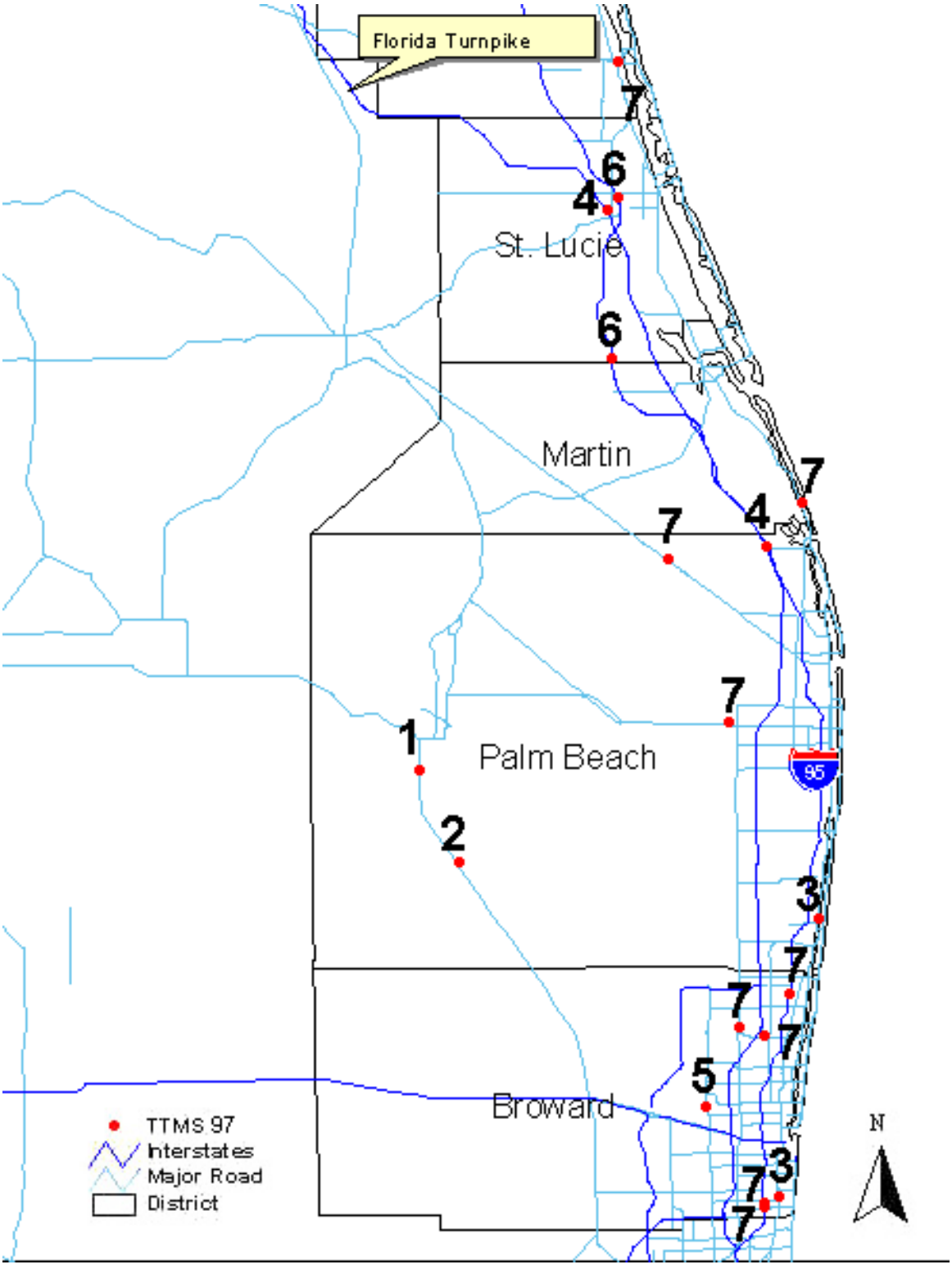


FIGURE 5 Seasonal Cluster Groups Determined by the MCQ Method (Year 1997).

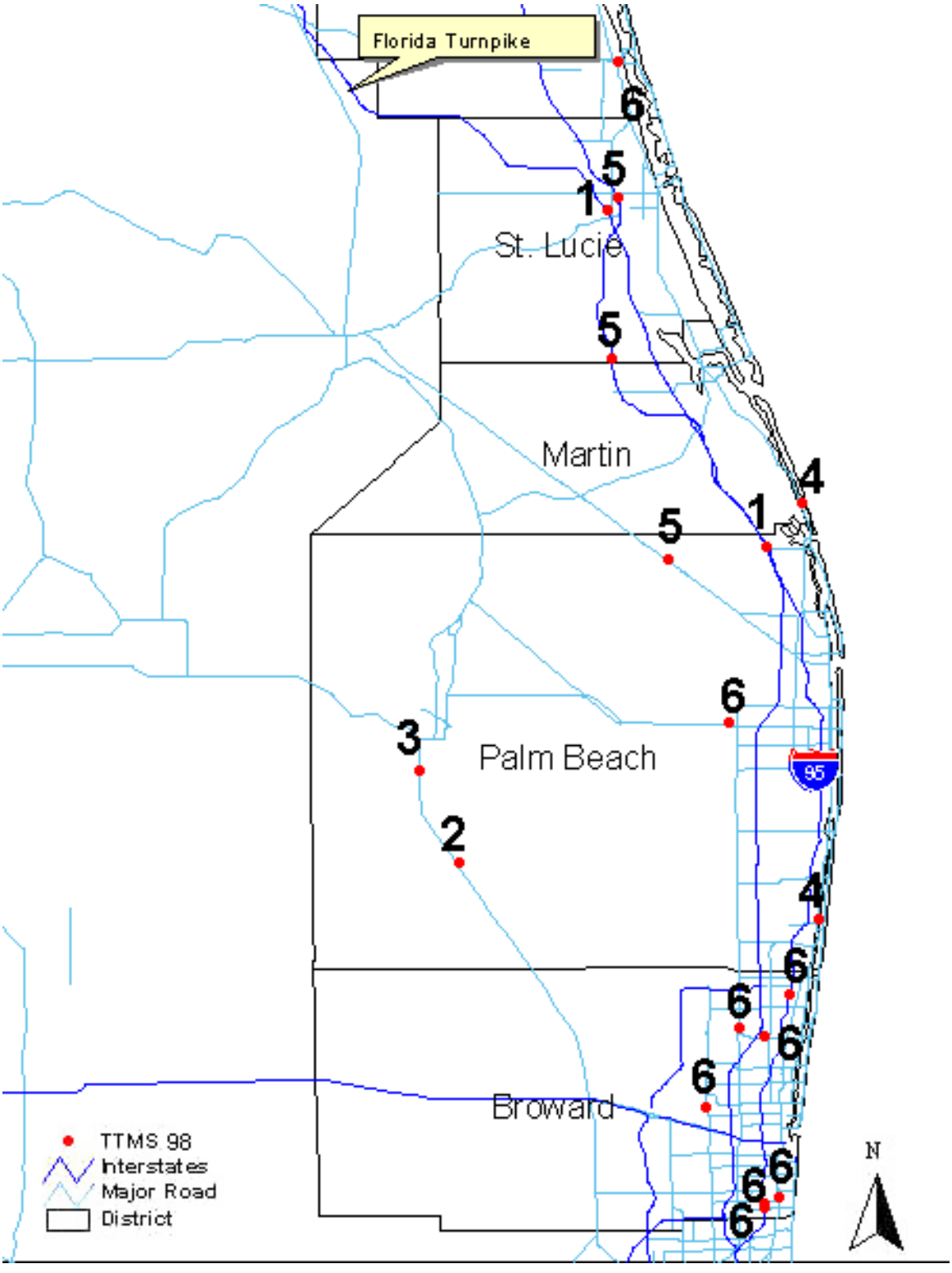


FIGURE 6 Seasonal Cluster Groups Determined by the MCQ Method (Year 1998).

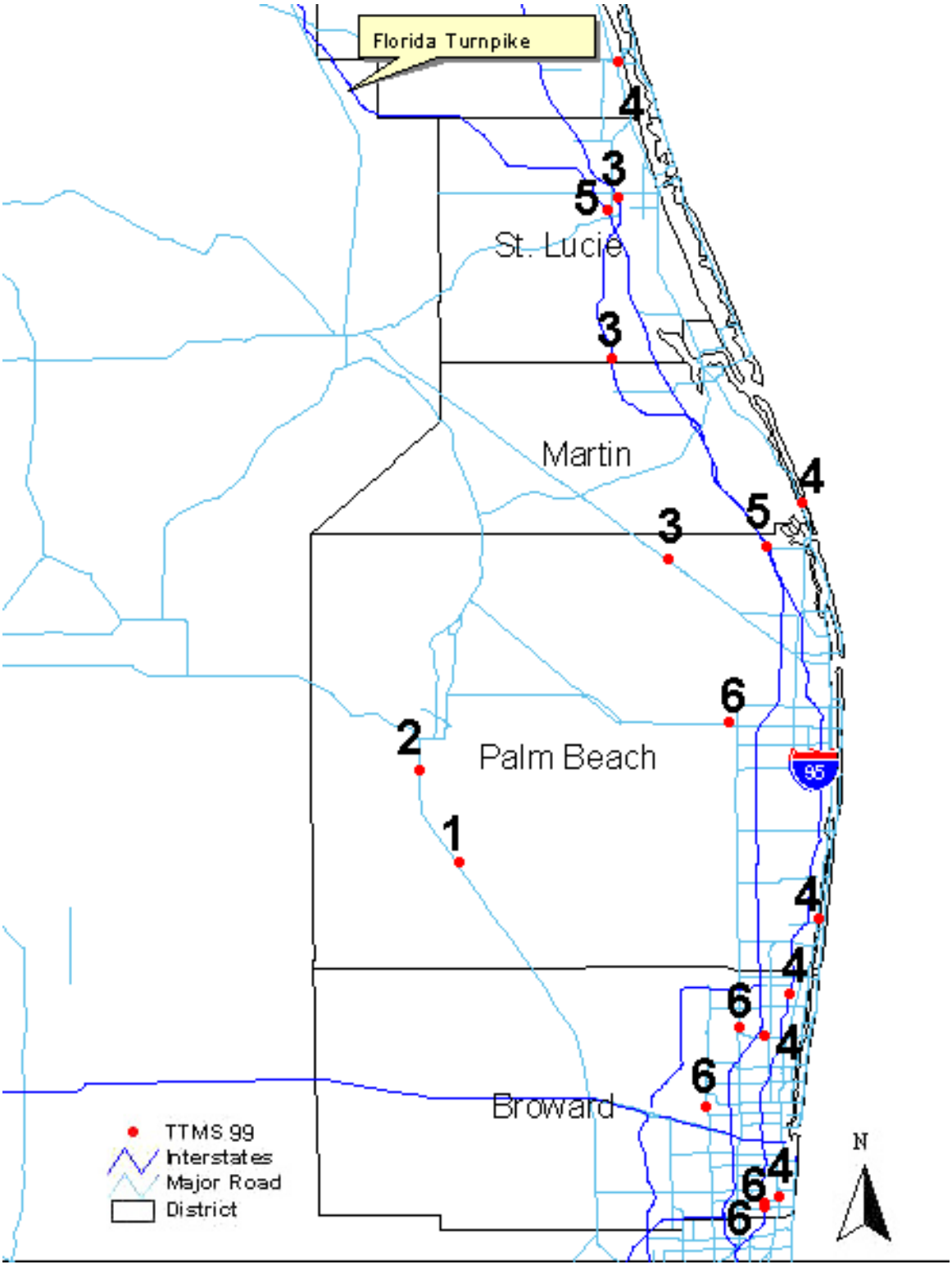


FIGURE 7 Seasonal Cluster Groups Determined by the MCQ Method (Year 1999).

Figures 5 to 7 show that the MCQ method did not consistently assign a TTMS to the same group from year to year. However, the change in the grouping patterns during the four-year period revealed a gradual shift in spatial clustering patterns from north-south direction to east-west direction. For example, although Figure 5 does not show any significant spatially clustered pattern among the TTMSs in 1997, Figure 6 shows that two relatively larger groups, i.e., Groups 5 and 6, were formed at the north and south sides of the district. Figure 7 shows that the spatially clustering patterns started to shift from north-south direction to east-west direction, similar to what was illustrated in Figure 4 for the year 2000 data. One of the possible reasons for such a shift could be due to traffic that circulates around the major activity centers within each county in/before 1997. The traffic going from one county to another became increasingly significant with time and the interstate freeways, including I-95 and Florida Turnpike, were then highly utilized by the cross-county traffic. Similar seasonal fluctuation patterns were thus generated for the TTMS along the freeways. After the freeways became congested around 1998 and 1999, traffic started to shift to local streets. Since the land uses and developments along the east and west sides of the I-95/Florida Turnpike were different, two distinguished seasonal patterns were thus formed.

CONCLUSIONS

A total of eight agglomerative clustering methods have been evaluated in this study. The average linkage, centroid, and single linkage methods were found to be more robust to outliers than the other methods. The study also found that the McQuitty's (MCQ) method performed better than the other methods on grouping TTMSs after outliers were eliminated. Although the results from analyzing the four-year MSF data with the MCQ method shows that the compositions of seasonal groups were not stable over time, the change in the spatially clustering pattern suggested that other variables should be included in the process of determining seasonal cluster groups over multiple years. The study also led to the finding that roadway functional classification did not seem to play an important role in determining seasonal groups; rather, it is the spatial location of a given TTMS that matters.

This paper has outlined a procedure that allows practitioners to construct seasonal factor groups from TTMS sites. Based on data from FDOT District 4, the PSF statistic was found to be a good measure to determine the number of clusters after possible outliers were excluded. Outliers need to be further studied to investigate possible causes of their unique monthly/weekly traffic patterns. It should be noted that the conclusions of this study could not be generalized to all other districts in Florida or to other states as they were drawn based only on the traffic stream conditions of a specific district in Florida. Similar studies may be performed in other areas to determine the robustness of the evaluation procedure that has been employed in this research.

ACKNOWLEDGEMENTS

This research was sponsored by a grant from the Florida Department of Transportation (FDOT). The opinions, findings, and conclusions expressed in this paper are those of the authors and not necessarily those of the FDOT.

REFERENCES

1. *Project Traffic Forecasting Handbook*, Florida Department of Transportation, Tallahassee, FL, April 2002.

2. *Traffic Monitoring Guide*, Federal Highway Administration, U.S. Department of Transportation, Washington D.C., May 2001.

3. Davis, G.A. and Y. Guan, "Bayesian Assignment of Coverage Count Location to Factor Groups and Estimation of Mean Daily Traffic", *Transportation Research Record 1542*, TRB, National Research Council, Washington, D.C., 1996, pp. 30-37.

4. Sharma, S.C. and A. Werner, "Improved Method of Grouping Provincewide Permanent Traffic Counters," *Transportation Research Record 815*, TRB, National Research Council, Washington, D.C., 1981, pp. 12-18.

5. Sharma, S.C., "Improved Classification of Canadian Highways According to Type of Road Use," *Canadian Journal of Civil Engineering*, No. 3, Vol. 10, 1983, pp. 497-509.

6. Sharma, S.C., "Design Hourly Volume from Road Users' Perspective," *Journal of Transportation Engineering*, ASCE, No. 4, Vol. 112, 1986, pp 435-440.

7. Flaherty, J., "Cluster Analysis of Arizona Automatic Traffic Record Data," *Transportation Research Record 1410*, TRB, National Research Council, Washington, D.C., 1993, pp. 93-99.

8. Aunet, B., "Wisconsin's Approach to Variation in Traffic Data," *North American Travel Monitoring Exhibition and Conference CD*, Wisconsin Department of Transportation, Madison, WI, August 2000.

9. Ritchie, S.G., "A Statistical Approach to Statewide Traffic Counting," *Transportation Research Record 1090*, TRB, National Research Council, Washington, D.C., 1986, pp. 14-21.

10. Faghri, A., M. Glaubitz, and J. Parameswaran, "Development of Integrated Traffic Monitoring System for Delaware," *Transportation Research Record 1536*, TRB, National Research Council, Washington, D.C., 1996, pp. 40-44.

11. *SAS-OnlineDOC, Version 8*, SAS Institute Inc., Cary, NC, 1999.

12. Johnson, R.A. and D.W. Wichern, *Applied Multivariate Statistical Analysis, Fifth Edition*, Prentice-Hall, Inc., Upper Saddle River, NJ, 2002.

13. Faghri, A. and J. Hua, "Roadway Seasonal Classification Using Neural Networks," *Journal of Computing in Civil Engineering*, No. 3, Vol. 9, July 1995, pp. 209-215.