A NON-LINEAR RELATIONSHIP BETWEEN CONTROLLER WORKLOAD AND TRAFFIC COUNT

Paul U. Lee
SJSU / NASA Ames Research Center
Moffett Field, CA, 94035-1000
plee@mail.arc.nasa.gov

Controller workload has been a focal topic in air traffic management research because it is considered a key limiting factor to capacity increase in air traffic operations. Because workload ratings are subjective and highly prone to individual differences, some researchers have tried to replace workload with more objective metrics, such as aircraft count. A significant caveat in substituting these metrics for workload ratings, however, is that their relationships are non-linear. For example, as the number of aircraft increases linearly, the controller's perceived workload jumps from low to high at a certain traffic threshold, resulting in a step-function increase in workload with respect to aircraft count, suggesting that controllers perceive workload categorically. The non-linear relationship between workload and aircraft count has been validated using data collected from a recent study on the En Route Free Maneuvering concept element (Lee, Prevot, Mercer, Smith, & Palmer, 2005). The results suggest that objective metrics, such as aircraft count, may not be used interchangeably with subjective workload. In addition, any estimation on workload should not be extrapolated from a set of workload measures taken from an experiment since the extrapolated workload is likely to significantly underestimate workload.

INTRODUCTION

Anticipated increase in future air traffic demand has led to a number of concepts aimed at improving efficiency and increasing capacity while maintaining a high level of safety. While some concepts or procedures, such as Reduced Vertical Separation Minima (RVSM), can significantly increase physical airspace capacity, the increased capacity cannot be fully utilized unless additional aircraft can be safely managed by the sector controller(s). Since workload is a limiting factor in en route and transition airspace capacity, it is a key metric that has been researched in the past (e.g. Hilburn, Bakker, Pekela, & Parasuraman, 1997, Stein, 1985).

Since workload is highly subjective and has been shown to have large individual differences, attempts have been made to replace subjective workload metrics with correlated objective metrics, such as peak aircraft count, traffic geometry, total time in sector, number of clearances, etc. Over the years, researchers have explored ways to quantify the underlying traffic and sector complexity that drive controller workload. Mogford and his colleagues (1995) proposed that air traffic control (ATC) complexity – defined as a combination of both sector and traffic complexity – affects controller workload through mediating factors, such as quality of equipment, cognitive strategies, and individual differences. Another term, "Dynamic Density", has been used to define the collective effect of all factors that contribute to ATC complexity (e.g. Kopardekar & Magyarits, 2002). One of the key motivations for Dynamic Density research is to find a set of metrics that can replace current day Monitor Alert Parameters (MAPs) to predict traffic complexity and associated controller workload.

A general approach to correlating workload and objective metrics has been to first identify and/or define objective metrics. Using those metrics, multivariate linear regression models were fitted to the data and the metrics that contribute little to the overall prediction were eliminated from the regression models (e.g. Kopardekar & Magyarits, 2002; Majumdar & Ochieng, 2002; Laudeman, Shelden, Branstrom, & Brasil, 1999).

A key shortcoming of such an approach, however, is that perceived workload is often non-linear. For example, controllers may report relatively low workload given a busy but manageable traffic. Once the traffic reaches a certain critical threshold, however, a significantly higher workload may result from a slight increase in traffic or an occurrence of an off-nominal event. Despite a general recognition in the research community that relationships between workload and complexity factors are probably non-linear, details of the relationships remain largely unknown (Athenes, Averty, Puechmorel, Delahaye, & Collet, 2002) and unsubstantiated by experimental data.

In this paper, a linear regression model of workload was compared to two alternative non-linear models to examine the relationship between workload and aircraft count. The regression models were fitted against high traffic data that were collected during an evaluation of En Route Free Maneuvering concept element under Distibuted Air-Ground Traffic Management (DAG-TM) (Lee, Prevot, Mercer, Smith, & Palmer, 2005). The resulting aircraft count was actually higher than current day MAPs due to advanced decision support tools that reduced overall task load per plane, thereby allowing the controllers to handle more planes. The aircraft count was selected as the objective metric to be compared to workload in this initial analysis because it has been shown to correlate well with workload in past research (e.g. Masalonis, Callaham, & Wanke, 2002; Manning, Mills, Fox, Pfleiderer, & Mogilka, 2001). Similar analyses are currently being planned for other ATC complexity metrics.

Each of the following regression models supports a different hypothesis of controller workload:

- Linear – workload increases linearly with increasing aircraft count (and presumably with other task load metrics). A linear model suggests that subjective workload directly reflects the number of tasks that a controller performs in a given traffic situation.
- Exponential – workload is relatively low when the traffic is low, but workload increases at a faster pace as the traffic increases. An exponential model suggests that workload is modest below a certain threshold but that it quickly becomes unmanageable after the threshold with each added aircraft.
- S-curve – workload is low and relatively constant when the traffic is low but when the aircraft count passes a certain threshold, the workload becomes high and then levels off with increasing traffic. An s-curve model suggests that subjective workload is categorical – i.e. the controller feels that workload is low, high, or ultimately unmanageable. As the aircraft count, traffic complexity, and/or task loads increase, workload remains relatively constant in one of the three categories (i.e. low, high, unmanageable) until a certain threshold is reached, at which point the workload "jumps" to the next level.

By comparing the traffic scenarios that were run in the 2004 DAG-TM study with the three regression models, we can evaluate which model fits best with the observed data. Following the results section, implications of the findings will be discussed.

METHOD

Participants

The experiment included 22 commercial airline pilots and 5 certified professional air traffic controllers. Four controllers staffed four radar positions (three high altitude sectors and one low altitude sector). One additional controller served as a tracker/supervisor, supporting the radar controllers during peak workload periods. Twenty one aircraft simulators were flown by participant pilots at NASA Ames and NASA Langley. All remaining aircraft in the simulation were flown by pseudo-pilots with autonomous agent support at NASA Ames and NASA Langley.

Airspace

The simulation airspace included portions of Albuquerque Center (ZAB), Kansas City Center (ZKC), Fort Worth Center (ZFW) and Dallas-Fort Worth TRACON (DFW) (Figure 1). Controller participants worked four test sectors in the northwest arrival corridor: three high altitude sectors (Amarillo in ZAB, Wichita Falls and Ardmore in ZFW), and one ZFW low altitude sector (Bowie). Three retired controllers handled the surrounding traffic that entered or exited the test sectors.

Arrivals transitioned Amarillo high and Wichita Falls high from the northwest and Ardmore high from the north. The two main streams of arrivals merged at the BAMBE meter fix in the Bowie low sector before entering the TRACON. The traffic mix in Amarillo consisted of arrivals and overflights in level flight. A significant portion of the Wichita Falls traffic was arrivals while Ardmore had arrivals, departures, as well as a significant number of overflights.

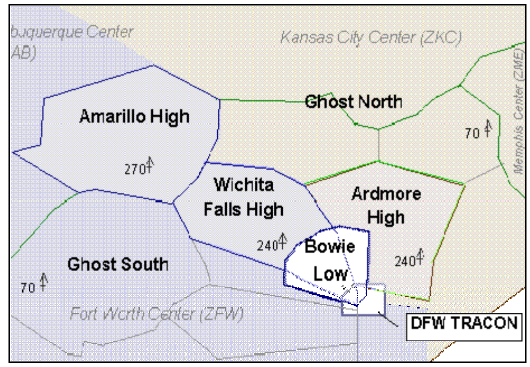

Figure 1. Simulated airspace

Ground Capabilities

To maximize the benefits of the advanced air and ground-side decision support tools (DSTs), they were integrated with Controller Pilot Data Link Communication (CPDLC) and the Flight Management System (FMS). This integration allows controllers and pilots to exchange 4-D trajectory information quickly and with low workload. The controller decision support tools have been integrated into a high fidelity emulation of the Display System Replacement (DSR) controller workstation. This DSR emulator is highly configurable to mimic both DSR workstations in the field today and future DSRs with advanced decision support tools. In order to support the concept, all aircraft were equipped with CPDLC, FMS, and automatic dependent surveillance-broadcast (ADS-B). The aircraft flown by the commercial pilot participants also had Cockpit Display of Traffic Information (CDTI) displays integrated with conflict detection & resolution (CD&R) and advanced required time of arrival (RTA) capabilities. More detailed descriptions of the ground capabilities are presented in Lee, et al. (2005).

Experimental Conditions

The study consisted of four experimental conditions. Each condition was run four times. The first two conditions were conducted at slightly above current day maximum traffic levels (Level 1), the former consisting of entirely managed aircraft and the latter having a mix of autonomous and managed aircraft. The next two conditions added increasing numbers of self separating overflights. The arrival traffic, while demanding, remained relatively constant throughout all scenarios. Consequently, the Bowie low sector, which only had the arrival traffic, maintained a relatively constant traffic volume across conditions.

The overall study lasted two weeks, which were roughly divided into one week training followed by a week of data collection runs. Due to constraints in training time, the controller participants did not rotate through different sectors during training or during data collection, maximizing their proficiencies in their assigned sectors. Hence, the workload ratings in each sector also represent the ratings of a particular controller who was assigned to that sector, confounding sector specificity with individual differences.

For the analyses in this paper, only the first condition (i.e. 100% managed airspace with high traffic), consisting of four runs per sector, is discussed. Details of the overall experimental results are presented in Lee et al. (2005). The traffic levels for the three high altitude sectors were established through an informal "traffic load test" prior to the final simulation (Lee, Mercer, Smith, & Palmer, 2005). In the load test, traffic scenarios with different levels of peak aircraft count were presented to controller participants. After managing each traffic scenario, they provided feedback on whether or not the traffic was manageable. Based on the feedback, the traffic levels were adjusted and the process was repeated until the maximum manageable traffic level was established in each sector. The traffic scenarios were subsequently re-adjusted during a series of dress rehearsals based on further controller feedback. The resulting traffic scenarios were quite challenging, often pushing the controllers up to, but not over, their workload limits.

RESULTS

Aircraft Count

The traffic scenarios gradually increased in traffic to its maximum during the first twenty minutes of the simulation. The maximum traffic was maintained during the next 30-35 minutes before tapering off for the last 5-10 minutes of the simulation run. Figure 2 illustrates the general traffic patterns in the high altitude sectors by plotting the maximum number of owned aircraft at each 5-minute time block vs. simulation run time. The maximum aircraft count in the three high altitude sectors was 26, 22, and 21 for Amarillo, Ardmore, and Wichita Falls, respectively. The MAP values for these sectors were 18. Since the workload was low for the Bowie sector controller, Bowie sector data were excluded from the following analyses.

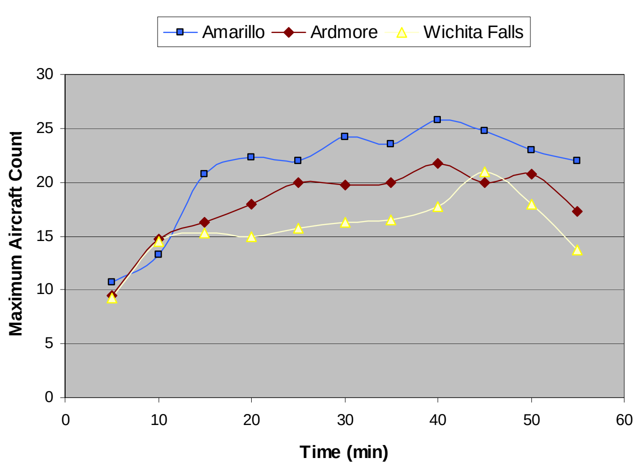

Figure 2. Aircraft count vs. simulation run time; aircraft count averaged across 4 simulation runs

Subjective Workload

Subjective workload assessments were collected from the controllers using the Air Traffic Workload Input Technique (ATWIT) (Stein, 1985). During the simulation runs, controllers were required to rate their workload on a scale of 1 to 7 at 5-minute intervals. Figure 3 shows the workload ratings over time for the three high altitude sectors. The pattern of workload ratings generally mirrors that of the aircraft count, as suggested by the similarities of the plots in Figure 2 and 3. However, regression analyses in the following section reveals that the correlation between workload and aircraft count are non-linear.

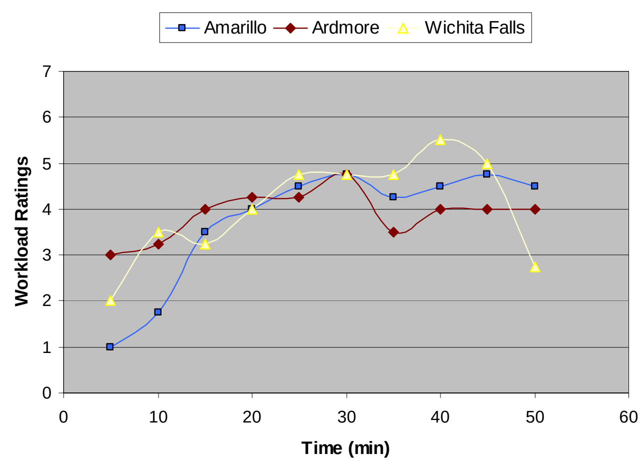

Figure 3. Workload ratings vs. simulation run time; ratings averaged across 4 simulation runs

An interesting and unexpected outcome of the workload ratings was that the controller participants rated the workload to be only moderately high, as their ratings rarely reached 6 or 7. These ratings were counter to both the over-the-shoulder observations and feedback from controller participants during the traffic load test and dress rehearsals. A possible factor for the lower-than-expected workload ratings may be that traffic scenarios which seemed challenging during the initial feedback stage may have become easier to handle as they became more familiar with the tools and the sector characteristics after the final training sessions. Another factor may be that controllers generally "saved" a 6 or 7 workload rating for extremely difficult traffic situations, such as ones caused by thunder storms, and they felt that high aircraft count alone did not warrant those maximum workload ratings during the experiment.

Regression Fit

Controller workload was plotted against the peak aircraft count during the five minute span prior to the workload assessment. For the Amarillo sector, a visual inspection of the plotted data suggests relatively low workload ratings between 1 and 2 for the aircraft count of 16 and below. Workload ratings jump to 4 and 5 when the aircraft count is greater than 20. A similar pattern of results emerged from the other two sectors, Ardmore and Wichita Falls.

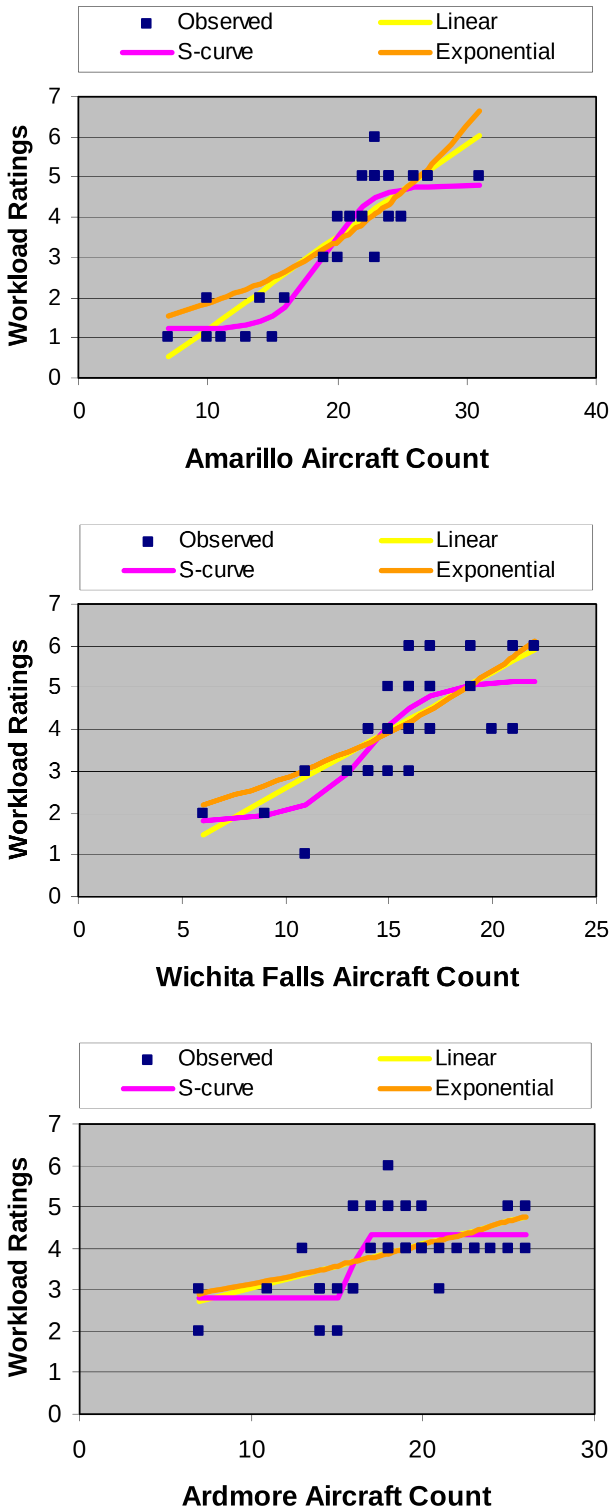

Figure 4. Workload vs. aircraft count: observed and regression fits for the high altitude sectors

A linear, exponential, and s-curve regression lines/curves were fitted against the plotted data. The R^2 values were then calculated for the regression lines/curves. The regression equations were defined as follows:

- W = predicted workload
- AC = aircraft count

- Linear: W = $a + b * AC$
- Exponential: W = $a * e^{b*AC}$
- S-curve: W = $\dfrac{a}{1 + e^{b*AC+c}} + d$

For the non-linear regression curves (i.e. exponential and s-curve), the parameters in the regression models were iteratively estimated to minimize the $SS_{error.}$ The iterative estimation/optimization of parameters (i.e. a and b for exponential; a – d for s-curve) was done using Microsoft Excel's Solver. For example, Table 1 shows the optimized parameter values that provided best fits for the regression models in the Amarillo sector.

Parameters	Linear	Exponential	S-curve
a	-1.102	1.002	3.592
b	0.231	0.061	-0.566
c	–	–	-10.73
d	–	–	1.207

Table 1. Optimized regression model parameters for the Amarillo sector

Table 2 summarizes the R^2 values for the regression curve fits in the three high altitude sectors. Across all three sectors, the s-curve regression model provided the best fit, followed by the linear model, then the exponential model. All models were significant, i.e. $p < 0.001$.

	Linear	Exponential	S-curve
Amarillo	0.77	0.70	0.84
Wichita Falls	0.54	0.50	0.61
Ardmore	0.27	0.25	0.52

Table 2. R^2 of linear, exponential and S-curve regression fit in the high altitude sectors; p-values < 0.001

The results showed that the Amarillo sector's workload data provided the best fit to the regression models, followed by the Wichita Falls data, then the Ardmore data, regardless of the type of regression model. The s-curve model provided a much better fit to the data compared to the linear and the exponential models in Ardmore sector, but the improvement was relatively small in Amarillo and Wichita Falls sectors.

DISCUSSION

A comparison of three possible models of controller workload as a function of aircraft count yielded the best fit with the s-curve model. This finding suggests that the subjective workload was categorical. A sudden jump in workload ratings from low to high was striking, especially for the Amarillo and Wichita Falls sectors. The workload ratings remained low and constant for aircraft count below a certain threshold (approximately 12 for Wichita Falls and 15 for Ardmore and Amarillo), then quickly increased for the next 3 – 8 additional planes, and then leveled off again for the high traffic counts. Since the traffic scenarios were designed to create high but manageable workload, we could not examine the amount of traffic that would increase the workload ratings from high to unmanageable.

It is unclear, however, where the transition from high to unmanageable workload would have occurred. Workload ratings that increased quickly around the threshold remained constant once the aircraft count exceeded 20 for Amarillo and 16 for Ardmore and Wichita Falls. Over-the-shoulder observations seemed to suggest that even though the controller participants reported a constant workload with increasing aircraft count above 20, the increased task load (e.g. handoffs, check-ins, etc.) associated with the greater number of aircraft seemed to reduce their availability to perform additional tasks. If the task load exceeded the time available to work the traffic, controllers likely would have reported the workload to be unmanageable. Besides aircraft count, the factors that contribute to the time demands include the number of potential conflicts, traffic complexity, and off-nominal events. The contributions by these and other potential factors have yet to be fully examined and will be addressed in the near future.

An interesting outcome of the regression models is that although the s-curve provided the best fit to the data, the linear regression captured a significant portion of the variance in the Amarillo and Wichita Falls sectors, as shown by similarities in R^2 values between the two models in those two sectors. The results suggest that a linear regression model may often be adequate to predict *interpolated* workload values between measured workload ratings under certain situations (e.g. certain controllers and/or sectors). However, a significant improvement in R^2 value for the Ardmore sector using an s-curve fit suggests that significant differences in workload characteristics between different controllers and/or sectors may be more consistently captured by an s-curve model.

Human-in-the-loop simulation data are often used as input parameters to feed fast-time models that explore various air traffic concepts. However, the data may be unsuitable for use in any fast-time modeling that explores increased airspace capacity by lowering controller workload. The results from this study indicate that none of the models adequately *extrapolate* workload from measured data. For example, if the workload data were taken only for traffic scenarios of 16 or less aircraft in the Amarillo sector, at an aircraft count of 21, a linear and an s-curve regression models would have yielded predicted workload values of 1.94 and 3.0, respectively, whereas the actual workload ratings at 21 aircraft were 4.0. Both models severely underestimated the workload, although the s-curve model still performed better than the linear model. Therefore, research based on workload extrapolations should be done with great caution and with conservative assumptions.

CONCLUSION

The relationship between workload and aircraft count has been examined using traffic scenarios with a high aircraft count. Linear and non-linear regression models were fitted to the observed data, which yielded the best fit with an s-curve model, suggesting that perceived workload is categorical. There are interesting implications to the non-linear relationship between subjective workload and traffic count. First, metrics such as traffic count or task loads should not be used interchangeably with subjective workload. Secondly, any estimation on workload should not be extrapolated from a set of workload measures taken from an experiment since the extrapolated workload is likely to significantly underestimate workload. Finally, the threshold that separates low, high, and unmanageable traffic levels needs to be determined in order to fully characterize the impact of traffic patterns and other situational factors on controller workload, but further research is needed to understand how to accurately account for these factors.

References

Athenes, S., Averty, P., Puechmorel, S., Delahaye, D., and Collet, C. (2002). ATC complexity and controller workload: trying to bridge the gap, HCI-Aero 2002: International Conference on Human-Computer Interaction in Aeronautics, Cambridge, MA.

Hilburn, B. G., Bakker, M. W. P., Pekela, W. D.,& Parasuraman, R. (1997). The effect of free flight on air traffic controller mental workload, monitoring and system performance. *In Proceedings of the 10th International CEAS Conference on Free Flight* (pp. 14–1/14–12) Amsterdam: CEAS.

Kopardekar, P. & Magyarits, S. (2002). Measurement and prediction of dynamic density, *5th USA/Europe Air Traffic Management R&D Seminar,* Budapest, Hungary.

Laudeman, I. V., Shelden, S. G., Branstrom, R., & Brasil, C. L. (1999). Dynamic density: an air traffic management metric, *NASA-TM-1998-112226.* NASA Ames Research Center, Moffett Field, CA.

Lee, P. U., Mercer, J., Smith, N., & Palmer, E. (2005). A Non-linear relationship between controller workload, task load, and traffic density: the straw that broke the camel's back, *Proceedings of the 13th International Symposium on Aviation Psychology*, Oklahoma City, OK.

Lee, P. U., Prevot, T., Mercer, J., Smith, N., & Palmer, E. (2005). Ground-side perspective on mixed operations with self-separating and controller-managed aircraft, *The 6th USA/Europe Air Traffic Management R&D Seminar ATM-2005,* Baltimore, MD.

Majumdar, A. & Ochieng, W. Y. (2002). Factors affecting air traffic controller workload. Multivariate analysis based on simulation modeling of controller workload, *Transportation Research Record, 1788,* pp. 58-69.

Manning, C., Mills, S., Fox, C., Pfleiderer, E., and Mogilka, H. (2001). The relationship between air traffic control communication events and measures of controller taskload and workload. The 4th USA/Europe Air Traffic Management R&D Seminar, Santa Fe, NM.

Masalonis, A. J., Callaham, M. B., & Wanke, C. R. (2002). Dynamic density and complexity metrics for real time traffic flow management, *5th USA/Europe Air Traffic Management R&D Seminar,* Budapest, Hungary.

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, NJ.

Stein, E. (1985). Air traffic controller workload: an examination of workload probe., *DOT/FAA/CT-TN84/24.* DOT/FAA Technical Center. Atlantic City, NJ.