

Aggregation Effects in Air Traffic Arrival Flows

James DeArmon

The MITRE Corporation, McLean, Virginia 22102

Air traffic management includes the function of traffic flow management or the balancing of air traffic demand and capacity. Since runway capacity is rather constant, at least in the short term, the flow manager typically adjusts air traffic demand when there is a predicted imbalance. Adjustment of air traffic demand takes two forms: in time, through delays, and in space, through reroutes. Local flow managers who regulate the arrival traffic bound for a capacity-constrained airport are continually faced with the question: How much flow control intervention is appropriate now? This paper attempts to structure that question using analytic methods. With refinements, this technique could become one of a suite of decision support tools.

Introduction

AIR traffic management has two components: separation assurance and traffic flow management (TFM). Separation assurance, also known as air traffic control (ATC), is the function of providing safe separation between pairs of aircraft. TFM, by contrast, manages groups or aggregations of flights, seeking to balance the demand for service with the available capacity. One of the standard problems in TFM is the structuring of arrival traffic flows in anticipation of their entering the approach control and then landing on a runway. As part of a research and development initiative funded by the Federal Aviation Administration, we are investigating TFM decision support tools that would help with this problem. A major theme in the field of TFM analysis and research is the effect of system uncertainty on air traffic flows. We couch the research question here in terms of TFM intervention versus uncertainty. We demonstrate how statistical methods may help in recommending an arrival schedule.

This paper explores a technique whereby statistical cluster analysis and applied probability methods are used to detect arrival flow "bunching," that is, a surge of demand bound for an airport. This technique should not be seen as competing with the substantial successes of arrival metering programs now active in the field. Rather, it should be seen as complementary, one of a suite of available decision support tools for the traffic flow manager.

The problem of smoothing or "metering" of inbound arrival flows has interested researchers and developers in both industry and academia. In the American National Airspace System, two major initiatives are active. The EnRoute Spacing/Arrival Sequencing Program automation has been available to traffic managers and controllers on the Host computer system for several years.¹ A more intricate algorithm/software system, the Center Terminal Radar Approach Control (TRACON) Automation System (CTAS), is being prototyped at several sites with good success.^{2–4} Researchers in Europe (Germany⁵ and the United Kingdom⁶) are similarly engaged in addressing the problem. Vandevenne and Andrews⁷ have examined the smoothing problem in light of "controllability," a concept similar to TFM intervention as discussed in the present paper.

The paper is organized as follows. The next section describes the test scenario used to investigate the subject decision support technique. This is followed by a description of cluster analysis applied to the detecting of bunching. The next two sections address the topic of uncertainty and the experimental setup. The final two sections present results and conclusions.

Test Scenario

As part of our research on TFM, we have developed scenarios to study specific flow problems. We hypothesize the following scenario, based on actual operational data, for investigating the subject decision support technique.

A local flow manager (i.e., the traffic management coordinator at the Traffic Management Unit at an Air Route Traffic Control Center) uses his or her decision support tool and finds an aggregate or clustering of inbound arrival traffic about 45 min from landing at Chicago's O'Hare airport. The Air Traffic Control System Command Center, the national flow management component, has intervened earlier to achieve a rough demand/capacity balance. However, the local flow manager is confronted with uneven demand at a finer granularity: an approximate 30-min period that begins about 45 min from the current time. See Fig. 1, where zeros indicate individual aircraft a given number of minutes from landing at Chicago. For the total time displayed, 41–68 min from Chicago, there is a demand/capacity balance: 30 arrivals within 28 min matches an airport acceptance rate of 64 per hour rather well. However, the demand structure within the 28-min window is uneven. A first-pass cluster analysis (cluster analysis is described in the next section) for a superset of the sampled time, 120 min of inbound arrival flights, yielded three clusters summarized in Table 1 within this 28-min window, indicated in Fig. 1 by the notation: a—b.

The question is: Given this situation, how much should the local flow manager intervene at this time? Intervention will lead to some flattening of demand across the 28-min window. How much flattening should be performed? The trade-off is obvious. Too little intervention will leave an unevenness of demand that may be



James DeArmon is a Lead Engineer with The MITRE Corporation. He has worked for the last several years in research and modeling of air traffic flow dynamics. He holds a Master's degree in operations research, and has taught graduate courses at the University of Maryland. He has written numerous papers in applied math and computer science.

Table 1 Summary of results of first-pass cluster analysis on original demand structure

Time range, (to landing), min	Aircraft count
41–50	17
52–58	10
65–68	3

Clustering	
Time	Aircraft
(41) a	0
(42) l	000
(43) l	000
(44) l	
(45) l	00
(46) l	00
(47) l	00
(48) l	00
(49) l	
(50) b	00
(51) l	
(52) a	0
(53) l	00
(54) l	0
(55) l	0
(56) l	00
(57) l	
(58) b	000
(59) l	
(60) l	
(61) l	
(62) l	
(63) l	
(64) l	
(65) a	0
(66) l	
(67) l	0
(68) b	0

Fig. 1 Results of first-pass cluster analysis on original demand structure.

more costly to flatten downstream in terms of fuel, controller stress, consumed degrees of freedom, etc. Too much intervention, i.e., too much delay assignment now, may result in missed arrival landing slots. Is it possible to apply analytical methods to aid the decision process or must the decisions be of an ad hoc nature, based solely on experience?

This paper reports on some analysis performed to answer, or at least to structure a framework for answering, that question. A theme want to entertain here is working with aggregates and not individual aircraft. There has been much conjecture that TFM should deal with flow aggregations as its control elements, leaving the separation assurance function to deal with individual aircraft as its control elements. The analysis here pursues that conjecture by working with aggregates as much as possible.

Another major idea here is the use of probability theory to model the variability inherent in an inbound air traffic flow. Because of uncertainties in actual arrival times, the best estimate of the demand structure at time of landing computed 45 min prior to landing (histogram in Fig. 1) may not represent the actual demand. Although the Fig. 1 demand structure is the current best estimate, the probability of the demand occurring in precisely that way is quite small. Rather than an aircraft's (trajectory-derived) time to landing being treated as a known, deterministic value, it will be modeled here as a random variable.

Cluster Analysis for Peak Detection

We must first establish a quantitative definition for "uneven" demand. To do so, we will attempt to translate into numbers what is depicted visually by the histogram in Fig. 1. Examining this histogram, we assert that the demand is too uneven—visually, there is too much demand in the beginning and too much unused capacity at the end of the 28-min window. This period was selected from several hours of scenario data as being an interesting case of demand unevenness.

One might question why the analysis is limited to this interval. Are there not demand–capacity situations just prior to and following the 28-min window whose inclusion could benefit our analysis? There are, but one needs to package the problem into feasible units of work

Table 2 Maximum observed arrival count for time intervals

Interval, min	Maximum, arrival flight count
1	3
2	5
3	6
4	8
5	9
6	10
7	11
8	12
9	14
10	16

Clustering	
Time	Aircraft
(41) a	0
(42) l	000
(43) b	000
(44) l	
(45) a	00
(46) l	00
(47) b	00
(48) a	00
(49) l	
(50) b	00
(51) l	
(52) a	0
(53) l	00
(54) l	0
(55) b	0
(56) a	00
(57) l	
(58) b	000
(59) l	
(60) l	
(61) l	
(62) l	
(63) l	
(64) l	
(65) a	0
(66) l	
(67) l	0
(68) b	0

Fig. 2 Result of second-pass cluster analysis on original demand structure.

appropriate to the flow manager's scope of interest. This issue could be an area of research: What is the right size of the work unit? Does it change? Can work units overlap?

We do not really know whether the demand depicted in this histogram is too uneven, but we can use operational data as a guide in making this determination. To translate our visual assessment of uneven demand into a quantification, we first perform a finer, second-pass cluster analysis on the 28-min window of inbound arrival flights, seeking, say, six clusters. [Several experiments varying the number of clusters were performed, and through visual assessment of results, six seemed like a good choice. The clustering algorithm we use is the Fisher-optimal partition procedure.^{8,9} The algorithm minimizes within-group distance (total sum of squared deviation from the group mean), given a set of observations and a desired number of clusters.]

To understand what may constitute an excessive peak in clustering, we set up a table of thresholds that defines the maximum number of landings over a time interval from operational data. To accomplish this, we analyzed actual arrival times for Chicago during a period when the airport acceptance rate was equal to 64 aircraft per hour. Since the cluster algorithm generates clusters with various time ranges, we must specify a full range of thresholds. Therefore, we computed the maximum observed 1-min arrival count, 2-min arrival count, etc., these maxima representing the thresholds of what the system can handle for various intervals. A portion of the results is presented in Table 2. Note that shorter time ranges allow for more peaking; for example, 2 flights in 1 min can probably be observed fairly frequently, whereas 20 in 10 min may be quite rare. We will define a high-density cluster (HDC) as a cluster of arrivals that exceeds the maximum observed number of arrivals.

The demand structure of Fig. 1 shown again in Fig. 2 following the second-pass cluster analysis, again with the notation a—b indicating

the clusters. The first cluster, time range 41–43, has seven aircraft and is a HDC. This situation hence has a demand structure that has too much peaking, based on our table of maximum observed rates from operational data. If the system were deterministic, i.e., if all 30 of these inbound arrival flights satisfied their current best estimate of landing time exactly, then the current assessment of overly uneven demand would hold 41–68 min hence, and TFM intervention (now) would be indicated.

At this point, we have defined a simple criterion for declaring a demand structure too uneven: The cluster analysis forming six clusters yields one or more HDCs when compared with operationally derived maximum rates. More sophisticated analyses of demand structure are worth pursuing in continuing research.

Uncertainty

A further step in the analysis is needed. Our conclusion that the demand structure is too uneven is based on the faulty assumption that the times to landing are deterministic. Rather, they should be considered as being drawn from probability distributions with a nonzero variance. What really happens is that the remaining flight time constitutes exposure to the vagaries or random effects in the system, and there is some chance that a flight will arrive somewhat earlier or later than the best arrival estimate now known. Random effects could also be called unmodeled effects, since most modern ATC automation systems do model some exogenous effects, such as winds, in building/maintaining the trajectory. These random effects create some natural dispersion that has the effect of diluting clusters or at least affecting clusters in some way (i.e., it could happen that random effects would cause the aggregation of flights not currently aggregated).

We need to assess the potential impact of random effects on our demand structure. For this purpose, we turn to applied probability and Monte Carlo sampling. First, a characterization of the probability distribution of the inbound arrival flights is required. Typically, a lognormal probability density function characterizes a “time until completion of a task” type of distribution. The basic shape of the distribution is shown in Fig. 3 and can be interpreted as follows: 1) zero probability of completing the task earlier than now (left-tail truncation below $x = 0$), 2) some probability of completing the task early relative to \bar{X} (area between $x = 0$ and hump of curve), and 3) higher probability of completing the task late (right-tail asymptotic damping).

The lognormal probability density function has two parameters (shape and scale), which can be derived from the mean and standard deviation of the closely related normal distribution. For the application at hand, we use the trajectory-derived time to landing as the mean, and for purposes of experimentation and research, we parameterize the standard deviation as some percentage of the mean. (In statistical terms, the ratio of the standard deviation to the mean is known as the coefficient of variation.)

In actual practice, these parameters should be observed and calculated based on actual data. In this paper, our goal is to understand the relationship between the degree of uncertainty and the amount of intervention necessary. Therefore, rather than using specific empirical values of uncertainty, we study the effect of this uncertainty and the degree of TFM intervention upon the outcome of actual (i.e., realized) demand at the runway.

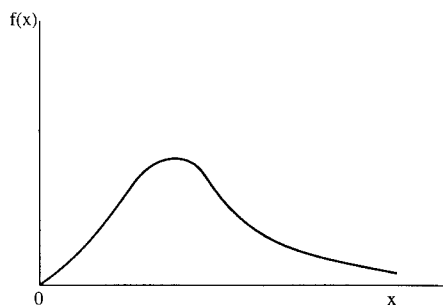


Fig. 3 Typical lognormal distribution.

To do this, we generate some large number (say, 1000) of “realizations” of the process. A realization is essentially a single sample of what the demand structure might look like at the time of landing. The 30 flights in the 28-min interval are subjected to two processes: 1) TFM intervention, subjecting certain flights to specific delays, and then 2) uncertainty, whereby flights are deviated as a result of random effects in the system. For each of the 30 subject flights, using the mean and standard deviation as described above, we draw one variate as one possible time at which the flight might land. (The variates are thus treated as independent, although this is a simplification. Actually, airspace events, wind errors, and relationships of proximate aircraft would imply correlated, as opposed to independent, times. In follow-on work, more accurate modeling should be pursued.)

These 30 new, randomized landing times are considered in concert as a single realization of the stochastic process. Each realization is evaluated to find HDCs. As a final step, we can state the probability of overly uneven demand (i.e., at least one HDC at landing and hence some potential need for TFM intervention now), given the current state of the demand structure and a specified amount of uncertainty (characterized by the log normal probability density function).

Note there is an important link between the function used to detect overly uneven demand and the stochastic process: If the thresholds for HDC detection were lower, then we would expect the probabilities derived from a sample of realizations to rise. If the concept described in this paper were to become part of a decision support aid, site adaptation and calibration would be needed to set the thresholds based on observed uncertainties.

Experimental Setup

Given the structure described above, we can perform experiments and learn something about the dynamics of the system. Two inputs are varied in the experiments: 1) the standard deviation of the probability distribution used for random sampling parameterized as a percentage of the mean and 2) the level of TFM intervention to be performed now, the units being the number of flights delayed, in a first-come, first-served discipline, in the 28-min window.

Again, each flight’s current trajectory-derived time to landing is used as the mean of a probability distribution. These means are inputs to the experiment but are not to be varied. There is one output from the model: the probability that at least one HDC exists in the demand structure at the time of landing.

As an example, for the first experiment we assume that the standard deviation is 3% of the mean and that we perform no TFM intervention now. In recent tests at Denver, estimated time of arrival errors being achieved with CTAS² are less than 3%. The result after 1000 samples is that 463 realizations show at least one HDC and 537 do not. In the language of probability and statistics, we would say that given the current demand structure, there is a 46% (463/1000, rounded) chance that at least one HDC will exist in the demand structure of these 30 aircraft at the time of landing. We must now ask: How sensitive is the outcome probability to variability in the flight times? How much would TFM intervention now help lower the outcome probability?

We try another standard deviation, 15% of the mean, which is probably more realistic for systems without accurate trajectory prediction systems similar to those in CTAS. (As analysis of Host-Z data for flights about 45 min away from Chicago showed a standard deviation-to-average ratio of about 15%. Admittedly, this real-world data is not exactly appropriate. It contains in it the delay due to flow control, the very impact we are modeling in this paper. It would be challenging to capture real-world data without this effect. The 15% value is hence approximate and for illustration.) Also, we try three TFM intervention strategies for flattening the current demand structure to varying degrees to reduce the outcome probability. These three interventions are labeled Delay-5 and Delay-12 for, respectively, the delaying (now) of 5 or 12 aircraft, and Full Intervention. Note that full intervention is the extreme of completely flattening the demand and might also be called Delay-26, since it requires 26 aircraft delays. (A broader range of levels of intervention was

tested, and these two intermediate levels, 5 and 12, were selected for presentation here.)

We assume for simplicity that maneuvers are available and feasible for ATC and the pilots to satisfy these strategies. Also, we assume that these reassigned times will follow the same lognormal distribution, although in reality, a new family of random effects is probably introduced. Not considered here is the fact that delay taken now at higher altitude will burn less fuel than delay taken later (even if it is due to “random effects” and not TFM directives) at lower altitude. This trade-off should be considered in follow-on research.

As an aside, in keeping with the theme in this paper, we relied on applied statistics to formulate the Delay-5 and Delay-12 strategies. We attempted to handcraft some first-come, first-served strategies (sets of delay assignments) at various levels of intervention. This was a daunting task and after some trials was not very successful. As an alternative, randomized selection was undertaken. That is, for the Delay-5 and Delay-12 strategies, we tried 100 random sets of delay assignments (each set sampled 1000 times for the lognormal probability draws that yield 1000 realizations) and the best [lowest $P(\text{HDC})$] was chosen. Expressed in pseudocode:

```
do  $i = 1 \dots 100$ 
  generate  $i$ th random first-come, first-served delay set
  do  $j = 1 \dots 1000$ 
    perform log normal sampling to generate  $j$ th arrival
    times realization
    evaluate HDC for  $j$ th realization
  end do
  calculate  $P(\text{HDC})$  for  $i$ th first-come, first-served delay set
  save best [lowest  $P(\text{HDC})$ ] first-come, first-served delay set
end do
```

This approach is rather abstracted, at some remove from the reality of ATC. In reality, the flow manager or air traffic controller would have appreciation of which flights were candidates for delay. On the other hand, it is not unreasonable to rely on computer automation to aid the decision process: Given that a certain level of intervention is required, say, five flights delayed first-come, first-served, which five flights delayed give the best response in lowering $P(\text{HDC})$? Automation could quickly generate good candidate solutions. This is another research area beyond the scope of the present discussion. Results of the experiment are sensitive to the details of an intervention strategy.

Results

The results of the two-standard-deviation, four-strategy experiments are presented in Table 3. These outcomes possess an analysis-of-variance model structure. Proceeding down rows (within a column), we see the TFM intervention effect; proceeding across columns (within a row), we see the uncertainty effect. Considerable refinement of the experiment would be necessary before these trends and the interaction of these main effects could be studied, especially before we could proceed down rows and across columns simultaneously.

The trends in the probabilities can be interpreted as follows. Proceeding down rows within a column, probabilities fall: As the TFM intervention effect increases, flattening the current demand structure, the probability of an HDC at the runway generally falls, as we would expect. Comparing strategies, Delay-5 and Delay-12 both yield improvements, as evidenced by the falling outcome probabilities. The decline is steeper for the 3% column.

Table 3 Probability of HDC of demand at time of landing

Strategy	Standard deviation ^a	
	3%	15%
No intervention	0.46	0.23
Delay 5	0.39	0.21
Delay 12	0.35	0.17
Full intervention	0.10	0.13

^a As percentage of mean.

The general trend within the first three rows is that as the uncertainty effect increases from a 3 to a 15% coefficient of variation, natural dispersion is helping to break clusters up, and the probability of HDC at the runway falls. But the trend across columns is different in the last row. With this level of intervention, the probability *increases* when uncertainty is greater, i.e., comparing 3% with 15%. For this row, if the probability density function standard deviation is 15% of the mean, then natural dispersion acts contrarily and in some cases reagggregates, to a slightly greater degree than 3%, the demand that was flattened through the full intervention strategy. Uncertainty is making the impacts of intervention less tenacious.

In the 15% column, the outcome probabilities make a rather smooth decline across the reported strategies/rows. The marginal improvement in moving from strategy Delay-5 to Delay-12, four percentage points, is the same as the marginal improvement in mov-

Clustering		
Time		Aircraft
(41)	a	0
(42)	l	000
(43)	b	000
(44)		
(45)	a	00
(46)	l	0
(47)	l	00
(48)	b	00
(49)	a	0
(50)	l	
(51)	b	00
(52)		
(53)	a	0
(54)	l	000
(55)	l	0
(56)	b	0
(57)	a	0
(58)	l	
(59)	b	000
(60)		
(61)		
(62)		
(63)		
(64)		
(65)	a	0
(66)	l	
(67)	l	0
(68)	b	0

Fig. 4a Original demand structure, a candidate Delay-12 strategy.

Clustering		
Time		Aircraft
(38)	a	0
(39)	l	0
(40)	l	0
(41)	l	
(42)	l	000
(43)	b	0
(44)		
(45)	a	0
(46)	l	0
(47)	l	
(48)	b	0
(49)		
(50)		
(51)		
(52)	a	0
(53)	l	0
(54)	l	
(55)	l	00
(56)	l	0
(57)	b	00
(58)		
(59)	a	000
(60)	l	0
(61)	l	0
(62)	l	
(63)	b	0
(64)		
(65)	a	00
(66)	l	
(67)	l	0
(68)	l	
(69)	l	0
(70)	b	0
(71)	a	0
(72)	l	
(73)	l	
(74)	l	
(75)	l	
(76)	b	0

Fig. 4b Sample realization of “no HDC at landing,” Delay-12 strategy, 15% standard deviation.

ing from strategy Delay-12 to Full Intervention. However, these intervention levels are hardly linear. Full Intervention might be called a "Delay-26" strategy, since it requires that many aircraft delays (first-come, first-served) to achieve the complete flattening of demand. The marginal return, i.e., the reductions in P(HDC) due to TFM intervention, of moving from row 2 to row 3, is much better than that of moving from row 3 to row 4. These results show numerically what many TFM researchers would intuit: It may be overcontrolling to completely flatten demand when 45–60 min of flight life remains until landing without accurate trajectory prediction algorithms such as being included in such systems as CTAS. Modeling the cases here analytically offers some promise that this technology might be used effectively as a decision aid, that decisions need not be solely ad hoc, based on long experience.

For more insight into these results, we show a sample realization from the simulation. Figures 4a and 4b show two histograms for the case of the 15% standard deviation/Delay-12 strategy. Figure 4a shows an initial distribution, i.e., the demand structure now, following the Delay-12 TFM intervention, for one of the 100 candidate first-come, first-served delay sets. The histogram in Fig. 4b is available only after the Monte Carlo samples have been generated it is one of the 1000 realizations. Figure 4b shows a sample realization of "no HDC at landing." There is a 0.83 chance ($1.00 - 0.17$) of realizing a demand structure in this set. We selected this histogram from the 1000 available as having a median time range value among all the "no HDC at landing" realizations of this experiment. Figure 4b is only exemplary; as it happens, the odds of realizing the demand structure in Fig. 4b precisely are quite remote.

It may be noted that there is a parallel here regarding aggregation versus single-element conceptualization. In exploring the concept of overly uneven demand, we have tried to deal in aggregates by working as much as possible with the output of cluster analysis. Now in our probabilistic assessment, we must similarly pool our realizations into two aggregate classes: 1) demand structure with one of more HDCs at landing and 2) demand structure with no HDCs at landing. Statistically, a single realization has virtually no significance.

Conclusions

We have pursued a framework for answering the question. How much should a local flow manager intervene, given a certain demand structure of inbound flights? We have used statistical cluster analysis as a technique for identifying aircraft groupings or aggregates and

randomized sampling to generate TFM intervention strategies and demand profiles of aircraft landings. We have demonstrated how uncertainty analysis might be used to structure the decision process. Further research is needed in cluster detection, reassignment strategies and their effects, operational data collection, and determination of the values for the appropriate probability distributions. Also of interest for follow-on investigation are appropriate HDC levels, time frames to be used and their relationship (e.g., nonoverlapping, overlapping, or sliding), and the effect of different lead times. We believe the technique of comparing flow interventions in light of uncertainty as presented here could be useful as part of a suite of TFM decision support tools.

Acknowledgments

The author acknowledges helpful discussions with Kerry Levin, George Swetnam, and Paul Ostwald of the Center for Advanced Aviation System Development.

References

- ¹Cherone, R. L., and Ostwald, P. A., "Traffic Management Coordinator User's Guide for ESP/ASP," MITRE Corporation, MTR 90W00111, McLean, VA, Sept. 1990.
- ²Erzberger, H., Davis, T. J., and Green, S., "Design of Center-TRACON Automation," *Proceedings of the AGARD Guidance and Control Panel 56th Symposium on Machine Intelligence in Air Traffic Management* (Berlin, Germany), AGARD, 1993, pp. 11-1-11-12 (AGARD-CP-538).
- ³McLaughlin, M. P., "Transition/Terminal Air Traffic Control Automation—An Analysis of Two Planning Aids," MITRE Corporation, MTR 88W00087, McLean, VA, 1989.
- ⁴Spencer, D., Welch, J., and Andrews, J., "An Experimental Examination of the Benefits of Improved Terminal Air Traffic Control Planning and Scheduling," *Lincoln Laboratory Journal*, Vol. 2, No. 3, 1989, pp. 527–536.
- ⁵Voelckers, U., and Schenk, H., "Operational Experience with the Computer Oriented Metering Planning and Advisory System (COMPAS), at Frankfurt, Germany," *Proceedings of the AIAA Guidance, Navigation and Control Conference* (Boston, MA), AIAA, Washington, DC, 1989 (AIAA Paper 89-3627).
- ⁶Magill, S., Tyler, A., Wilkinson, E., and Finch, W., *Computer-Based Tools for Assisting Air Traffic Controllers With Arrivals Flow Management*, RSRE Report No. 8801, Royal Signals and Radar Establishment, Great Malvern, England, Sept. 1988.
- ⁷Vandevenne, H. F., and Andrews, J. W., "Effects of Metering Precision and Terminal Controllability on Runway Throughput," *Air Traffic Control Quarterly*, Vol. 1, No. 3, 1994, pp. 277–297.
- ⁸Hartigan, J. A., *Clustering Algorithms*, Wiley, New York, 1975.
- ⁹DeArmon, J. S., "Air Traffic Demand Assessment Through Cluster Analysis," *Air Traffic Control Quarterly*, Vol. 1, No. 1, 1994.