A QUANTITATIVE MODEL FOR EN ROUTE ERROR RATE ANALYSIS

Daniel Murphy, Federal Aviation Administration, Washington, DC

Gerald Shapiro, The CNA Corporation, Alexandria, Virginia

Abstract

In this paper we demonstrate that the rate of en-route operational errors increases faster than the increase in traffic. We propose a model that explains this relationship, which allows us to compare operational error rates for periods with different traffic levels. We use this model to show that URET, a controller decision support tool recently deployed by the FAA, has significantly reduced the rate of operational errors. These results have important implications for analyzing the safety and capacity impact of new en-route automation systems being proposed as part the next generation air transportation systems for the United States and Europe.

Key Words

URET, Operational Errors, controller workload

Introduction

In the en-route environment, capacity limits are determined by safety standards rather than by some physical constraint on the number of aircraft a piece of airspace (a sector) can hold. Therefore, as next generation air transportation systems are being designed in the United States and in Europe to increase en-route capacity, it is critical to understand the relationship between safety and traffic. It is also critical to understand how new automation systems affect this relationship.

We will use operational errors (OE) as a proxy for safety. An OE is the violation of minimum separation standards as the result of a failure on the part of air traffic control.

In this paper we first demonstrate that en route operational error rates grow at least as fast as the square of the number of aircraft under control at the time of the error. This result alone is significant, as the rarity of operational errors have made the relationship of operational errors and traffic less than obvious. Although some researchers have pointed out that increasing traffic is a correlate of increased OE risk, other researchers have characterized operational errors as occurring at low to moderate workload levels. The latter statement implies that a low level of mental stimulation can result in OEs. Although this explanation is plausible from our daily experiences, the data show that it is not the general case in air traffic control. Risk grows, not decreases, with traffic. We propose a model which captures this relationship.

Secondly, we use our model to show that automation systems which alleviate controller workload can reduce the rate of operational errors. The particular tool we discuss is the User Request and Evaluation Tool (URET), which has substantially reduced operational error rates. Although the fundamental relationship between traffic and error rates remains unaltered, i.e. error rates grow faster than traffic, URET greatly reduces the magnitude of the effect.

Analyses of this type will be important in future system evaluations. En route operational errors are a measure of safety in the en route system. We must ensure that capacity-enhancing system modifications are also safety-enhancing. If operational risks increase with an increase in traffic, then air traffic managers will be forced to limit traffic in order to preserve safety margins. Our analysis allows us to evaluate how the fundamental relationship between errors and traffic is modified by a system change, independently of the concurrent changes in traffic levels.

We conclude with observations on the means by which URET may be achieving safety benefits, and the implications for designers and developers of the next generation of air traffic management tools.

Background

Previous researchers have observed that the number of en route operational errors per unit time grows like the square of flights handled per unit time. ([1], [2]).Graham and Orr developed a model that implies that terminal-area near midair collisions grow with the square of terminal area operations [3]. The model was validated by reported near midair collision data.

Other researchers have speculated that en-route operational error rates are correlated with the number of aircraft under control at the time of the error, ([1], [2], and [4]). Rogers, et al. [4] provide a summary survey of research supporting this idea. In this paper we develop an explicit mathematical formulation of the relationship between operational errors and the number of aircraft under the control of a sector at the time of the error.

The number of aircraft under control of a sector is only one of the factors contributing to the risk of an operational error. Our model is parsimonious, in that it summarizes all of these factors into two parameters, but due to that austerity, we must rely on other research in our attribution of causality. Hansen and Zhang [1] survey studies investigating factors contributing to operational error risk, and conduct an empirical investigation of their own. A similar study was conducted by Love [5]. Gosling [2] and Rogers, et al. [4] also provide valuable summaries of sector, workload, and traffic factors affecting OE risk. Rogers, et al. analytically examine the significance of various factors contributing to OEs in Atlanta Center.

Regarding our case study, Bolic and Hansen [6] identify the features of URET that are commonly used. We relate these features to the risk reduction factors in the literature, as explanation of the error rate reduction that we see when URET is introduced.

Our operational error data comes from Federal Aviation Administration (FAA) databases, and includes all categories of OEs.

A Parametric Model for Operational Error Rates

Let us assume that an air traffic controller, or sector controller team, in any short period of time is equally likely to make a mistake with each aircraft handled. We further assume that this mistake rate is the same for each sector, regardless of the particular controllers present or the number in the sector controller team. Under these assumptions, the chance of an air traffic control mistake with an aircraft at a given instant in time is proportional to the number of aircraft *n* in the sector at that instant:

$$P(M \mid n) = \alpha\, n$$

An example of a mistake might be that the controller assigns an aircraft an incorrect altitude or heading, or that the controller fails to recognize a pilot read-back error. Such a mistake by itself does not constitute an operational error, unless the mistake results in the loss of separation with a second aircraft. Assuming that the chance that such a second aircraft will be in the airspace where the first aircraft was misdirected is also proportional to the number of aircraft in the sector, then the chance of an operational error is

$$P(OE \mid n) = \beta \times n \times P(M \mid n) = \gamma\, n^2$$

It is likely that the non-linearity is greater than suggested by this argument. Our first assumption was that the rate of mistakes is independent of the number of aircraft in the sector. Intuitively, it is probable that mistakes are more likely when a controller is managing more aircraft. We can model this assumption by assuming that

$$P(M \mid n) = \alpha\, n^q$$

where q is some exponent greater than or equal to 1.

Our observable is the number of operational errors, and therefore we can fit the probability of an operational error to the following

$$\rho_n = P(OE \mid n) = \gamma\, n^{1+q} \qquad (1)$$

Evidence Supporting Model

To directly test the theory outlined above, we need to know how many aircraft a sector was responsible for when each operational error occurred, a number which the FAA's Air Traffic Organization tracks as part of its review of each operational error. The solid line in Figure 1 shows the distribution of operational errors versus the sector load at the time of the operational error. The data are all 3,057 operational errors in the CONUS for fiscal years 2002 through 2005.

As pointed out in [1], [2], and [4], just the count of operational errors by sector load is not sufficient to test Equation (1). The error count by sector load does not account for the fact that sector loads occur with different frequencies. Formally, the operational error database contains the joint distribution of operational errors and sector load, but we desire to evaluate the conditional distribution of an operational error given the sector load. To do this, we need, in addition to the joint distribution, the distribution of the number of aircraft in a sector. The Host Aircraft Management Exec (HAME) data provides us a source for this distribution. HAME records when each aircraft is "handed off" by a controller to the next sector. HAME data are available from August 2000 through the end of April 2006 for all of the twenty CONUS centers. For each sector in every CONUS en-route center, we have computed the number of aircraft being handled in one-minute bins; the resulting distribution for the entire CONUS for the same time period is also shown in Figure 1.

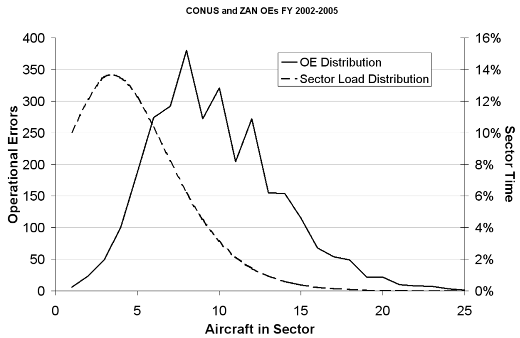


Figure 1 – Distribution of Operational Errors and Aircraft Handled

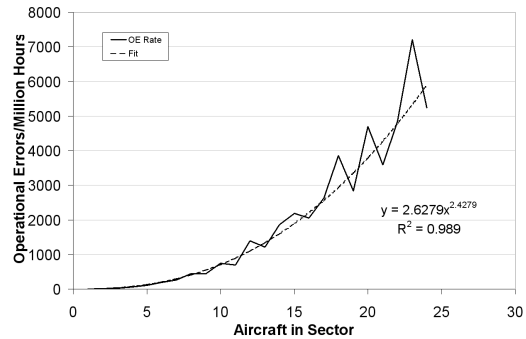


Figure 2 – OE Rate vs. Instantaneous Count

Notice that the peak of the instantaneous sector load distribution occurs at approximately four aircraft, while the peak of the operational error (OE) distribution occurs near eight aircraft. This shift in the peak indicates that the relation between operational errors and traffic is non-linear. To show this explicitly, we divide the OE distribution by the sector load distribution to obtain the conditional distribution, and plot the result in Figure 2. This plot shows the rate, or probability, that an operational error will occur at each sector load level, and therefore is directly comparable to Equation (1). The smooth line in Figure 2 is a power-law fit to the OE rate, which shows that the exponent is approximately 2.5, confirming the theory that operational errors should vary as at least the square of instantaneous sector load. We note that the vertical scale in Figure 2 has been adjusted from the scale of Figure 1, OEs per 4 years, to OEs per million center-hours (hours per year times twenty).

Using the Model: Estimating URET Impact on Operational Errors

The original impetus of this research was tracking OEs to ensure that deployment of URET was not having an adverse safety impact. In fact, the results point to the opposite conclusion, that URET decreases operational error rates. Figure 3 shows NAS operational error rates from fiscal year 2000 through fiscal year 2006, the time period during which URET was deployed. Operational error rates at centers without URET roughly rose and fell in consort with the overall changes in traffic levels. Those centers with URET, on the other hand, showed a decreasing error rate.

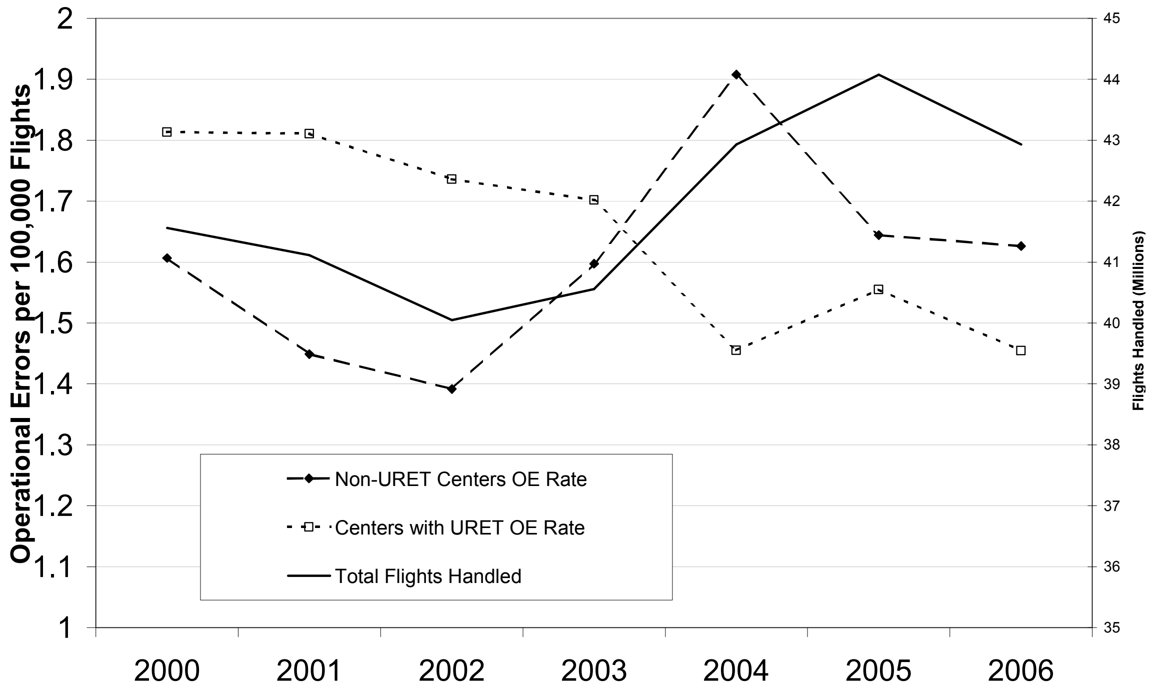


Figure 3 – URET vs. non-URET Operational Error Rates

Although the data in Figure 3 is highly suggestive of URET inducing a reduction in operational error rates, it is not probative. To further explore the impact of URET, we analyzed seven centers, those for which there were at least 20 months of HAME data pre-and post-URET by March of 2006. For each of those centers, we estimated the parameters γ and q from Equation (1) by minimizing the weighted sum-of-squared errors. The weighting factor used was the number of sector minutes observed for each load, N_n, a factor that compensates for the unequal variances at each observed point. (Assuming that the number of observed errors at each load n is a binomial random variable with probability ρ_n, then the variance of the observed rate is proportional to $1/N_n$.) Observed data with fewer than three, or greater than nineteen aircraft in the sector were not included in estimating γ and q, as points in this range did not fit the model very well. This does not seem unreasonable. Errors when handling one or two aircraft are likely due to different causes than errors when three or more aircraft are involved. Furthermore, handling twenty or more aircraft is quite a large traffic load, and may represent a threshold of complexity that changes the relationship between traffic and error rate. For two cases additional points were excluded as outliers, (points n=3, 18 and 19 for Denver, and points n=15-19 for Chicago), in order to avoid instability in the solver program.

Power law models were estimated for each center over three time frames: pre-URET, post-URET, and the combined times pre-URET and post-URET. Data from each center's transition period, when URET was not in use in all areas of the center, are excluded from our analysis. Appendix A contains plots of the observed pre-URET and post-URET error rates, along with the estimated power law curves for those data.

It is clear from the plotted power curves that URET does not appear to have impacted error rates in either Cleveland or Ft. Worth Centers, either positively or negatively, and we will not analyze those centers further.

For the remaining centers, there does appear to be a lower operational error rate post-URET. We evaluate the statistical significance of this difference by comparing the likelihood ratio, Λ, of the data using separate pre- and post- power law rates against the model using a single power law for the combined time frames. If URET has no impact on error rates, then Λ will not (when compensated for the additional degrees of freedom in the more complex model) differ substantially from unity. On the other hand, if URET does influence the power law, then Λ will differ substantially from 1.0. The details of the likelihood ratio derivation are contained in Appendix B.

The models we are comparing here are hierarchical. That is, the model with the same rates pre- and post-URET is a special case of the model with separate rates. For such models, the statistic $-2\ln\Lambda$ has a chi-square distribution. The degrees of freedom in the chi-square distribution for the null hypothesis of no difference is the number of extra parameters in the more complex model, here two (γ, q pre- and post- URET, versus a single pair).

Table 1 shows the p-value for the likelihood ratio test for the five centers where there appears to be a difference pre- and post-URET. The p-value gives the chance that random data variation is responsible for the observed difference in predictive power between the separate models and the combined model. A small p-value indicates that the difference is significant.

Table 1 – p-values for Hypothesis that URET Reduces OE Model Parameters

Center	**Pre-URET *q***	**Post-URET *q***	**p**
Chicago	1.4	1.7	0.0007
Washington	2.2	1.3	0.000006
Denver	1.7	1.0	0.036
Kansas City	3.0	1.7	0.002
Minneapolis	1.5	1.1	0.010

For our data we see that the statistical significance of URET reducing operational error rates is extremely high for Washington, Chicago, and Kansas City Centers. Minneapolis and Denver Centers have a very high significance level. We note that in all cases but Chicago, the parameter q in the OE rate model decreases post-URET. In Chicago Center, the difference in pre-URET and post-URET rates was captured by the parameter γ, which decreased by a factor of three. We also note that in all cases, the parameter q is at least one, as predicted by the model.

We only speculate here why URET has a great impact on some centers, but virtually no impact on others. Additional aspects of the center's traffic or operations may have changed, besides the introduction of URET. Another possible reason is that URET is used differently in different facilities. [6] Hopefully field personnel familiar with their respective environments can offer suggestions for avenues of investigation of this question.

Why Does URET Reduce OE Rates?

URET's contributions to improved situational awareness and workload reduction make it a valuable tool in reducing en route operational errors.

Rogers, et al. [4] survey a number of studies investigating factors that influence operational error rates. Lack of situational awareness and high controller workload are, not surprisingly, consistently identified as factors that lead to higher rates of operational errors. URET's conflict alert functionality contributes to situational awareness for some controllers, but anecdotal evidence indicates that this is not a universally used feature [6]. On the other hand, Bolic and Hansen [6] report that the task-simplification aspects of URET are quite popular and widely used. Perhaps the most significant set of tasks URET automates concerns flight information. Prior to URET, flight information was presented to controllers on paper strips, which were annotated and organized by hand. URET provides flight information on its display, and for most sectors controllers are no longer required to use paper strips. This aspect of URET reduces the workload associated with data entry and retrieval, effectively reducing the workload associated with any given traffic level. Haase and Boone conducted a time-motion study that showed eliminating paper strip manipulation saved controllers an average of four minutes out of every twenty minute observation cycle [7]. While the savings varied based on traffic levels and conditions, and on sector staffing levels, it is clear that there is a substantive workload savings. It is not surprising then, that OE rates drop as URET is deployed and utilized.

Conclusion

In air traffic management today we limit traffic, and hence workload, assigned to an air traffic control sector to prevent operating in an unsafe fashion. *De facto* it is safety considerations that limit capacity. We have shown that operational error rates increase faster than the rate of traffic, which implies that en-route capacity may be more constrained than is commonly recognized.

We have also shown that a particular automation tool, URET, has reduced the rate of operational errors by reducing controller workload. Other new tools or procedures that reduce controller workload associated with a given traffic level may enable capacity to be increased without decreasing safety margins. Designers and planners of future ATM systems must consider how to further replace routine, non-planning ATC functions with automation alternatives.

Planners of the next generation automation systems currently being developed must consider this relationship between safety, capacity, and traffic in order to quantify the benefits of the proposed programs. We provide in this paper a model and technique for evaluating the effectiveness of safetyaltering system changes, in a way that is independent of fluctuations in traffic levels.

References

[1] Hansen, Mark and Yu Zhang, 2005, *The Link Between Operational Performance and Operational Errors in the National Airspace System*, ATM 2005, http://atmseminar.eurocontrol.fr/all-seminars/6th-usa-europe-atm-2005-r-d-seminar/paper_094

[2] Gosling, Geoffrey, 2003, *Analysis of Factors Affecting the Occurrence and Severity of Air Traffic Control Operational Errors*, Transportation Research Record, 1788.

[3] Graham, Walton and Robert H. Orr, 1970, *Terminal Air Traffic Flow and Collision Exposure,* Proceedings of the IEEE, vol 58, no. 3, pp 328-336.

[4] Rodgers, Mark, Richard Mogford, and Leslye Mogford, 1998, *The Relationship of Sector Characteristics to Operational Errors*, DOT/FAA/AM-98/14, Washington, DC, Office of Aviation Medicine, Federal Aviation Administration.

[5] Love, Dwight, 2006, *Effect of Sector Characteristics on the Frequency of Operational Errors -- Updated Regression Analysis (unpublished)*, F084-B06-029, McLean, VA, the MITRE Corporation

[6] Bolic, Tatjana, Mark Hansen, 2005, *User Request Evaluation Tool (URET) Adoption and Adaptation, Three Center Case Study*, ATM 2005, http://atmseminar.eurocontrol.fr/all-seminars/6th-usa-europe-atm-2005-r-d-seminar/paper_100

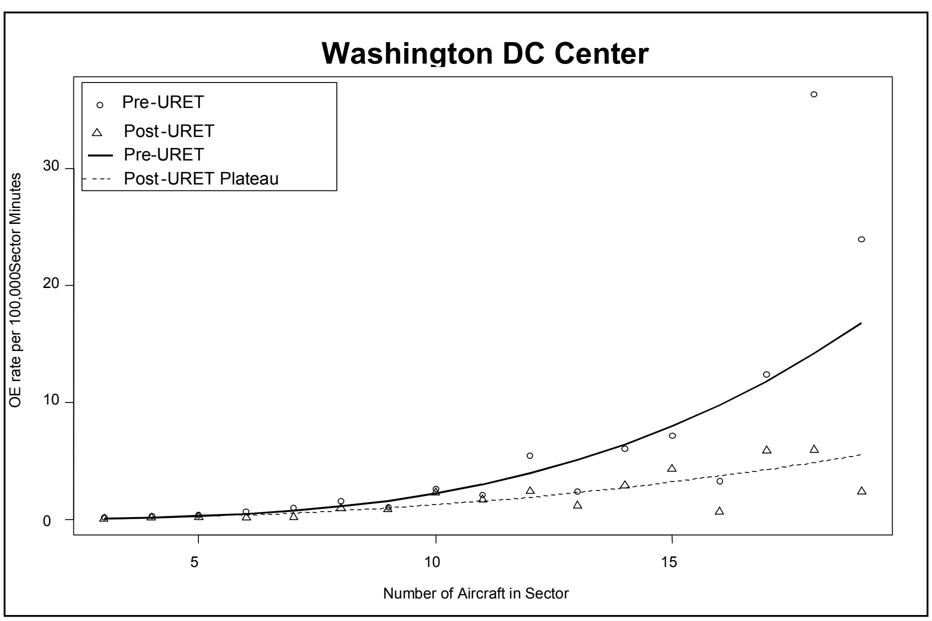
[7] Haase, Nikki and Brenda Boone, 2007, *FINAL REPORT: URET Electronic Flight Data Time and Motion Study (in preparation),* Washington, DC, Human Solutions, Inc.

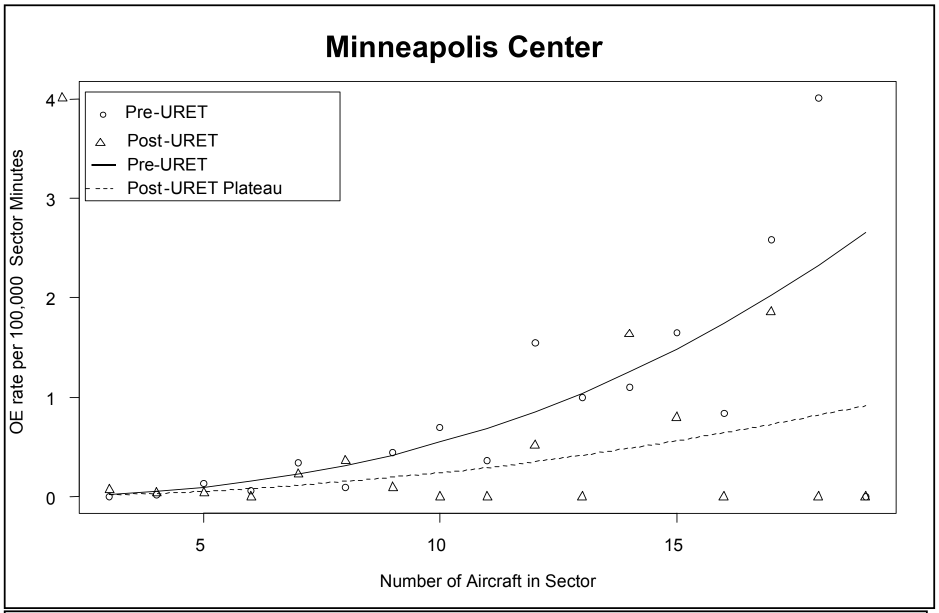
Biographies

Daniel Murphy manages the Business Outlook group within the Air Traffic Organization of the FAA. He has a PhD and MA in Physics from Duke University, and AB in Physics from Washington University in St. Louis.

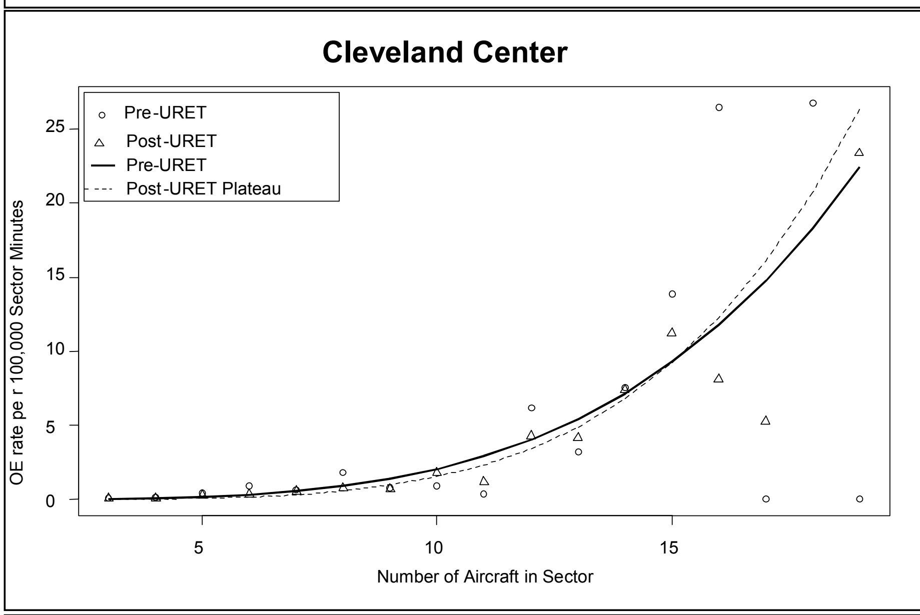
Gerald Shapiro is a Research Analyst with the CNA Corporation. He has worked in air traffic modeling and analysis on behalf of the FAA, NASA, Nav Canada, Calgary, Zurich, and numerous US airport authorities. He holds a PhD in Operations Research from North Carolina State University, and a BS in Systems Science and Mathematics from Washington University in St. Louis.

Appendix A – Fitted Models

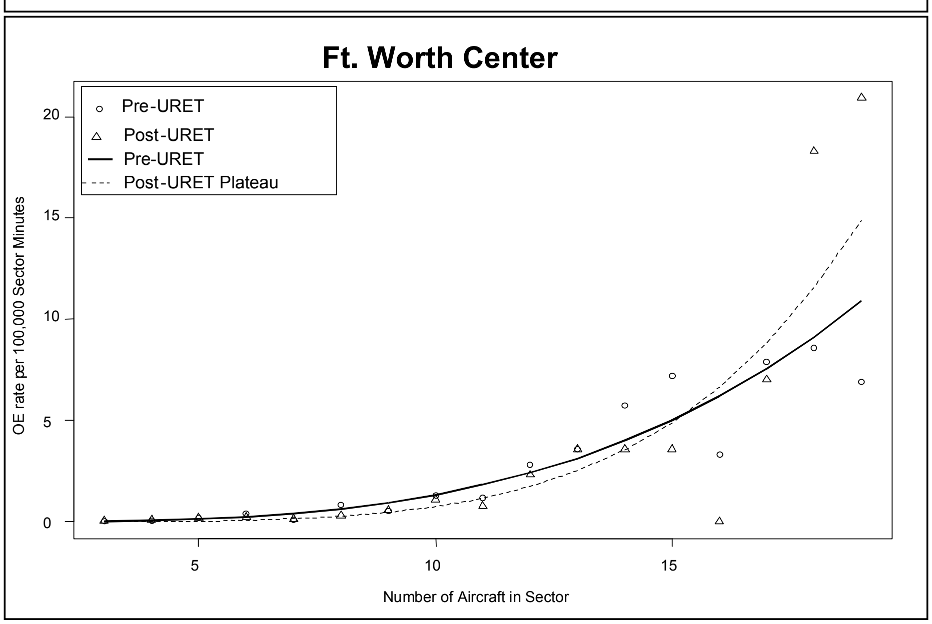




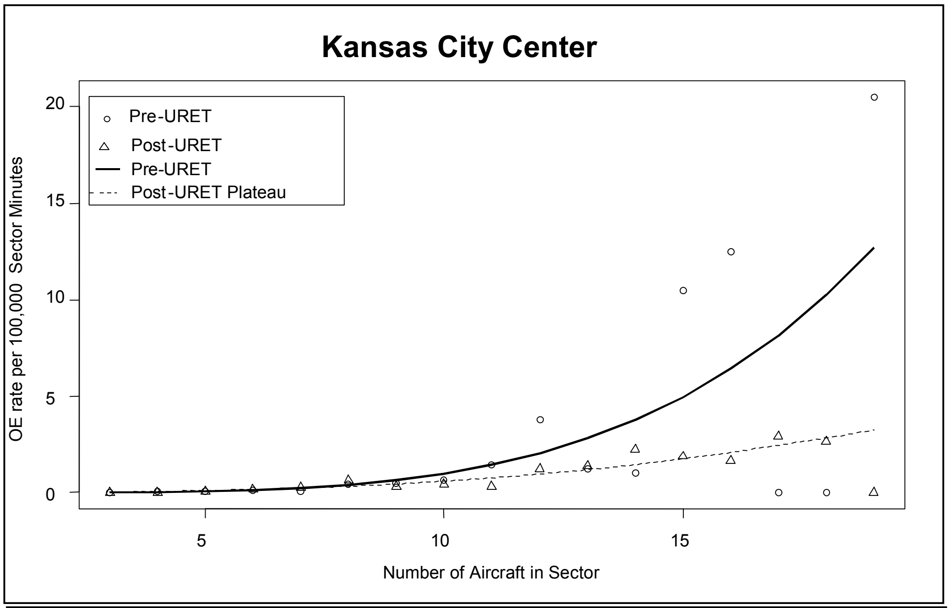
Minneapolis Center
Pre-URET
Post-URET
Pre-URET
Post-URET Plateau
OE rate per 100,000 Sector Minutes
0
1
2
3
4
5
10
15
Number of Aircraft in Sector

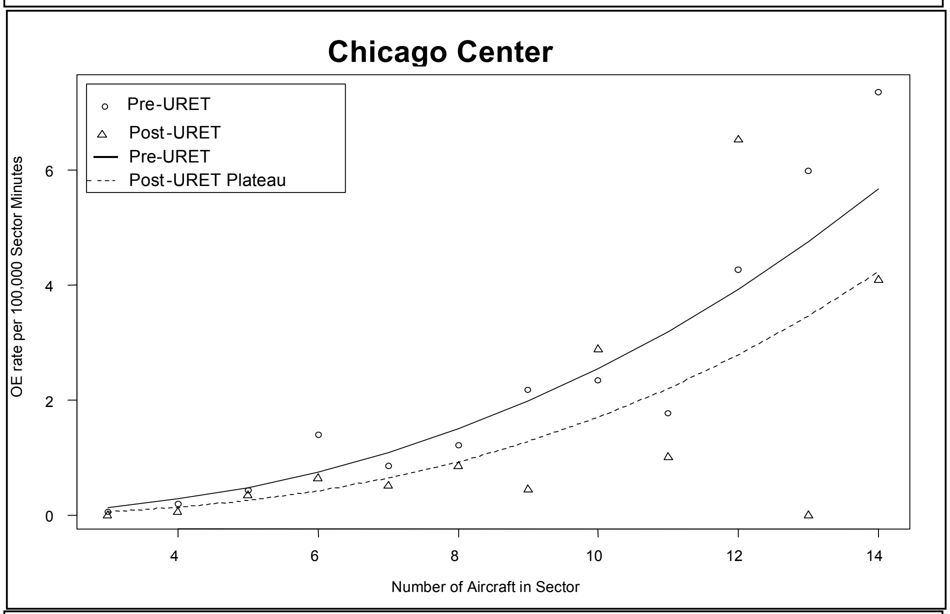


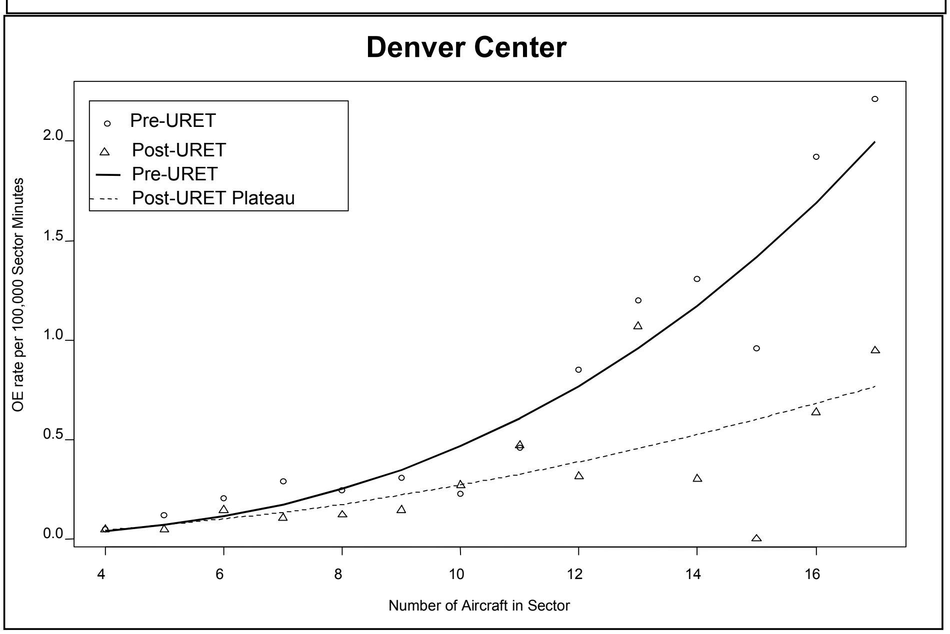
Cleveland Center
Pre-URET
Post-URET
Pre-URET
Post-URET Plateau
OE rate pe r 100,000 Sector Minutes
0
5
10
15
20
25
5
10
15
Number of Aircraft in Sector



Ft. Worth Center
Pre-URET
Post-URET
Pre-URET
Post-URET Plateau
OE rate per 100,000 Sector Minutes
0
5
10
15
20
5
10
15
Number of Aircraft in Sector


Kansas City Center
Pre-URET
Post-URET
Pre-URET
Post-URET Plateau
OE rate per 100,000 Sector Minutes
Number of Aircraft in Sector
0
5
10
15
20
5
10
15


Chicago Center
Pre-URET
Post-URET
Pre-URET
Post-URET Plateau
OE rate per 100,000 Sector Minutes
Number of Aircraft in Sector
0
2
4
6
4
6
8
10
12
14


Denver Center
Pre-URET
Post-URET
Pre-URET
Post-URET Plateau
OE rate per 100,000 Sector Minutes
Number of Aircraft in Sector
0.0
0.5
1.0
1.5
2.0
4
6
8
10
12
14
16

Appendix B – Likelihood Ratio

Let us define some notation.

R_n the number of observed operational errors when there are n aircraft in the sector
N_n the number of minutes sector is observed with a load of n
ρ_n the modeled error rate when the sector load is n

We distinguish the pre-URET, post-URET plateau, and combined time frame parameters by the superscripts p, u, and c, respectively.

Under the assumption that the number of observed errors follows a binomial distribution for each value of n, the likelihood function for the observed pre-URET data is

$$L^p = \prod_n \binom{N_n^p}{R_n^p} \left(\rho_n^p\right)^{R_n^p} \left(1-\rho_n^p\right)^{N_n^p - R_n^p}$$

Similarly, for the post-URET data the likelihood function is

$$L^u = \prod_n \binom{N_n^u}{R_n^u} \left(\rho_n^u\right)^{R_n^u} \left(1-\rho_n^u\right)^{N_n^u - R_n^u}$$

The likelihood functions for the pre-URET and post-URET time frames using the parameters estimated from the combined time frame are

$$L^{cp} = \prod_n \binom{N_n^p}{R_n^p} \left(\rho_n^c\right)^{R_n^p} \left(1-\rho_n^c\right)^{N_n^p - R_n^p}$$

$$L^{cu} = \prod_n \binom{N_n^u}{R_n^u} \left(\rho_n^c\right)^{R_n^u} \left(1-\rho_n^c\right)^{N_n^u - R_n^u}$$

The likelihood ratio comparing the hypothesis of separate pre- and post- parameters against the hypothesis of a single set of power curve parameters is

$$\Lambda = \frac{L^{cp} L^{cu}}{L^p L^u}$$