MEASUREMENT AND PREDICTION OF DYNAMIC DENSITY

Parimal Kopardekar, Ph.D., Federal Aviation Administration, NASA Ames Research Center, Moffett Field, CA, USA

Sherri Magyarits, Federal Aviation Administration, Atlantic City, NJ, USA

This paper describes results of a multi-year, multi-organizational research initiative related to the measurement and prediction of sector level complexity called Dynamic Density (DD). The researchers first identified a number of candidate DD measures. They then identified eighteen 30-minute traffic samples from each of four selected en route Air Route Traffic Control Centers (ARTCCs). At each ARTCC, they collected complexity ratings at two-minute intervals for each traffic sample from approximately 70 air traffic controllers and supervisors. Using the traffic and sector data, various DD variables were computed. Using a linear regression method, the relationships between different DD variables and complexity ratings were determined. A unified DD metric composed of variables from several organizations performed the best. The results indicated that DD represents instantaneous sector complexity better than aircraft count, which is the currently used method. The results also indicated that the prediction of complexity using DD is somewhat better than the prediction using aircraft count most likely due to the inherent inaccuracy of predicting aircraft count.

Background

Dynamic Density (DD) is analogous to complexity or difficulty of an air traffic situation. RTCA Task Force 3 report [1] defines DD as "the essential factors affecting conflict rate in both en route and terminal airspace." The broader definition of DD considers as ATC taskload, which is the basis of controller subjective workload. It is a "measure of control-related workload that is a function of the number of aircraft and the complexity of traffic patterns in a volume of airspace" [2]. For this research, the term DD is defined as the collective effect of all factors, or variables, that contribute to the sector level air traffic control complexity or difficulty at any given time [3].

Operational Need for Dynamic Density

In order to accommodate user preferences, aviation researchers and developers are exploring initiatives such as collaborative decision-making, dynamic resectorization, user-preferred routes, shared separation, and free flight. The core element of all of these concepts is the ability to measure and predict sector-level complexity. Changes in traffic flows will be better managed if an accurate measurement and prediction of sector-level complexity is available.

The current ATC system uses the monitor alert parameter of the Enhanced Traffic Management System (ETMS) to measure sector level activity and the corresponding air traffic controller taskload. It is widely recognized, however, that aircraft count, and hence the monitor alert parameter, has significant shortcomings in its ability to accurately measure and predict sector level complexity [4]. Although the term Dynamic Density is relatively new, the factors that contribute to sector level traffic complexity have been of interest to researchers for a long time. In a literature review of sector complexity, Mogford, Guttman, Morrow, and Kopardekar (1995) found that aircraft count, sector geometry, traffic flows, separation standards, aircraft performance characteristics, and weather are the most common factors that contribute to air traffic complexity or difficulty. The following provides a comprehensive list of factors that contribute to air traffic complexity [5]:

1. Number of aircraft,
2. Aircraft density or traffic volume,
3. Aircraft handled in prior time interval (e.g., last hour),
4. Number of arrivals,
5. Number of departures,
6. Number of emergencies,
7. Number of special flights,
8. Coordination,
9. Traffic mix (arrivals, departures, and overflights),
10. Number of airport terminals,
11. Traffic distribution,
12. Staffing,
13. Weather conditions,
14. Equipment status,
15. Number of communications with aircraft,
16. Number of communications with other sectors,
17. Presence of conflicts,
18. Number of path changes,
19. Preventing conflicts (crossing or overtake),
20. Number of handoffs and printouts,
21. Handling pilot requests,
22. Traffic flow structure,
23. Clustering of aircraft,
24. Control adjustments involved in merging and spacing,
25. Mixture of aircraft types,
26. Climbing and descending aircraft,
27. Number of intersecting flight paths,
28. Number of required procedures,
29. Number of military flights,

30. Airline hub location,
31. Weather and its severity,
32. Aircraft routing,
33. Special use airspace,
34. Sector geometry,
35. Sector size,
36. Requirements for longitudinal and lateral spacing,
37. Radar coverage,
38. Frequency congestion,
39. Number of altitudes used, and
40. Others.

Researchers have been interested in examining how the effect of the above complexity factors can be measured using quantifiable variables. Since 1995, multiple exploratory studies have aimed at identifying the variables that contribute to DD.

DD Research Partners

In the year 2000, the FAA developed a Research Management Plan (RMP) to promote coordination with all parties interested in conducting DD research [6]. The RMP brought several organizations together to effectively use resources and eliminate duplication of effort. The FAA WJHTC's ACB-330 led the effort and Titan Systems, NASA Ames Research Center, Metron Aviation (formerly Wyndemere), and Mitre CAASD participated as research partners. The WJHTC developed one DD metric (with 10 variables), Metron Aviation developed one DD metric (with 10 variables), and NASA Ames Research center developed two DD metrics (NASA1 with 16 variables and NASA2 with 9 variables). Mitre CAASD brought the Collaborative Routing and Coordination Toolset (CRCT) to the DD research effort to provide a means of ingesting ETMS raw data and producing DD output.

Description of DD Metrics

Each DD metric consisted of multiple variables. Only a high level description of the variables is provided in this paper due to space limitations. For detailed formulas, computations, and descriptions of all metrics, please refer to a review article by Kopardekar [7].

WJTHC/Titan Systems Metric

The WJHTC/Titan metric description, rationale, and formulas are provided in Kopardekar [7].

AD1 Aircraft density 1 - number of aircraft/occupied volume of airspace

AD2 Aircraft density 2 - number of aircraft/sector volume

CRI Convergence recognition index – measure of the difficulty of detecting converging aircraft with shallow angles

SCI Separation criticality index - proximity of conflicting aircraft with respect to their separation minima

DOFI Degrees of freedom index – based on maneuver options in a conflict situation

CTI1 Coordination taskload index 1 - based on aircraft distance from the sector boundary prior to hand-off

CTI2 Coordination taskload index 2 - different formula based on the same principle as CTI1

SV Sector volume

ACSQ Square of aircraft count

In addition to the above quantitative variables, the WJTHC/Titan metric also contained categorical variables such as facility and sector types (i.e., high/low).

NASA Metric 1

The NASA-1 metric consisted of 16 variables. For details of the calculations, readers are encouraged to refer to Chatterji [4].

C1 Number of aircraft
C2 Number of climbing aircraft
C3 Number of cruising aircraft
C4 Number of descending aircraft
C5 Horizontal proximity metric 1
C6 Vertical proximity metric 1
C7 Horizontal proximity measure 2
C8 Vertical proximity measure 2
C9 Horizontal proximity measure 3
C10 Vertical proximity measure 3
C11 Time-to-go to conflict measure 1
C12 Time-to-go to conflict measure 2
C13 Time-to-go to conflict measure 3
C14 Variance of speed
C15 Ratio of standard deviation of speed to average speed
C16 Conflict resolution difficulty based on crossing angle

NASA Metric 2

The NASA-2 metric consisted of 8 variables. Laudeman, Shelden, Branstrom, and Brasil [8] and Sridhar, Sheth, and Grabbe describe these variables in detail [9]. The metric consisted of:

N Traffic Density
NH Number of aircraft with Heading Change greater than 15º
NS Number of aircraft with Speed Change greater than 10 knots or 0.02 Mach

NA	Number of aircraft with Altitude Change greater than 750 feet

S5	Number of aircraft with 3-D Euclidean distance between 0-5 nautical miles excluding violations

S10	Number of aircraft with 3-D Euclidean distance between 5-10 nautical miles excluding violations

S25	Number of aircraft with lateral distance between 0-25 nautical miles and vertical separation less than 2000/1000 feet above/below 29000 ft

S40	Number of aircraft with lateral distance between 25-40 nautical miles and vertical separation less than 2000/1000 feet above/below 29000 ft

S70	Number of aircraft with lateral distance between 40-70 nautical miles and vertical separation less than 2000/1000 feet above/below 29000 ft

Metron Aviation Metric

The Metron metric consisted of 10 variables. For further details, refer to Wyndemere [10].

WACT	Aircraft count within a sector

WDEN	Aircraft count divided by the usable volume of sector airspace.

WCLAP	Number of aircraft with predicted separation less than a threshold value (e.g., 8 miles) at a particular time.

WCONVANG The angle of converge between aircraft in a conflict situation

WCONFLICTNBRS Count of number of other aircraft in close proximity to a potential conflict situation (e.g., within 10 miles laterally and 2000 feet vertically).

WCONFBOUND Count of predicted conflicts within a threshold distance of a sector boundary (e.g., 10 miles).

WALC	Count of number of altitude changes above a threshold value with the sector.

WHEADVAR Count of number of bearing changes above a threshold value with the sector.

WBPROX Count of number of aircraft within a threshold distance of a sector boundary (e.g., 10 miles).

WASP	The squared difference between the heading of each aircraft in a sector and the direction of the major axis of the sector, weighted by the sector aspect ratio.

Method

The metrics developed by the FAA WJHTC / Titan Systems, NASA Ames Research Center, and Metron Aviation were evaluated in a series of studies due each organization's inclusion in the RMP. These were the first validation exercises that examined all DD metrics using the same common data set to identify their applicability, strengths, and weaknesses. The DD research activities were performed in three phases. The first two phases involved developing and refining the DD metrics, selecting traffic samples, and collecting subjective complexity ratings from controllers and supervisors at multiple Air Route Traffic Control Centers (ARTCCs) across the country on the complexity of those traffic samples. Phase III focused on data analysis, including the programming of the proposed metric variables into CRCT, generation of DD variable and metric values, and comparing DD output to the complexity ratings.

Phase I – Pilot Study

Phase I was a pilot study designed to refine experimental procedures for collecting complexity ratings from controllers and supervisors. Phase I was performed at Denver Center (ZDV) in October 1999. The pertinent findings of the study were as follows:

- ❖ Controller and supervisor complexity ratings were significantly and positively correlated, with supervisors providing consistently higher complexity ratings than controllers, and
- ❖ Controller complexity ratings were significantly and positively correlated with one another, as were supervisor complexity ratings.

The study involved both controllers and supervisors because supervisors typically analyze traffic situations from an over-the-shoulder perspective to make staffing decisions (e.g., appropriate number of controllers on position). Controllers were included because they have a hands-on perspective of sector complexity, being closest to the operations at hand.

Phase II – Data Collection Study

During Phase II, operational traffic data from four ARTCCs were collected. The four ARTCCs were chosen based on subject matter input to include a variety of traffic characteristics. These ARTCCs were Atlanta Center (ZTL), Cleveland Center (ZOB), Denver Center (ZDV), and Fort Worth Center (ZFW). The DD researchers collected a total of 72 thirty-minute samples of traffic data from a total of 36 high and low sectors.

Three supervisors and three controllers individually provided complexity ratings every two minutes for each of the 72 traffic samples. The traffic samples were replayed Using Systematic Air Traffic Operations Initiative (SATORI). This resulted in about 6480 complexity ratings.

Phase III- DD Metric Validation

The goals of Phase III were to determine the following:

- Can the DD metric(s) accurately measure complexity?
- Are the DD metric(s) reliable/persistent for predicting complexity starting 2 hours out?

The researchers extracted the proposed DD variables from ETMS data since it is the data source currently used for monitor alert predictions. The researchers recruited the support of MITRE CAASD, developed CRCT, to compute the DD variables from the ETMS data. CRCT was selected due to its capability to use ETMS data, its trajectory modeler (an essential element to the computation of many variables), and its deployment at several operational facilities, which could facilitate the deployment of DD.

The specific objectives of the data analysis were as follows:

Objective 1. Determine how accurately the DD metrics represent the subjective complexity ratings. In particular,

- Develop a DD model or weights for different variables that constitute different DD metrics
- Compare different DD metrics
- Select a 'best-fit' DD metric
- Test DD metric(s) for accuracy

Objective 2. Determine how stable the predictions are over time for the selected metric. Specifically, examine DD metric prediction performance starting from 2 hours prior to traffic sample intervals.

Figure 1 shows the difference between the instantaneous and predicted DD.

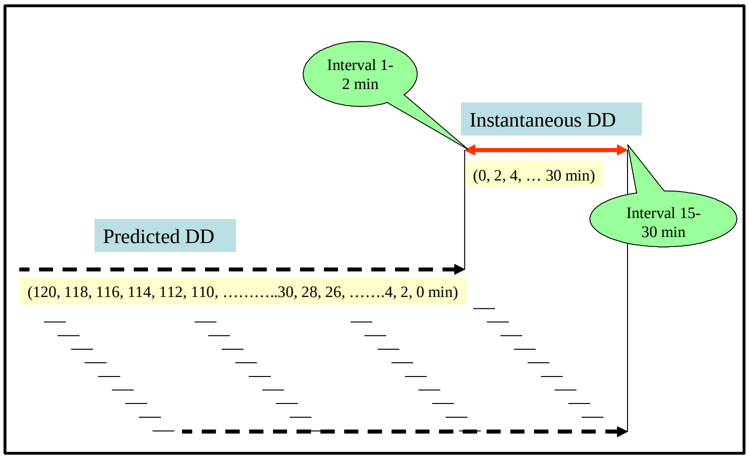

Figure 1. Instantaneous and Predicted DD

Data Analysis Approach

The researchers used regression analysis to establish weights for the different DD variables and to identify the significant DD variables in each DD metric. The 72 30-minute traffic samples were divided into two groups. The first group of 60 traffic samples was used for building the DD equation (i.e., establish variable weights) for each metric. The remaining 12 traffic samples were used to test the DD metrics. The traffic samples were randomly divided into the groups.

The researchers performed two different DD calculations for each of the traffic samples. The first focused on the instantaneous DD calculations to meet the first objective. For instantaneous DD calculations, the values of the DD variables were computed at 2-minute intervals, which corresponded to the same intervals that complexity ratings were provided for all traffic samples. The second calculation involved the predicted DD calculation to address the second objective. For the predicted DD calculations, the values of DD variables were computed at 2-minute intervals up to 120 minutes prior to actual traffic sample times.

DD metrics were developed for the WJTHC/Titan variables, two NASA metric variables, and Metron variables. In addition, a unified metric consisting of all variables from all metrics was also developed. This model development activity was conducted using the first group of 60 traffic samples. A regression method was used to develop the weights of individual DD variables that constituted the DD metric(s).

Results

DD Metrics Development

The regression analysis results were as follows: (See Tables 1 through 5):

- All four DD metrics represented complexity better than currently used aircraft count (as indicated by coefficient of variation R^2).
- All DD metrics performed the best for ZDV and the worst for ZOB. This implies that facility differences were not completely captured. In general, different DD metrics performed better for different facilities.
- WJTHC/Titan metric performed best for all facilities combined.
- A unified DD metric (i.e., variables from all four proposed metrics) provided the best results in all conditions.

The fact that the unified DD metric performed better that all individual DD metrics implies that the RMP process of collaboration among research organizations was highly successful and resulted in a better utilization of resources.

Tables 1 through 5 indicate the R^2-values of the regression for ZDV, ZTL, ZFW, ZOB, and all facilities combined. The yellow highlighted cells indicate the highest R^2 and the blue highlighted cells indicate the second highest R^2 value.

<u>Note</u>: S refers to supervisor complexity ratings, C refers to controller complexity ratings, L refers to low altitude sectors and H refers to high altitude sectors.

Table 1. Regression Results (R^2 values) for ZDV

Metrics	S&C, H&L	S&C, L	S&C, H	C, H&L	S, H&L	C, H	S, H	C, L	S, L
Aircraft Count	0.53	0.55	0.55	0.51	0.57	0.63	0.57	0.49	0.63
Tech Center	0.55	0.65	0.56	0.56	0.55	0.55	0.62	0.62	0.74
NASA-1	0.53	0.64	0.60	0.52	0.56	0.59	0.61	0.61	0.73
Metron	0.63	0.67	0.75	0.63	0.64	0.77	0.75	0.67	0.74
NASA-2	0.13	0.27	0.26	0.11	0.16	0.28	0.26	0.22	0.36
Unified	0.67	0.74	0.72	0.69	0.70	0.78	0.72	0.75	0.84

Table 2. Regression Results (R^2 values) for ZTL

Metrics	S&C, H&L	S&C, L	S&C, H	C, H&L	S, H&L	C, H	S, H	C, L	S, L
Aircraft Count	0.21	0.04	0.15	0.21	0.20	0.18	0.14	0.03	0.05
Tech Center	0.46	0.11	0.26	0.48	0.45	0.31	0.24	0.11	0.15
NASA-1	0.46	0.13	0.29	0.47	0.46	0.34	0.29	0.15	0.15
Metron	0.39	0.27	0.22	0.41	0.39	0.25	0.38	0.33	0.27
NASA-2	0.11	0.03	0.12	0.10	0.12	0.15	0.13	0.03	0.04
Unified	0.57	0.50	0.44	0.60	0.59	0.51	0.47	0.56	0.55

Table 3. Regression Results (R^2 values) for ZFW

Metrics	S&C, H&L	S&C, L	S&C, H	C, H&L	S, H&L	C, H	S, H	C, L	S, L
Aircraft Count	0.29	0.16	0.26	0.23	0.38	0.17	0.39	0.19	0.14
Tech Center	0.32	0.27	0.30	0.29	0.41	0.42	0.39	0.42	0.24
NASA-1	0.36	0.21	0.40	0.31	0.45	0.36	0.49	0.26	0.25
Metron	0.31	0.36	0.33	0.32	0.35	0.33	0.38	0.51	0.38
NASA-2	0.15	0.19	0.18	0.13	0.19	0.15	0.23	0.20	0.23
Unified	0.46	0.46	0.53	0.47	0.54	0.53	0.62	0.64	0.52

Table 4. Regression Results (R^2 values) for ZOB

Metrics	S&C, H&L	S&C, L	S&C, H	C, H&L	S, H&L	C, H	S, H	C, L	S, L
Aircraft Count	0.13	0.10	0.05	0.12	0.16	0.05	0.07	0.08	0.13
Tech Center	0.20	0.15	0.16	0.19	0.25	0.14	0.24	0.17	0.22
NASA-1	0.21	0.22	0.22	0.21	0.27	0.23	0.29	0.22	0.35
Metron	0.18	0.19	0.20	0.19	0.21	0.18	0.28	0.26	0.20
NASA-2	0.06	0.11	0.09	0.05	0.10	0.09	0.23	0.09	0.17
Unified	0.32	0.40	0.37	0.34	0.43	0.40	0.49	0.45	0.59

Table 5. Regression Results (R^2 values) for All Facilities

Metrics	S&C, H&L	S&C, L	S&C, H	C, H&L	S, H&L	C, H	S, H	C, L	S, L
Aircraft Count	0.23	0.20	0.20	0.20	0.27	0.18	0.26	0.14	0.18
Tech Center	0.33	0.33	0.26	0.31	0.38	0.26	0.29	0.26	0.43
NASA-1	0.29	0.24	0.31	0.27	0.34	0.29	0.35	0.22	0.30
Metron	0.23	0.19	0.26	0.22	0.26	0.25	0.29	0.19	0.22
NASA-2	0.10	0.11	0.15	0.08	0.12	0.13	0.17	0.09	0.15
Unified	0.39	0.41	0.35	0.40	0.43	0.39	0.37	0.39	0.51

The regression equation output for the unified DD metric is presented in Table 6 since the unified metric performed better than other DD metrics. The table shows the significant variables and their corresponding weights (beta values), t-value, and p-value. A 0.05 was chosen as a level of significance. For further analysis, only the unified DD metric was considered.

Table 6. Regression Equation Output

	Unstandardized Coefficients		Standardized Coefficients	t	p-value
	B	Std. Error	Beta		
(Constant)	1.699	.207		8.188	.000
1=low sector, 2= high sector	.695	.084	.254	8.306	.000
ac_count_sqrd	-4.406E-03	.001	-.263	-6.371	.000
TECH CENTER_AD2	683.138	75.150	.242	9.090	.000
sector volume/aircraft	-2.269E-04	.000	-.648	-16.182	.000
TECH CENTER_DOFI	-1.057E-02	.003	-.056	-3.038	.002
sector volume	1.865E-05	.000	.643	14.170	.000
NASA-1_C1	1.161	.229	.215	5.073	.000
NASA-1_C4	.302	.132	.044	2.279	.023
NASA-1_C6	.123	.055	.028	2.247	.025
NASA-1_C8	4.819E-02	.017	.040	2.799	.005
NASA-1_C9	6.075E-03	.002	.039	2.537	.011
NASA-1_C14	5.104E-03	.002	.118	2.713	.007
NASA-1_C15	-1.907	.634	-.161	-3.008	.003
NASA-2_NH	.128	.035	.047	3.672	.000
NASA-2_NA	-5.408E-02	.020	-.072	-2.739	.006
NASA-2_S25	8.303E-02	.028	.045	2.962	.003
MET_ac_count	.101	.023	.161	4.392	.000
MET_density	-6.411E-03	.001	-.168	-7.943	.000
MET_conflict neighbors	-.118	.020	-.260	-6.007	.000
MET_conflict near boundary	8.733E-02	.015	.430	5.792	.000
MET_heading variation	-1.867E-02	.006	-.039	-2.888	.004
MET_boundary proximity	-.321	.078	-.112	-4.118	.000
MET_airspace structure	3.257E-02	.003	.215	10.506	.000

DD Metrics Testing

Results for Instantaneous DD Model

The second group of data was used to conduct a performance assessment of the unified DD model. All analyses reported in this section are based on the second group data that was not used to build the model. Figure 2 shows that the unified DD model followed the complexity ratings better than the model based only on the aircraft count. Additionally, the R^2 value for the first group of data was higher for the unified DD model than the aircraft count based model.

Note: CSRATING refers to controller and supervisor complexity ratings.

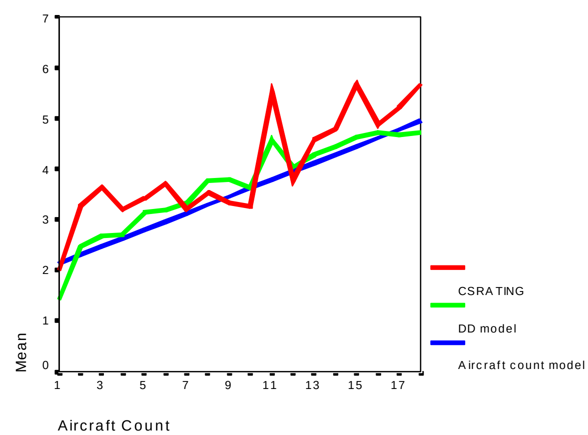

Figure 2. Performance of Unified DD Metric

Note: Model 1: CS rating = 1.970 + 0.165*Ac_count, R^2 = 0.23, Model 2: CS rating = unified DD equation, R^2 = 0.39.

Table 7 shows the difference between the output of the unified DD model and the actual complexity ratings. About 63% of the second group data points were within a 1 unit difference from the actual complexity ratings and about 23% of the data points matched the ratings exactly. Less than 10% of the differences were greater than 2 units.

Table 7. Difference Between DD and Complexity Ratings

Value	Percent	Cumulative Percent
-4	0.8	0.8
-3	1.9	2.7
-2	12.5	15.2
-1	27.8	43
0	28.6	71.5
1	21.1	92.7
2	7.1	99.8
3	0.2	100
Total	100	

Table 8 shows the performance comparison of the unified DD based model and the aircraft count based model. The yellow highlighted cells indicate the lowest errors. The results indicated that the mean absolute difference (MAD), Root Mean Square (RMS) difference, and standard deviation of difference between the actual complexity ratings and model-based predictions were all smaller for the unified DD based model. This indicates that the unified DD based model seems to be better than the aircraft count based model in representing the complexity.

Table 8. Performance Measures for Unified DD Based and Aircraft Count Based Model

	Minimum	Maxmimum	Mean	Std Deviation
Complexity ratings	1.00	7.00	3.69	1.41
DD based model	1.00	5.00	3.25	0.91
MAD	0.00	4.05	0.99	0.71
RMS	N/A	N/A	1.22	N/A
Aircraft count based model	2.13	4.94	3.09	0.62
MAD	0.00	4.70	1.23	0.88
RMS	N/A	N/A	1.51	N/A

Figure 3 indicates that the MAD between complexity derived by the unified DD based model and actual complexity ratings was the lowest when the complexity ratings were closer to 3. The MAD increased at the higher and lower ends of complexity ratings (1, 6, or 7). One possible explanation for this is that the data used to build the DD model contained a higher percentage of 2, 3, 4, and 5 complexity ratings and a much smaller percentage of 1, 6, and 7 ratings.

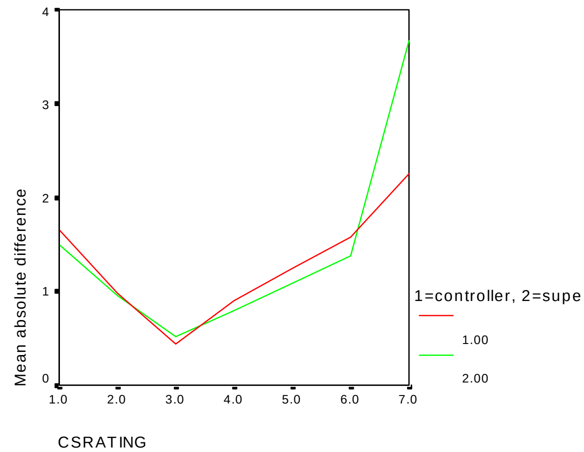

Figure 3. Mean Absolute Difference for Different Complexity Ratings

Factor Analysis for Instantaneous DD Model

The unified model consisted of several variables. Therefore, it was examined for possible interdependencies. The researchers therefore analyzed the relationships among the different variables using Principle Components Analysis (PCA). The correlations and significance values indicated that a number of variables were significantly correlated with each other. Using an eigen value of 1.0 as the threshold, 12 components were identified thatdescribed DD. Using a a factor-loading threshold of 0.50, the variables with high loadings on each component were identified. After examining these variables, the researchers identified the following potential factors.

1. Component 1: Overall monitoring
2. Component 2: Conflict detection
3. Component 3: Transitioning aircraft
4. Component 4: Communication and coordination
5. Component 5: Aircraft mix
6. Component 6: Time to resolve conflict
7. Component 7: Vertical separation monitoring
8. Component 8: Horizontal separation monitoring (for mid-to-large separations)
9. Component 9: Horizontal separation monitoring (closer to separation minima)
10. Component 10: Arrivals
11. Component 11: Hand-offs
12. Component 12: Facility and sector geometry

It must be noted that extracting the factors and assigning them descriptive names based on their potential attributes is a highly subjective process. The curious readers are encouraged to read the details of the complete factor analysis.

Results of DD Prediction Model

After the researchers established that the DD model could better gauge the instantaneous complexity than solely aircraft count, the next obvious interest was to determine how accurately DD could be predicted ahead of time. If the DD predictions were also accurate at larger look-ahead times, they could assist in planning traffic flow changes, sector level staffing needs, dynamic resectorization, and other operational decisions. Therefore, another DD equation on the first group of data was developed. The purpose of this equation was to predict DD up to 120 minutes ahead of an actual instance. Therefore, in addition to the variables included in the unified DD model, another variable called "look-ahead" time was included in the regression equation. As before, the analysis was conducted only on the second group of data because the first group of data was used to develop the weights.

The accuracy of the DD equation with look-ahead time, the original DD equation, the aircraft count based model, and the complexity ratings were compared. Figure 4 indicates that the model based on DD with look-ahead time appeared to better follow the complexity ratings (i.e., CSRating) trends. The original unified DD model used for instantaneous DD did not perform as well as the DD with look-ahead time and thus further validates the inclusion of the look-ahead time as a variable. It clearly implies that the DD predictions are dependent on look-ahead time. In general, this is a logical finding because more and perhaps better information is available about flights and weather as the prediction horizon narrows down.

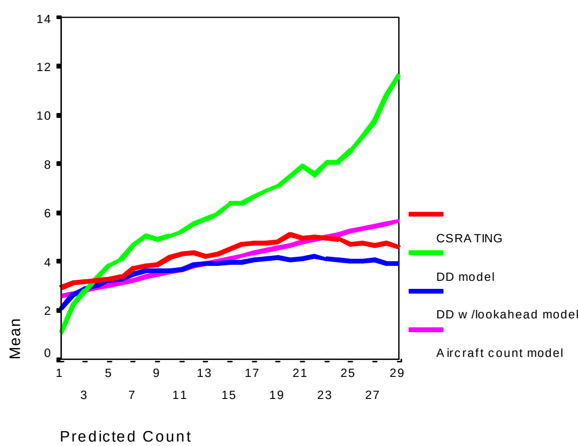

Figure 4. Predicted DD

Note: CS rating = 2.481 + 0.110*Predicted Aircraft Count, R^2= 0.17, CS rating = DD equation + look-ahead time, R^2 = 0.40

Figure 5 depicts the stability of the predicted DD values across look-ahead time. The DD model with look-ahead time appeared to provide fairly stable predictions. Interestingly, the model based on only predicted aircraft count also seemed to perform well. However, the raw predicted aircraft count varied considerably with look-ahead time.

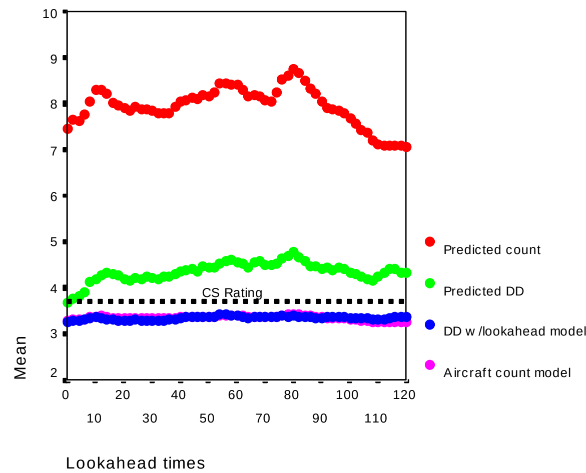

Figure 5. Stability of Predicted DD

Tables 9 and 10 show the performance comparisons of the DD model with look-ahead time, the instantaneous DD based model, and DD based on predicted aircraft count only. The MAD, standard deviation of absolute difference, and the RMS error were all smaller for the DD model with look-ahead time (as shown by yellow highlighted cells). Hence, the DD model with look-ahead time appears to better predict the complexity up to 120 minutes ahead of an instance. However, it must be noted that performance was very similar between the DD model with look-ahead time and the DD based on predicted aircraft count.

Table 9. Performance of Predicted DD

	Minimum	Maxmimum	Mean	Std Deviation
Predicted count	0	29	7.42	5.17
Complexity ratings	1.00	6.17	3.79	1.23
DD with look-ahead time model	1.25	5.16	3.35	0.73
Instantaneous DD model	0.16	16.94	4.36	2.03
DD based on predicted aircraft count only	2.48	5.67	3.30	0.57
Absolute difference for DD with look-head time	0.00	2.94	0.93	0.57
Absolute difference for instantaneous DD model	0.00	14.61	1.44	1.38
Absolute difference for model based on predicted aircraft count	0.01	2.92	1.05	0.69
RMS for DD with look-ahead time model	N/a	N/a	1.09	N/a
RMS for instantaneous DD model	N/a	N/a	1.99	N/a
RMS for model based on predicted aircraft count	N/a	N/a	1.25	N/a

Table 10. Performance of Predicted DD across Different Look-ahead times

Look-ahead Time (min)	MAD for instantanous DD model	MAD for model based on predicted aircraft count	MAD for DD model with look-ahead time	RMS for instantanous DD model	RMS for model based on predicted aircraft count	RMS for DD model with look-ahead time
0	0.91	1.06	0.96	1.31	1.11	1.11
20	1.08	1.01	0.92	1.24	1.35	1.07
40	1.33	1	0.89	1.22	1.85	1.04
60	1.63	1.03	0.95	1.25	2.17	1.11
80	1.86	1.05	0.96	1.24	2.61	1.11
100	1.64	1.08	0.93	1.26	2.23	1.11
120	1.62	1.12	0.9	1.28	2.36	1.07

The DD with look-ahead time model (i.e., prediction intervals built into equation) performed better than the model based on instantaneous DD only. DD appeared to be more stable over time than predicted number of aircraft and DD appeared to be more accurate over time than predicted number of aircraft.

It must be noted that the DD model with look-ahead time seemed to perform only slightly better (in terms of MAD and RMS) than the model based on predicted aircraft count only for predicting complexity up to 120 minutes. This is an interesting finding considering that the DD model for instantaneous complexity seems to a lot better than the aircraft count only. One quick possible explanation is that the DD model does not perform better as expected. However, careful scrutiny of the results shows that the complexity that is predicted up to 20 minutes ahead using the DD model with look-ahead time is still using the number of predicted aircraft count as one of the measures. Hence, it was highly dependent on the predicted number of aircraft. It has been shown that these predictions of number of aircraft are not very good [11]. Because the predicted DD values used these less accurate estimates of the predicted aircraft count, it is completely plausible that the predicted DD values were also inaccurate. Therefore, it is not the DD model that was inaccurate, rather, the researchers believe that it was the inherent inaccuracy in the predicted aircraft count that made the predicted DD inaccurate. Further, the researchers believe that the instantaneous DD seemed to be more accurate than the predicted DD. In the case of instantaneous DD, there was no such aircraft prediction inaccuracy, whereas in the predicted DD there was inherent prediction inaccuracy.

Overall Conclusions

- The DD metrics have promise, most notably as a unified metric with contributing variables from the FAA WJHTC/Titan Systems, NASA, and Wyndemere/Metron metrics.
- The DD metrics perform better than aircraft count, which is the basis of the presently used complexity gauge.
- The models can be further developed and tested with techniques such as neural networks, genetic algorithms, and non-linear regression.
- The current study used ETMS as the raw source of traffic data. However, using more frequently updated data, such as System Analysis and Recording, Center TRACON Automation System, or a combination of the above could further increase the accuracy of aircraft positions.
- The researchers recommend using the DD metric in the simulation environment and plan to continue fine tuning the variables and their weights. Subsequently, an operational prototype could be deployed at a test site for hands-on evaluations.
- The performance of the predicted DD with look-ahead time is marginally better than the predicted aircraft count, quite possibly due to the inherent inaccuracy in the aircraft count prediction rather than the DD model itself.

References

[1] RTCA Task Force 3 Free Flight Implementation Report, 1995, RTCA: Washington DC.

[2] Laudeman, I.V., Brasil, C. L., & Branstrom, R., 1996, *Air Traffic Control in Restructured Airspace: A Study in Tactical Decision-Making*, Power point presentation.

[3] Federal Aviation Administration, 2001, *The Measure of Air Traffic Control Sector Complexity for the En Route Environment: Phase II Experiment Plan*, FAA WJHTC Internal Document.

[4] Chatterji, G.B. & Sridhar B., 2001, *Measures for Air Traffic Controller Workload Prediction*, Proceedings of the First AIAA Aircraft Technology, Integration, and Operations Forum, Los Angeles, CA.

[5] Mogford, R.H, Guttman, J.A., Morrow, S. L., & Kopardekar, P., 1995, *The complexity construct in Air Traffic Control: A review and Synthesis of the Literature*, DOT/FAA/CT-TN-95/22, FAA Technical Center: Atlantic City.

[6] Federal Aviation Administration, 2002, *Dynamic Density Metric Development and Validation Research Management Plan*.

[7] Kopardekar, P., 2000, *Dynamic Density: A Review of Proposed Variables*, FAA WJHTC Internal Document.

[8] Laudeman, I.V., Shelden, S.G., Branstrom, R., & Brasil, C.L., 1999, *Dynamic Density: An Air Traffic Management Metric*, NASA-TM-1998-112226.

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

[10] Wyndemere, 1996, *An Evaluation of Air Traffic Control Complexity*, Final Report, Contract Number NAS2-14284 (NASA contract).

[11] Volpe National Transportation Systems Center, 1996, *Traffic Management System (TMS)- ETMS Monitor Alert Analysis Report*, U.S. Department of Transportation, Cambridge, Massachusetts.

Biographical Sketches

Parimal Kopardekar holds Ph.D. and M.S. degrees in Industrial Engineering. He works as an Engineering Research Psychologist with the Federal Aviation Administration and is located at the National Aeronautics and Space Administration's Ames Research Center. His experience and interests include examination of advanced aviation concepts, human performance assessment, usability assessments, design of experiments, simulation and modeling, and project management. He can be contacted at pkopardekar@mail.arc.nasa.gov.

Sherri Magyarits has a M.S. degree in Aeronautical Science and a B.A. degree in Psychology. She is a Statistician/Research Psychologist in the Simulation and Analysis Branch (ACB-330) at the FAA William J. Hughes Technical Center in Atlantic City, NJ, and Project Lead for DD. She has over ten years of experience in the examination and analysis of human performance, design of experiments, and conduct of air traffic control simulations. She can be contacted at sherri.magyarits@faa.gov.