DYNAMIC DENSITY AND COMPLEXITY METRICS FOR REALTIME TRAFFIC FLOW MANAGEMENT

Anthony J. Masalonis, Michael B. Callaham, Craig R. Wanke
Center for Advanced Aviation System Development (CAASD), The MITRE Corporation, McLean, VA, USA

Abstract

This study's goal was to model airspace Dynamic Density and complexity (and hence controller workload) using traffic characteristic metrics. The focus was on metrics that could eventually enable Traffic Flow Management (TFM) personnel to strategically prevent overloads using triggers other than predicted sector traffic count.

Potential metrics from past studies were assessed in terms of how well they could be predicted at time horizons required for TFM decision support (up to 120 minutes), and their face validity. Also, proportional odds logistic regression determined the metrics' usefulness for predicting subjective complexity ratings collected in an FAA-NASA study.

Based on these analyses, a subset of 12 metrics was chosen (from the original 41). Further multiple regression analyses were conducted with this reduced model, to determine which metrics provided unique contributions to the prediction of subjective complexity, and to see the extent to which the same complexity factors related to subjective workload in different airspaces.

Structured interviews with a sample of eight Traffic Management Coordinators were used to cross-check the quantitative findings.

Specific aircraft proximity, density, and airspace structure metrics were found potentially useful for real-time TFM decision support. Many of the useful metrics were normalized and smoothed measures from an algorithm developed by Wyndemere, Inc. Also, it was found that different metrics related to subjective complexity in different centers, but the differences were small enough that a generalized set of complexity metrics might be applicable to multiple airspaces, at least in the near term. Future work could determine the viability of airspace-adapted complexity algorithms.

The fact that multiple types of metrics are useful suggests that a multidimensional visual representation of predicted workload might be useful in TFM, as opposed to combining all relevant factors into a single metric.

Introduction

The Air Traffic Management (ATM) system features ever-increasing traffic volume, in addition to more dynamic traffic flows and the need to accommodate more airspace user preferences. Aviation researchers and developers have therefore been exploring such concepts as collaborative decision making, dynamic resectorization, and shared separation authority. An issue cutting across all of these concepts is the need for operationally useful sector workload measures.

Significant research interest has been generated in the concept of "Dynamic Density" (DD) and complexity, identified as integral issues in the "Free Flight" concept [1, 2]. DD, it was envisioned, would be used to define situations where the traffic was complex enough to require reversion to centralized control for that place and time. The motivation to study DD arose from the notion that the number of aircraft in a sector may not adequately reflect the difficulty of working that sector, which remains an important issue even before the implementation of Free Flight procedures.

The primary measure of sector volume currently used in the U.S. for operational decisions about staffing and TFM initiatives, such as rerouting flights out of an overloaded sector, is the peak aircraft count -- the most aircraft that will be in the sector during any minute of a 15-minute period. This metric is imperfect for several reasons. First, it fails to reflect the duration for which the peak load will continue, and so is insensitive to sustained high workload periods. In addition, it is very sensitive to minor fluctuations in the sector entry or exit time of a few flights, which can change the peak count without changing the amount of sustained workload. Finally, the same flight count can represent a vastly different level of controller workload, depending on the complexity of the traffic mix and flows, and the weather. A number of DD equations have been proposed, all involving a single workload metric derived from weighted combinations of multiple sector workload factors. Various taxonomies have been proposed for workload metrics; among the most useful is that of Histon et al. [3], who divide workload metrics into airspace design factors, dynamic traffic characteristics, and operational (e.g., procedural) factors, and list a number of each.

The majority of the factors in DD equations are dynamic traffic characteristics, which would generally be the most useful for realtime decision support. Measures of actual controller activity, e.g., communications and physical interaction with the workstation, are more difficult to record, and may in any event be less useful in workload models applicable to realtime decision support. Manning et al. [4, 5] found that communications measures correlated with subjective workload, but did not provide incremental predictive benefit over other factors. They also found that certain physical workload parameters, such as route displays and strip requests, were not well correlated with controller performance and mental workload. These authors provide some evidence that other physical workload parameters (e.g., data entries) are important in workload models; however, most of the useful factors in their models are traffic characteristics.

Many researchers have found traffic measures other than count to be relevant to Air Traffic Control (ATC) workload. Significant relationships have been demonstrated between complexity-type measures and subjective workload, stronger than the relationship between peak count and subjective workload [6]. Similar results were found in [5], as mentioned above. Delahaye and Puechmorel [7] propose airspace complexity models based on factors including relative aircraft position and velocity, and crossing traffic angles. Further evidence that traffic count may be insufficient to reflect workload comes from [8], wherein several characteristics of sectors other than peak count were found to be related to operational errors, and from [9], a human-in-the-loop simulation study suggesting that operational errors may be more likely to occur after rather than during a peak in traffic count. Chaboud et al. [10] reviewed two separate studies on using complexity metrics to assess a facility's performance and usual workload; in both studies, number of aircraft changing altitudes (and traffic volume) emerged as important factors.

There is clear evidence that DD measures, other than traffic count, may reflect sector workload. However, the weighted combinations used in DD equations are less actionable; i.e., it may not be clear how to resolve the high-workload situation. If the traffic count exceeds the operationally-defined threshold by three aircraft, the solution is to remove three flights. In contrast, the DD equations derive abstract numbers from which the solution is not obvious. Therefore, one purpose of the present study was to explore metrics more reflective of workload than simple traffic count, that still provide some indication to the decision maker as to what factor(s) is/are predicted to cause the problem, and thus provide improved decision support. One method for doing this would be to select a small number of metrics that could be presented individually.

Much of the past research on sector workload has focused on the opinions, behavior, and subjective and objective performance of controllers at the sector level, since it is the controller who is responsible for safety and whose workload is ultimately of interest. However, the personnel with the first opportunity to impact controller workload are Traffic Management Coordinators (TMCs), who make decisions affecting how much traffic a controller will have to deal with, as well as traffic complexity. Thus, it is appropriate to predict and represent traffic load in a way that can help the TMC decide on actions affecting controller workload. With only a few exceptions, e.g., [11], in which researchers collected subjective difficulty data from a limited number of TMCs at one ATC facility, the TFM perspective on sector complexity has not been well studied.

The overall goals of this work were to suggest indicators of sector workload that could be used operationally for realtime TFM decision support. As mentioned above, rather than using observed metrics of physical work such as data entries and radio communications, the study focused on the traffic characteristics that ultimately lead to increases in actual work. These are more readily predictable before-the-fact (i.e., using trajectory modeling) and also may be related to cognitive aspects of workload that are not be directly observable.

The quantitative characteristics of potential metrics were assessed using data from a joint FAA-NASA DD study [12]. Additionally, to assess metrics' utility from an operational perspective, feedback was collected from current TMCs using a structured interview.

Quantitative Analysis of Complexity Indicators

To be operationally useful, TFM complexity metrics must be closely correlated with controller-perceived workload, and must be better predictors of such workload than sector count, since sector count is probably the most intuitive workload metric. Also, these metrics must be predictable using available data sources and prediction algorithms. Predictability of DD measures was studied in [13]; the present study extends that work to a larger number of airspaces and a wider traffic sample.

MITRE worked with the Simulation and Analysis Branch (ACB-330) of the FAA to systematically evaluate four different algorithms for predicting controller-perceived air traffic complexity. The complete results of this study are available in [12] and [14]. The complexity model presented here is an outgrowth of the previous work, and draws from all of the previously proposed complexity algorithms.

Complexity Metrics

The four DD algorithms included two developed at NASA Ames Research Center [6, 15], one developed by the FAA [16], and one developed at Wyndemere [17]. A summary of all four algorithms can be found in [12]. Each algorithm uses predicted trajectories to compute a variety of complexity-related metrics, and combines those metrics into a single value for the predicted complexity (or DD). The four algorithms together contain 41 individual metrics, spanning many types of traffic features, including most of those discussed previously. Because many of the metrics have similar definitions, and based on the desire for model parsimony, a subset of the 41 metrics was derived. Parsimony is important both for operational reasons – i.e., ATC personnel can more readily be trained on the meaning of each metric if the model has fewer metrics – and to reduce computational load, especially important if the model will be implemented in a realtime decision support system. The next two subsections describe the analyses.

Subjective Workload Ratings

In order to provide "truth" data against which to compare the complexity metrics, subjective traffic complexity ratings were collected at Fort Worth (ZFW), Atlanta (ZTL), Cleveland (ZOB), and Denver (ZDV) Air Route Traffic Control Centers (ARTCCs). A total of 72 30-minute traffic samples were taken, evenly divided between the ARTCCs and representing a range of sector types. For each 30-minute traffic sample, controllers and supervisors (up to three of each) were shown a traffic replay and asked to rate the situation at two-minute intervals. Two ratings were given: a Complexity rating from 1 ("very easy") to 7 ("very hard"), and a Number of Controllers rating (1, 2, or 3, indicating how many controllers would be required to handle the traffic situation) [12].

The yardstick for determining the metrics' success at predicting the subjective ratings was the geometric mean, over all observed ratings, of the probability predicted by the model for the rating observed. This will be referred to as Geometric Mean of Probability (GMP), and can be calculated as:

$$GMP = \exp((edf - (AIC/2))/N)$$

where: <u>edf</u> = equivalent degrees of freedom (the number of model parameters estimated from the data), and <u>AIC</u>, referring to Akaike's Information Criterion [18] is given by the following formula [19]:

$$AIC = 2\ edf - 2\ \ln(L)$$

where <u>L</u> is the likelihood of all observations (ratings), given the model, and <u>N</u> is the number of observations.

In the GMP formula, the expression edf - (AIC/2) is the natural logarithm of likelihood of model parameters, given the observations. That likelihood, by definition, is the joint probability of all observations, given the parameterized model. GMP can vary from 0 to 1, inclusive. It would equal 1 if and only if the model predicted a probability of 1 for each observed rating, and would be 0 only if the model predicted a 0 probability for each observed rating.

Applying GMP analysis to the traffic metrics and complexity ratings, a proportional odds logistic regression model [19] was built to predict the Complexity or the Number of Controllers rating from one or more of the DD metrics. The metrics' values in a given traffic sample were used to predict the probability of each subjective rating (e.g., 1 to 7), and GMP was derived by comparing the probabilities predicted by the model, to the actual ratings of the study participants.

Predictability of Complexity Metrics

Computing complexity metrics requires knowledge of current and projected aircraft position and velocity vectors, which in turn requires a trajectory prediction algorithm and a system to synthesize the track reports, flight plans and wind data required for trajectory prediction. A prototype developed for the FAA/MITRE Collaborative Routing Coordination Tools (CRCT) program was used for this purpose [20].

In the past, the CRCT prototype has been used for developing TFM decision-support capabilities, focusing on human-in-the-loop evaluation. However, for this study only the algorithmic core was needed. Therefore, a modified version of the prototype was used containing none of the human-computer interface elements, but featuring a modified traffic analysis module to compute the complexity metrics. The metrics were computed by replaying archived traffic data from the Enhanced Traffic Management System (ETMS), and wind data. ETMS data contains flight plans, 1-minute position reports, ground delay program information, and airline collaborative decision making information. The CRCT prototype trajectory modeler used ETMS and wind data, and adapted ATC restrictions to create trajectories of all aircraft included in a 30-minute scenario. For each minute of a simulation run, values of all DD metrics were predicted from 0 to 120 minutes into the future in 1-minute time intervals. Therefore, in order to cover the 30-minute subjective data collection period, the replay was started two hours prior to the start of the period and allowed to run until the period was over.

Quantitative Results

To arrive at the smaller subset of metrics, each of the 41 metrics, as well as several basic characteristics of the sector or traffic (including traffic count) was assessed on the following criteria:

- GMP and AIC. Performance, on the GMP measure described earlier, of a univariate linear model attempting to predict the Complexity and Number of Controllers ratings from that metric, and whether the metric was retained in a multivariate model built to minimize AIC, which balances predictive power and parsimony.

- Predictability. Least number of minutes of look-ahead time at which the correlation between the prediction and the actual value falls below 0.3, calculated via the methods described above.

- Face validity. Metrics were considered more desirable if they met the following quantitative characteristics, defined in [14], that any metric should satisfy in order to make operational sense:

    1. Adding another aircraft should not reduce complexity.

    2. Shrinking the geometry of the airspace, or increasing the speeds of all aircraft in the airspace, should not reduce complexity.

    3. Repositioning one aircraft so that it is now farther from every other aircraft should not increase complexity.

    4. The metric should be independent of the orientation and origin of the coordinate system.

- Redundancy. The metric, assuming it is desirable on the other characteristics, should have a low correlation with others that capture the same traffic characteristic (e.g., aircraft proximity or speed). This criterion shows whether the metric provides a unique representation of the situation not captured by other metrics that attempt to measure similar characteristics of the traffic.

The most desirable metrics according to this analysis are listed in Table 1. Details of the analysis are found in [14].

Table 1. Complexity Metrics in Reduced Model.

Metric	Definition
NUM	Sector aircraft count
MAP	Monitor/Alert Parameter (operationally defined threshold)
SECTVOL	Sector Volume, cubic nautical miles (nm)
SC	Speed Change; number of aircraft with an airspeed change greater than 10 knots or 0.02 Mach during a 2-minute interval
WACT	A normalized measure of the aircraft count per sector
WDEN	A normalized measure of the aircraft density per sector
WCLAP	A measure incremented by aircraft pairs with less than 8-nm horizontal distance, and to a lesser extent by pairs with less than 13-nm horizontal distance
WCONVANG	A measure of the convergence angle for aircraft pairs within 13 nm of each other
WCONFLICTNBRS	A measure of the number of aircraft in close proximity to an aircraft pair projected to be in conflict
WCONFBOUND	A measure of the number of aircraft pairs in conflict with each other and close to a subsector boundary
WALC	A measure of the number of aircraft with an altitude change greater than 500 feet per minute
WASP	A measure of the distribution of aircraft relative to sector structure

Most of these metrics (those beginning with "W") are from the Wyndemere algorithm [17]. It is likely that these metrics perform better at the aforementioned criteria because they are corrected by a normalization factor. In addition, some are smoothed, beginning with the maximum of the given parameter over some time interval, rather than an instantaneous value. This aids predictability, because if a high value of a metric is predicted to begin at a given minute and actually occurs several minutes later, the analysis used in this study will record a prediction error for the original predicted minute, though the prediction is still operationally valid. Other metrics in the subset are SC from the NASA algorithm [6], and NUM, MAP, and SECTVOL, which are basic traffic or sector characteristics that are not specific to any of the algorithms, but that are component terms in some of the DD metrics.

Airspace-Adapted Models

One might expect that a model relating traffic and subjective metrics would better reflect the characteristics of a given airspace if *adapted* to the airspace, by adjusting the coefficients of predictors in the equation, and/or by adjusting the set of metrics that are retained for the model. Adaptation models can be used to infer whether a model fit to data for multiple airspaces exhibits an adequate fit for a particular airspace.

Analyses were conducted to examine the adaptability of models predicting the subjective ratings from values on the aforementioned DD metrics. Adaptability was assessed across the four aforementioned ARTCCs. To assess each metric's relative importance in each adapted model, it was determined how much a model's GMP would decrease if each regressor were dropped from the model, one at a time. These results are shown in Figures 1-2, which show the decrement in GMP relative to the GMP for that center's full model, which adjusts for the slight differences between centers in the ability of the full model to predict the subjective ratings. A higher value means that the metric is more important to the model, i.e., the model would be less predictive of the subjective rating if that metric were dropped.

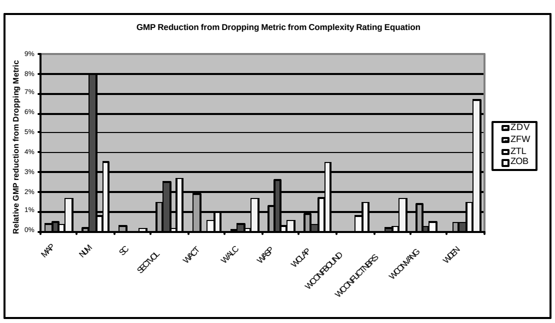

Figure 1. Contribution of Each Metric to Model for Complexity Rating.

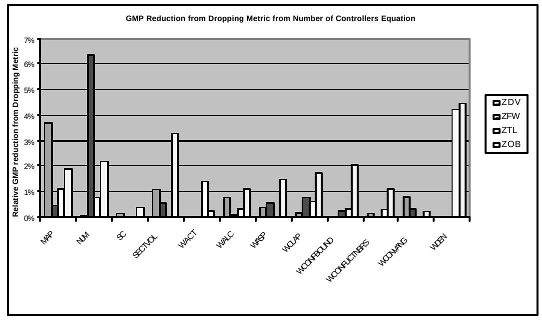

Figure 2. Contribution of Each Metric to Model for Number of Controllers Rating.

The figures show that some metrics do not affect the models' ability to predict subjective complexity ratings for any center or either subjective rating. These include SC, WALC (relating to speed and altitude changes, respectively), WCONFLICTNBRS ("conflict neighbors" factor), and WCONVANG ("convergence angle" factor). Even though these metrics individually exhibit the aforementioned desirable quantitative characteristics, they do not provide much unique contribution to a regression model explaining the subjective ratings. Therefore, if the goal is as parsimonious as possible a set of metrics, these may be candidates for dropping, though as discussed later, some of these metrics may still be operationally desirable.

It can also be seen that different metrics provide a high unique contribution to the prediction of subjective complexity, depending on the center. For example, the WDEN measure contributes more to the predictive power of the models in ZOB and (at least for the Number of Controllers rating) in ZTL. This may be because density varies more in these centers. A number of other differences exist as well, suggesting that different factors may contribute to perceived complexity and difficulty, in different centers and altitudes.

It can be concluded that there is some difference in the complexity "profile" depending on the airspace. Additional analysis assessed the magnitude of these inter-center differences. The 12-metric model was used to predict the most likely Complexity and Number of Controllers ratings, which were compared with the participants' actual ratings to determine the percent of ratings predicted correctly. For the Complexity ratings, the prediction was permitted to be within one rating point of the actual rating, due to the wide range of possible responses and the fact that with Likert scales having this many points, there is no great semantic difference between responses near each other on the scale. Figures 3-4 show the results.

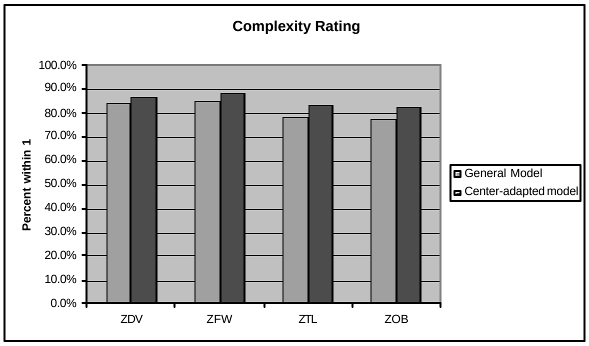

Figure 3. Predicting Complexity Rating from Traffic Characteristics with General and Adapted Models.

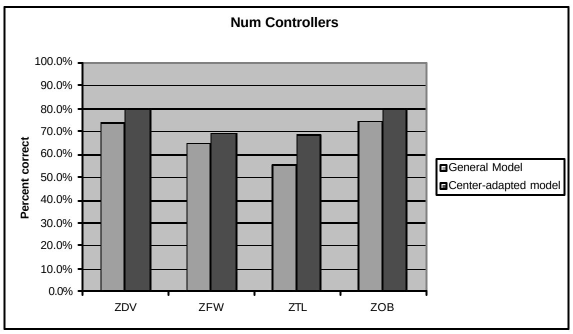

Figure 4. Predicting Number of Controllers Rating from Traffic Characteristics with General and Adapted Models.

Naturally, a slight benefit is always observed when the model is adapted to the altitude band. According to a chi-square analysis, adopting a criterion of $\underline{p}<.01$ for statistical significance, the number of correctly-predicted ratings is significantly increased by adapting to the center's data for all four centers, with one exception, the complexity rating at ZDV, which approaches significance ($\underline{p}\sim.10$).

It can be concluded that the traffic parameters associated with subjective complexity differ between ARTCCs. Whether these statistically significant differences are operationally significant is a question for future research. The fact that percentage differences are relatively small suggests the viability of using the same complexity metrics for all airspaces, at least as a start.

Operational Interview

A structured interview was administered to 5 TMCs from ZKC (Kansas City ARTCC) and 3 from ZID (Indianapolis ARTCC). For each of 16 workload factors, TMCs rated, from 1 ("not at all important") to 5 ("very important") the overall importance of the factor to TFM decision making. In addition, several illustrated questions were included regarding how information about sector workload should be displayed. Each workload factor was also accompanied by graphics to ensure consistent understanding of the question's meaning. The workload factors chosen for the interview were based on a review of controller workload literature and on input from MITRE/CAASD personnel, many with ATC experience. The factors do not map exactly to the DD metrics used in the quantitative analyses, but do represent the constructs captured by those metrics. For example, WALC equals 1.60355 times the number of flights climbing at greater than 500 feet per minute. The information of interest from the operational personnel was not the appropriateness of the parameters 1.60355 and 500, but whether the arrival and departure banks producing altitude changes are important for TFM decision support.

Each participant was interviewed by MITRE/CAASD facilitators. Participants were reminded at the outset that future capabilities were being studied, and that they should therefore focus on what information *could* be useful, as opposed to what they use today. In addition, participants were asked not to consider how predictable they believed each factor to be when making judgments about its utility, so that responses regarding metrics' operational value could be studied independently of their quantitative characteristics. Results of the ratings with regard to overall operational importance of each factor are seen in Table 2.

Table 2. Rated Operational Importance of Factors.

Factor	Mean Importance
Predicted peak number of aircraft	4.63
Percent sector capacity unavailable due to severe convective weather	4.63
Weather at busy destination airports of flights traversing sector	4.63
Traffic at same altitudes on merging/crossing flows	4.50
Impact of severe convective weather on other sector(s)	4.38
Total number (occupancy)	4.25
Departure push near sector	4.25
Arrival push near sector	4.13
Traffic at same altitudes, close lateral proximity	3.88
Amount of time above MAP	3.75
Number of aircraft to enter sector	3.06
Merging/crossing point near sector boundary	2.63
Merge or cross at narrow angles	2.43
Mix of aircraft type	2.38
Traffic off airways, structured routes, or usual routes	2.25
Total time in sector across all aircraft	1.50

What Decisions Would Complexity Factors Support?

As part of the structured interview, TMCs rated which of the 16 workload factors could and could not be used in executing each of 7 common decisions or tasks:

- Volume Management
- Ensuring Capacity is Utilized Efficiently
- Miles In-Trail
- Ground Stops
- Ground Delay Program
- Lateral Rerouting
- Altitude “Rerouting”, e.g., capping.

Figure 5 illustrates the percent of "yes" responses for each task for the six factors from Table 2 that relate directly to traffic complexity.

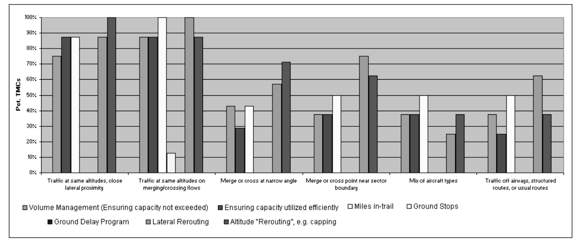

Figure 5. Percent TMCs Saying Each Factor Could be Used for Each TFM Decision.

The complexity measures that TMCs believe they would use for making decisions about most initiatives are traffic at the same altitudes on merging or crossing flows, and traffic at the same altitudes in close lateral proximity. The decisions for which complexity factors are deemed most useful are volume management and ensuring capacity is utilized efficiently.

However, there is a wider variety of factors that the majority of TMCs could use for lateral rerouting decisions. Merging near sector boundaries or at narrow angles, and traffic off usual routes, were rated by a majority of TMCs as useful for lateral rerouting decisions despite lower overall importance ratings for these factors. Several TMCs commented that while a rerouting initiative would not be implemented simply to *avoid* complexity factors like narrow-angle crossings or merge points near a boundary, some TMCs would take care that an initiative they put out did not *introduce* these types of situations.

A larger variety of complexity factors are relevant to altitude rerouting as well. For example, one TMC reported that the aircraft type mix factor comes into play for altitude rerouting, because props (propeller aircraft) could be capped in a lower sector if it would prevent the complexity resulting from a mix of props and jets in the overlying sector.

Operational Viability of Multidimensional Workload Predictions

Potential drawbacks of using a single weighted DD sum for realtime decision support were discussed earlier. To obtain operational input on this issue, the proposed display of multidimensional complexity information seen in Figure 6, was presented to TMCs for comment.

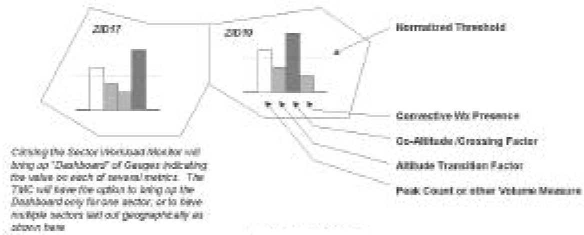

Figure 6. Bar Chart Showing Multidimensional Workload Predictions.

The bar chart presents four different workload elements, and each is predicted to have a certain value, which is color coded according to whether it exceeds some operationally defined threshold, and by how much. The design of the chart is inspired by previous MITRE/CAASD work, and by the principles of emergent features and proximity compatibility [21]. Details of the interface, and the factor names and thresholds, were not discussed; the goal was feedback on the general concept of multidimensional workload.

The above display was considered less useful than a single DD-type measure of workload by some participants, partly because many TFM decisions today are based on the unidimensional measure of peak count, and TMCs are therefore used to looking a single number. Also, having too many factors to analyze may create information overload, drawing attention from and slowing the decision process.

However, the results also show willingness to use multidimensional information. Some of the TMC comments indicate that the display in Figure 6 could help to decide whether an alerted sector was "really red", as one participant put it (i.e., really required initiative(s) to reduce the flow). Specifically, if all the factors were high, multiple initiatives might be required, while if only one factor was high, no initiative would be needed.

Pros and cons were stated for both single and multidimensional workload representation. One possible solution, shown in Figure 7, would allow the TMC to view predicted sector demand color-coded for count, composite DD, or the "dashboard" of Figure 6. This option received favorable comments from TMCs.

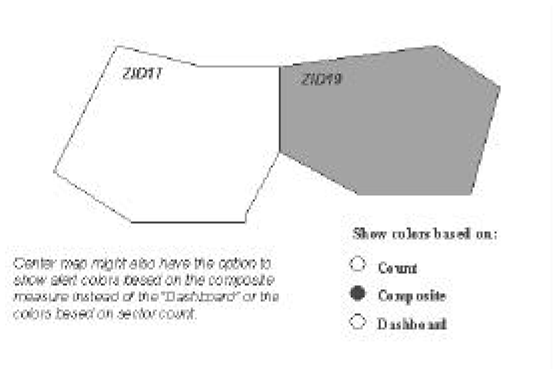

Figure 7. Option to Display Predicted Demand by Count or Complexity.

Responses to the interview questions regarding workload factors can be combined with the quantitative results to determine which workload factors and metrics might be operationally important in TFM as well as exhibiting appropriate quantitative characteristics. More detail on the results of the interview can be found in [22].

Conclusions

It was determined from the operational interview that aircraft at the same altitudes on crossing flows and/or in close lateral proximity are important factors for TFM decision making. DD metrics that were found to capture this factor include WCONFLICTNBRS and WCLAP. It was found that WCONFLICTNBRS does not provide much incremental benefit to the models predicting subjective difficulty from traffic characteristics. The WCLAP measure may be a more promising proximity-related workload measure.

Other complexity-oriented metrics, including merging and crossing points near a sector boundary (which could be captured by WCONFBOUND), are less important overall in TFM. However, they may represent useful parameters to present in what-if tools for determining the potential impact of TFM initiatives before deciding to implement them.

The impact to en-route sectors of arrival and departure pushes emerged in the interview as another complexity-oriented factor that influences decisions about TFM initiatives. Appropriate metrics from the DD equations that relate to this factor include WALC and SC, since arrival and departure banks increase the number of flights changing speed and altitude. Although these two metrics did not provide a unique contribution to the prediction of subjective workload, they represent an operationally different traffic characteristic than other metrics, and therefore may be useful for TFM decision making.

In addition to complexity factors, it is obvious that TFM requires some measure of volume. The interview investigated potential volume measures other than peak count. Of the alternate volume measures presented in the interview, amount of time above MAP may be the most promising for inclusion in an operational system. The "occupancy" factor also received high ratings, but some of the TMCs' comments supporting the occupancy factor were really about the length of time for which the high workload would continue, e.g., "gives an idea whether it's busy for a minute or is it really going to be busy", and "gives an idea of how busy (sector will be) over a sustained period". The quantitative characteristics of these and other alternate volume measures have yet to be studied systematically.

Adapting a set of metrics to a specific center generally provides a statistically significant improvement to the prediction of subjective difficulty ratings from traffic characteristics. However, most of these significant differences are relatively small from a practical perspective. If complexity metrics were implemented in the TMU in the relatively near term, it would probably be unnecessary to use different models for each airspace. If a set of complexity metrics did in fact have different operational importance in different centers, local protocols for usage of the information could be adapted without actually adapting the metrics themselves. For the longer term, future work with a larger sample of airspaces could determine the operational and computational viability of airspace-adapted complexity and DD equations.

Future work could also focus on metrics other than those studied herein. During the interview, TMCs informally identified additional ways of numerically representing complexity and workload. These include the predicted load on a subsector (i.e., a specific posting fix), projected arrival rate at an airport of interest as compared to the airport's acceptance rate, and, when large-scale rerouting is underway, information regarding the percentage of rerouted flights on each potential reroute corridor.

Other potential workload measures not addressed in this study are the number of different routes through a sector that aircraft are on, and the number of restrictions and constraints currently underway in the sector, such as miles-in-trail and altitude restrictions. These factors arose in informal conversations with TMCs and with MITRE/CAASD personnel experienced in controller workload research [e.g., Al McFarland and Melvin Zeltser, personal communication, 2002].

The mixed operational feedback regarding the display options points to the need for further concept development and experimentation to determine the relative merits of using the multivariate representation, the single DD number, or both.

Improved measures of complexity would find applications outside TFM decision support, including realtime decision support for the operational supervisor. In particular, the relationship of the DD metrics to the Number of Controllers ratings, is relevant to the study of supervisor automation capabilities, as this type of staffing decision is an important role of the operational supervisor. Applications of these results also exist for post-event analysis and airspace design -- see [14] for a brief discussion -- and airspace slot allocation.

A variety of complexity and other workload factors were deemed operationally important and were shown to relate to subjective workload. Therefore, it can be concluded that multiple aspects of workload, not limited to volume measures, are relevant to TFM decision making. It is also concluded that the concept of using complexity information in TFM has some viability and that the DD metrics mentioned in this section, among others, may be promising for individual or combined presentation to help TMCs make decisions to enhance efficiency and safety. Conclusions are necessarily preliminary, because data were only collected from a small number of ARTCCs. Human-in-the-loop simulation and further algorithmic analysis are required to derive more definitive recommendations regarding which complexity metrics can aid TFM decision making, and regarding specific procedures for the use of this information.

References

[1] RTCA, Inc. (1995). *Report of the RTCA Board of Directors' Select Committee on Free Flight.* Washington, DC: RTCA.

[2] RTCA, Inc. (2000). *National Airspace System Concept of Operations. Addendum 4: Free Flight Phase 2.* Washington, DC: RTCA.

[3] Histon, J., G. Aigoin, D. Delahaye, R. Hansman, and S. Puechmorel (2001). Introducing Structural Considerations into Complexity Metrics. *4th USA/Europe Air Traffic Management R&D Seminar*, Santa Fe, NM.

[4] Manning, C., S. Mills, C. Fox, E. Pfleiderer, and H. Mogilka (2000). Investigating the Validity of Performance and Objective Workload Evaluation Research (POWER).

3rd USA/Europe Air Traffic Management R&D Seminar, Napoli, Italia.

[5] Manning, C., S. Mills, C. Fox, E. Pfleiderer, and H. Mogilka (2001). The Relationship Between Air Traffic Control Communication Events and Measures of Controller Taskload and Workload. *4th USA/Europe Air Traffic Management R&D Seminar*, Santa Fe, NM.

[6] Chatterji, G., and B. Sridhar (2001). Measures for Air Traffic Controller Workload Prediction. *First AIAA Aircraft Technology, Integration, and Operations Forum*, Los Angeles, CA.

[7] Delahaye, D., and S. Puechmorel (2000). Air Traffic Complexity: Towards Intrinsic Metrics. *3rd USA/Europe Air Traffic Management R&D Seminar*, Napoli, Italia.

[8] Rodgers, M., R. Mogford, and L. Mogford (1998). *The Relationship of Sector Characteristics to Operational Errors*. Report No. DOT/FAA/AM-98/14. Washington, DC: Office of Aviation Medicine.

[9] Murphy, L., and K. Smith (2001). Heart-Rate Variability is a Robust Measure of Response to Task Demand: A Study of Operational Errors in Air Traffic Control. *Human Factors and Ergonomics Society 45th Annual Meeting*, Minneapolis, MN.

[10] Chaboud, T., R. Hunter, J. Hustache, S. Mahlich, and P. Tullett (2000). *Investigating the Air Traffic Complexity: Potential Impacts on Workload and Costs*. Eurocontrol/NATS Project GEN-40E2. http://www.eurocontrol.fr/public/reports/eecnotes/2000/11.pdf.

[11] Mogford, R., E. Murphy, and J. Guttman (1999). Using Knowledge Exploration Tools to Study Airspace Complexity in Air Traffic Control. *International Journal of Aviation Psychology, 4(1)*, 29-45.

[12] Kopardekar, P., and S. Magyarits (2002). Dynamic Density: Measuring and Predicting Sector Complexity. *21st Digital Avionics Systems Conference*, Irvine, CA.

[13] Sridhar, B., K. Sheth, and S. Grabbe (1998). Airspace Complexity and its Application in Air Traffic Management. *2nd USA/Europe Air Traffic Management R&D Seminar*, Orlando, FL.

[14] Holly, K., Y. Cabeza, M. Callaham, D. Greenbaum, A. Masalonis, and C. Wanke (2002). *Feasibility of Using Air Traffic Complexity Metrics for TFM Decision Support.* Report No. MTR02W055. McLean, VA: MITRE.

[15] Laudeman, I., S. Shelden, R. Branstrom, and C. Brasil (1998). *Dynamic Density: An Air Traffic Management Metric*. Report No. NASA/TM-1998-112226. Moffett Field, CA: NASA.

[16] Kopardekar, P. (2000). *Dynamic Density, A Review of Proposed Variables.* Egg Harbor Twp., NJ: FAA NAS Advanced Concepts Branch, ACT-540.

[17] Wyndemere, Inc. (1996). *An Evaluation of Air Traffic Control Complexity, Final Report.* Contract Number NAS2-14284. Boulder, CO: Wyndemere.

[18] Akaike, H. (1973). Information Theory and an Extension of the Maximum Likelihood Principle. In B. Petrov and F. Csaki, Eds., *Second International Symposium on Information Theory*, 267-281. Budapest: Akademiai. Kaidó.

[19] Venables, W., and B. Ripley (2002). *Modern Applied Statistics with S, 4th edition.* New York: Springer-Verlag.

[20] Wanke, C., D. Berry, J. DeArmon, S. Miller, N. Taber, L. Rhodes, and M. Yee, (2002). Decision-Support for Complex Traffic Flow Management Actions: Integrated Impact Assessment. *2002 AIAA Guidance, Navigation, and Control Conference*, Monterey, CA.

[21] Wickens, C. (1992). *Engineering Psychology and Human Performance, Second Edition*. New York: HarperCollins.

[22] Masalonis, A., M. Callaham, Y. Figueroa, and C. Wanke (2003). Indicators of Airspace Complexity for Traffic Flow Management Decision Support. *12th International Symposium on Aviation Psychology*, Dayton, OH.

Author Biographies

Dr. Anthony Masalonis is a Senior Engineer at MITRE/CAASD, specializing in Human Factors, operational evaluations, and statistical analysis. He holds a Ph.D. in Applied Experimental Psychology from the Catholic University of America, an M.A. in Industrial Psychology from the University of Connecticut, and a B.S. in Psychology, minor in Computers and Info Systems, from King's College,.

Dr. Michael Callaham is a Senior Operations Research Analyst at MITRE/CAASD, where he has been employed since 1995. Dr. Callaham holds an Sc.D. degree in Instrumentation, an S.M. degree in Aeronautics and Astronautics, and an S.B. degree in Mathematics, all from MIT. He is a Senior Member of the American Institute of Aeronautics and Astronautics, a Member of the Institute of Electrical and Electronics Engineers, and a Member of the International Institute of Forecasters. At MITRE, he uses air traffic simulation to forecast benefits of proposed improvements in air traffic management systems and procedures.

Dr. Craig Wanke is a Principal Engineer at MITRE/CAASD, specializing in simulation and modeling. He has worked for the past 15 years on decision support systems for cockpit and air traffic management applications. He holds S.B, S.M., and Ph.D. degrees in aeronautical engineering from MIT.