AD-A205 489

(2)

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION: Unclassified

1b. RESTRICTIVE MARKINGS: DTIC FILE COPY

2a. SECURITY CLASSIFICATION AUTHORITY:

2b. DECLASSIFICATION / DOWNGRADING SCHEDULE:

3. DISTRIBUTION/AVAILABILITY OF REPORT: Distribution Unlimited

4. PERFORMING ORGANIZATION REPORT NUMBER(S):

5. MONITORING ORGANIZATION REPORT NUMBER(S):

6a. NAME OF PERFORMING ORGANIZATION: Air Force Operational Test & Evaluation Center (AFOTEC)

6b. OFFICE SYMBOL *(If applicable)*: HQ AFOTEC

7a. NAME OF MONITORING ORGANIZATION:

6c. ADDRESS *(City, State, and ZIP Code)*: Kirtland AFB NM 87117-7001

7b. ADDRESS *(City, State, and ZIP Code)*:

8a. NAME OF FUNDING/SPONSORING ORGANIZATION: Same

8b. OFFICE SYMBOL *(If applicable)*: Same

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER:

8c. ADDRESS *(City, State, and ZIP Code)*:

10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.

11. TITLE *(Include Security Classification)*: AFOTEC Technical Journal, Volume 2, Number 1, January 1988

12. PERSONAL AUTHOR(S):

13a. TYPE OF REPORT: Journal

13b. TIME COVERED: FROM ______ TO ______

14. DATE OF REPORT *(Year, Month, Day)*: 880100

15. PAGE COUNT: 13

16. SUPPLEMENTARY NOTATION: Periodical (2 Issues/Year)

17. COSATI CODES

FIELD	GROUP	SUB-GROUP
14	02	

18. SUBJECT TERMS *(Continue on reverse if necessary and identify by block number)*:
Test and Evaluation
Operations Research

19. ABSTRACT *(Continue on reverse if necessary and identify by block number)*

Accession For
- NTIS GRA&I ☒
- DTIC TAB ☐
- Unannounced ☐
- Justification ______

By ______
Distribution/ ______
Availability Codes

Dist	Avail and/or Special
A-1	

DTIC COPY INSPECTED 1

DTIC ELECTE 28 FEB 1989 S D E

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT: ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS

21. ABSTRACT SECURITY CLASSIFICATION: Unclassified

22a. NAME OF RESPONSIBLE INDIVIDUAL: K R Newsom

22b. TELEPHONE *(Include Area Code)*: (505)846-2574 (AV246-)

22c. OFFICE SYMBOL: HQ AFOTEC/RS

DD Form 1473, JUN 86 — *Previous editions are obsolete.* — SECURITY CLASSIFICATION OF THIS PAGE

AIR FORCE OPERATIONAL TEST AND EVALUATION CENTER

VOLUME 2 NUMBER 1 JANUARY 1988

Comments from the Chief Scientist

This is the second AFOTEC Technical Journal. We have tried to improve the format of the journal and at the same time allow space for more articles. On the first go-around, we received a number of articles which were used in the first journal or are being included in this one. We still have a few more left to publish, but need others. Think about it, and let us know if you have an article, an idea or a question.

The journal has two basic functions: To let people within AFOTEC, and within the rest of the T&E community, know about techniques and procedures developed to support test and evaluation, and to allow a forum to pose questions on AFOTEC policy. Of course, 55-1 and all the other regulations contain all of the policy information, and are always complete and current, so there should be very few questions about policy on test plans, test conduct or report format. In addition, the analytical techniques for small sample size, guidelines and for achieving the desired degree of statistical significance in OT&E, and all the other potential technical issues are well defined, so there should be no questions there. However, in the event that there is some degree of doubt about some small detail, drop a note. Anonymous letters are accepted.

As a matter of interest: A new Test and Evaluation executive position is being established in SAF/AQ.

A study of the AFOTEC organization has been completed by the Air Force Academy and some management experts and has been briefed to the AFOTEC commander. The prospects for an OT&E range are getting better. If more detail on any of these areas is desired, you might write or call.

Dr. Marion L. Williams

AFOTEC Technical Journal

The AFOTEC **Technical Journal** is a funded Department of the Air Force field command periodical published twice a year for more than 1,000 people assigned or attached to the Air Force Operational Test and Evaluation Center at Kirtland AFB, N.M., 87117-7001, at a three people per copy ratio. Opinions expressed herein do not necessarily represent those of the United States Air Force.

Commander.......Maj. Gen. Cecil W. Powell
Vice Commander.....Col. Joseph E. Merrick
Director of Public Affairs..Capt. Garrett T. Mason
Editor....................Dr. Marion L. Williams
Editorial Assistant.............SSgt. True Carr

Directorate of Analysis Technical Paper Series

By Mike Stolle

The OA Technical Paper Series was developed to provide a hands-on, working reference for some of the OT&E issues commonly encountered by analysts. The papers cover a wide variety of subjects ranging from statistical concepts to the development of operational scenarios. We are in the process of revising some of the papers. Second versions of the Statistical Concepts (No. 1.1) and Circular Error Probable (No. 6.1) papers were recently published; and a revised Sample Size (No. 2.1) paper. that includes Sequential Sampling (Former Tech Paper No. 9), is due out by the end of the year. These papers have been prepared by HQ analysts with inputs from the AFOTEC test teams, detachments and OLs. We are continually searching for new topics, papers or ways to improve our current set of papers. If you have any ideas, contact Mr. Mike Stolle, HQ AFOTEC/OA, AV 244-0321. The following paragraphs provide a short description of our current Technical Papers. Note that paper numbers 4.0, 9.0, and 10.9 are not assigned. If you would like copies of any of the current papers send your request to HQ AFOTEC/OA, Kirtland AFB, NM 87117-7001.

1.1 Statistical Concepts.

This paper presents a review of a common set of statistical concepts frequently used in the planning, testing and reporting of OT&E. For those versed in statistics, it is a good refresher. For the newcomer, it is an easily understood "how to" guide. The paper provides a framework within which all statistical testing can be viewed. It also warns against some common errors and misuses of statistics in the context of OT&E. This is the second version of this paper which has been extensively revised to frame the process of doing statistics in the mold of the scientific method.

2.1 Sample Size.

Operational testing is motivated by the need to know how well a production article will perform in the field. Since we cannot test all production articles, we estimate a systems field performance from testing a few "representative" samples. In statistical terms, this process is known as making statistical references from a sample. The purpose of this paper is to allow the tester to determine the resources (sample size) required to obtain a given confidence, or conversely to determine the confidence in the results given a limited amount of resources. Statistics provide a systematic, consistent means of making these judgments. The paper covers normal and non-normal distributions sequential sampling, and microcomputer routines. This is the first revision of this paper and now includes former Technical Paper No. 9, Sequential Sampling.

3.0 Test Design.

This paper is intended as an introduction to the art of test design. It is a comprehensive, theoretical paper on statistics. The emphasis here is on the practical application of test design principles with lots of numerical examples. The goal is to provide the analysts a framework for building a good test design and some knowledge of the sampling techniques that are available.

4.0 Not Assigned.

5.0 Billy-Bob Statpak Users Guide.

The Billy-Bob statistics package provides general statistics for the Z-100 microcomputer. The package is divided into five programs ONEWAY and TWOWAY (written in MS-Fortran), QUIKTEST, PROBS and MEANS (all written in Turbo-Pascal). ONEWAY provides analyses on one or two value columns described by one category column. TWOWAY analyzes one value column described by two category columns and is capable of investigating interaction effects. QUIKTEST allows the analysts to do statistical tests with data parameters rather than data elements. PROBS and MEANS are programs for sequential sampling and sample size. This paper described the input file structure, output formats, command line syntax, and gives a brief description of the routines available in ONEWAY and TWOWAY.

6.1 Circular Error Probable (CEP).

CEP is a statistical technique most commonly used in the areas of navigation and system accuracy. It is often misunderstood and misused. This paper discusses, step-by-step, how to examine your data set and which tests to use to determine if CEP is the appropriate measure. If your data will not support a calculation of CEP, the paper discusses some alternative "accuracy" measurements. This paper uses a significant number of examples.

7.0 Questionnaire Handbook.

This handbook was created to provide a set of reasonable guidelines for designing, collecting, and evaluating subjective data. The chapters were written generically to provide a useful reference for many types of research and opinion surveys/studies; however, it is primarily oriented toward operational test and evaluation. The examples and data used were derived from the testing arena. The handbook was written for the engineer or operations analyst with approximately one course in basic statistics.

8.0 Service Report (SR) Prioritization. This paper discusses several available methods of ordering or prioritizing service reports. It provides a brief introduction to the SR and the SR process (with references), and then discusses three methods of ordering them. Each method section contains a description of methodology, advantages and disadvantges, and recommendations. Finally, an approach is suggested that combines the advantages of each method.

9.0 Not Assigned.

10.0 Not Assigned.

11.0 Developing Operational Threat Scenarios. Operational scenarios are estimates of how an adversary might attack the system we are testing. A carefully constructed operational scenario helps the analyst decide which threats to exclude from a survivability assessment and how to prioritize the threats that are retained. This paper describes a procedure that folds together the system mission timeline and the potential threats to produce an operational scenario. Included are a hypothetical system example and a comprehensive list of data sources.

Circular Error Probable (CEP)

Note: This article is a summary of an AFOTEC Technical Report on CEP analysis which includes all the appropriate statistical tests, equations, tables and numerical examples of CEP.

by
Capt. James F. Sheedy
AFOTEC/OAC

CEP is a statistical technique most commonly used in the areas of navigation and defense system accuracy and is often misunderstood and misused. By definition,

> "CEP is a circle, centered about the mean, whose boundary is expected to include exactly 50% of the population within it."
>
> [1:1-2]

The key word in the definition is population, not sample. CEP may not contain exactly 50% of the sample data points, but it is expected to include 50% of the "true" population.

Since CEP is a parameter of "circular bivariate normal (CBN) distribution," it's fairly obvious that if your sampling distribution is not CBN, then CEP should not be used to estimate a 50% circle. Many people, however, calculate and utilize CEP when it is not justified. To determine if you're sampling from a CBN distribution, the following assumptions must be tested:

1. The X and Y components are statistically independent.
2. The distribution of X and Y are both normal.
3. The distribution is circular.
4. The mean point of impact (MPI) is at the target.

These assumptions are seldom tested because if they were, you would find CEP to be inappropriate.

Prior to using any step-by-step procedure for statistical analysis, it should be understood that all tests have underlying assumptions. If these assumptions do not hold, you can no longer be sure of your results. The guideline is that you must be satisfied in your own mind, or test for validity, any assumptions made. Failure to do so may result in an invalid analysis or incorrect conclusions. The following is a general guide on how to complete an analysis using classical CEP techniques with the option for performing noncircular analysis.

1. LOOK AT THE DATA: Histograms, scatter-plots, etc., are presentations of the data which help you determine whether you should continue with a CEP analysis, or try a different approach to measuring system accuracy. Visual inspection is not a substitute for detailed tests of the assumptions, but if a distribution appears non-circular, there is no reason to use CEP.

2. TEST THE ASSUMPTIONS - The HQ

AFOTEC/OA Technical Report shows how to use the following tests:

		Parametric Test (normal distribution)	Non-Parametric Test (distribution free)
Assumption #	1	t-test	Spearman's Rho
	2	Chi-Square	Lillifors
	3	F-test	Seigel-Tukey
	4	t-test	Wilcoxon -Signed Rank

3. ANALYSIS:

If all assumptions hold, you can calculate an estimate for the true CEP [CEP] using one of the following equations:

$$\hat{CEP} = .6152\ S_y + .5640\ S_x$$

$$\hat{CEP} = (.8209K - .0007)S_y + .6745\ S_x, \text{ if } K < .3$$

or

where,

$$K = S_y/S_x \text{ if } S_x \geq S_y \text{ or } K = S_x/S_y \text{ if } S_y > S_x$$

S_x, S_y = sample standard deviations of x and y, respectively

CEP by itself, however, is not the entire story. You must calculate a confidence interval [CI] and report it with the CEP. You can calculate either an upper CI or a two-sided CI on CEP. The upper CI is calculated as follows,

$$\text{Upper CI} = \hat{CEP}\sqrt{\frac{(1 + K^2)(N-1)}{\chi^2_{[1-\alpha,\ \text{Int}[(1 + K^2)(N-1)]]}}} \qquad (1:11)$$

Note that Int$[(1+k^2)(n-1)]$ is the integer part of $[(1+k^2)(n-1)]$.

The following equations, however, are used to calculate the two-sided CI.

$$\text{Outer Value} = \hat{CEP}\sqrt{\frac{(1 + K^2)(N-1)}{\chi^2_{[1-\alpha/2,\ \text{Int}[(1 + K^2)(N-1)]]}}}$$

$$\text{Inner Value} = \hat{CEP}\sqrt{\frac{(1 + K^2)(N-1)}{\chi^2_{[\alpha/2,\ \text{Int}[(1 + K^2)(N-1)]]}}}$$

Just as there was a misunderstanding of CEP, there is also a misunderstanding of the CI surrounding CEP. By definition, if you calculate a 95% upper CI on CEP, then you are 95% confident that the "true" CEP is within this interval. This CI does not mean that 95% of the "sample" data points will fall within this interval.

If you cannot assume normality, you can still estimate CEP. Since the median is also a measure of location and 50% of the data values are on each side of it, the median radial error (MRE) is a good measure of CEP. If you convert the (X,Y) coordinates into radial distances ($r = \sqrt{X^2 + Y^2}$) and take the median of the resultant distribution, you then have MRE. Fifty percent of the sample distances will fall within a circle of radius MRE, and you will have a 50% circle of expected miss distances.

Once again, the CI on CEP is just as important as CEP itself. You can get the upper CI value using a cumulative binomial table with p = .5 since you want a 50% circle. You enter the binomial table with your sample size (n), p =.5, and extract the appropriate upper value based upon your confidence level. For example, if you have a sample size of 13 projectiles, then after ordering the projectiles by increasing MRE the 99% CI extends to the projectile with the eleventh largest MRE (.01 probability in the upper tail). My indepth report on CEP includes both the proper binomial tables and an example of this procedure.

If the distribution is not circular, a modified Range Error Probable (REP) and Deflection Error Probable (DEP) can be used to calculate Box Error Probable (BEP) which is a rectangle that is expected to contain 50% of the population. To develop this box, the spreads in X and Y should each include 70.7% of the projectiles (.707 x .707 = .4998 $\approx$ .50).

Using the normal distribution, 70% of the probability for REP and DEP lies between ± 1.052 S_y and ± 1.052 S_x, respectively.

If you want to find a 90% CI for a particular rectangle (see figure), you must have a $\sqrt{.9}$ or approximately a 95% CI in each direction. (4.140

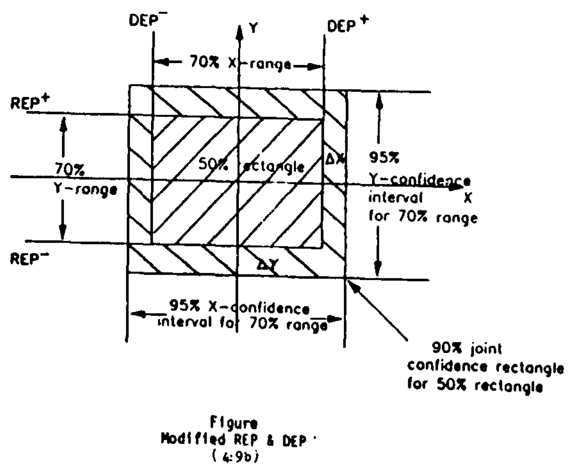

Figure
Modified REP & DEP
(4:9b)

The 90% CI can be developed using the chi-square statistic. You define ΔY and ΔX by the following formulae,

$$\Delta Y = 1.052 \sqrt{\frac{(N-1)\,S_Y^2}{\chi^2_{.10\,N-1}}} - 1.052\,S_Y$$

$$\Delta X = 1.052 \sqrt{\frac{(N-1)\,S_X^2}{\chi^2_{.10\,N-1}}} - 1.052\,S_X \qquad (4:15)$$

ΔY and ΔX represent the difference between the sides of the original rectangle and the sides of the larger CI rectangle (See Figure).

The main objectives of this article were to point out what CEP actually represents, what is actually meant by a CI around CEP, how you calculate CEP and CIs, and that CEP (or BEP) has limited value without CIs. I hope that this presentation of CEP will help fellow AF personnel solve defense system accuracy problems.

BIBLIOGRAPHY

1. Burgess, Gregg M., 'CEP — What It Is and How To Use It,' USAF Academy CO, no date. 2. Directorate of Analysis, 'Circular Error Probable [CEP],' HQ AFOTEC Technical Paper 6.0, Kirtland AFB NM, July 1987. 3. Roege, William H., 'CEP and Other Accuracy Measures,' HWY AFOTEC Technical Paper 6.0, Kirtland AFB NM, January 1986. 4. 'Evaluating CEP,' no author, no date. 5. Hines and Montgomery, 'Probability and Statistics in Engineering and Management Science,' 2nd Ed, 1980. 5. Math 358 Class Notes, USAF Academy Co, Spring 1986.

Today's model is more than a pretty face

By Dr. Pat Sanders

Once upon a time, when someone spoke of a "model" the image conjured up was one of a face and figure adorning the cover of a fashion magazine. But these days around AFOTEC when one speaks of a "model" it's likely that an entirely different meaning is being attached to that word.

While many different definitions of "model" are possible, the term as we use it here refers to a mathematical representation of a system or part of a system and/or some part of its environment. The model is thus a set of equations, which may be solved manually or by means of a computer, to determine the behavior of the system under a specific set of conditions or during a series of sets of conditions.

OT&E is generally more effective if it can consist of real exercises involving real systems operated by real personnel in a real environment. However, when this is not entirely possible, a model can make a contribution to the OT&E process. Some situations where their use is appropriate might be:

—In test planning, models can be valuable in identifying sensitive areas of system performance so that tests can concentrate on these areas and avoid the prohibitive cost of extensive testing in areas where system performance is insensitive to changes. They can also be beneficially used to identify meaningful MOEs and associated criteria. A simple "back of the envelope" resolution model is being used to determine the sensitivity of aircraft in a battlefield area interdiction scenario to ECM system parameters in order to help develop criteria for the Airborne Self Protection Jammer (ASPJ) IOT&E. Models of the electromagnetic environment are used to predict areas where electromagnetic interference (EMI) might be a problem for a system. These conditions can then be approximated or duplicated in test to determine actual performance.

—Models can sometimes be used to extrapolate from data gathered in field testing to system performance in an environment or under conditions which could not be replicated in the actual live tests. For IR MAVERICK and LANTIRN, we use the TAC REPELLER model to extend test results to a survivability assessment in a multi-threat environment. This use of models enables us to interpret test data in more operationally meaningful terms. A digital simulation called TAC JAMIT is being used with the Joint Tactical Information Distribution System (JTIDS) to take developmental test data on jam-to-signal ratios and graphically portray areas and times within a representative operational scenario when communications could take place. This performance can then be verified in a limited number of costly flight tests. Modeling is being used to interpret results of testing the B-1B defensive avionics system against representative technologies in terms of predicted effectiveness against the actual threat.

—Another payoff from the use of models is in predicting system level performance when only a portion of the system is available for test. Our IOT&E of the Ground Wave Emergency

Network (GWEN) will be conducted on the "thin line system"—a representative portion of the full operational capability. The SIMSTAR network model will extend these results to estimate expected performance of the complete system. Similarly, for the Over-the-Horizon Backscatter (OTH-B) radar, we will test the East Coast radar only but use a model to predict the capability of the entire system including West, Central and Alaskan segments.

—When an early operational assessment is required to make decisions about systems during developmental stages when no operational hardware exists, one way to make judgments is through the use of digital representatives of the system. We plan to incorporate a model of the Joint Surveillance Target Acquisition Radar System (JSTARS) into the Warrior Preparation Center in Germany to help assess the impact of information provided by that sensor on the battle commanders capability to conduct the war. This will be accomplished during exercises before the system is ever fielded in the theater.

Models will play a critical role in our early approach to assessing the operational utility of the Stratetic Defense Initiative (SDI) programs.

Because models are such valuable tools for enhancing our capability to perform comprehensive operational evaluations, AFOTEC is investing in the development of a significant inhouse capability to maintain, model, run and, in some cases, build, these mathematical representations. We are seeking to have useful credible appropriate models available to calibrate these models with field test data, to validate the threat portions of the models with the intelligence community, and to have the expertise to use these models effectively in support of our mission.

A model on the cover of a fashion magazine may be beautiful to behold, but a good computer model is more than a pretty face. It can work for us to truly enhance our capability to perform our mission of operational test and evaluation.

The challenge of Operational Test and Evaluation for Command, Control, Communications and Intelligence Systems

By Major Robert F. Baltz
OL-AW, Onizuka AFS, CA

The acquisition of Command, Control, Communications, and Intelligence Systems (C3I) presents special challenges to the operational tester. These systems are often one-of-a-kind and hence do not fit the classic acquisition strategy of exploration, demonstration, validation, full-scale development, and production and deployment. Unlike planes, tanks, missiles, and other weapons systems; command posts, missile warning systems, communications networks and other C3I systems do not have production models available for the operational tester. The tester must test on the finished and deployed article. Further complicating the tester's job are the large scale interfacing requirements usually required for these systems.

In reading the regulations governing Air Force operational test and evaluation, we find a most curious fact. Almost all the discussion revolves around the testing of weapons systems which will be produced in some quantity. Air Force Regulation (AFR) 80-AFR 55-43 mentions these programs in only two paragraphs. However, each regulation does state that operational test and evaluation for these systems should draw from standard OT&E procedures.

Standard OT&E procedures are designed to provide decisions makers with evaluations that will assist them in deciding whether or not to proceed with developing, buying, modifying or deploying systems. (4:1) However, in the case of the systems we have defined, the development, purchase and deployment of the modification can be identified through normal day-to-day operations. Given these facts, I have heard operational users of these systems ask: Why do we need to do initital operational testing, we have to use the system anyway? First let me point out that the question raised is valid and should not be dismissed lightly. However, the major premise on which the objection rests is that the need for necessary modifications can be developed by observing day-to-day operations. In the balance of this short paper I, will attempt to answer that question.

Day-to-day operations of C3I systems provide only a limited set of operational data on which a decision maker can base his decisions to modify the system. The set of operational data is limited because daily operations become stereotyped and do not vary much. Therefore, it is only when something unusual occurs and the system fails to act adequately that action can be taken to correct the problem. However, it now may be to late. A simple example can serve to illustrate the point.

During normal operations, a command post keeps track of the status of assigned forces and is responsible for deploying them as the situation requires. During a wartime situation, this command post must analyze all available information and deploy assigned forces quickly. If the C3I system supplies the decision makers in the command post with too much, too little, or the wrong information, critical decisions can be delayed. This delay might render the decision useless because the situation has changed. Unfortunately it is now too late to modify your system to correct the information flow; your command post has been overrun. A number of writers have pointed out this problem. Their articles have indicated that one of the major problems in any future war is that military commanders will be called upon to make decisions at extremely high speed in the face of a flood of data electronically delivered at a volume unprecedented in warfare. (2:162) One study has shown that for the commanders in a NATO Central Command bunker to keep up with the flow of information and orders coming over the communications system they would have to read 790 words per minute round the clock. (5:52) Therefore, it is necessary to operate the system under a variety of different operational scenarios to determine if the system will require modification in order to effectively deal with any changes in the operational environment. These changes should be those that might be expected to occur during the life of the system.

Critics of IOT&E may agree that developing operational scenarios is necessary and that using these for testing may provide valuable insight into system weaknesses. However, they might also argue that this can be done through normal exercise and rehearsals that are conducted as a part of routine operations. Examples of these types of exercises vary in scope from NORAD's Vigilant Overview Exercises to individual unit scenarios such as the Air Force Satellite Control Facilities launch rehearsals.

But these exercises are typically one-shot events where commanders are primarily interested in providing training in proven tactics for the forces or individuals involved. "Operational testing requires measurement and replication of activities or scenarios". (6:210) This is so the tester can vary specific actions or events to see how the outcome is effected.

Another hindrance to trying to do this type of scenario testing with an operational system is the amount of time that can be spared from normal operations to support the test. While I was working with the Ground Electro Optical Deep Space Surveillance System (GEODSS), an experiment was designed to measure the photometric signature of a satellite from two separate stations simultaneously. The purpose of the experiment was to see if better techniques could be developed to aid in using these signatures for space object identification. While problems were encountered with weather and visibilities, the primary reason the test was cancelled was that the stations could not be spared from normal operations for sufficient time to gather the required data.

One more problem that affects control centers of all types is simulating events while trying to conduct real world operations. NORAD's Vigilant Overview Exercises are prime examples of the difficulties of testing while trying to support normal operations. During these exercises, participants must react to simulated events as if they were real. However, real work events take precedence over scenario events. The dilemma for the tester should be clear. The events which he would be trying to measure will be influenced by real world events which he can neither predict nor control. Further duplication of test events to confirm findings will be highly unlikely.

When can the appropriate scenario testing be done? If we rule out attempting the test after the system is turned over to the operator, the only time available is after the system is finished development, and this is exactly when IOT&E is done.

IOT&E then provides the operator with the opportunity to test various operational scenarios against his system. These scenarios, if properly designed, will demonstrate the strengths and weaknesses of the system in various threat or stressed environments. "Once the base data on technical and human performance is obtained, we can analyze the data to determine the appropriateness of our employment of the system." (1:11) Once the system is put into use for normal operations, this chance will be lost. Given then the mission C3I systems perform waiting for real world events to see if we have made the right decision may prove to be catastrophic. (3:36) Clearly then IOT&E for these systems has a definite and large value that the operator should not waste.

BIBLIOGRAPHY

1. Baltz, Robert F. "Test and Evaluation for Command, Control, Communications and Intelligence Systems," Research Report, National Defense University, 1986.

2. Barnaby, Frank, Future War, New York: Facts on File Publications, 198[illegible]

3. Carter, Ashton B. "The Command and Control of Nuclear," Scientific American, (Jan. 1985).
4. Department of the Air Force. AF Regulation 55-43, Management of Operational Test and Evaluation. 28 June 1985.
5. Fallows, James, National Defense, New York: Random House, 1981.
6. Stockfisch, J.A. Plowshares into Swords, New York: Mason & Lipscomb, 1973.

Logistic Studies and Analysis Division HQ AFOTEC/LG4

By Lt. Col. Roger D. Hartman

Recently, a senior AFOTEC officer asked, "In non-technical terms, what does LG4 do?" It was somewhat of a shock because we thought our stuff was so important that everyone knew. Later, we realized that others might have the same question; so we would like to share our reponse with you—in non-technical terms.

The primary purpose of the Logistics Studies and Analysis Division is to integrate operational suitability data into a system view of a new weapon system's capability to function in its intended operational environment. This means analyzing and melding such factors as reliability, maintainability, manpower, spares support, test equipment, support equipment, operations and maintenance concepts, and mission scenarios to estimate mature system suitability and its influence on system capability. While we are part of each system's Test Planning Group, and we contribute technically throughout AFOTEC's involvement, all of the LG4 activity normally culminates in two major evaluation areas: Availability and Mission Reliability.

Put very simply, Availability is the measure of a system's readiness for commitment to the mission. Mission Reliability is the probability that, once committed to the mission, the system performs that mission without a critical equipment failure (Ref AFM 55-43 and AFR 800-18 for details). These major objectives are supported through various activities, the most salient being test design, data collection, and analysis of results.

Before we move on to those three points, one problem that is becoming more and more prevalent is the lack of realistic user requirements. Like all of OT&E, our part begins with the user's requirements. In many cases, the requirements have not been developed and/or are not well documented. If an SPO is already on contract for a system with a mean time between failure of 100 hours, and the user finally states that his requirement is for 500 hours mean time between maintenance, there is a problem. Often, we are the ones who identify the problem and try to help both sides reach some agreement. All of this hopefully takes place before detailed test design begins.

TEST DESIGN

In test design, we scope the using command's requirements and the system's critical issues to determine how much and what kind of data we need to collect during test to evaluate objectives at a specific level of confidence. For example, we might determine that 250 captive carry test hours are needed in order to have 80% confidence that a missile will meet the required captive carry reliability. Of course, for you more technically oriented people, this assumes that our test articles are representative of the production population. The composite of all such analyses becomes part of AFOTEC's required test resources and events.

One "problem" that we are beginning to face is the greatly increased reliability of equipment. When we were testing items with a mean time between failure (MTBF) of 100 hours, we could ask for (and receive) test time well in excess of the MTBF. But when new equipment comes along with a 1,000-hour MTBF, we can ask for 5 or 6,000 hours of test time, but probably not get it, since it would take months to accumulate. Therefore, we usually ask for enough test time to have 80% confidence in our results, but each program has to be realistically (pragmatically) analyzed to do what's smart.

The whole idea of test design is to zero in on the quantity and types of data needed to support an evaluation. It's a plan to purchase a quantity of information which, like any other commodity, may be acquired from varying sources at varying prices. We may trade off the cost of OT&E data by using acceptable DT&E data or other comparable data as long as it is operationally relevant. In the end, the "decision makers" may decide the cost of "good" information is too high—that something less is acceptable considering cost and program risk. Once the thrashing is completed, we do our best with what we can get and strive to ensure that the best quality data is collected.

DATA COLLECTION

Test teams collect suitability data. LG4's responsibility is to ensure that procedures for gathering reliability and maintainability (R&M) data and properly managing the data base are in place and followed. Proper data recording and management are critical because all R&M conclusions and recommendations are based on data-oriented calculations. Overall management control is critical because HQ AFOTEC is responsible for making judgments and recommendations to senior Air Force leaders. To assure understanding and acceptability, these judgments and recommendations must be based on data definitions, assumptions, and procedures common to the R&M community. This commonality is prescribed in various DOD, Air Force, AFSC and AFLC directives and documents.

Several years ago, AFSC was tasked to implement, in conjunction with AFLC, AFOTEC and the using commands, a data system for DT&E and IOT&E efforts. They selected the System's Effectiveness Data System (SEDS), a computer processing system which has become the primary R&M data system for OT&E. However, for those AFOTEC efforts for which we cannot obtain SEDS support (i.e., most FOT&E), we usually use the Air Force Maintenance Data Collection (MDC) system. However, there are several nuances from program to program that must be worked out by HQ AFOTEC, test teams, program offices and contractors. For example, the program office may plan to use a contractor data system in lieu of SEDS. Many small programs use manual systems. Each option must be explored and the best selected based on many factors. Our primary concern is that the data system is common to all users and that the operationally relevant data can be identified for our use. We also emphasize unique program requirements and standardization during early development of the charter for each program's Joint Reliability and Maintainability Evaluation Team (JRMET). The JRMET is composed of personnel from the implementing, testing, supporting and using commands as well as involved contractors and is the formal forum for classifying/categorizing R&M test data and resolving issues in this area. Standard terms and common data are the keys to proper understanding, clear communication, and credible results. R&M data forms the basis for all our work. If it's good, our analysts can help field a better system for the Air Force. If it's bad, no amount of analysis can ensure a good evaluation.

ANALYSIS OF RESULTS

As mentioned at the beginning of this article, most of our system's analysis focuses on availability and mission reliability. We tackle it thorugh integrating the previously mentioned data, operations and maintenance concepts, etc., into models of the system. Models are necessary to assess operational expectations because the test environment usually falls far short of the operational environment. For example, a test environment may have one pre-production aircraft, dedicated contractor logistics support, and be conducted at Edwards AFB, California. The true operational environment may be hundreds of aircraft with total Air Force logistics support at many bases around the world. Simulation models give us the capability to give significance and utility to test results beyond the narrow conditions of the test; to answer the questions of availability and mission reliability. Models also provide a structure for qualifying, in mission terms, the impacts of deficiencies, changes in policy/concepts, and proposed improvements. They give us the best insight into the expected readiness of the system and how to improve it.

Within LG4, we start thinking about a system's availability model as soon as we see the statement of operational need for the new system. The early idea of the system forms the basis for test design, and the model concept is developed with inputs from the program office and using command. As the idea is further defined, we design and build the simulation model, matching as closely as possible the planned operations and maintenance concept. Naturally, much of the early model is based on plans, tentative plans at best. Still, the model is useful to highlight areas requiring further investigation or test concentration. As further information about a system surfaces, such as logistics support analysis data from the contractor, we use that data to "check out" the model and further identify potential problem areas. If available, we use existing data from similar fielded systems to try to get as much "operational" as possible into an operational analysis. As the test period approaches, we task the using command to validate the model, certifying that the flow and logic do indeed reflect what the user intends to do with the system. As test data are gathered and run through the JRMET for classification and agreement, we exercise the model. These preliminary exercises serve to identify areas that need more detailed examination as well as provide a "how goes it" to the test team. Ultimately, we analyze model results to evaluate the new system's capability to meet its expected mission tasking. In some cases, at the end of OT&E, we transfer the model to the using command to further refine estimates over time and provide a means to assess other changes in the system or its concept of employment. Our overall methodology has served AFOTEC and the Air Force well by providing the means to assess the mission impact of logistics factors and address the operational benefits of changes to the system, in a quantitative manner.

Now to keep this non-technical, we won't go into modeling details, reliability growth projections, mathematical models vs. Monte Carlo simulations, simulation languages, etc. However, you should realize these are tools of our trade and that the melding of the many actors into a system model is not an easy task accomplished on the job at AFOTEC, but our people are fast learners and perform well. For areas where we do not have the experience or expertise, we call on the other AFOTEC directorates and divisions for advice. We especially use the hands-on experience and expertise of the Logistics Evaluation Division (LGM). Most of the people in LG4 have advanced degrees in operations research or mathematics, and what they lack in Air Force operations or maintenance experience, they make up for with enthusiasm. Just ask for a briefing on the rites and rituals of TOADs if you don't believe me.

As far as we know, AFOTEC has the largest collection of people dedicated to operational analysis of availability, reliability, and maintainability in the DOD. Our contributions do not come cheap, but the benefits of those contributions have been well-recognized throughout the Air Force and DOD.

SUMMARY

In summary, we like to think of our job as helping to build the suitability test design "box", ensuring that the test fills that box with the right quantity and quality of data, and tying the ribbon around the box through availability and mission reliability analysis. We don't do this in a vacuum and we don't do it alone. We depend on the expertise of the members of the test support group and the test team to scope, define and refine our efforts. As a team, we provide the answers to two of the most critical questions asked about a new system: Will this system be available when we need it, and will it complete the mission?

Operational Availability: The forgotten 'Ility'

The Air Force Operational Test and Evaluation Center (AFOTEC) is charged with evaluating the operational effectiveness and operational suitability of new systems in their intended operational environment. An operational suitability evaluation nearly always includes test objectives dealing with reliability (both mission and logistics), maintainability, logistics supportability, and operational availability. Mission reliability is a measure of the ability of a system to complete its planned mission or function. Logistics reliability measures the system's ability to operate as planned under the defined operational and support concepts using specified logistics resources. Maintainability measures the ability to retain an item in or restore it to a specified condition, using prescribed resources. Logistics supportability includes all elements of support, such as support equipment, training, technical data, spare parts, manning, and so on. Operational availability determines whether a system will be ready when needed.

During operational test and evaluation (OT&E), the above measures are evaluated in operational terms, considering all impacts of the intended logistics support concept. Let's look more closely at these measures and at the analysis techniques used.

R&M—A Generic Term

Recently, there has been increased activity and discussion regarding the importance of reliability and maintainability (R&M) in the weapon system acquisition process. In the past, R&M received widely varying emphasis during the different phases of a system's life cycle. The publication of the R&M 2000 Action Plan in February 1985 should help stabilize the attention that R&M receives, but there is still confusion regarding what R&M really involves. In fact, AFR 800-18 acknowledges that "R&M" is a general term and includes such items as availability and readiness. R&M cannot be considered as two distinct disciplines; R&M are part of the overall operational suitability, and must be considered in concert with logistics supportability and operational availability. There is even the grammatical question: should it be "R&M are..." or "R&M is..."? For this paper, I'll use "R&M," in quotes, to mean the generic term, referring to all of the support disciplines.

While the renewed emphasis is welcome, some very important aspects of suitability may be neglected by concentrating too heavily on R&M. The impacts of poor logistics supportability on a system are well known (since they are usually obvious), and this helps promote management visibility. However, operational availability is suffering from some neglect. For the moment, let's continue that neglect and look at the interrelationship of reliability, maintainability, and logistics supportability.

Picture the interrelationship of reliability, maintainability and logistics supportability as three intermeshed gears, the reliability gear including both mission and logistics reliability.

Mission reliability and logistics reliability are good measures of probability of mission success and of frequency of required maintenance, without full consideration of the scope of the maintenance. Mission reliability is usually expressed in terms of weapon system reliability (WSR)—a probability. Logistics reliability is usually expressed as mean time between maintenance (MTBM). MTBM is further subdivided into corrective maintenance due to inherent failures, induced failures, for which no failure is found, and total corrective maintenance. MTBM for preventive or scheduled maintenance can also be specified and evaluated.

WSR is a probability that a system (or piece of equipment) will perform its mission or function for the required period of time when called upon, given that it was initially capable of performing the mission. To determine WSR, the mission scenario and the individual reliabilities of the critical subsystems or components are needed. Once the failure distribution of the equipment is determined, the WSR for any mission length can be calculated. If different subsystems operate for different times, or during different phases of a mission, the mathematical relationship must be carefully established. As an example, the refueling boom on a tanker is certainly critical to accomplishing the mission, but its operating time is a great deal less than that of an engine. If the boom and engine have the same failure rate, the boom will still have a much higher WSR, since its mission length is much shorter. Thus, the boom and engine must be considered separately to evaluate the overall WSR of the tanker.

MTBM is a better measure of the required logistics support than WSR, since MTBM includes all maintenance actions, regardless of the impact on the mission. One of the engines on a tanker may experience a loss of thrust; not seriously enough to abort the mission, but certainly requiring maintenance when it returns to its home base. The WSR is unaffected by this failure mode, but a demand is placed on the support structure—probably including trained engine technicians, support equipment, technical orders (TOs), and supply.

This demand on the support structure brings maintainability into the scenario. Maintainability is usually based on time: time to accomplish a repair action, downtime, or man-hours. The most frequently used operational terms are mean downtime and maintenance man-hours per flying hour. Mean down time is the average amount of time a system is inoperable due to a maintenance action, including the administrative and supply delays associated with dispatching the technician, traveling to the system, ordering and waiting for parts, and so on. Maintenance man-hours per flying hour measures the amount of manpower required to support a system, based on the usage of the system. Other operational maintainability terms include mean repair time (for hands-on maintenance time), mean time to troubleshoot, manpower slots per aircraft, or any other derivative of the basic "time and people" aspect of maintenance.

Logistics supportability evaluations during OT&E encompass all aspects of integrated logistics support, except the time and people involved. The quantity and effectiveness of support equipment, adequacy of TOs, training programs for maintainers, and range and depth of supply support all affect the maintainability of a system. In fact, poor logistics supporability can cause reliability problems and greatly increase the maintenance time required. As an example, poorly written TOs can result in a maintenance technician inducing failures in the system and then requiring an extreme amount of time to troubleshoot.

Thus, these three parts of operational suitability affect each other and drive the support resources required to field a system. Logistics reliability determines the demand, maintainability determines the time and manpower required to restore the system, and logistics supportability determines the support elements that must be in place to return the system to operation. But these three are not enough. Three intermeshed gears are locked in place. They cannot revolve. To break the gridlock and truly evaluate the suitability of a system, operational availability is essential.

Operational availability is a measure of the degree to which an item is in an operable and commitable state when the mission is called for at a random point in time. Operational availability includes the effects of design, quality, installation, environment, operation, maintenance, software, repair, funding and management policy. Operational availability is the only operational term of "R&M" that encompasses the entire logistics support structure.

Operational availability usually uses such terms as "fully mission capable rate," "sortie generation rate," or "uptime ratio" to define requirements and evaluate progress. Occasionally, other meaningful terms, such as mission capable or alert rate, are used. In fact, if treated as a measure of the interaction of operations and logistics concepts, operational availability is the ideal term to truly define the cost effectiveness of a system. Availability measures can include the effects of spares procurement, manning levels, training provided, deployment concepts, designed R&M, and any other constraints unique to the system.

A statement of required system operational availability includes the effects of all the "R&M" elements, including those listed below.

Logistics Reliability. How often maintenance is required.

Maintainability. Time and manpower required for corrective or preventive maintenance.

Manning. Number of people and workload for both on- and off- equipment maintenance.

Spares and spare parts. Range and depth of items available and their physical location.

Support Equipment. Operational effectiveness and suitability of support equipment.

Transportation. Effects of delays, special handling required, packaging.

Software. Effects of software-induced problems and efficiency of diagnostic routines.

Administration. Effects of management decisions such as task prioritization, safety considerations, documentation.

Analysis Techniques

To adequately consider and evaluate the impacts of all these factors and their interactions, a simulation model is an extremely effective tool. A model can be constructed very early in the life of a system, using estimates for the various input parameters. These estimates can be based on historical data from similar systems, engineering analyses, preliminary concepts, or desired characteristics. Then, as more and more detailed information becomes available regarding the true system design characterists or support resources, the model can be exercised to evaluate the system's supportability in the field. Later in the life of the system, the model can be used to evaluate alternative support concepts, effects of modifications, changing tactics, and many other "what if..." questions.

AFOTEC develops availability models for analysis of nearly every type of system being tested. AFOTEC has created models for subsystems, missiles, munitions, ground electronics, and space systems, each tailored to the unique features of the system under test and its planned operations and support concepts.

Models are used to project test results into the true planned operational environment and to evaluate a system against the using commands' mature requirement.

More significantly, analysis of operational availability provides all involved agencies or managers with useful information on the system. Basically, operational availability answers the simple question: Will the system be ready when it is needed? By using simulation models, the answer can consider the interrelationships of "R&M" and the overall support posture, and the answer can be provided early in the life of a system to assist in finalizing the system support posture. An availability model provides the means to answer nearly any "what if..." question.

If the operational ability of a system is known or estimated, the remaining key question is: Will the system do its job when called upon? In other words, what is the WSR?

The Mission Continuum

The relationship of operational availability and WSR can be graphically depicted on a time line as shown.

```
                      Mission              Mission
O                      Start              Complete
I----------------------I-----------------------I
Operational                         Weapon System
Availability                          Reliability
```

Operational availability determines if a system will be ready when called upon, and WSR determines if the system can complete the mission. Considering these two together covers the full mission continuum and allows analysis of impacts of changes to the operation or support concepts.

Conclusions

To ensure thorough and meaningful evaluations using analysis techniques, using commands should determine operational availability and WSR requirements early in the life of a system and include them in the statement of need process. Program offices should develop simulation models as systems are designed and the support posture is planned and developed. A simulation model can be an extremely effective part of the logistics support analysis process. Operational availability and WSR requirements and results should be analyzed and fed back to the design or modification process throughout the life of a system. Naturally, coordination between the using, developing, testing, and supporting commands is essential.

R&M are very important aspects of a system, but the interrelationships of all the terms and measures of "R&M" make a total understanding of the operational suitability of a system difficult. The two key measures of operational availability and WSR can provide a total picture of system operational suitability. Operational availability cannot be neglected, it is a key measure of the overall suitability of any system. With proper emphasis on all the "Ilities," the Air Force will field more cost- and combat-effective weapon systems.