RELATIONSHIP OF THE AIRCRAFT MIX INDEX WITH PERFORMANCE AND OBJECTIVE WORKLOAD EVALUATION RESEARCH (POWER) MEASURES AND COMPLEXITY RATINGS

Elaine M. Pfleiderer, FAA, Oklahoma City, OK

Abstract

Aircraft mix (i.e., the mix of aircraft with different performance characteristics in a sector) has been repeatedly cited as a complexity factor in en route air traffic control. However, little attention has been focused on examining this relationship statistically. The present study is the third in a series of investigations designed to define, quantify, and assess the validity of aircraft mix as a contributor to traffic complexity. Eighteen 30-minute samples of System Analysis Recording (SAR) data were collected from the Fort Worth and Atlanta en route centers. Performance and Objective Workload Evaluation Research (POWER) measures and the Aircraft Mix Index were computed in 6-minute intervals for each of the 36 samples. Principal Components Analysis (PCA) of the combined data sets produced four components with eigenvalues >1 accounting for approximately 71% of the variance. The Aircraft Mix Index was most closely associated with Component 1, which was composed of variables generally associated with traffic complexity. These variables were used as predictors against a criterion of controllers' subjective "Complexity" ratings in multiple regression analyses of low- and high-altitude sector samples. The Aircraft Mix Index failed to contribute significantly to the explained variance in the both the low-altitude (R=.69; R^2=.47) and high-altitude (R=.57; R^2=.33) sector models. In the aggregate, the results suggest that although aircraft mix appears to be associated with traffic complexity, it may not be as influential as other complexity factors in the en route environment.

Introduction

Aircraft mix has been proposed as one of the traffic characteristics contributing to sector complexity in en route air traffic control [1], [4], [5], [7], [9], [15], [19]. However, its reputation has primarily been based on anecdotal evidence with little attention focused on examining this relationship statistically. The present study is the third in a series of investigations designed to define, quantify, and assess the validity of aircraft mix as a contributor to traffic complexity in en route air traffic control.

First in the series was an investigation of the salient features of aircraft mix as it relates to aircraft performance characteristics [16]. For this analysis, 30 Certified Professional Controllers (CPCs) from several en route centers across the United States provided average speed, climb, and descent rate estimates for a sample of 30 distinct aircraft types. A matrix of squared Euclidean distances was derived from summary estimates (i.e., means of speed, climb, and descent rates) and used to construct a multidimensional scaling (MDS) model of controllers' perceptions of the aircraft performance characteristics of the aircraft. Interpretation of the two-dimensional MDS model suggested that Dimension 1 was related to engine type, and Dimension 2 was associated with weight class. The results were interpreted as evidence of performance-based prototypes (see [16] for further explanation). However, it was also evident from the position of the elements (i.e., aircraft types) in the derived stimulus space that it might be possible to develop a measure of aircraft mix using these two easily obtained variables.

Pfleiderer [17] continued that line of investigation in a study designed to determine whether controllers' perceptions of aircraft performance were comparable to the actual recorded performance of aircraft in a sample of live air traffic data. In general, controllers' perceptions of aircraft performance characteristics were similar to the actual performance of the aircraft in the recorded data. However, weight class was far less salient as a separate dimension in the model derived from the matrix of System Analysis Recording (SAR) data than in the model based on controller estimates. The relationship between weight class and engine type in the SAR data model was a clear reminder that weight class is a correlate of engine type (i.e., most piston-driven aircraft are Small, most turboprops are Large, all Heavy aircraft are jets). This result led to the conclusion that engine type alone was an appropriate and sufficient dimension for calculating the Aircraft

Mix Index. (See [17] for a complete description of Aircraft Mix Index calculations.)

Based on the apparent success of the first phase of the investigation, a second phase was initiated that focused on testing the Aircraft Mix Index for its ability to discriminate between altitude strata. After all, if the index had sufficient variability and precision it should be able to discriminate between high- and low-altitude sectors. This was based on the assumption that high-altitude sectors should have a lower incidence of aircraft mix due to the relatively low service ceilings of some aircraft, whereas low-altitude sectors should have a much higher incidence of aircraft mix because all aircraft must climb and descend through low-altitude airspace at some point in their flight. For this analysis, the Aircraft Mix Index was calculated in 15-minute intervals for all active sectors within a 1-hour sample of air traffic data recorded at the Kansas City en route center (15 high-altitude and 13 low-altitude sectors). As anticipated, values of the Aircraft Mix Index tended to be higher in low-altitude sectors than in high-altitude sectors. A comparison of the two groups using the distribution-free Mann-Whitney U statistic [12] revealed that the Aircraft Mix Index was reliably different between high- and low-altitude sectors.

Because the Aircraft Mix Index was able to discriminate between sector strata, it passed the "minimum test" to be considered as a possible addition to the suite of Performance and Objective Workload Evaluation Research (POWER) variables. POWER refers to a set of measures developed for quantifying en route air traffic controller activity and task load (see [14] for a detailed description of POWER measures and methodology). In the first phase of the present study, I conducted an evaluation of the relationship between the Aircraft Mix Index and existing POWER measures using Principal Components Analysis (PCA). PCA is a statistical technique often used to reveal patterns of correlations among variables. Values in the component loading matrix produced by PCA represent the correlation of individual variables with the underlying dimension the component describes. If the Aircraft Mix Index was redundant with traffic volume (as opposed to describing some aspect of the complexity associated with that traffic) it should load onto the same component as the total number of controlled aircraft. On the other hand, if the Aircraft Mix Index provided information about the complexity associated with the presence of aircraft with different performance characteristics then it should load onto a component with others that relate to traffic complexity. Moreover, its loading should be of sufficient magnitude to suggest that this is a reliable relationship.

Though PCA offers insight into the relationship of the Aircraft Mix Index relative to the other POWER variables, it cannot tell us whether the information it provides is unique. More importantly, it does not address the larger question regarding the relative contribution of aircraft mix to traffic complexity. Therefore, contingent upon the results of the PCA, a multiple regression analysis was conducted using a subjective criterion of "Complexity" provided by controllers from the each of the en route centers sampled. The predictor variable set consisted of those variables identified by the PCA as being most closely related to the "Complexity" dimension/construct. The results of the multiple regression analysis should tell us whether aircraft mix (as measured by the Aircraft Mix Index) contributes a significant amount of unique information to the prediction of controllers' perceptions of sector complexity.

Method

System Analysis Recording data and subjective complexity ratings were generously provided by researchers associated with the Dynamic Density project [10]. Traffic samples selected for the analyses were collected at the Fort Worth and Atlanta en route centers. The Fort Worth data consisted of samples from six high-altitude and three low-altitude sectors. The Atlanta data were from five high-altitude and four low-altitude sectors. Two 30-minute samples were collected from each of the selected sectors (a total of 36 samples). Traffic sample descriptions are provided in Tables 1 and 2.

Three controllers individually viewed Systematic Air Traffic Operations Research Initiative (SATORI) re-creations [20] and rated the complexity of the traffic situation on a scale from one to seven (lowest to highest) at 2-minute intervals throughout the 30-minute sample time frame. For the current study, means of the individual controller ratings were averaged over 6-minute intervals to create a total of 180 observations.

POWER measures were computed in 6-minute intervals for each of the traffic samples, producing a total of 180 observations for each POWER measure. Variables selected for the PCA (shown in Table 3) consisted of five POWER measures that have consistently demonstrated a relationship with controller activity and task load [13] [14], and five thought to relate to traffic complexity. The selected Traffic Complexity/Proximity variables are relatively new additions to the POWER suite of measures. (Consequently, this analysis also represents a serendipitous opportunity to examine whether or not these variables do, in fact, appear to describe a separate dimension.) The reasons for including the Aircraft Mix Index have already been described in some detail, but the rationale behind the other variables in this group deserves some explanation.

Table 1. Atlanta En Route Center Samples

High-Altitude		**Low-Altitude**	
Sector	**Time (local)**	**Sector**	**Time (local)**
03	1945-2015	04	2005-2035
03	2030-2100	04	2012-2042
06	1918-1948	19	1240-1310
06	1940-2010	19	1830-1920
20	1730-1800	38	1645-1715
20	1935-2005	38	1815-1845
22	1725-1755	41	1330-1400
22	1918-1948	41	1950-2020
39	1450-1520		
39	2240-2310		

Table 2. Ft. Worth En Route Center Samples

High-Altitude		**Low-Altitude**	
Sector	**Time (local)**	**Sector**	**Time (local)**
28	0035-0105	29	1240-1310
28	1815-1845	29	1845-1915
46	1520-1550	75	1555-1625
46	1505-1535	75	2235-2305
47	1550-1620	96	1255-1325
47	1555-1625	96	1325-1355
48	1223-1253		
48	2235-2305		
49	1505-1535		
86	1245-1315		
86	1855-1925		

Table 3. POWER Variables Selected for Principal Components Analysis

Activity/ Task load	Number of Controlled Aircraft
	Number of R-side Entries
	Number of R-side Entry Errors
	Number of RA-side Entries
	Number of RA-side Entry Errors
Proximity/ Complexity	Aircraft Mix Index
	Mean Lateral Distance
	Mean Vertical Distance
	Number of Altitude Changes
	Number of Heading Changes

There is little doubt that the number of aircraft within a sector affects controller workload. It is also doubtful that this measure alone sufficiently captures all aspects of the complexity associated with that traffic [8]. One of the traffic complexity issues that should be addressed is the relative position of the aircraft. For the suite of POWER measures, we have opted to incorporate summary measures of aircraft proximity (i.e., Mean Lateral Distance, Mean Vertical Distance). Though not as elegant as some measures, such as clustering techniques developed by Delahaye and Puechmorel [2], they do have the advantage of reflecting the dimensions controllers use to evaluate aircraft separation.

The number of climbing and descending aircraft is well established as a contributor to traffic complexity [1], [9], [3], [7], [19], [21]. The number of altitude changes provides more information than a count of the number of aircraft in transition. Altitude changes have been shown to correspond well with the number of altitude clearances and may provide some indication of the amount of workload associated with monitoring the response to and the effectiveness of the issued clearance [18].

Heading changes have the potential to profoundly impact the complexity of an air traffic situation, whether they occur as part of the scheduled flight plan or in response to a clearance. It is not surprising, therefore, that heading changes have been shown to contribute significantly to sector complexity [11]. It should be noted that POWER only counts heading changes greater than or equal to 10º that persist for a minimum of 36 seconds. These computer-detected heading changes have been shown to correspond well with the number of issued heading clearances [18].

Results and Discussion

Principal Components Analysis (PCA)

As shown in Table 4, the distributions of most of the variables selected for the PCA approximated normality. However, distributions of the Aircraft Mix Index, R-side Entry Errors, and Radar Associate

Entry Errors deviated significantly. Although assumptions regarding normality are not generally in effect when PCA is used descriptively, in this particular application it is important to remember that PCA is sensitive to the sizes of correlations. To the extent that normality fails, the solution may be degraded and this should be considered when interpreting the results of the analysis [22].

Table 4. Principal Components Analysis Descriptive Statistics (N = 180)

Variable	Mean	*S.D.*	Skew.[1]	Kurt.[2]
Aircraft Mix Index	10.85	14.78	2.23	6.62
Mean Lateral Distance (nm)	48.65	13.84	.25	-.48
Mean Vertical Dist. (ft/100)	54.59	17.77	.74	.53
Altitude Changes	6.86	3.71	.37	-.13
Heading Changes	4.41	3.03	.71	-.14
Controlled Aircraft	14.82	4.02	.25	.41
R-side Entries	33.99	11.74	.31	-.46
R-side Entry Errors	2.23	2.35	1.78	4.13
RA-side Entries	6.64	5.97	.98	.41
RA-side Entry Errors	.66	1.21	3.04	12.31

[1]S.E. Skew. = .181; [2]S.E. Kurt. = .360

Principal components analysis is generally used in the exploratory stages of research when the exact number and nature of the dimensions are not known. Although the selected variables were hypothesized to represent elements of two dimensions, extraction of one or more additional components would not be entirely unexpected. Therefore, a minimum eigenvalue of 1.00 (as opposed to a specified number of components) was selected as the criterion for component extraction, thus allowing for true exploration of the data.

Varimax rotation was selected for the analysis because it increases the interpretability of the solution. As the name suggests, varimax (*vari*ance *max*imizing procedure) simplifies components by maximizing the variance of the loadings within components. Simply stated, varimax makes small loadings smaller and large loadings larger. This simplifies interpretation of the components by making it more obvious which variables are associated with them [22].

PCA with varimax rotation converged in eight iterations and produced four components with eigenvalues > 1. These components accounted for approximately 71% of the variance in the data set. As shown in the rotated component matrix in Table 5, all variables loaded onto at least one component with a loading of .40 or greater.

Component 1 had an eigenvalue of 2.40 and accounted for approximately 24% of the variance in the data set. Without exception, the variables associated with this component were selected to represent various aspects of traffic complexity. Values in the loading matrix describe the correlation of each variable with the underlying dimension the component represents. Notice that the Aircraft Mix Index has one of the highest loadings on this component.

Component 2 had an eigenvalue of 2.06 and accounted for about 21% of the variance. The two variables with the highest loadings, the number of controlled aircraft and the number of Radar controller computer entries, are straightforward activity measures. Radar controller entry errors tend to increase as controller activity increases. Therefore, a conservative loading of .54 on this component makes sense within the context of the other variables.

Component 3 had an eigenvalue of 1.35 and accounted for approximately 13% of the variance in the data set. Generally, components described by only two variables are considered to be unreliable and are not interpreted. However, this component emerged in a previous analysis [14] suggesting that it may, in fact, be reliable. The extraction of "Radar Associate Activity" as a separate component may reflect the unique relationship that Radar Associate Entries and Errors share with other activity measures. When activity is relatively low, a Radar controller working alone has time to make entries on the Radar Associate's console (because some entries can only be made from that console). As the traffic situation becomes more demanding, the Radar controller no longer has time to make entries from the RA-side console. During peak hours, a Radar Associate controller is assigned to the sector and entries made from the RA-side console become more frequent. It is probably the distinctive "J-shaped" distribution of RA-Entries and their relationship to RA-Entry Errors that distinguishes these variables as a separate component. Therefore, Component 3 might be viewed as a subset of general activity.

Component 4 had an eigenvalue of 1.27 and accounted for approximately 13% of the variance. The same caveat regarding two-variable components applies to Component 4, only in this case it may be more justified. Components defined by only two variables may be reliable if the variables are highly correlated with one another (i.e., $r > .70$) and are relatively uncorrelated with others in the variable set. These variables fail to meet these criteria in that the bivariate correlation between them ($r = .35$) is less than .70 and they correlate significantly with a number of the other variables in the set. It is also important to note that the distributions of both these variables are severely positively skewed and leptokurtotic. Therefore, it is likely that the communality described by this component reflects a similarity of distribution rather than of meaning. On the other hand, data entry errors tend to increase with the number of entries so this component might also be viewed as a subset or correlate of general activity.

Table 5. Principal Components Analysis Rotated Component Matrix

	Component			
Variable	**1**	**2**	**3**	**4**
Aircraft Mix Index	.71			
Mean Lateral Distance (nm)	-.67			
Mean Vertical Dist. (ft/100)	.61			
Altitude Changes	.80			
Heading Changes	.64			
Controlled Aircraft		.85		
R-side Entries		.79		
R-side Entry Errors		.54		.69
RA-side Entries			.89	
RA-side Entry Errors			.47	.79

* Component loadings < .40 not shown.

The results of the PCA demonstrate that the Aircraft Mix Index was consistently associated with other variables thought to relate to traffic complexity. Moreover, the magnitude of its loading (.71) suggests that this is a reliable relationship. This leads us to the next phase of the experiment: Multiple regression analysis using the variables associated with the "Complexity" dimension (Aircraft Mix Index, Mean Lateral Distance, Mean Vertical Distance, Number of Altitude Changes, and Number of Heading Changes) to predict controllers' subjective "Complexity" ratings.

Multiple Regression Analysis

Perhaps the most important assumption of a regression analysis is that the observations are sampled from the same population. Although preliminary tests indicated that the data from the two facilities were similar enough to justify pooling for a PCA, the results of comparisons of the Aircraft Mix Index in high- and low-altitude sectors in a previous study [18] suggested that high- and low-altitude sectors might constitute heterogeneous samples. Therefore, initial data screening was conducted by visually examining scatterplots of each predictor variable against the criterion with observations color-coded according to altitude strata. It was immediately apparent that high- and low-altitude sectors should be analyzed separately. Unfortunately, splitting the sample resulted in a sample size of 70 for the low-altitude sectors (i.e., a 14:1 case to independent variable ratio). Ideally, we would want a ratio of 20 cases for every predictor to ensure sufficient statistical power to detect small effect sizes and to accommodate measurement error. Nevertheless, a 14:1 ratio exceeds the absolute minimum requirement of five cases for every predictor [22].

Descriptive Statistics

Descriptive analyses were conducted separately for the high- and low-altitude samples. The criterion variable (i.e., Complexity ratings provided by controllers) originally consisted of discrete values representing anchor points along an underlying continuum of the controllers' perceptions of traffic complexity. However, the Complexity ratings used in this analysis represent the means of ratings taken every 2 minutes, summarized over 6-minute intervals. As such, these ratings were normally distributed in both sample sets (see Table 6).

In the low-altitude sector sample, the distribution of the Aircraft Mix Index deviated by as much as five standard deviations in both skewness and kurtosis. Square root transformation of the Aircraft Mix Index in this sample reduced deviations to less than one standard deviation from normal. In the high-altitude sector sample, transformation of the Aircraft Mix Index was contraindicated because transformation would create unacceptable deviations in a distribution that was acceptable in its natural state. (Indeed, all the selected variables were normally distributed in the high-altitude sample.)

Table 6. Multiple Regression Analysis Descriptive Statistics

Variable	Mean	S.D.	Skew.[1]	Kurt.[2]
Low_Altitude Sample (*N* = 70)				
Complexity Ratings	2.65	1.05	.35	-.93
Aircraft Mix Index	22.91	17.53	1.43	2.95
Square Root Aircraft Mix	4.44	1.79	.35	-.02
Mean Lateral Distance (nm)	36.23	7.54	.20	-.57
Mean Vertical Dist. (ft/100)	66.46	17.99	.37	.25
Altitude Changes	8.83	3.79	.05	-.05
Heading Changes	5.10	3.10	.55	-.29
High-Altitude Sample (*N* = 110)				
Complexity Ratings	3.84	1.07	.13	-.26
Aircraft Mix Index	8.63	3.88	-.16	-.21
Mean Lateral Distance (nm)	56.55	10.80	.28	.06
Mean Vertical Dist. (ft/100)	47.04	12.88	.59	.37
Altitude Changes	5.60	3.07	.26	-.47
Heading Changes	3.97	2.91	.83	.10

[1] Low-Alt.: *S.E. Skew.* = .287; High-Alt.: *S.E. Skew.* = .230
[2] Low-Alt.: *S.E. Kurt.* = .566; High-Alt.: *S.E. Kurt.* = .457

Tests of Assumptions

Multicollinearity refers to a very strong linear relationship between sets of predictor variables that renders the regression coefficients unstable [6], [22]. It is important to note that it is not the bivariate correlations among the predictors that creates multicollinearity, but rather the multiple correlation of the regression of a particular predictor on the others. Therefore, the best way to test for multicollinearity in the predictor set is to conduct a series of regressions with each of the predictors taking turns as the criterion and examining the squared multiple correlations for perfect or near perfect values (which would indicate multicollinearity). The results of these tests when conducted on the selected set of independent variables revealed no indication of multicollinearity in either the high- or low-altitude samples.

No univariate outliers (i.e., cases with values greater than three standard deviations from the mean) were detected in either the low-altitude or high-altitude sector samples. Mahalanobis distances using $p < .001$ failed to uncover any multivariate outliers (i.e., cases with an unusual pattern of values) in either data set.

The assumption of linearity and the assumption of constant variance of Y for all values of X can be easily tested by visually examining a plot of residuals against predicted values. If both assumptions are met, there will be no systematic pattern in the plots. Studentized residuals against the predicted values were randomly distributed in a horizontal band around zero, indicating that the assumptions were met.

One of the simplest ways to test whether errors of prediction are normally distributed is by visual examination of a cumulative probability plot of the observed distribution of residuals against that expected of a normal distribution. If the two distributions are identical, a straight line results. Cumulative probability plots demonstrated that the assumption of normally distributed errors was met.

Because the scenario data were collected sequentially, they were screened to determine whether the time series had produced systematic variance in errors. Unfortunately, statistical procedures for testing sequential correlation of adjacent error terms (e.g., the Durbin-Watson) were not designed to test discrete groups of sequential data. Therefore, Studentized residuals were plotted against the sequence variable and visually examined. Non-independence of prediction errors in these data would manifest itself in “zigzag” or “herringbone” patterns. No patterns were detected, thus indicating that the assumption of independence of errors had been met.

Regression Model: Low-Altitude Sample

Standard multiple regression analysis of the low-altitude sector sample produced a multiple *R*=.69 which was significantly different from zero, $F(5,62) = 10.97$, $p < .01$. As shown in Table 7, the regression model derived from the selected variables accounted for approximately 47% of the variance in Complexity ratings. Table 7 also contains unstandardized regression coefficients (*b*), their standard errors (*S.E.*), standardized regression coefficients (*ß*), and squared semipartial correlations (sr^2) for each of the predictors. In standard multiple regression, sr^2 represents the unique contribution of a predictor to the total variance explained. It is clear that the Number of Heading Changes was the only variable that accounted for a significant amount of unique variance (22%). The difference between R^2 and the sum of sr^2 for all predictors in the variable set represents shared variance. Thus, 23% of the variance described by R^2 was unique whereas 24% was shared.

Regression Model: High-Altitude Sample

Because the value of the Aircraft Mix Index is set to “system missing” in the absence of any aircraft with differing performance characteristics within a given sector, there were a considerable number of missing values in the high-altitude data set. Nevertheless, standard multiple regression analysis produced a multiple *R*=.57, which was significantly different from zero, $F(5,40) = 3.94$, $p < .01$. As shown in Table 8, the only variable to account for a significant amount of unique variance in this data set was Mean Vertical Distance (16%). The Number of Altitude Changes (7%), the Number of Heading Changes (5%), and Mean Lateral Distance (3%) added to the total 31% unique variance explained, but these contributions were not statistically significant. Again, the Aircraft Mix Index failed to contribute any unique information to the prediction of controllers' Complexity ratings.

Table 7. Regression of POWER Complexity Variables on ATC Complexity Ratings: Low-Altitude Sample

Model Summary	R	R^2	Adj. R^2	*S.E.* Est.
	.68**	.47	.42	.79
Variable	***b***	***S.E.***	***ß***	sr^2
Transformed Aircraft Mix Index	.003	.072	.005	.00
Mean Lateral Distance (nm)	.001	.015	.009	.00
Mean Vertical Distance (ft/100)	.001	.006	.024	.00
Number of Altitude Changes	.032	.033	.113	.01
Number of Heading Changes	.216	.042	.620	**.22** **

Significance levels derived from *t*-tests: ** $p < .01$; * $p < .05$

Table 8. Regression of POWER Complexity Variables on ATC Complexity Ratings: High-Altitude Sample

Model Summary	R	R^2	Adj. R^2	*S.E.* Est.
	.57**	.33	.25	.80
Variable	***b***	***S.E.***	***ß***	sr^2
Aircraft Mix Index	.000	.031	-.002	.00
Mean Lateral Distance (nm)	-.017	.013	-.175	.03
Mean Vertical Distance (ft/100)	-.034	.011	-.502	**.16** **
Number of Altitude Changes	.123	.063	.321	.07
Number of Heading Changes	.108	.061	.250	.05

Significance levels derived from *t*-tests: ** $p < .01$; * $p < .05$

Conclusions

It is important to note that the list of predictors included in the regression analyses was not intended to be exhaustive. It is certainly possible, even probable, that other measures might account for additional variance in controllers' complexity ratings. Therefore, the results should not be interpreted as evidence that a single variable is sufficient to describe complexity in the en route environment. The focus of the analyses was simply to assess the relative contribution of aircraft mix to sector complexity. The fact that the Aircraft Mix Index failed to explain a significant amount of unique variance in controllers' Complexity ratings was disappointing, particularly with respect to the low-altitude sample, but not entirely unanticipated. Historically, the evidence supporting aircraft mix as a complexity factor has been anecdotal rather than statistical, no doubt because aircraft mix was considered to be "non-quantifiable" [4]. Certainly controllers' verbal representations/reports constitute a valuable heuristic, but with such evidence comes the risk of misattribution. For example, a specific factor might be particularly salient to the controller during periods of perceived increases in "workload" or "complexity" and yet be a mere correlate of the factor that is actually driving their subjective experience. Thus, it is particularly important to make every attempt to quantify and assess each proposed complexity factor to determine its relative influence and relationship with other factors.

This is not to say that aircraft mix should be automatically discounted based on the results of a single analysis. It is entirely possible that aircraft mix does not, in fact, share a linear relationship with complexity and therefore cannot be captured using a linear regression analysis. Hilburn and Flynn [8] have proposed that linear regression is ill suited for the study of air traffic complexity because "complexity factors combine in a non-linear way. Though the constellation of factors might well apply across contexts, the relative importance of each differs by context" (p. 200).

Another potential reason that aircraft mix failed to describe a significant amount of variance in controllers' complexity ratings is that it might only be a relevant factor in a few sectors, but in those sectors it is a major contributor to traffic complexity. Every sector is unique. This presents a challenge when attempting to build models that will generalize. It is, therefore, vital to investigate the potential of aircraft mix and other prospective complexity factors using data collected at multiple facilities with a number of different statistical strategies before drawing any firm conclusions.

In that same vein, as gratifying as it may be that all the variables selected to represent "Traffic Complexity/Proximity" in the principal components analysis formed a single dimension, there is no guarantee that this would be the case in all data sets. Neither do these variables represent a comprehensive list of factors that might relate to sector complexity. A considerable amount of work has been done in this area (see [3] for an excellent review of the literature) and many of the proposed measures will be considered as possible additions to the POWER variables. Each candidate will be tested with the same rigor as the Aircraft Mix Index using similar methodology (i.e., examining the validity of each measure individually, testing its performance within the framework of the POWER variable suite, then examining its contribution relative to an external criterion). Each iteration of this process brings us closer to developing a set of measures that might comprehensively describe the sector environment to better understand the nature of sector complexity and its effects on controller workload and performance.

References

[1] Christien, R., A. Benkouar, 2003, *Air traffic complexity indicators and ATC sectors classification*. Paper presented at the 5th USA/Europe Air Traffic Management R&D Seminar, Budapest, Hungary.

[2] Delahaye, D., S. Puechmorel, 2000, *Air traffic complexity: Towards intrinsic metrics.* Paper presented at the 3rd USA/Europe Air Traffic Management R&D Seminar, Napoli, Italy.

[3] EUROCONTROL 2004, *Cognitive complexity in air traffic control: A literature review*, EEC Note No. 04/04, Brétigny-sur-Orge, France, Author.

[4] Federal Aviation Administration, 1984, *Establishment and validation of en route sectors,* FAA Order No. 7210.46, Washington, DC, Author.

[5] Federal Aviation Administration, 1999, *Position classification standard for air traffic control: Series ATC – 2152 terminal and en route,* Washington, DC, Author.

[6] Fox, J., 1991, *Regression diagnostics: An introduction*, Sage University Paper series on Quantitative Applications in the Social Sciences, 07-079, Newbury Park, CA, Sage.

[7] Grossberg, M., 1989, Relation of sector complexity to operational errors, *Quarterly report of the Federal Aviation Administration's Office of Air Traffic Evaluations and Analysis*, Washington, DC, Federal Aviation Administration.

[8] Hilburn, B., G. Flynn, 2004, *Toward a non-linear approach to modeling air traffic complexity*. Paper presented at the 2nd Human Performance Situation Awareness and Automation Conference, Daytona Beach, FL.

[9] Histon, J.M., G. Aigoin, D. Delahaye, R.J. Hansman, S.Puechmorel, 2001, *Introducing structural considerations into complexity metrics.* Paper presented at the 4th USA/Europe Air Traffic Management R&D Seminar, Santa Fe, NM.

[10] Kopardekar, P., S. Magyarits, 2003, *Measurement and prediction of dynamic density.* Paper presented at the 5th USA/Europe Air Traffic Management R&D Seminar, Budapest, Hungary.

[11] Laudeman, I.V., S.G. Sheldon, R. Branstrom, C.L. Brasil, 1998, *Dynamic density: An air traffic management metric,* NASA/TM-1998-112226, Washington, DC, National Aeronautics and Space Administration.

[12] Mann, H.B., D.R. Whitney, 1947, On a test of whether one of two random variables is stochastically larger than the other, *Annals of Mathematical Statistics, 18,* pp. 50-60.

[13] Manning, C.A., S.H. Mills, C. Fox, E.M. Pfleiderer, H. Mogilka, 2001, *The relationship between air traffic control communications events and measures of controller taskload and workload.* Paper presented at the 72nd Annual Scientific Meeting of the Aerospace Medical Association, Reno, NV.

[14] Mills, S.H., E.M. Pfleiderer, C.A. Manning, 2002, *POWER: Objective activity and taskload assessment in en route air traffic control,* Report No. DOT/FAA/AM-02/02, Washington, DC, Office of Aerospace Medicine.

[15] Mogford, R.H., E.D. Murphy, R.J. Roske-Hofstrand, G. Yastrop, J.A. Guttman, J.A., 1994, *Application of research techniques for documenting cognitive processes in air traffic control: Sector complexity and decision making,* Report No. DOT/FAA/CD-TN94/3, Atlantic City, NJ, Federal Aviation Administration Technical Center.

[16] Pfleiderer, E.M., 2000, *Multidimensional scaling analysis of controllers' perceptions of aircraft performance characteristics,* Report No. DOT/FAA/AM-00/24, Washington, DC, Office of Aviation Medicine.

[17] Pfleiderer, E.M., 2003a, *Development of an empirically-based index of aircraft mix,* Report No. DOT/FAA/AM-03/8, Washington, DC, Office of Aviation Medicine.

[18] Pfleiderer, E.M., 2003b, *Relationship between computer-detected altitude, heading, and speed changes with controller clearances in en route air traffic control*. Paper presented at the 74th Annual Scientific Meeting of the Aerospace Medical Association, San Antonio, TX.

[19] Robertson, A., M. Grossberg, J. Richards, 1979, *Validation of air traffic controller workload models,* Report No. DOT/FAA/RD-79/83, Cambridge, MA, U.S. Department of Transportation Research and Special Programs Administration – Volpe National Transportation System Center.

[20] Rodgers, M.D., D.D. Duke, 1993, *SATORI: Situation assessment through the re-creation of incidents,* Report No. DOT/FAA/AM-92/12, Washington, DC, Office of Aviation Medicine.

[21] Stein, E.S., 1985, *Air traffic controller workload: An examination of workload probe,* Report No. DOT/FAA/CT-TN84/24, Atlantic City, NJ, Federal Aviation Administration Technical Center.

[22] Tabachnik, B. G., L.S. Fidell, 1989, *Using multivariate statistics* (2nd ed.), New York, HarperCollins.

Keywords

Aircraft Mix, ATC Workload, ATC Task load, Sector Complexity

Biography

Elaine Pfleiderer is a graduate of the University of Central Oklahoma with a Masters degree in Experimental Psychology. Ms. Pfleiderer has been a research associate at the Civil Aerospace Medical Institute since 1995, working primarily on the development of objective measures of air traffic controller task load using routinely-recorded National Airspace System (NAS) data.