Logistics Management Institute

Predicting Wartime Demand for Aircraft Spares

AF501MR2

F. Michael Slay
Craig C. Sherbrooke

April 1997

Predicting Wartime Demand for Aircraft Spares

AF501MR2

F. Michael Slay
Craig C. Sherbrooke

Logistics Management Institute
2000 Corporate Ridge
McLean, Virginia 22102-7805

Predicting Wartime Demand for Aircraft Spares

Executive Summary

Recent revisions in the Air Force's *War and Mobilization Plan (WMP)* revealed a long-standing flaw in the way demands for spare parts have been forecast. The major regional conflict scenarios projected in the 1993 *WMP* called for longer fighter sorties, causing the war materiel requirement to rise dramatically. But after Operation Desert Shield/Desert Storm, where sorties were long yet parts demands low, the increase in the requirement seemed unwarranted and not credible. Recognizing this discrepancy, the Deputy Chief of Staff for Logistics imposed a moratorium on computing wartime spares from the new *WMP* until a better demand forecasting method could be found.

Our study found that the dramatic rise in the requirement was rooted in the assumption that parts fail strictly on a per-flying-hour basis; for example, that twice as many parts would break on a 2-hour sortie as on a 1-hour sortie. Analysis of maintenance data for more than a quarter of a million sorties showed that 2-hour fighter sorties cause about 10 percent more parts to break than do 1-hour sorties. We also developed a practical method to adjust for this phenomenon in wartime demand forecasting. The Air Force incorporated this method — called decelerated demand forecasting — in the 1995 fighter requirements computation and lifted the moratorium. The wartime spares requirement resulting from implementing this new forecasting method, along with the new war plans, changed only modestly, and unit capability assessments became more accurate. Decelerated demand forecasting prevented an overstatement of $1.1 billion in the gross war reserve requirement for fighters.

We found similar results for other aircraft, although they differed quantitatively. We recommend decelerated demand forecasting for most other weapon systems, with factors ranging from 20 to 60 percent. At the low end, bomber deceleration should be on the order of 2-hour sorties demanding 20 percent more parts than 1-hour sorties. At the high end, deceleration should not be used at all for some helicopters.

The Air Force implemented decelerated demand forecasting for bombers in 1996. Implementation is still in progress for the other aircraft. Research continues — particularly to try to derive separate deceleration factors for various subsystems. However, the deceleration factors already implemented are a significant step forward, greatly reducing the requirements while sharpening the assessments.

Contents

Contents (Continued)

Illustrations

Tables

Tables (Continued)

CHAPTER 1

Background

INTRODUCTION

U.S. fighter aircraft required surprisingly few spare parts in Operation Desert Shield/Desert Storm despite flying long hours. Although the sorties flown were much longer than their peacetime counterparts, demands per sortie remained about the same. This rekindled the long-held suspicion that parts fail on the basis of sorties flown — not hours flown — even though the standard Air Force planning systems had been forecasting demands on the basis of projected flying hours.

In 1993, the *War and Mobilization Plan, Volume 5 (WMP-5)* [1] was revised to take into account the longer sortie durations that are likely to result from responding to regional contingencies. Had the Air Force continued to use demands per flying hour as the basis for wartime demand prediction, the wartime spares requirement would have increased dramatically. Furthermore, unit capability assessments would have been too low to be credible. Since this situation would be unacceptable, the Air Force imposed a moratorium on implementing the 1993 *WMP-5* until a better demand forecasting method could be found.

Because wartime demand is predicted from peacetime data, and because predictions drive inventory investment and capability assessments, it is critical to know whether spares demand is driven by the number of sorties, by flying hours, or by some combination of them. The Logistics Management Institute was asked to study the matter and, indeed, our analysis confirmed that demand is much more closely related to sorties than it is to flying hours.

We have developed a new model for forecasting wartime demand that the Air Force is now incorporating in its computation of wartime spares, avoiding a $1.1 billion overstatement in the gross fighter requirement. The requirement for wartime spares resulting from implementing this new method of computation — along with the new war plans — changes only modestly, and unit capability assessments are more realistic. In the pages that follow, we review the previous research and examine some Operation Desert Shield/Desert Storm findings. Chapter 2 presents the new demand model for fighter aircraft, along with a risk analysis showing that even if actual demands differ significantly from those predicted by our model, the kit is robust enough to handle the difference without significant loss of capability. Chapter 3 details how the Air Force has implemented our model and explains how to use the model with current USAF Readiness Spares Package (RSP) computation and assessment models. In Chapter 4, we extend the model to non-fighter aircraft. Finally, Chapter 5 discusses other extensions and applications for wartime demand forecasting.

Previous Research

Quantitative Studies of Maintenance Removals Versus Sortie Duration

Comparing the influence of flying hours on demands with that of sorties is equivalent to studying the impact of sortie duration on demands per sortie, since average sortie duration is just the ratio of flying hours to sorties. A number of studies performed in the 1960s and 1970s used regression analysis to model maintenance removals per sortie, Y, as a linear function of the sortie duration in hours, x. Because spares demand data were unavailable, unscheduled maintenance removals were used instead. Shaw, who performed many of these analyses, chose to express the relationships as a constant term representing maintenance removals arising from a 1-hour mission plus a variable for the additional removals for durations beyond 1 hour as

$$Y = a + b(x - 1),$$

where a and b are regression coefficients that vary by aircraft. [2]

Unfortunately, it is hard to compare various types of aircraft using this formulation, because the values for a and b vary greatly. To enable comparison between aircraft with different failure rates, we have factored out the coefficient "a" to obtain a normalized slope. Thus, the regression model becomes

$$Y = a\left[1 + \frac{b}{a}(x-1)\right].$$

The normalized slope (b/a) is the fractional increase in maintenance removals per additional hour of sortie duration. Table 1-1 gives this normalized slope from various previous studies. For example, the bottom line in Table 1-1 shows that the normalized slope for the B-52D aircraft is 20 percent. In other words, for each hour of sortie duration in excess of 1 hour, the number of unscheduled maintenance removals increases by 20 percent of the baseline, 1-hour rate.

Table 1-1.
Regressions of Maintenance Removals on Sortie Duration From Previous Studies

Aircraft	Systems	Normalized slope (%)	No of sorties examined	Author	Date
C-5A	All	5	79,181	Shaw	1980
C-5A	Engine	8	79,181	Shaw	1980
C-141	All	28	835,000	Shaw	1980
C-141	All	22	73,000	Shaw, Howell	1980
C-130E	All	33	45,000	Shaw, Howell	1980
B-52D	All	20	10,809	Boeing	1970

What can we conclude from Table 1-1? A normalized slope of zero percent would indicate that maintenance removals are purely sortie driven. On the other hand, a 100 percent slope would indicate that removals are purely flying-hour driven. The slopes in Table 1-1 fall between these extremes, averaging about 19 percent. They are much closer to zero percent than to 100 percent, suggesting that demand is much more closely related to the number of sorties than it is to the number of flying hours.

However, the studies cited are of only limited relevance to current tactical fighter aircraft, for three major reasons:

- The average sortie durations in those studies are much longer than typical fighter sortie durations. While typical fighter sorties are 1 to 2 hours long, the transport aircraft studied had average sortie durations of about 4.5 hours; the B-52D, about 8 hours.
- Each aircraft was flying only one type of mission, while tactical missions tend to be different from one another.
- The data are over 15 years old.

Another problem with data such as these that have not been collected in a controlled experiment is that most of the sortie durations were near the average. For example, 80 percent of the transport sorties were between 3 and 6 hours. The only exception is the B-52D data, which were collected from three bases flying combat missions in the 1960s. Since average durations for the three bases were 4, 8, and 11.2 hours, the dispersion was particularly good for studying the impact of sortie duration on maintenance removals.

Other Studies of Sortie Duration Effect on Maintenance Removals

Other studies also found that maintenance actions were related more to the number of sorties than to the number of flying hours. Donaldson and Sweetland found that unscheduled flight-line man-hours were only slightly related to sortie length (B-52, F-100, F-102, F-4C, F-5A). [3] The C-130 showed a fairly constant man-hour/flying-hour relationship, but only for those missions requiring multiple sorties between maintenance stops. Boeing found that after 4 hours of a 12-hour B-52 mission, 50 percent of the failures and 47 percent of the abort-causing conditions had occurred; at 8 hours, the percentages were 80 percent and 93 percent, respectively. [4] Little concluded that sortie length and number of landings per sortie had no apparent effect on maintenance man-hours for the C-5. [5] Hunsaker et al. reported that F-4 sortie duration, which varied between 0.8 and 1.8 hours, had little effect on the equipment failure rate per sortie. [6] Casey observed that a C-5 sortie tends to result in a given number of maintenance write-ups regardless of the sortie's length. [7] Goldfarb and Smiley found that an increase in sortie duration was accompanied by a much smaller increase in the demand for a selected group of Air Force spares (the increase in demand was only about 13 percent of the increase in sortie duration). [8]

Howell found more flying-hour dependence than did other researchers. [9] For the B-52D, roughly half of total maintenance removals per sortie were found to be independent of sortie length, while the other half were related to sortie duration. For the C-141A, C-130E, and Boeing 727, most of the maintenance removals per sortie depended on sortie length only. This held true for each major aircraft system as well.

Effects of Mission Type

Considering mission type, Kern and Drnas found that one component during a 12-month period had field mean times between failures (MTBFs) ranging from 107 to 917 hours across six different aircraft types. [10] The MTBFs for avionics equipment on subsonic bombers and transports were 2 to 4 times higher than those for similar equipment installed on high-performance tactical or training aircraft. Hunsaker et al. noted that the type of mission flown by the F-4 has a direct impact on the number of maintenance write-ups within specific work unit codes (WUCs). [6]

Location Effects

Donaldson and Sweetland measured dramatic differences in the number of aircrew-reported malfunctions at two bases operating under very similar conditions. [3] Despite the reported differences, the difference in mission capability, as measured by on-aircraft electronic evaluators, was negligible. Interviews with base maintenance officers indicated that the reported differences were most likely due to differences in policies concerning malfunction reporting. In a base-to-base comparison of three equipments on one aircraft type operating from nine different bases, Kern and Drnas found MTBF variations of as much as 5 to 1 from base to base. [10] Between the two best and two worst bases, there was, on average, a 2-to-1 difference in reported MTBF. Tetmeyer noted that hydraulic leaks appear to be related to temperature variations, certain avionics failures to wet climates, and weather radar unscheduled maintenance to thunderstorm activity. The hydraulic power system on the B-52D showed a distinct sortie effect. [11]

Utilization Rate Effects

Greater utilization reduces the demand per sortie and per flying hour. For example, Boeing observed that B-2 maintenance man-hours per flying-hour decrease as utilization increases and sortie length is held constant. [4] Kern and Drnas noted that aircraft utilization rates (flying hours per month per aircraft) varied as much as 3 to 1 between different types of aircraft. [10] With military avionics typically operated for only a limited time each month, the non-operating period may be more significant than previously recognized. For one equipment, the data indicated that 40 percent of reported failures had occurred during non-operational periods. Berman et al. found that the probability of engine removals on the C-141 is a function of engine age (operating hours since overhaul) and utilization. [12] As utilization increased, principally because of longer sortie durations, demand per flying hour decreased.

Operation Desert Shield / Desert Storm

During the first 30 days of Desert Storm, one F-15C squadron we examined flew 236 percent of its *WMP-5* planned flying hours but only 85 percent of the sorties (Table 1-2). Observed demand rates were much lower than those expected from pure flying-hour-based demand forecasts. On an item-by-item basis, 214 of the items were better estimated by a pure sortie-based forecast, 58 by a pure flying-hour-based forecast. Similar results were obtained for the F-16C/D. These results are consistent with the literature review and suggest that longer sorties do not result in a proportional increase in demand.

Table 1-2.
Desert Storm Data

Data category	F-15C	F-16C
Desert Storm as a percent of planned activity		
30-day number of flying hours	236%	142%
30-day number of sorties	85%	91%
Accuracy of forecast item demands per flying hour		
Over-predicted by more than 25 percent	84%	81%
Within +/ – 25 percent	7%	10%
Under-predicted by more than 25 percent	9%	9%
Number of items predicted better by:		
Flying hours	58	23
Sorties	214	117

CHAPTER 2

Fighter Analysis

A regression model of parts failures as a function of sortie duration, like that presented in Chapter 1, should be reasonably suitable for fighter aircraft. However, many questions remain.

Which model is best? Are the data available in standard USAF systems sufficient for selecting and building such a model? What impact will each model have on the predicted spares requirements and the specific spares packages (kits)? How resistant are these kits to changes/errors in the model — is there any risk? To answer these questions, we begin with a closer look at the F-15 aircraft.

AN F-15C/D CASE STUDY

We analyzed two types of data from the Core Automated Maintenance System (CAMS): operational (tail number, sortie length, time, location, and mission type) and maintenance (tail number, start time, WUC, how malfunctioned, when discovered, and action taken).[1] We used unscheduled maintenance removals as a surrogate for demands on supply. Detailed results are presented for the F-15C/D at Langley Air Force Base (AFB) from January through late September 1993, followed by summaries of those for other tactical aircraft with *WMP-5* wartime tasking. We excluded the 12 tail numbers that deployed to Southwest Asia in the May – June 1993 period, since their utilization was quite different from that of the others.

[1]The maintenance history records of interest are those for on-aircraft removals, excluding cannibalizations and those items removed to facilitate access to other items. We excluded maintenance removals with "how malfunctioned" codes indicating "no defect." WUCs 01 – 09 were excluded because they are aircraft servicing codes. Tech order compliance items were excluded since they are not due to activities from the previous sortie. We excluded time change items as well, since these depend on number of hours or sorties, not on activities from the previous sortie. We linked the maintenance start time to the previous sortie, except that we excluded sortie aborts ("when discovered" = C) because, by their very nature, they generally reflect shortened sorties. (There were 16 air aborts in the Langley data.) "When discovered" = K, M, or Q records were excluded because they reflect hourly post-flight or special inspections, were few in number, and are not usually related to activities during the previous sortie.

Identifying Confounding Effects

Sortie Number

For the remaining 68 aircraft in our database, aircraft that flew only once on a particular day had the most demands per sortie. When an aircraft flew multiple sorties during a day, demands per sortie were much lower and tended to decline slightly with each succeeding sortie — except for the final sortie of the day, when the rate was almost as high as for an only sortie of the day (Table 2-1). While sortie number has never been identified as an important variable in any previous study, it has emerged as the most significant variable in our current analyses of tactical aircraft spares demand.

Table 2-1.
Impact of Sortie Number on Langley F-15C/D Demand

Sortie number of day	Number of sorties	Average length (hours)	Average demands/sortie
Only sortie of day	1,857	1.54	0.62
1 of multiple	2,804	1.35	0.17
2 of multiple	796	1.22	0.14
3 of multiple	418	1.15	0.12
4 of multiple	178	1.12	0.10
5 of multiple	45	1.00	0.11
6 of multiple	1	0.90	0.00
Final of multiple	2,820	1.33	0.52
Overall total/average	8,919	1.36	0.37

Mission Type

Differences in mission type also caused large differences in demand rates (Table 2-2). During aerial combat training sorties, the aircraft are heavily stressed and may pull as much as 8 Gs. In contrast, cross-country training sorties tend to be longer and less stressful, as are training deployment sorties. Because the shorter sorties tend to be more stressful missions, this table shows higher demand rates associated with *shorter* sorties. If we are not careful to account for the effect of mission type, any positive relationship between demands and sortie length could be overwhelmed.

Table 2-2.
Impact of Mission Type on Langley F-15C/D Demand

Mission type	Number of sorties	Average length (hours)	Average demands/sortie
Aerial combat training	7,247	1.32	0.39
Cross-country	498	1.47	0.15
Deployment	973	1.64	0.27
Other	201	1.23	0.56
Overall total/average	8,919	1.36	0.37

Location

In analyzing demand rates for aircraft based at multiple sites, we found pronounced location effects. For example, in the case of the A-10 (see analysis later in this chapter), there were six bases with an average of 0.29 demands per sortie but eight other bases with an average of only 0.12 demands per sortie. The difference appears to be attributable to the fact that the latter bases are primarily Air National Guard sites in urban areas without a target range. The high-demand-rate bases had a longer average sortie duration of 1.83 hours as opposed to 1.56 hours for the low-demand-rate bases. If we are not careful to account for location effects, demand sensitivity to sortie length could be greatly overstated.

Modeling Demands Versus Sortie Duration

We analyzed as a group the 7,108 aerial combat training missions that took off and landed at Langley. This group comprises most of the sorties in the data while eliminating the impacts of mission type and location. We noted that one of the highest demand rates was for the group of 177 sorties lasting less than 0.8 hours. We excluded these sorties because of their high prevalence of air aborts and functional check flights. Functional check flights are very short, very high-failure post-maintenance test flights. Since they were often not correctly coded as such in the mission-type data field, the only way to exclude them was to drop all short sorties. Theoretically, air aborts should be included, but they should be tabulated according to the planned sortie duration, not the actual, shortened duration resulting from the air abort. Unfortunately, information on planned sortie duration is not available.

Linear Regression

The resulting regression was for sortie durations between 0.8 and 7.3 hours and includes 7,020 sorties. The regression has a slope of about 18 percent and is statistically significant at the 95 percent level (i.e., there is less than a 5 percent likelihood that such a large slope was caused by chance instead of by a real relationship between sortie duration and demand).

As noted earlier in our discussion of Table 2-1, the impact of accounting for only/last sorties is very large. It is statistically more significant than sortie duration even after the short sorties are eliminated. We found that most of the difference in demand rates between earlier sorties of the day and the last sortie of the day results from deferred maintenance, not grounding breaks. Thus, the demand rate after earlier sorties is understated, and the demand rate after the last sortie is overstated. Since we are trying to relate the actual demand to each sortie, we define an early/last sortie variable that assumes a value of −1 on the earlier sorties, a value of 1 on the last of multiple sorties, and a value of 0 on the only sortie of the day. We estimate the magnitude of this deferred maintenance by a regression assuming that the amount of overstatement of the last of multiple sorties equals the combined amount of understatement of all the earlier sorties.

When sortie duration and this variable for earlier/last sortie are used as independent variables in a multiple regression, the slope for demand as a function of sortie duration drops to 13 percent (still statistically significant). The smaller slope results because the last sortie of the day, which has more demand, tends to be slightly longer, as can be seen in Table 2-1.

Curvilinear and Piecewise Linear Models

But is a linear regression model the right functional form for expressing the relationship of demand to sortie duration? In Figure 2-1, the solid horizontal line has a zero percent slope, where demand depends only on the number of sorties, and the solid sloping line has a 100 percent slope, where demand depends only on the number of flying hours. The "truth" is presumably somewhere between these extremes, where the theoretical effect of sortie duration is as shown by the dashed line. In Figure 2-1, this truth is a curved line. In actuality, the truth line could be straight, piecewise linear, or something else.

Tactical aircraft seldom fly combat or combat training missions exceeding 2.5 hours. For example, in the 7,020 Langley F-15C/D sorties, only 183 were 2.5 hours or longer and only 62 were 3.7 hours or longer. Thus, it is very hard to estimate whether the truth has any curvature, as suggested in Figure 2-1, and it is risky to extrapolate the data to longer missions without trying to validate those data with independent data sets.

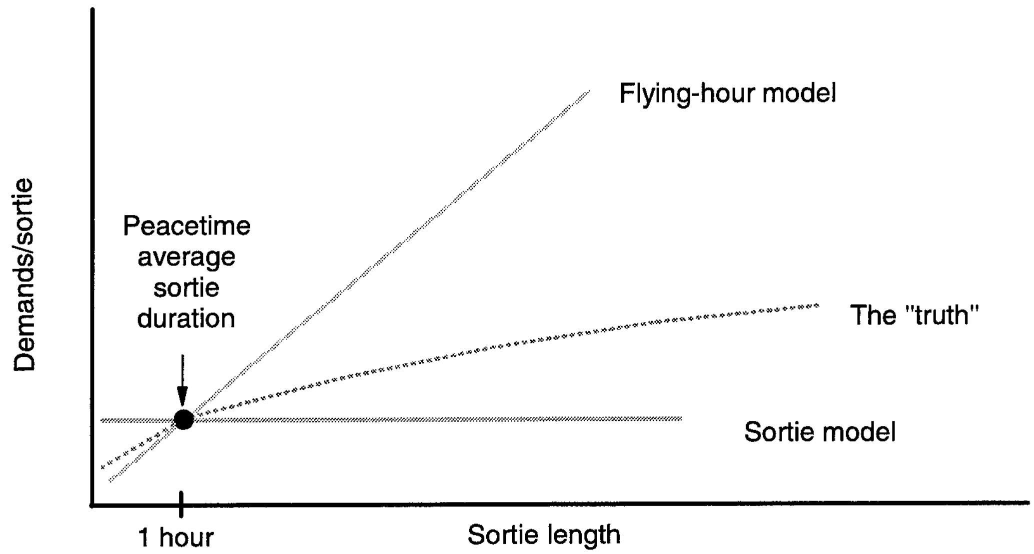

Figure 2-1.
Forecasting Demands

Figure 2-2 is a scatter plot of three normalized F-15C/D data sets:

- the 1993 Langley training missions (from the CAMS data discussed earlier)
- the 1993 Dhahran Southern Watch missions (from CAMS data for 1,224 Saudi Arabia sorties with an average duration of 3.29 hours)
- all 1994 F-15C/D sorties (data for 20,060 sorties with an average sortie duration of 1.57 hours, from the Reliability and Maintainability Information System [REMIS], which is a worldwide roll-up of base-specific CAMS data).

Each point on the scatter plot represents all the sorties at a particular sortie length. For example, the extreme right-hand plus sign represents seven 6.0-hour sorties. Unfortunately, there were few Langley training sorties above 3 hours and few Dhahran sorties above 4 hours.

Thus, many of the points on the right-hand side of the scatter plot represent only a few sorties. Even though we could not show the number of sorties for each point in the figure, all our regressions and mean-square error calculations weight those points by their number of sorties.

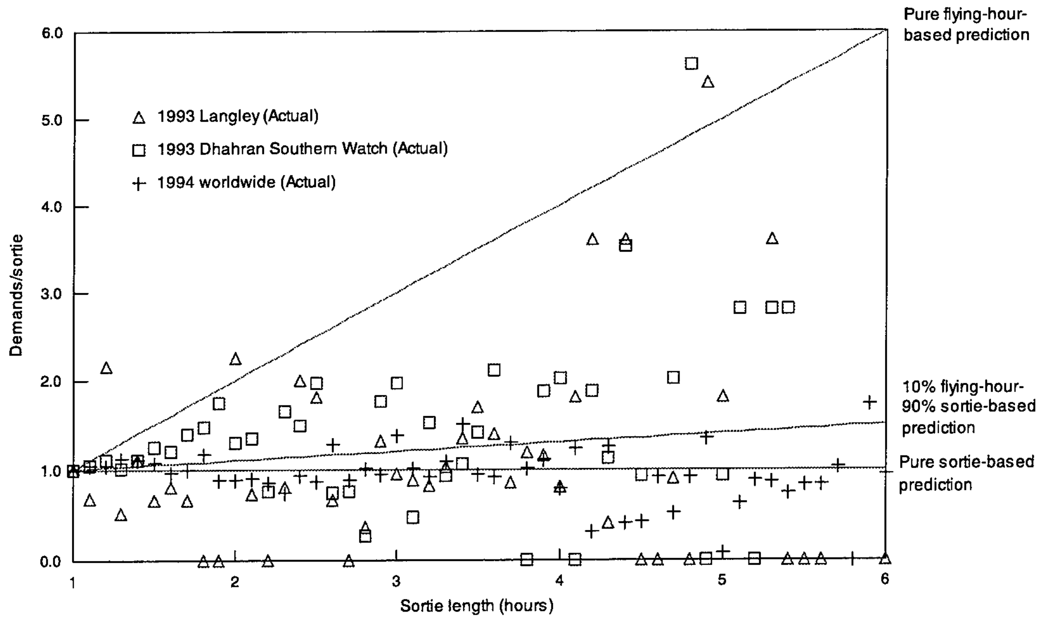

Figure 2-2.
Predicted Versus Actual Demands for the F-15C/D

Mean-Square Error Validating Data Set

Visually, the data in Figure 2-2 seem closer to the zero percent slope than to the 100 percent slope, but we analyzed the 1994 REMIS worldwide data to confirm this impression.

We calculated the mean-square error of the demands per sortie for the flying-hour model, the sortie model, and various intermediate models, including a piecewise linear model (Table 2-3).

Table 2-3 assesses five candidate models against the 1994 F-15C/D REMIS data, showing the mean-square error for each. In addition to the models based on pure flying hours, pure sorties, and 10 percent flying hours/90 percent sorties, we also considered a nonlinear model and a piecewise linear model. Clearly, the data do not support using a pure flying-hour-based forecast.

The parabolic model is a compromise between the pure sorties model and the pure flying hours model, and assumes that demand is proportional to the square root of sortie length. This nonlinear function produces a parabolic plot of demands versus sortie length similar to the truth line in Figure 2-1.

The 40 percent; 6 percent flying-hour model is a piecewise linear model that assumes that demands are 40 percent flying-hour/60 percent sortie dependent for sorties up to 1.5 hours and 6 percent flying-hour/94 percent sortie dependent above 1.5 hours. This piecewise linear, or "kneed," function was the best fit for the 1993 Langley CAMS data. Yet when this model was used to compute a RSP, the spares mix was nearly identical to one computed using the simple 10 percent flying-hour model.

Table 2-3.
Model Evaluation Versus 1994 F-15C/D REMIS Data

Model	Mean-square error
Pure flying hours	0.1001
Parabolic	0.0234
10 percent flying hours	0.0085
40 percent; 6 percent flying hours	0.0101
Pure sorties	0.0048

Analysis of Other Tactical Aircraft Data Sets

Other REMIS worldwide data sets showed the same characteristics. As with the Langley F-15C/D data, in each case we found a distinct only/last sortie of the day effect. Unlike the Langley case, this analysis was limited to training missions between 0.9 hours and 2.5 hours in duration. As explained earlier, very short sorties often represent functional check flights or air aborts and thus are unusable. Long flights tend to involve landing at remote locations, a situation that rarely permits maintenance and even more rarely results in a record of it. We had to drop longer sorties here, because REMIS data do not include landing location. Thus we could not filter out those long sorties with remote landings. (Had we not dropped these longer sorties, the regression slopes would have been even lower.) This is explained in detail in the companion special report by Craig Sherbrooke. [13]

Table 2-4 shows the normalized slope of demand as a function of sortie length obtained by regression for each data set. The first column of numbers is the regression slope, where the independent variable is sortie length and the dependent variable is demands. The second column is the slope taking into account the impact of earlier sortie versus last sortie of multiple sorties during the day. As with the Langley F-15C/D, accounting for last sortie effects usually lowers the slope, because the last sortie often has a high demand rate (as a result of deferred maintenance) and is longer.

Half of the entries in both columns of slopes are zeroes, because the slope from regression was negative, and negative slope is ruled out by the physics of the problem. (If an aircraft has a given number of demands after a certain number of hours in a sortie, the number of demands cannot be reduced by extending the length of the sortie). The number of sorties in the last column pertains to the number of observations used in each regression, but there were over 400,000 sorties in total from which these particular training sorties were extracted.

Table 2-4.
Summary of Slopes (Percent)

System	Slope before adjustment for sortie number (%)	Slope after adjustment for sortie number (%)	Number of sorties
A-10, OA-10	11*	6	33,081
F-15A	0	0	10,903
F-15C/D Langley	18*	13*	7,020
F-15C/D 1993	0	0	15,071
F-15C/D 1994	0	0	15,514
F-15C/D 1995	0	0	20,329
F-15E	0	0	11,623
F-16C/D 1994	21*	13*	61,499
F-16C 1995	11*	6*	60,166
F-16D 1995	7	0	12,349
F-111F	0	0	2,631
F-117A	0	0	8,794
Overall average/total	12	7	258,980

*Statistically significant at 95 percent confidence level.

The weighted average slopes, at the bottom of Table 2-4, are 12 percent before adjustment for early/last sortie and 7 percent after adjustment. While we believe that the slopes after adjustment for early/last sortie of the day are the more meaningful, those before adjustment are also shown, because they are comparable to the slopes found in other studies such as those in Table 1-1, where early/last sortie was not considered.

Aggregated Analysis

Aggregating all the sorties of a particular duration together distills the data set down to a manageable size, permitting new analyses. With only one data point for each sortie duration, a scatter plot is possible — Figure 2-2 was constructed using such aggregation.

Visual analysis of a scatter plot can be revealing. We have already noted that very short sorties are often functional check flights and should be excluded from the regressions. Similarly, we noted that very long flights are often remote landings with abnormally low recorded maintenance and should also be excluded.

A plot of the aggregated data makes all of this manifest. We can see the demands per sortie shooting up for very short sorties, and we can see the remarkably low demands for very long sorties. We can view the entire data set in such a way as to distinguish between useful input and misleading input, so that the data set can be cropped to retain only the "good" data. Figure 2-3 shows our largest data set — the 1994 F-16C/D.

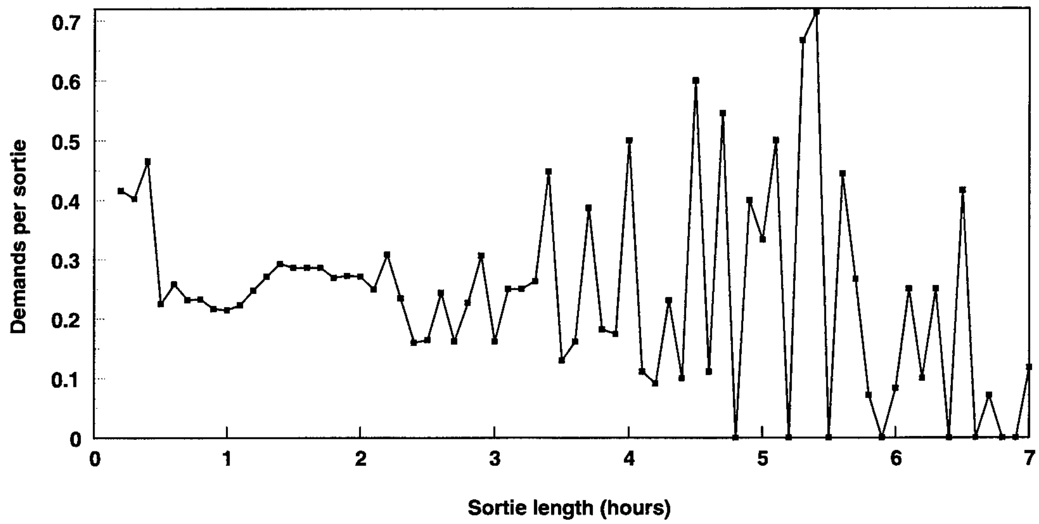

Figure 2-3.
1994 F-16 C/D Demands per Sortie

Notice the high demand rate for very short sorties and the low and volatile demand rate for very long ones.[2] On the basis of aggregated data such as those used to construct Figure 2-3, and conversations with USAF operational personnel, we selected 0.9 – 2.5 hours as the normal range of training sorties for fighters. This crop was used in the analysis (see Table 2-5).

Aggregating the data also permits types of analysis that would be impossible otherwise. For example, aggregating the data allows us to generate more regression statistics. More to the point, we have derived a method for analyzing the aggregated data that corrects for one major problem in the data — the number of failures per sortie is too large relative to the number of sorties with no failures at all.

The Probability of Failure Versus the Number of Failures

Aggregating the data highlighted a particularly troubling inconsistency. The number of demands per sortie is incompatible with the probability of no demands. For example, one aircraft might average one demand per sortie, even though 60 percent of the sorties had no demands at all. That means that the 40 percent of sorties that had demands averaged 2.5 demands each. This should

[2]The volatility for the longer durations is the result of small sample sizes, not something inherently wrong.

not happen. If component failures are independent, the distribution of failures on a single sortie should be approximately Poisson.[3] Barring battle damage, component failures are generally understood to be independent. Thus, we suspect that the high number of failures per sortie combined with the high probability of no failures has some human cause.

A number of possible explanations have been offered:

- Fault isolation is sometimes experimental.
- Repairs are not always successful.
- Minor repairs are often deferred until major maintenance is needed.

The tendency for multiple removals to have similar WUCs[4] supports the first two hypotheses, but deferred maintenance may be very significant. It is often inconvenient to service an aircraft and, when not an issue of safety, repairs are often deferred to a more convenient time or place. This practice is particularly common with cargo aircraft, where some landings are in remote locations with little maintenance capability. While less common with fighter aircraft, deferred maintenance is still a significant factor.

Regardless of the cause, this problem with the data causes difficulties with demand forecasting, because it concentrates the maintenance actions on fewer broken aircraft, decoupling the probability of maintenance from the number of maintenance actions when there is maintenance. In particular, the number of maintenance actions when there is maintenance becomes nearly constant, which can reduce the apparent sensitivity.

Aggregated Results

Table 2-4 repeated the results from the special report. [13] In Table 2-5, we examine the sensitivity to the choice of cropping and show the impact of the surrogate slope method.

---

[3] This has nothing to do with whether the demand process is Poisson or not; it assumes only that the demands come from a large number of independent components. Just as a binomial distribution approaches Poisson as the cardinal number becomes large, the sum of a number of component failures (independent Bernoulli trials) will approach Poisson as the number of components grows. As long as no individual component has a large percentage of the total demands, a Poisson distribution may be used for the number of demands in a single sortie.

[4] In the 1994 F-16C/D data, over 30 percent of the sorties with multiple demands had two or more demands for the same two-digit WUC.

Table 2-5.
Summary of Slopes (Percent) for Aggregated Analysis

System	Sortie lengths (hours)	Slope (%) (demands)	Slope (%) (surrogate demands)	Number of sorties total/cropped
A-10, OA-10**	0.9 – 2.5	11*	15*	45,428/33,081
F-15A	0.9 – 2.5	0	0	15,134/10,902
F-15C/D Langley	0.9 – 2.5	39*	44*	7,069/6,520
F-15C/D 1993	0.9 – 2.5	0	0	45,770/15,071
F-15C/D 1994	0.9 – 2.5	0	0	23,072/14,035
F-15C/D 1995	0.9 – 2.5	6	11	34,922/19,286
F-15E	0.9 – 2.5	0	4	20,195/11,623
F-16C/D 1994	0.9 – 2.5	23*	25*	99,988/67,533
F-16C 1995	0.9 – 2.5	1	8	92,811/60,166
F-16D 1995	0.9 – 2.5	9	15	17,169/12,349
F-111E	2.0 – 3.1	34*	35*	2,597/1,601
F-111F	1.2 – 3.0	1	0	4,805/3,260
F-117A	1.1 – 2.8	0	0	10,902/8,876
Overall average/total		10	13	419,682/264,303

* Significant at 95 percent confidence level.

** Multiple regression with dummy variable for Air National Guard/Reserve units.

The slopes here differ from those in Table 2-4 because of the cropping. Also note that using surrogate demands tends to raise slopes, in contrast to the effect of adjusting for first/last sortie, which lowered slopes in Table 2-4. Nevertheless, the overall slopes are similar, supporting our recommendation of 10 percent decelerated demand forecasting.

Note that the F-111 and the F-117, which had much longer peacetime sorties, were exceptions to the 0.9 – 2.5 hour cropping. For the F-111E, F-111F, and F-117A, only those sortie durations for more than 55 sorties were retained. For the F-111 total, which was a larger data set, we set a higher cutoff of 75 or more sorties. This is why the cropped numbers of sorties for the F-111 do not seem to add up.

Why the Surrogate Demands Have Higher Slopes

Consider this example. Suppose that maintenance is deferring minor repairs until a major repair is needed. For simplicity, assume that exactly two minor repairs will be needed every time the aircraft is worked on (in addition to the "real" repairs).

Suppose further that for 1-hour sorties, the aircraft has, on average, 0.5 removals per sortie. That translates to a particular distribution of the number of "real" removals: $p(0)=e^{-.05}$, $p(1)=0.5e^{-.05}$, $p(2)=0.5^2e^{-.05}/2$, $p(3)=0.5^3e^{-.05}/3!$, etc. Suppose further that demands are purely flying-hour dependent and thus 2-hour sorties generate, on average, 1.0 removals per sortie. That translates to $p(0) = e^{-1}$, $p(1) = e^{-1}$, $p(2) = e^{-1}/2$, $p(3) = e^{-1}/3!$, etc.

However, the deferred maintenance adds two removals whenever an aircraft is worked on. Thus, P(0) is unchanged, but p(1) is moved to p(3), p(2) to p(4), etc. Table 2-6 shows the resulting probability density functions (PDFs).

Table 2-6.
Impact of Deferred Maintenance on the PDFs of Removals

	Without deferred maintenance		With deferred maintenance	
	1-hour sorties	2-hour sorties	1-hour sorties	2-hour sorties
p(0)	0.6065	0.3679	0.6065	0.3679
p(1)	0.3033	0.3679	0	0
p(2)	0.0758	0.1839	0	0
p(3)	0.0126	0.0613	0.3033	0.3679
p(4)	0.0016	0.02	0.0758	0.1839
p(5)	0.0002	0.0031	0.0126	0.0613
p(6)	0	0.0005	0.0016	0.0153
p(7)	0	0.0001	0.0002	0.0031
p(8)	0	0	0	0.0005
p(9)	0	0	0	0.0001
p(10)	0	0	0	0
Mean	0.5	1.0	1.29	2.26

With deferred maintenance the mean for 2-hour sorties is no longer twice that for 1-hour sorties. The slope has dropped from 100 percent to 76 percent, because the number of removals per *broken* aircraft has lost some of its *relative* sensitivity. The probability of maintenance rises from 0.3935 for 1-hour sorties to 0.6321 for 2-hour sorties, both with and without deferred maintenance. However, without deferred maintenance, the expected number of removals per *broken* aircraft rises from 1.2706 for 1-hour sorties to 1.5820 for 2-hour sorties. With deferred maintenance, these numbers are 3.2706 and 3.5820, respectively. The addition of two repairs has compressed the relative growth in these numbers and thus reduced the slope. Since p(0) is unaffected, we can solve for the "true" demand rate as a function of p(0). Solving, we get $-ln[p(0)]$ as the surrogate demand rate. Using this surrogate demand rate we would get the true slope of 100 percent.

Note that this example is intended only to show how the clustering of removals can reduce the apparent slope of the regression. In the example, we give a specific cause for the clustering for clarity's sake and not because it is a reasonable physical mechanism. The real data simply *are* clustered and therefore need correction regardless of why they are clustered.

As shown earlier, deferred maintenance must be part of the problem, since the effect of the last sortie of the day has a significant impact on our regressions. Because last sorties tend to be longer than other sorties, their higher demand rates skew the regressions. Thus, the slopes after adjustment for sortie number in Table 2-4 are more valid.

But should we also be using $-ln[p(0)]$ or something like it? If the distribution of removals shows an inconsistency between p(0) and the mean, some correction is required. The number of demands when there is a failure is dubious, making the actual demand rate less trustworthy than the p(0).

Table 2-5 shows that the surrogate demand rate gives slightly higher relative slopes. Table 2-4 shows that directly accounting for deferred maintenance by adjusting for sortie number generally lowers the relative slopes. What if we did both?

Adjusting for Sortie Number in the Aggregated Analysis

Combining the surrogate demand rate with the adjustment for sortie number requires a large data set. The surrogate demand rate formula fails whenever p(0) is zero (i.e., every sortie of a particular duration has a removal), since the implied demand rate is infinite. Such a failure is likely to occur only when there are just a handful of sorties at a particular duration. Within the cropped region of our fighter data sets, there are enough sorties at each duration to preclude this problem from arising. However, in the multiple regression that corrects for sortie number, the data are separated into three groups: only sorties of the day, last sorties of multiple sorties of the day, and early sorties of multiple sorties. Generally, the longer sorties tended to be only sorties of the day, and the data for the other two groups were thin for sortie durations above 1.5 hours. Thus we could analyze only the largest data sets — the A-10/OA-10, the 1995 F-15C/D, and the 1994 F16C/D.

Table 2-7 shows that correcting for sortie number significantly reduces the slope, whether using regular or surrogate demands.

Table 2-7.
Summary of Slopes for Aggregated Analysis with Both Methods of Correcting for Deferred Maintenance

System	Sortie (%) (demands)	Slope (%) (surrogate demands)	Slope (%) (adjusted for first/last sortie)	Slope (%) (both adjusted and surrogate)
A-10/OA-10	9	14	4	9
F-15C/D 1995	6	11	0	0
F-16C/D 1994	23	25	15	18

Model Selection

While the linear regression model may not be perfect, it is a good choice for several reasons:

- It is the simplest model, with only two parameters (slope, intercept).
- We have no physical reasoning to guide us in selecting the sortie duration at which a knee between two linear segments might occur or the amount of curvature in a nonlinear segment.
- Our experiments with kneed regressions yielded only a negligible difference in the spares computed for the kit in comparison with those computed using a linear regression.
- A linear regression model is readily implementable with our existing data systems and inventory models.

However, it is not easy to decide what slope to use. The data are inconsistent, and we must crop out very short and very long sorties to get any meaningful results. When we adjust for sortie number, the slope falls. When we use $-ln[p(0)]$, the slope rises. We can tilt the results any way we choose. With an eye toward conservatism, we chose a slope of 10 percent.

We believe that a 10 percent slope provides a reasonable overall planning factor for the impact of sortie duration on demands by tactical aircraft. Using this factor implies that for each hour in excess of a 1-hour sortie, the expected demand will increase by 10 percent. That is, a 2-hour sortie has an expected demand of 110 percent of that of a 1-hour sortie, and a 3-hour sortie has 120 percent, etc., which is equivalent to saying that demands are 10 percent flying-hour-driven and 90 percent sortie-driven. This is known as demand deceleration. Figure 2-4 gives a visual representation of this deceleration.

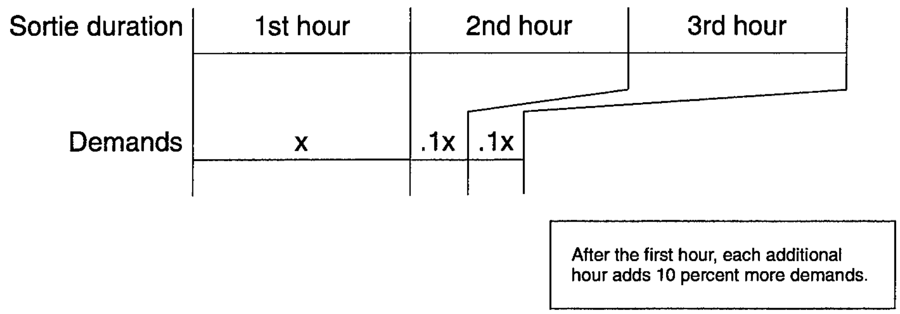

Figure 2-4.
Demand Deceleration

As a simple solution that can easily be implemented across all tactical aircraft, a 10 percent slope represents a reasonable overall model of the effect of sortie duration on demands. Of course, there are doubtless some components whose demands have a greater relationship to flying hours than a 10 percent slope would indicate and others whose slope is less. However, we could not identify them, and we encourage others to try. Whatever the case, the overall demand is clearly much more related to sorties than it is to flying hours.

There are two caveats to note concerning this conclusion. First, there were only a limited number of long-duration sorties, so that any extrapolation to very long durations has limited validity. Second, our data were maintenance data, and there is some difference between demands on supply and remove/replace maintenance actions.

Practical Considerations

Theoretically, the proposed model is sound, but practical questions remain. How much will using this method change the Mobility Readiness Spares Packages (MRSPs) or kits? How robust will the new MRSPs be? What are the risks if wartime demands turn out to be significantly different from those anticipated?

Cost Impact

This project began as a response to problems caused by using traditional, flying-hour-based demand forecasting methods with the new *WMP-5*. For almost every fighter in the Air Force, MRSP costs rose significantly. Table 2-8 shows the gross costs for some representative MRSPs as computed by the Aircraft Sustainability Model (ASM). (The ASM is the model used in the official production system to compute these costs.)

Table 2-8.
Cost Impact of WMP-5 Changes

Mission design series	Primary aircraft authorization	1986 *WMP-5* cost ($ millions)	1993 *WMP-5* cost ($ millions)
A-10A	24	4.0	5.5
F-4G	12	33.0	60.3
RF-4C	18	11.0	13.3
F-15C/D	18	14.7	41.2
F-16C/D	18	8.6	9.0
F-111F	18	81.0	109.3
F-117A	18	25.4	35.0

Not only were the costs for the new *WMP-5* unaffordable, but given Desert Shield/Desert Storm experience, they were not credible. In response to this problem, the Air Force declared a moratorium on using the 1993 *WMP-5* to compute MRSP requirements and assessments until a better demand forecasting method could be found. An exception was made for the F-15E, because its kit costs went down even using traditional demand forecasting methods.

While this research was being conducted, the Air Force was implementing the 2-level maintenance initiatives, which altered the kit computation significantly. Therefore our study focused on comparing the "old" MRSP cost (computed using the 1986 *WMP-5* and the traditional, pure flying-hour-based demand forecasting method) with the "new" MRSP cost (computed using the 10 percent deceleration method and the 1993 *WMP-5*, and incorporating the new 2-level maintenance initiatives). We used representative contingency kits and recomputed them both before and after all the changes. The resulting kit costs, and their associated weights and volumes (or cubes), are shown in Table 2-9.

The "after" figures include all the effects of using 10 percent decelerated demand forecasting, applying 2-level maintenance, and using the 1993 *WMP-5*. Note that the 1993 *WMP-5* is changed from the 1986 *WMP-5* in more ways than just the flying program. In particular, many Direct Support Objectives (DSOs) were revised. The DSO is the target number of aircraft that must be operational (that is, *not* not mission capable-supply [NMCS]). For example, in the 1986 *WMP-5*, 16.27 of the 18 F-16s had to be operational during the surge period of the war (on average). In the 1993 *WMP-5*, only 14.5 F-16s had to be operational during the surge. Thus, the ASM computes the F-16 MRSP with a target number of aircraft NMCS during the surge of 1.73 or 3.5, depending on which *WMP-5* is being used. Doing so has a significant impact on the MRSP cost and is largely responsible for the drop in the F-16 kit cost from $8.6 million to $3.6 million.

Table 2-9.
Cost, Weight, and Cube Impact of New Method and New WMP-5

Mission design series	Primary aircraft authori-zation	Before			After		
		Cost ($ millions)	Weight (lbs)	Cube (ft^3)	Cost ($ millions)	Weight (lbs)	Cube (ft^3)
A-10A	24	4.0	23,000	1,900	4.4	23,000	2,000
F-4G	12	33.0	28,400	2,700	32.2	26,500	2,400
RF-4C	18	11.0	19,600	2,000	6.7	13,900	1,400
F-15C/D	18	14.7	22,000	2,000	16.5	22,000	2,000
F-15E	18	23.6	19,000	1,500	19.7	15,000	1,200
F-16C/D	18	8.6	20,000	1,900	3.6	10,000	800
F-111F	18	81.0	68,000	7,300	83.8	69,000	7,300
F-117A	18	25.4	32,000	3,200	27.4	37,000	3,800

Risk Analysis

However comforting the statistical results and cost stability may be, the mix of parts in the kit is changed. These are new kits, built for a new tasking, using a new method. Any confidence in them has to be earned. They need to be tested for robustness, particularly with respect to flaws in the new forecasting technique. In other words, will the new kits be adequate if wartime demands turn out to be different from those predicted by our research? What will happen if the differences are substantial?

To answer these questions, we looked at kit robustness from two perspectives. First, we used a Monte-Carlo simulation model to measure how many sorties would be lost under various demand scenarios. Second, for each of those same scenarios, we used a specially modified version of the ASM to calculate how much materiel would have to be delivered to a squadron to restore it to its DSO (the minimum number of available aircraft that can support the flying program).

The simulation used 400 replications, because doing so yields three digits of precision, yet it is practical to run that many on a fast personal computer. With 400 replications, most of the results were accurate to within 0.03 percent. However, as a result of the inflated variance inherent in the backorder function, a few cases were accurate only to within 0.5 percent.

We then evaluated each kit's performance under a selection of rigorous scenarios. Since demands are modeled as 10 percent flying-hour-driven (90 percent sortie-driven), how will the kits perform if demands are actually 20 percent flying-hour-driven (80 percent sortie-driven)? That is, what is the impact if the model is off by a factor of two? How will the kits hold up if the model is totally wrong and demands turn out to be 100 percent flying-hour-driven, as the old model assumed? Table 2-10 summarizes the results.

Table 2-10.
Risk Assessment — Percentage of Planned Sorties That Would Actually Be Flown If Demands Are Driven 10 Percent, 20 Percent, or Purely by Flying Hours

		Percent of total sorties flown		
Mission design series	Kit cost ($ millions)	Demands purely flying-hour-driven	Demands 20 percent flying-hour-driven	Demands 10 percent flying-hour-driven
A-10A	4.4	99.3	99.9	100
F-4G	32.2	90.0	99.5	100
RF-4C	6.7	94.5	99.9	100
F-15C/D	16.5	80.8	99.2	100
F-15E	19.7	98.3	99.8	100
F-16C/D	3.6	98.9	99.9	100
F-111F	83.8	100.0	100.0	100
F-117A	27.4	96.6	99.9	100

Note: The F-111F has a high DSO. Thus, the kit has the reserve capacity to fly the full program even under extreme conditions.

Clearly, the risk of lost sorties is very low in the 20 percent flying-hour case. Even if our model is wrong by a factor of two, few sorties will be lost. The pure flying-hour analysis represents an extremely unlikely case. Yet, even then, all the aircraft, except the F-15C/D, lose fewer than 10 percent of their sorties.

Next, in each of these same cases, how much materiel would need to be delivered Desert Express[5] style to restore a squadron to its DSO? To answer this question, we modified the ASM to calculate only those backorders that caused the number of available aircraft to drop below the DSO. Using each backordered national stock number's (NSN's) weight and cube, the total 30-day Desert Express requirements were computed (Table 2-11).

[5]In Operation Desert Shield/Desert Storm, the Military Airlift Command (MAC), now the Air Mobility Command (AMC), addressed weaknesses in the priority system by setting up a special airlift route, "Desert Express," to move critical parts to the Gulf quickly. By the end of October 1990, a MAC cargo aircraft flew daily to the theater from Charleston AFB, South Carolina, with the most critical parts needed for wartime readiness.

Table 2-11.
Risk Assessment — Materiel Required to Restore Squadron to DSO If Demands Are Driven 20 Percent or Purely by Flying Hours

Mission design series	Kit cost ($ millions)	30-day Desert Express requirement: Demands purely flying-hour-driven lbs (ft^3)	30-day Desert Express requirement: Demands 20 percent flying-hour-driven lbs (ft^3)
A-10A	4.4	133 (18)	17 (3)
F-4G	32.2	803 (64)	28 (2)
RF-4C	6.7	1,248 (124)	30 (3)
F-15C/D	16.5	5,540 (421)	171 (12)
F-15E	19.7	613 (63)	53 (6)
F-16C/D	3.6	157 (17)	5 (1)
F-111F	83.8	447 (61)	63 (8)
F-117A	27.4	1,380 (201)	104 (15)

Note that these figures represent the total for 30 days — not the daily requirement. Thus, even if the 10 percent model is wrong by a factor of two, it would take very little in the way of express deliveries to the squadron to make up for the shortfall.

These results give us confidence that the 10 percent deceleration method will not put units at risk. The next task is to determine how the method can be implemented in the standard Air Force requirements and assessments systems.

CHAPTER 3

Implementation

OVERVIEW

We favor the 10 percent deceleration method partly because of its simplicity and ease of implementation. With this method, no reprogramming of the standard Air Force systems is needed. The sortie duration is simply decelerated and the inputs to the models (i.e., flying hours) are adjusted accordingly.

This is simple enough to use if one knows how to adjust certain parameters for the requirements computation — specifically the sortie durations and the total flying hours. Unfortunately, the mathematics of computing the decelerated sortie durations and flying hours involves normalizing the decelerated sortie length relative to the decelerated peacetime average sortie duration. The requisite data would not be available to all the users, even if they were willing to perform all the algebraic calculations. It would be better to intercept the data at the source and add decelerated figures to the existing numbers. The users would then use the decelerated numbers where appropriate, simplifying the process and assuring uniformity.

The source in this case is the RSP Authorization Document (the *Blue Book*). All USAF requirements and assessments are computed on the basis of the figures in that document. By adding a few columns to the appropriate tables in the *Blue Book* and instructing the users to use the figures in those columns for all requirements and assessments, we implemented demand forecasting. The revised *Blue Book* has been published. Of course, it would be unwise to change the way kits are assessed until the new kits are fielded, so users will be using the old *Blue Book* (based on the 1986 *WMP-5*) to assess kits until the new kits are computed and fielded.

DECELERATION ALGEBRA

Correcting for the decelerated impact of additional hours in a sortie involves adjusting the sortie length. Thus the models — which compute demands on a strict per-flying-hour basis — are "tricked" into correctly forecasting demands for long sorties. For models with sortie rate and sortie length as separate inputs, the sortie length is simply adjusted. For models that use flying hours, the adjusted sortie length is multiplied by the sortie rate multiplied by the number of aircraft to compute the adjusted flying hours.

Suppose an aircraft with a peacetime average sortie duration (ASD) of 2 hours has 3-hour wartime sorties. To use the pure flying-hour method, input 3.0 as the sortie length for the wartime computation. To switch to a pure sortie model, simply input the peacetime ASD (2.0) in place of the wartime figure.

But what is the correct input for 10 percent deceleration? 2.1? No, the matter is a bit more complicated than that, although 2.1 is close. The 3-hour wartime sortie length decelerates to 1.2, but we must also take into account the fact that the peacetime ASD decelerates to 1.1. Thus all the demand rates — computed demands per flying hour based on a peacetime ASD of 2 — are off by a factor of 2/1.1 for use with decelerated wartime sorties.

Thus, the wartime sortie duration, adjusted for deceleration, is

$$1.2 \times \frac{2}{1.1} = 2.182.$$

This formula was used to generate the decelerated sortie durations and flying hours in the *Blue Book*. It is also automated as the deceleration option in ASM 4.0 and the Initial Spares Aircraft Availability Calculation (ISAAC). However, users of deceleration need not know anything about the formula.

Using Deceleration

The foreword to the *Blue Book* calls for using the decelerated table of flying hours, when available, for all spares requirements. The decelerated figures are in specially marked columns. Decelerated hours are also appropriate for assessing any kit that was built using deceleration. Any kit based on the old demand forecasting method should not be assessed using deceleration.

Requirements determinations and assessments computed using deceleration are performed in the same way they were before, and the results are interpreted in the same way as previously. The process is changed at the beginning only.

Should a researcher want to experiment with deceleration, the algebra described above would need to be applied to the input data. While a pocket calculator is sufficient for the task, we have produced a simple menu-based program to simplify the calculation. This program is small enough to be sent by electronic mail and can be obtained by contacting the authors of this report at the Logistics Management Institute.

CHAPTER 4

Extensions to Other Aircraft

When the use of deceleration was approved for calculating requirements for fighters, the moratorium on computing RSPs using the 1993 *WMP-5* was lifted. The *Blue Book* now uses 1993 *WMP-5* figures for all aircraft except the C-130. Other aircraft were not as problematic as fighters, for two reasons:

- The changes in the *WMP-5* principally affected fighters. Other aircraft saw little budget impact from the new *WMP-5*.
- Only fighters have wartime ASDs that differ greatly from the corresponding peacetime ASDs. Thus, for other aircraft, the impact of using decelerated demand forecasting is less.

Nevertheless, if decelerated demand forecasting is correct, then it should be used on all applicable weapon systems. Having already solved all the implementation problems, it remains only to test the method on other aircraft and choose the appropriate deceleration factors.

Note: the following three sections are excerpted from the companion special report by Craig Sherbrooke. [13] For more detail, see that report.

BOMBERS

The most striking aspect of the bomber analysis is that implementing the 1993 *WMP-5* causes kit costs to drop, even without deceleration. Given the long peacetime ASD for bombers, decelerated demand forecasting may not reduce the kit costs much, but that reduction will be on top of the reductions caused by the new *WMP*. Thus, kit cost reductions resulting from deceleration may not translate into procurement savings. For many parts, the spares are already in the RSPs, and any reductions will only generate excess.

The bomber data in Table 4-1 comprise fewer than 11,000 sorties. In both data sets, the slope is larger after adjustment for sortie length, in contrast to our results on fighters. While the slope of 20 percent for the B-52H is quite large, the sample of approximately 3,000 sorties is very small.

Table 4-1.
Summary of Slopes (Percent) for Bombers

System	Slope (%) before adjustment for sortie number	Slope (%) after adjustment for sortie number	Number of sorties
B-1B	2*	8*	7,754
B-52H	14*	21*	3,197
Overall average/total	6	12	10,951

*Statistically significant at 95 percent confidence level.

RECONNAISSANCE, TANKERS, AND AIRLIFTERS

Table 4-2 displays the results for the E-3 reconnaissance aircraft, the KC-135 tanker, and several airlifters.

Table 4-2.
Summary of Slopes (Percent) for Reconnaissance, Tankers, and Airlifters

System	Slope (%) before adjustment for sortie number	Slope (%) after adjustment for sortie number	Number of sorties
E-3	11*	11*	2,438
C-130H	15*	6*	27,919
EC-130E	12	6	1,927
EC-130H	3	6	1,299
HC-130P	35*	20*	5,473
MC-130E	42*	26	1,716
MC-130H	37	21	4,944
KC-135	10*	12*	17,504
C-141	5*	1	12,501
C-5	5	5	2,825

*Statistically significant at 95 percent confidence level.

Helicopters

Table 4-3 displays the results for helicopters.

Table 4-3.
Summary of Slopes (Percent) for Helicopters

System	Slope (%) before adjustment for sortie number	Slope (%) after adjustment for sortie number	Number of sorties
UH-1N	39*	8	9,779
HH-60G	27*	15*	8,426
MH-53J	44*	23*	1,392
MH-60G	68*	15*	929

*Statistically significant at 95 percent confidence level.

Aggregated Analysis

Table 4-4 shows the slopes for both the demands per sortie and the surrogate demands per sortie. The slopes are for the cropped data. In order to avoid bias, we cropped all systems to the same range. The range of 1 to 6 hours was chosen on the basis of the number of sorties at those lengths, particularly for the cargo aircraft. The bombers had to be exceptions to this rule, because they had very few short sorties. For the B-1B, we retained only those sortie durations associated with 65 or more sorties. For the B-52H — a very sparse data set — we retained only those durations that had 20 or more sorties. Even with this low cutoff, there were some points in the middle, which we retained, that did not quite meet this criterion.

Model Selection

For the bombers, we recommended that a 20 percent deceleration factor be used. The Air Force has implemented this method. This factor should also be appropriate for the B-2 when its kits are fielded.

For the airlifters, the E-3, and the KC-135, a deceleration factor of 25 percent was chosen. This choice has encountered some opposition from Air Mobility Command (AMC) staff because of unrelated problems with their kit computations. The method for computing the C-5, the C-17, and the C-141 kits is currently under review. Until that review is completed, no changes to AMC kit computations are expected. However, the C-130, the E-3, and the KC-135 should not have to wait for AMC to solve problems that do not affect them.

Table 4-4
Summary of Slopes (Percent) for Aggregated Analysis

System	Sortie lengths (hours)	Slope (%) (demands)	Slope (%) (surrogate demands)	Number of sorties total/cropped
B-1B	2.5 – 6.7	7*	6*	7,303/4,666
B-52H	3.0 – 9.5	21*	22*	3,768/2,664
E-3	1.0 – 5.9	20*	15*	4,197/2,456
C-130	1.0 – 5.9	17*	11*	29,782/21,079
C-5	1.0 – 5.9	10*	11*	13,249/6,733
C-17	1.0 – 5.9	33*	23*	2,140/1,278
C-141	1.0 – 5.9	7*	7*	36,429/22,004
KC-135	1.0 – 5.9	11*	12*	24,200/16,440
AMC average/total	1.0 – 5.9	11*	12*	76,018/46,455
AC-130	1.0 – 5.9	31*	40*	4,197/2,456
EC-130	1.0 – 5.9	44*	35*	5,395/3,044
HC-130	1.0 – 5.9	46*	46*	12,053/7,524
MC-130	1.0 – 5.9	43*	51*	9,371/6,586
Special mission average/total	1.0 – 5.9	42*	45*	31,016/19,610
H-53J	1.0 – 5.9	56*	85*	5,812/1,196
MH-60G	1.0 – 3.9	53*	129*	1,702/740
HH-60G	1.0 – 3.9	25*	32*	13,472/7,726

* Statistically significant at 95 percent confidence level. Note that not only were all these regressions significant at the 95 percent confidence level, all but one were significant at the 99 percent confidence level. The exception, the C-141 slope (regular demands), was significant at the 98.7 percent confidence level.

For special mission C-130 aircraft, 60 percent deceleration is appropriate. It provides a considerable conservative margin, which is appropriate given the unpredictable nature of special missions.

For most of the helicopters, the evidence is insufficient to justify using deceleration. However, the HH-60G exhibits decelerated demands. We recommend 40 percent deceleration for the HH-60G.

Practical Considerations

We face the same questions here as with the fighters. How much will using this method change the MRSPs? How robust will the new MRSPs be? What are the risks if wartime demands turn out to be significantly different from those anticipated?

Cost Impact

Table 4-5 shows the impact of decelerated demand forecasting on some representative kits. Because the moratorium on using the 1993 *WMP-5* was lifted after we finished the fighter analysis, all these figures are for the new *WMP*.

Table 4-5.
Cost, Weight, and Cube Impact of Deceleration

Mission design series	Primary aircraft authorization	Undecelerated			Decelerated		
		Cost ($ millions)	Weight (lbs)	Cube (ft^3)	Cost ($ millions)	Weight (lbs)	Cube (ft^3)
B-1B	6	65.1	25,000	2,500	57.5	22,000	2,200
B-52H	6	97.1	144,000	22,200	88.8	132,000	20,000
C-130H	12	9.6	33,300	3,400	8.7	32,000	3,300
E-3B	3	46.2	38,800	3,200	44.5	38,300	3,150
KC-135R	12	9.7	29,200	3,160	9.3	28,600	3,080
C-5A	54	150.1	259,000	26,000	125.6	234,000	23,300
C-17A	12	196.9	49,400	7,230	184.1	46,500	6,830
C-141B	107	104.3	176,000	14,200	82.3	148,000	11,300
AC-130H	4	45.1	49,800	4,750	44.3	48,800	4,650
EC-130H	5	8.8	26,000	2,610	8.5	25,700	2,580
HC-130N	4	8.1	27,300	3,220	7.8	26,900	3,170
MC-130H	4	32.9	33,300	3,310	31.1	32,100	3,180
HH-60G	4	4.4	13,100	1,410	4.0	12,500	1,330

Risk Analysis

Our risk analysis is identical to that done for the fighters except that there is no lost-sorties analysis. Fighter RSPs are computed using a DSO that is based on the maximum turn rate for the aircraft. Thus, for fighters, when the number of mission capable aircraft falls below the DSO, the number of sorties that can be flown at the maximum turn rate will be below that called for in the plans, and we can compute a meaningful lost-sorties statistic. But, for other aircraft, the DSO is set by policy and thus is incompatible with computing sorties from the maximum turn rate. (This fact caused the lost sorties results for the F-111 and F-117 in Chapter 2 to be null; they, by policy, use bomber DSOs.) Thus, we cannot compute meaningful lost-sorties figures here.

The risk analysis is grouped according to the deceleration factor. First we show the bombers — which had 20 percent deceleration. Then we present the tankers, the airlifters, and the E-3 — which had 25 percent deceleration. Next we show the special mission C-130s — which had 60 percent deceleration. Lastly we present the HH-60G, which had 40 percent deceleration.

Bombers

Since bomber demands are modeled as 20 percent flying-hour-driven (80 percent sortie-driven), how will the kits perform if demands are actually 50 percent flying-hour-driven (50 percent sortie-driven)? Also, how will the kits hold up if demands are 100 percent flying-hour-driven?

As we did with fighters, we wanted to know how much materiel would need to be delivered Desert Express style to restore a squadron to its DSO. We used the same modified ASM cited in Chapter 2 to calculate the total Desert Express requirements (Table 4-6). As with the fighters, the B-52H Desert Express requirement is the total for 30 days. However, the B-1B support period is 14 days and the Desert Express requirement is for that period.

Table 4-6.
Risk Assessment — Materiel Required to Restore Bomber Squadrons to DSO If Demands Are Driven 50 Percent or Purely by Flying Hours

Mission design series	Kit cost ($ millions)	Desert Express requirement	
		Demands purely flying-hour-driven, lbs (ft^3)	Demands 50% flying-hour-driven, lbs (ft^3)
B-1B	57.5	59 (5)	30 (3)
B-52H	88.8	82 (15)	66 (11)

Note that it would take very little in the way of express deliveries to the squadron to make up for the shortfall.

Reconnaissance, Tankers, and Airlifters

Demands for reconnaissance aircraft, tankers, and airlifters are modeled as 25 percent flying-hour-driven (75 percent sortie-driven). How will the kits perform if demands are actually 50 percent flying-hour-driven (50 percent sortie-driven)? Purely flying-hour-driven? Specifically, how much materiel would need to be delivered Desert Express style to restore a squadron to its DSO? Once again, we used the modified ASM to calculate the total Desert Express requirement (Table 4-7). For the E-3, the C-130, and the KC-135, the Desert Express requirement is the total for 30 days. For the C-5, the C-17, and the C-141, the support period is 45 days and the Desert Express requirement is for that period.

Table 4-7.
Risk Assessment — Materiel Required to Restore Reconnaissance, Tanker, and Airlifter Squadrons to DSO If Demands Are Driven 50 Percent or Purely by Flying Hours

Mission design series	Kit cost ($ millions)	Desert Express requirement	
		Demands purely flying-hour-driven, lbs (ft^3)	Demands 50% flying-hour-driven, lbs (ft^3)
E-3B	44.5	0 (0)	0 (0)
C-130H	8.7	23 (2)	8 (1)
KC-135R	9.3	7 (1)	5 (1)
C-5A	125.6	1,300 (120)	300 (30)
C-17A	184.1	10 (2)	10 (1)
C-141B	82.3	1,800 (180)	300 (30)

The expedited weight and cube are small relative to the total kit size, although there are some substantial numbers for the C-5 and C-141. However, even these figures pale in comparison to the total weight and cube of the kits in Table 4-5.

Special Mission C-130s

Since special mission C-130 demands are modeled as 60 percent flying-hour-driven (40 percent sortie-driven), our only risk analysis here is for demands 100 percent flying-hour-driven. Table 4-8 summarizes the results. Because the support period is 30 days for all special mission aircraft, these figures are for that period.

Table 4-8.
Risk Assessment — Materiel Required to Restore Special Mission C-130 Squadrons to DSO If Demands Are Driven Purely by Flying Hours

Mission design series	Kit cost ($ millions)	30-day Desert Express requirement
		Demands purely flying-hour-driven, lbs (ft^3)
AC-130H	44.3	14 (1)
EC-130H	8.5	17 (2)
HC-130N	7.8	0.4 (0.1)
MC-130H	31.1	17 (2)

Once again, the express delivery requirement is small.

HELICOPTERS

Only one type of helicopter — the HH-60G — has decelerated demands. Since its demands are modeled as 40 percent flying-hour-driven (60 percent sortie-driven), our only risk analysis here is for demands 100 percent flying-hour-driven. Because the HH-60G support period is 30 days, the Desert Express requirement is for that period. Figure 4-9 shows the result.

Table 4-9.
Risk Assessment — Materiel Required to Restore HH-60G Squadrons to DSO If Demands Are Driven Purely by Flying Hours

Mission design series	Kit cost ($ millions)	30-day Desert Express requirement
		Demands purely flying-hour-driven, lbs (ft^3)
HH-60G	3.99	1.2 (0.2)

Note that it would take very little in the way of express deliveries to the squadron to make up for the shortfall.

CHAPTER 5

Future Issues

NOP

For many components, traditional flying-hour-based demand forecasting is so obviously inappropriate that the Air Force currently ignores it, setting the requirements levels manually. These NSNs, called non-optimized (NOPed), are part of every RSP kit and constitute nearly half of the gross RSP requirement. With the advent of decelerated demand forecasting, these NSNs should be reconsidered for regular computed requirements.

Not all NSNs should be computed. For example, halon bottles, which have virtually no failures, yet are needed for swapping when quick-turning the aircraft, should always be NOPed. But most parts exhibit normal failure patterns — random failures driven by use, whether sorties, flying hours, rounds fired, or some other factor.

Many of these parts can and should be computed. Not only will doing so improve requirements, but assessments would be sharpened as well. Currently, NOPed items are included in the assessments in an awkward and inaccurate way. A "backed out" demand rate (that which is implied by the requirements level) is computed for each NOPed item so that it can be included in the assessments. This method was developed to solve problems caused by totally excluding the NOPed NSNs from the assessments. While a step in the right direction, this method makes no claims to accuracy.

Some true NOPed items (such as halon bottles) *should* be excluded from assessments. Others that have failures and whose failures can ground aircraft are prime candidates for computed requirements.

SYSTEM-SPECIFIC DECELERATION

Just as demand data are collected on an NSN-specific basis, it would be desirable to collect NSN-specific data on demands as a function of both sorties and flying hours. Unfortunately, doing so is not as easy as it may seem. If one simply keeps track of an NSN's total demands, total sorties, and total flying hours, one cannot do the regressions on the sensitivity of demand to sortie duration. Three additional quantities are required: the sum of the squares of the sortie lengths, the sum of the squares of the number of demands from each sortie, and the sum of the product of sortie duration times the number of demands from each sortie. But, adding more data fields to the already crowded Air Force data collection systems would be difficult. Furthermore, even if all the desired data were to be collected, the NSN-specific slopes might be so erratic that some sort of smoothing or grouping technique would be required. At best, the needed procedures would take years to implement.

It may be necessary to aggregate NSNs by work unit code (WUC), by Federal Supply Group (FSG), or in accordance with some other classification scheme. We did aggregate F-16C data by WUC for 1994 and 1995. The F-16C was selected because we had the most data for it, about 125,000 sorties. The six WUCs shown in Table 5-1 were those with the largest number of demands.

Table 5-1.
F-16C by WUC Groups — Slopes After Adjustment for Sortie Number

WUC	System	Slope (%) 1994 data	Slope (%)1995 data
13	Landing gear	0	0
14	Flight control	5	15
42	Electrical system	7	0
74	Fire control	21*	24*
75	Weapons	4	0
76	Electrical warfare systems	31*	33*

* Statistically significant at the 95 percent level.

It is comforting that the landing gear, which had the largest demand rate, showed no relationship with sortie length in either of the two time periods. The only systems that did show a significant positive relationship were fire control and electronic warfare systems, and the slopes were quite consistent for the two time periods on each of these two systems. However, we do not know whether the larger slopes for these two WUCs are really meaningful. In the special report [13] some other aircraft were examined. The 1995 F-15C/D WUCs 74 and 76 again stood out as significant with similar slopes.

The new demand forecasting technique is but one of many initiatives in Air Force wartime supply policy. Changes in the DSO and in how long the kit must operate without resupply (the support period) are currently being considered. However, the 10 percent deceleration technique is appropriate for all such scenarios. While no demand forecasting method will ever be perfect — that is why the Air Force buys safety stock — the new method represents a vast improvement over the old one. It is simple, conservative, and easy to implement.

Glossary

AFB	=	Air Force Base
AMC	=	Air Mobility Command
ASD	=	average sortie duration
ASM	=	Aircraft Sustainability Model
CAMS	=	Core Automated Maintenance System
DSO	=	Direct Support Objective
EWS	=	electronic warfare systems
MAC	=	Military Airlift Command
MRSP	=	Mobility Readiness Spares Package
MTBF	=	mean time between failures
NMCS	=	not mission capable-supply
NOPed	=	non-optimized
NSN	=	national stock number
PDF	=	probability density function
REMIS	=	Reliability and Maintainability Information System
RSP	=	Readiness Spares Package
WMP	=	*War and Mobilization Plan*
WMP-5	=	*War and Mobilization Plan, Volume 5*
WUC	=	work unit code

References

[1] Department of the Air Force, *USAF War and Mobilization Plan, Volume 5 (WMP-5): Basic Planning Factors and Data,* Washington, DC: HQ USAF/XOXOW, March 1993.

[2] Colonel C. C. Shaw, *Saber Sustainer Briefing,* AF/SAGM, March 1981.

[3] T. S. Donaldson and A.F. Sweetland, *The Relationship of Flight-Line Maintenance Man-Hours to Aircraft Flying-Hours,* Santa Monica, CA: The RAND Corporation, August 1968.

[4] Boeing Military Airplane Systems Division, *B-52D Operations in Southeast Asia vs. CONUS,* D162-10015-1, Seattle, WA: The Boeing Company, September 1970.

[5] Captain Terry R. Little, *A Maintenance Manhour Sensitivity Model for Cargo Aircraft,* AFIT Report No. GSA/SM/72-9, March 1972.

[6] H. D. Hunsaker, F. Conway, and E.R. Doherty, *Logistics Impacts of Sortie Oriented Air Crew Training,* Technical Report No. 76-7, AFLS, April 1977.

[7] D. L. Casey, *Logistics Impacts of Longer C-5 Missions,* AFLC Technical Report No. 76-12, May 1977.

[8] Oscar A. Goldfarb and William A. Smiley, *An Evaluation of Program Activity as a Demand Prediction Tool,* Wright-Patterson Air Force Base, OH: Air Force Institute of Technology, 1967.

[9] Captain Lawrence D. Howell, *A Method for Adjusting Maintenance Forecasts to Account for Planned Aircraft Sortie Lengths,* ASD-TR-78-26, August 1978.

[10] G. A. Kern and T. M. Drnas, *Operational Influences on Reliability,* Report No. RADC-TR-76-366, Hughes Aircraft Company, December 1976.

[11] Colonel Donald C. Tetmeyer, *Forecasting Wartime Resource Requirements,* USAF Logistics Capability Assessment Symposium, 1982, IV122- IV136.

[12] M. B. Berman et al., *Predicting Aircraft Component Removal Rates with Varying Flying Programs: TF33-7 Engine Case Study,* R-2303-AF, Santa Monica, CA: The RAND Corporation, 1984.

[13] Craig C. Sherbrooke, *Using Sorties vs. Flying Hours to Predict Aircraft Spares Demand,* Report AF501LN1, Logistics Management Institute, April 1997.

REPORT DOCUMENTATION PAGE

Form Approved
OPM No.0704-0188

1. AGENCY USE ONLY *(Leave Blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	Apr 97	Final

4. TITLE AND SUBTITLE

Predicting Wartime Demand for Aircraft Spares

5. FUNDING NUMBERS

C DASW01-95-C-0019

PE 0902198D

6. AUTHOR(S)

F. Michael Slay
Craig C. Sherbrooke

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Logistics Management Institute
2000 Corporate Ridge
McLean, VA 22102-7805

8. PERFORMING ORGANIZATION REPORT NUMBER

LMI– AF501MR2

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

HQ USAF/LG
DCS/Logistics
Room 4E260, The Pentagon
Washington, DC 20330

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

A: Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

U. S. fighter aircraft demanded surprisingly few spare parts in Desert Storm despite flying long hours. Although the sorties flown were much longer than their peacetime counterparts, demands per sortie remained about the same. This simple observation raised suspicions that parts fail on the basis of sorties flown, not hours flown, even though Air Force planning systems forecast demands on the basis of projected flying hours. This method of forecasting demands proved inadequate in 1993 when the war plans were modified to include the longer sorties that typify regional conflict scenarios. We found that demands are, for most aircraft, much more closely related to sorties flown than to flying hours. We developed a demand forecasting method that incorporates our results into the wartime spares requirements computation. Called Decelerated Demand Forecasting, this method has been implemented by the Air Force for fighters and bombers, avoiding a $1.1 billion overstatement in the gross requirement.

Core Automated Maintenance System (CAMS), demand prediction, flying hours, requirements, REMIS, RSP kits, sorties, spares, wartime spares, MTBF

15. NUMBER OF PAGES

52

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

NSN 7540-01-280-5500

Standard Form 298, (Rev. 2-89)
Prescribed by ANSI Std. 239-18
299-01

Logistics Management Institute
2000 Corporate Ridge
McLean, Virginia 22102-7805