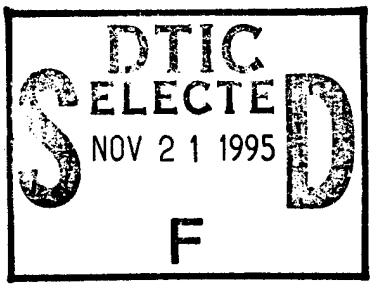
IDA PAPER P-3078

STATISTICAL AND MODELING UNCERTAINTIES IN THE THERMAL TARGET ACQUISITION MODEL IMPROVEMENT PROGRAM (TAMIP) PREDICTIONS



James D. Silk

September 1995

Prepared for
Advanced Research Projects Agency

INSTITUTE FOR DEFENSE ANALYSES
1801 N. Beauregard Street, Alexandria, Virginia 22311-1772

IDA Log No. HQ 95-46730

DEFINITIONS

IDA publishes the following documents to report the results of its work.

Reports

Reports are the most authoritative and most carefully considered products IDA publishes. They normally embody results of major projects which (a) have a direct bearing on decisions affecting major programs, (b) address issues of significant concern to the Executive Branch, the Congress and/or the public, or (c) address issues that have significant economic implications. IDA Reports are reviewed by outside panels of experts to ensure their high quality and relevance to the problems studied, and they are released by the President of IDA.

Group Reports

Group Reports record the findings and results of IDA established working groups and panels composed of senior individuals addressing major issues which otherwise would be the subject of an IDA Report. IDA Group Reports are reviewed by the senior individuals responsible for the project and others as selected by IDA to ensure their high quality and relevance to the problems studied, and are released by the President of IDA.

Papers

Papers, also authoritative and carefully considered products of IDA, address studies that are narrower in scope than those covered in Reports. IDA Papers are reviewed to ensure that they meet the high standards expected of refereed papers in professional journals or formal Agency reports.

Documents

IDA Documents are used for the convenience of the sponsors or the analysts (a) to record substantive work done in quick reaction studies, (b) to record the proceedings of conferences and meetings, (c) to make available preliminary and tentative results of analyses, (d) to record data developed in the course of an investigation, or (e) to forward information that is essentially unanalyzed and unevaluated. The review of IDA Documents is suited to their content and intended use.

The work reported in this document was conducted under contract DASW01 94 C 0054 for the Department of Defense. The publication of this IDA document does not indicate endorsement by the Department of Defense, nor should the contents be construed as reflecting the official position of that Agency.

IDA PAPER P-3078

STATISTICAL AND MODELING UNCERTAINTIES IN THE THERMAL TARGET ACQUISITION MODEL IMPROVEMENT PROGRAM (TAMIP) PREDICTIONS

James D. Silk

September 1995

INSTITUTE FOR DEFENSE ANALYSES
Contract DASW01 94 C 0054
ARPA Assignment A-162

PREFACE

This paper has been prepared for Mr. Thomas Hafer, Deputy Director Advanced Systems Technology Office, ARPA, in partial fulfillment of IDA task order on Analysis and Model Development. Additional cognizance and direction have been provided by Mr. John Brand and Mr. Eugene Patrick, U.S. Army Research Laboratory (ARL), S^3I Special Projects Office; and Mr. John D'Agostino, U.S. Army Night Vision and Electro-optics Systems Directorate (NVESD), Visionics Division.

These analyses would not have been possible without the high quality target acquisition performance data obtained by the Visionics Division of NVESD in their Phase I and Phase IV target acquisition tests.

CONTENTS

FIGURES

EXECUTIVE SUMMARY

As a participant in the Army's Target Acquisition Model Improvement Program (TAMIP), IDA has helped the Army's Night Vision and Electro-optics Systems Directorate (NVESD) improve models of human performance in target acquisition tasks using infrared sensors. One product of NVESD work is a model that predicts target detection probability based on measurable properties of the target and the background scene.

As is true of any model that computes probability, this model cannot precisely predict the result of any given experiment. Any finite sample of data will give an imprecise estimate of the probability of a given event. The expected departure of the data from the actual probability can be predicted on statistical grounds. If the actual disagreement between the data and the model exceeds this expected value, then there must be some residual modeling uncertainty.

In this paper we assess the quantitative agreement of the NVESD model with the available test data. Because of the good statistical reliability of the observer tests that were performed by NVESD in support of TAMIP, the experimental variability is small enough that the model uncertainty can be reliably measured. Our analysis shows that when the prediction that is computed by the model is suitably transformed, the model uncertainty is *unbiased* in the sense that it is numerically independent of the true probability. This property allows us to evaluate the remaining model uncertainty in a simple way.

Our final result is a quantitative description of the modeling uncertainty that is both accurate and easy to use. In particular, it is very easy to numerically simulate the uncertainty. In a forthcoming work,[1] we will exploit this property and demonstrate how to incorporate modeling uncertainty (as well as variation among observers) into wargaming simulations.

---

1 James D. Silk, *Modeling the Observer in Target Acquisition*, Institute for Defense Analyses, IDA Draft Paper P-3102, in preparation.

I. INTRODUCTION

A. BACKGROUND

No model that predicts probabilities, however accurately, can be expected to precisely match a given set of data. There will always be unpredictable "statistical fluctuations" which depend on the size of the sample. Our purpose herein is to describe quantitatively the degree of departure of the Army's Thermal Target Acquisition Model Improvement Program (TAMIP) Model predictions from the available data. If the model predictions were exactly correct, then the degree of departure would be consistent with well-known binomial error estimates. To the extent that the predictions exceed those expected from the binomial analysis, the model predictions are imprecise.

B. SCOPE

Note that this dichotomous categorization of errors requires that we define the term "modeling uncertainty" in a very broad sense. It includes a multitude of effects that are not shortcomings of the model *per se*. One simple example is the determination of the size of the target in a scene; whether determined from geometry or imagery, it is susceptible to measurement uncertainty. The model prediction will then reflect the error in the input. A complete list of the various sources of error that are expected to play a role is available elsewhere.[1] For the purpose of this paper, we have taken the point of view of the model user (rather than the developer) in that we are adopting the most inclusive interpretation possible of "modeling uncertainty."

C. OVERVIEW

We find in Section II that the variation of the current data from the model is inconsistent with the expected statistical error; therefore, the model predictions are not precise. We then determine a quantitative, unbiased measure of the model uncertainty in Section III. In Section IV we demonstrate that the uncertainty estimate is accurate for a data

---

[1] John D'Agostino et al., "Final Technical Report for FY93: TAMIP Thermal Modeling Program," NVESD Report, May 1994.

set that was not used in the quantitative formulation of the model. Finally in Section V we review the results and consider the scope of their validity.

II. STATISTICAL UNCERTAINTY

The Thermal TAMIP product predicts detection probability on the basis of a single composite statistic,

$$P_D = \frac{X^E}{1 + X^E} \quad , \qquad \text{(Eq. 1)}$$

where the exponent E = 3. The statistic is determined from three variables, in the form

$$X = \text{CONSTANT} \times \frac{PSS\sqrt{AREA}}{SV} \quad . \qquad \text{(Eq. 2)}$$

We prefer the natural logarithm of this predictor variable, x = ln X (see below). Then

$$x = \text{constant} + \ln(PSS) + \frac{1}{2}\ln(AREA) - \ln(SV) \qquad \text{(Eq. 3)}$$

and

$$P_D = \frac{\exp(Ex)}{1+\exp(Ex)} \quad . \qquad \text{(Eq. 4)}$$

Figure II-1 compares the prediction based on this formula with the data from the NVESD Phase 1 observer tests. (This is the data set that was used in the development of the Thermal TAMIP Model.[2]) We observe in Fig. II-1 that the data clearly follow the trend represented by the model but that there is some departure from the prediction. Note also that by using the logarithmic predictor, x, as the independent variable, we have made the "horizontal scatter" in the data fairly uniform for all P_D. This property will be very important later because it allows us to construct an unbiased model of the prediction uncertainty.

For a given fixed probability and a given number of samples, simple application of the binomial formula yields the frequency distribution that the test results would be expected to follow. Twenty-two observers participated in this test. The plot in Fig. II-2 illustrates the frequency distributions expected for this size sample for several fixed probability values.

---

2 Barbara L. O'Kane, Clarence P. Walters, John D'Agostino, Mel Friedman, "Target Signature Matrices for Performance Modeling," *Proceedings of IRIS Symposium on Passive Sensors*, Vol. 2, p. 161, 1993.

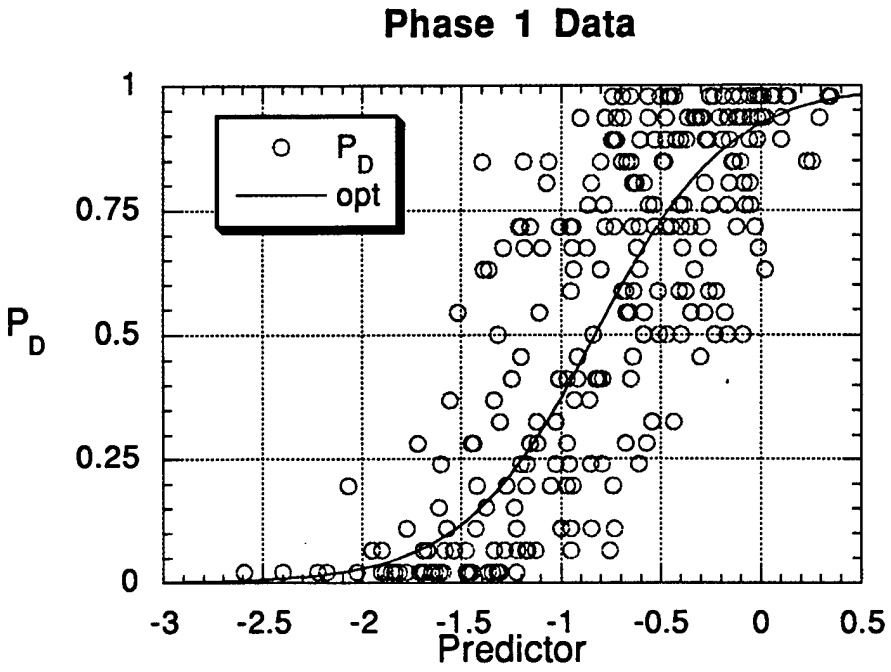


Figure II-1. The optimal (or maximum likelihood) fit, superimposed on the NVESD Phase 1 data.

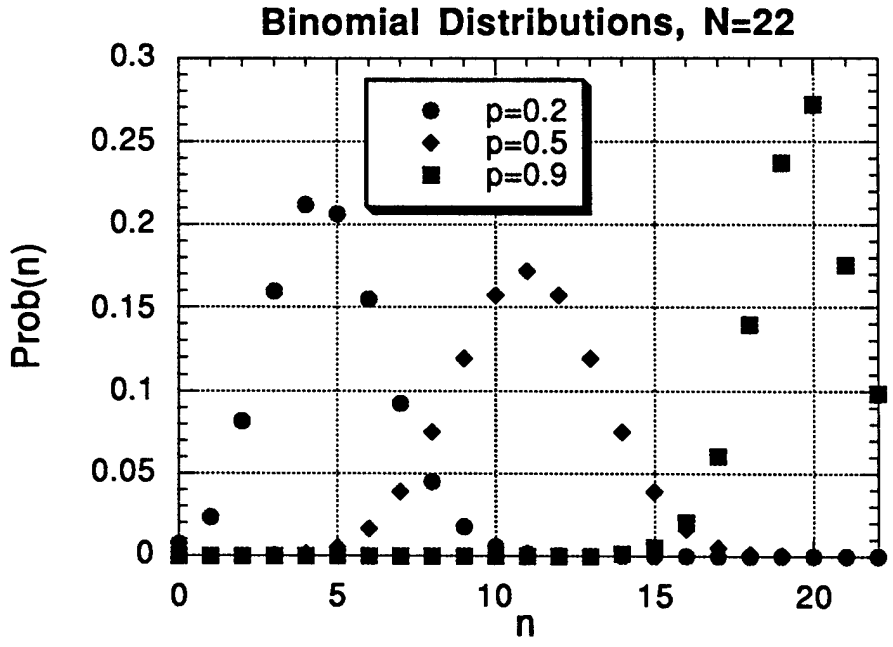


Figure II-2. Expected distribution of experimental results for 22 observers, assuming various known probabilities.

The question that we wish to answer is whether the scatter in the data of Fig. II-1 is primarily due to binomial statistical errors, or to an imperfect model. Our strategy is to construct an error envelope around the prediction (the curve in Fig. II-1) corresponding to some confidence interval. If the fraction of the data points enclosed by the error interval is consistent with the specified error interval, then the uncertainties for a single case would be primarily statistical, and the errors in the model could be assumed to be relatively small. This is in a sense the inverse of the simpler problem that we solved in the preceding paragraph. That is, given a set of test data, we need to compute the "error bar" associated with the corresponding probability estimate.

The determination of binomial error intervals is, unfortunately, not completely model free. The usual approach is Bayesian, and it is therefore necessary to specify a prior distribution of the probabilities that are to be estimated. We leave the details of the computation of the confidence interval to the Appendix, and show the results in Fig. II-3. We note two aspects of that computation. First, the choice between the two most common models of prior distributions do not make any appreciable difference here; we choose the one that shows most consistent with the test conditions. Second, this model gives the maximum likelihood estimate of probability as

$$P_D = \frac{n+1/2}{N+1}, \text{ not } \frac{n}{N} \qquad \text{(Eq. 5)}$$

so we use this prescription in Fig. II-3.

The confidence interval that we have displayed is the 10–90 percent interval. Therefore 80 percent of the data points should fall between the solid curves in Fig. II-3. Instead, that region encompasses only 105 out of 275 points, or 38 percent. We conclude that the residual errors in the model prediction are more significant than the statistical uncertainties at the precision of the NVESD test.

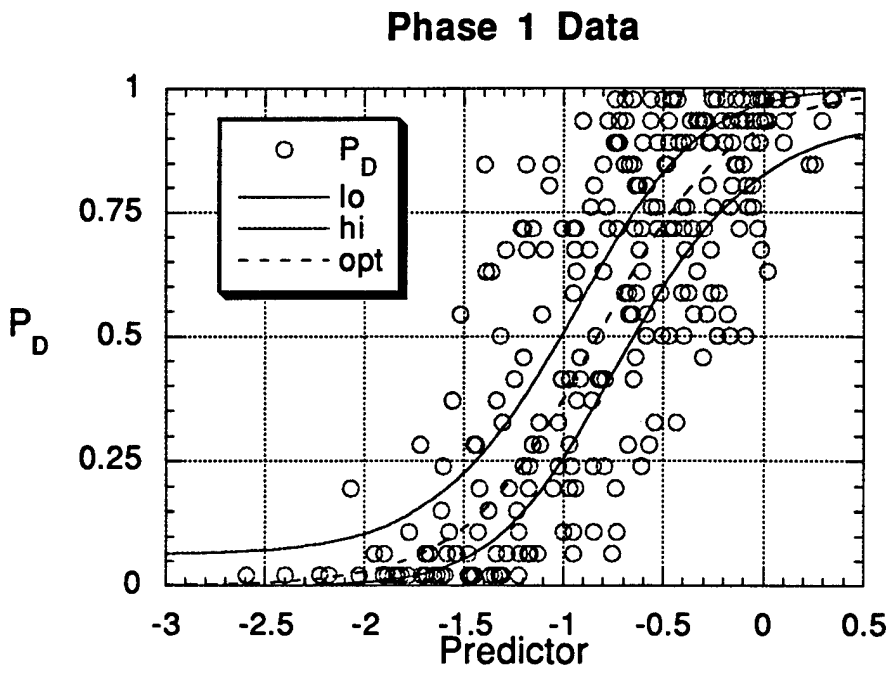


Figure II-3. As Figure II-1, with binomial error envelope about the maximum likelihood fit.

III. MODEL CONFIDENCE INTERVAL

We have established that the variation of probability model predictions from the estimates deduced from the NVESD tests exceeds that expected from the finite statistical samples. Therefore, a description of the model uncertainties must measure the uncertainty associated with our model predictor, x. In other words, referring to Fig. II-3, the "vertical errors" do not suffice to explain the data, so we must quantify the "horizontal errors."

Our presumption is that the model predictor x, given by Eq. 2, is only an approximation to the true predictor, x', which is presumed to exist but is still unknown. We shall also presume (since the data seem to support it) that the model predictor is an unbiased estimate of the true predictor. That is,

$$x = x' + \eta \, . \qquad \text{(Eq. 6)}$$

Before determining the error envelope for x, we need to review the procedure by which the exponent E in Eq. 1 was determined. The choice of E represents a maximum likelihood estimate of the prediction of the detection probability. In effect, the choice represents a minimization of the vertical component of the variations in Fig. II-1. On the other hand, we have now determined that the actual source of the error is in the estimate x of x', and conjectured that the error is independent of x'. Therefore we need to determine a new fit which minimizes the horizontal departure of the data from the fit from this baseline; then residuals in the x coordinate can be determined (in an unbiased manner) and associated with confidence intervals.

Figure III-1 shows two fits. The first, shown as a solid line, is the same one shown in Figure II-1. It is based on a maximum likelihood fit of the predicted P_D to the data, and therefore in a sense attempts to minimize the scatter in the vertical direction. The second fit, shown as a dashed line, is steeper than the first. It was obtained using the same functional form as the first (Eq. 4), but with a different value of E, which is chosen to minimize the mean square departure in the horizontal direction. Inspection verifies that the horizontal scatter from the steeper curve is quite independent of the position along the curve.

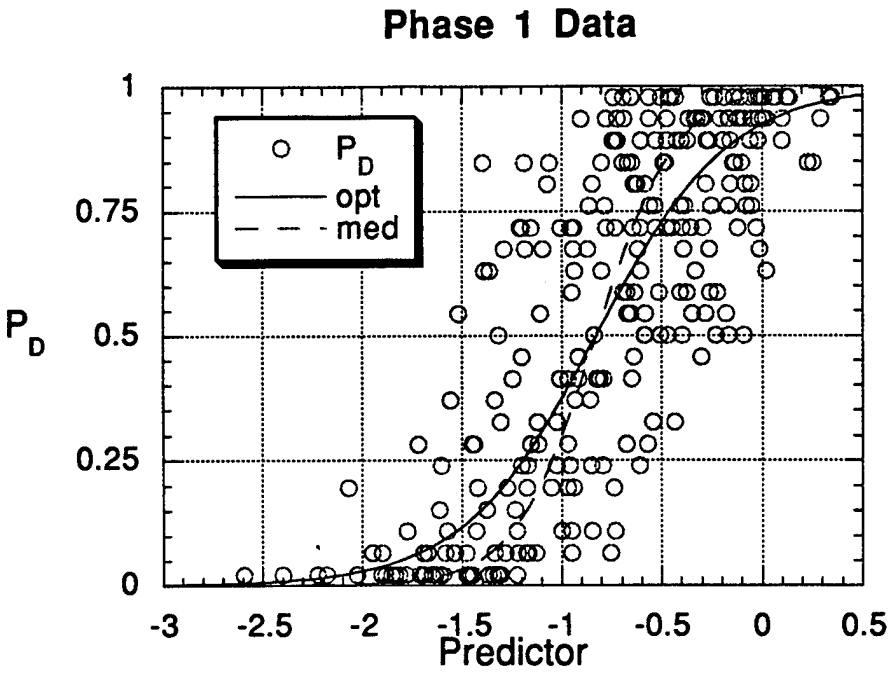


Figure III-1. As Figure II-1, with additional fit that minimizes the horizontal variance between the fit and the data.

The cumulative probability plot in Figure III-2 shows the distribution of the residual in the predictor based on the second fit. By residual, we mean here the horizontal distance between the data and the new fit. In the context of our assumptions about x and x', this plot should reflect the statistics of the random variable η. Since the percentile coordinate of the graph is normalized, the fact that the graph is nearly straight line means that the error is approximately Gaussian. It is easy to pick off the 10 percent and 90 percent confidence limits, which correspond to an 80 percent confidence interval $x \pm \delta x$ where $\delta x = 0.45$. (This cumulative distribution can in principle be used to determine any desired confidence interval. For various reasons to be discussed later, we recommend that this 10–90 percent interval be used generally.)

This value is used to generate the 80 percent confidence envelope in Figure III-3. It is easy to see that this envelope is unbiased, as the data that fall outside the envelope are uniformly distributed in P_D. This completes the specification of the model uncertainty.

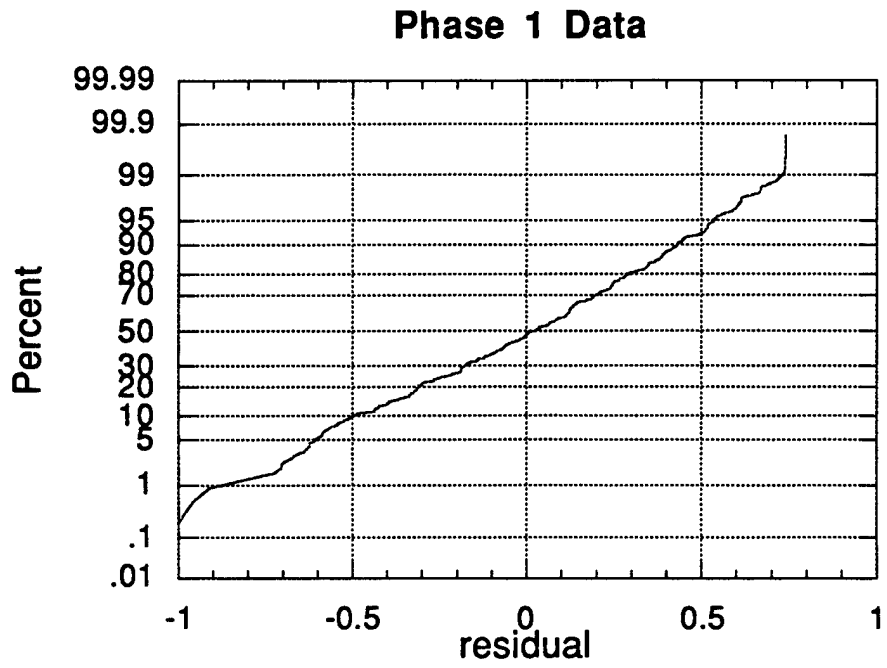


Figure III-2. Cumulative distribution of the residuals corresponding to the random contribution to the predictor variable.

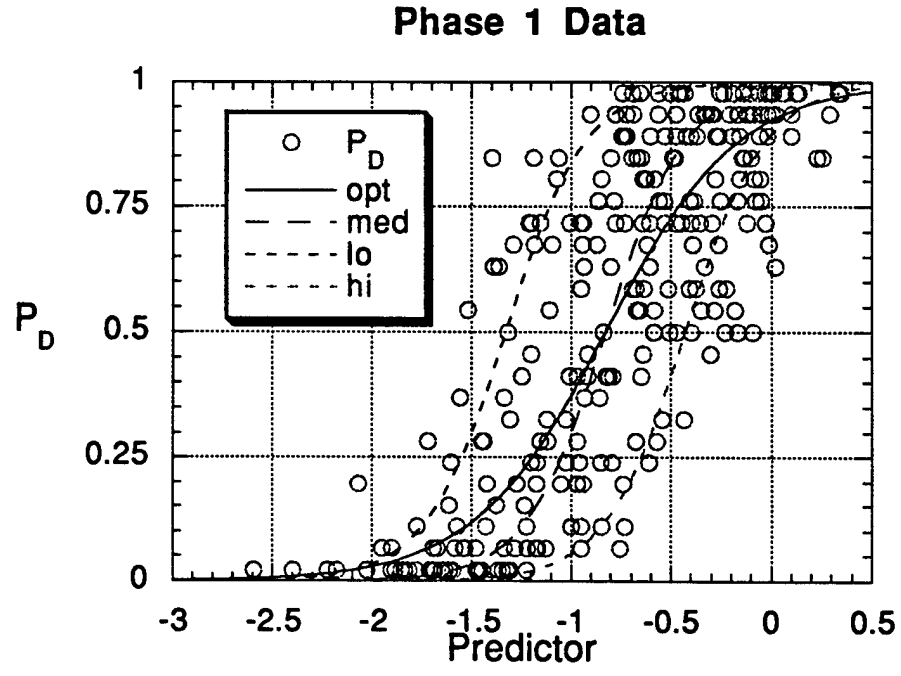


Figure III-3. As Figure III-1, with the 80 percent confidence envelope.

IV. VALIDATION USING PHASE 4A DATA

The NVESD Phase 1 data set formed the foundation for the analyses which supported the Thermal TAMIP model development. Due to the need for an extensive, well-controlled test database it relies on simulated sensors and model targets and backgrounds. The Phase 4 data set is superior in the sense that it was conducted using real thermal imagery collected in the field. It is perforce more limited (due to expense) and less controlled, and therefore the ideal resource for model validation.

Figures IV-1 and IV-2 are the analogs of Figures III-2 and III-3, but these come from the Phase 4 data set. Note that the 80 percent error interval and the least squares fit value of E = 5 for this data set are essentially identical to the Phase 1 values. (The overall constant of Eq. 3 was, however, chosen to optimize the fit. This reflects the effects of the difference between the simulated and real sensors.)

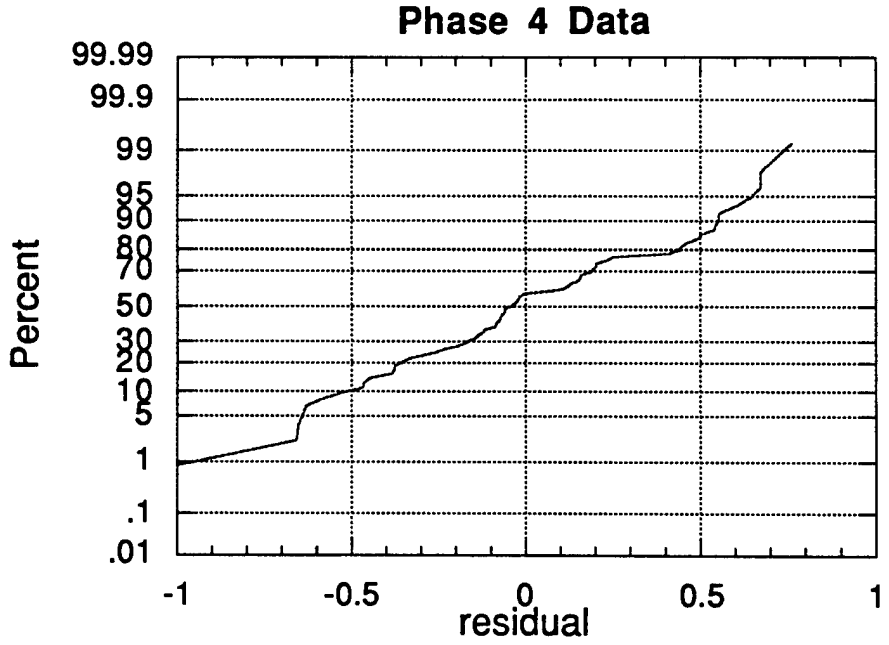


Figure IV-1. Analogous to Figure III-2, but for the Phase 4 data set. Recall that in Fig. III-2 the 10 and 90 percent points were at ± 0.45.

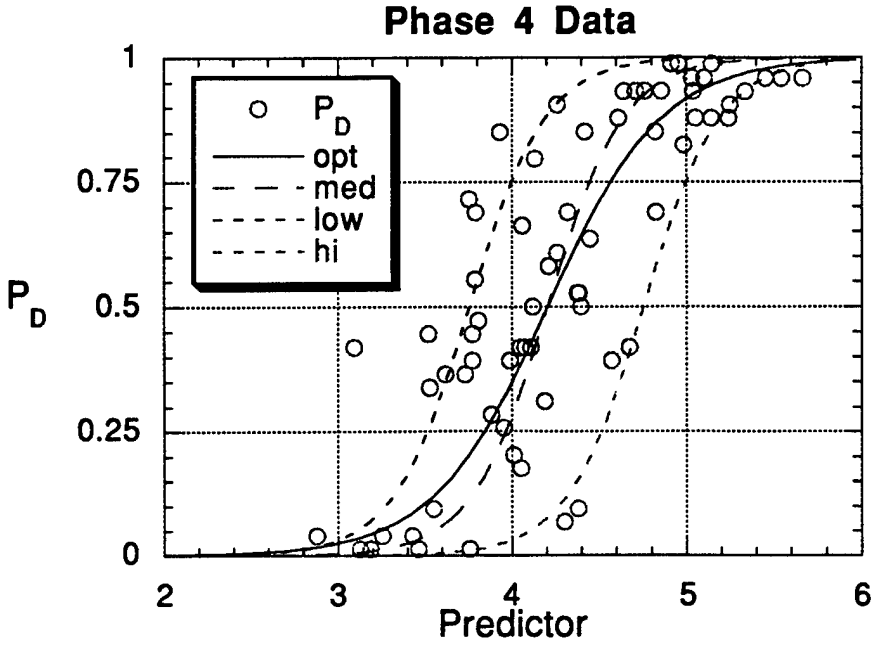


Figure IV-2. Analogous to Figure III-3, but for the Phase 4 data set.

V. SUMMARY AND DISCUSSION

The statistical errors in the present data samples, which are based on cohorts of 22 and 36 observers, have been shown to be small compared to the model uncertainty. We estimate that the crossover point, where the statistical errors are comparable to the model uncertainty, is in the vicinity of eight observers.

To establish unbiased confidence intervals for the model predictions, we have introduced a new exponent for use in Eq. 4. The new formula is used to generate confidence estimates for the predictor, x. We have, in so doing, introduced the notion that there is a "true predictor" x', for which the measured parameter x is an unbiased estimate, and that P_D is precisely determined via Eq. 4 using the new exponent. We find that a value of E = 5 removes the bias from the predictor residuals, and that the 10–90 percent confidence interval for the predictor corresponds to $\pm \delta x = 0.45$.

Intuitively, the steeper curve may look like a better fit than the shallower one. Nevertheless, it would be wrong to use the E = 5 curve to predict performance based on the current formulation of x. The resulting predictions will grossly underestimate P_D at low x, and grossly overestimate it at high x. The value of E = 3 is the only one that will give predictions of P_D that are trustworthy. The point here is that E = 5 applies only to the "true" predictor x'—which is at best still unknown, and may not even exist.

The confidence interval itself must be chosen judiciously. Clearly, since there are 275 trials in the sample, it doesn't make sense to push the confidence interval much past 5–95 percent. Moreover, since the unbiased confidence envelope has an unfortunate property in that it crosses the maximum likelihood prediction at extreme values, it is unwise to push the envelope much tighter than roughly 20–80 percent.

APPENDIX

COMPUTATION OF BINOMIAL UNCERTAINTIES

APPENDIX
COMPUTATION OF BINOMIAL UNCERTAINTIES

The Mathcad™ script shown on the following page performs the computation of the binomial confidence intervals. The method is based on the book by Martz and Walker.[1]

---

1 Harry F. Martz and Ray A. Walker, "Bayesian Reliability Analysis," John Wiley & Sons, 1982.

This file computes error bars on binomial success probabilities.
The estimate is Bayesian with equal prior probability.
See Martz & Waller, Bayesian Reliability Analysis.
Run time for N=40 is about half an hour.

$$f(p,n,N) := \frac{\int_0^p x^n \cdot (1-x)^{N-n}\,dx}{\int_0^1 x^n \cdot (1-x)^{N-n}\,dx}$$

Define the PDF. f is the probability that the true success probability is less than p, having measured n successes out of N trials.

$$f(p,n,N) := \frac{\Gamma(N+2)}{\Gamma(n+1)\cdot\Gamma(N-n+1)} \cdot \int_0^p x^n \cdot (1-x)^{N-n}\,dx$$

This form runs faster, but will eventually overflow.

Given $f(p,n,N) = g$

$P(g,n,N,p) := \text{Find}(p)$

Now wish to invert the PDF to findp=P in terms of f=g. P is defined as a solve block. Note initial guess is passed as an argument.

Now set up inputs, the number of samples and the desired confidence interval:

$N := 22$ $\quad CI := .80$

Symmetrize the confidence limits and loop over n.

$g_{lo} := \frac{1-CI}{2}$ $\quad g_{hi} := \frac{1+CI}{2}$

$n := 0..\frac{N}{2}$ $\quad PL_n := P\left(g_{lo}, n, N, \frac{n}{N}\right)$ $\quad PH_n := P\left(g_{hi}, n, N, \frac{n}{N}\right)$

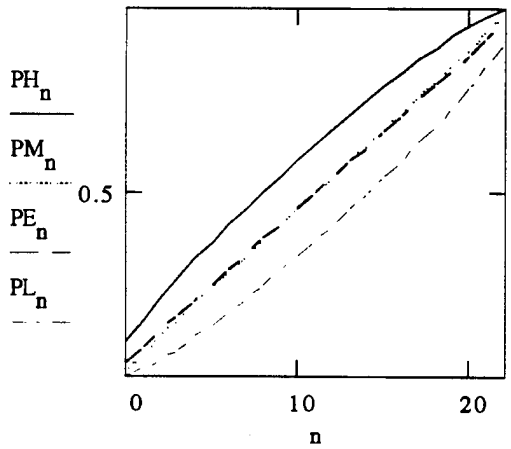
Use symmetry of limits to save time.

$PL_{N-n} := 1 - PH_n$ $\quad PH_{N-n} := 1 - PL_n$

$n := 0..N$

$PM_n := \frac{n+0.5}{N+1}$

$PE_n := \frac{n+1}{N+2}$



Results.

$N = 22$

$g_{lo} = 0.1$ $\quad g_{hi} = 0.9$

$m := 0..3$

PL_m	PH_m
0.005	0.095
0.023	0.159
0.049	0.215
0.078	0.268

$M^{<0>} := PM$

$M^{<1>} := PL$

$M^{<2>} := PH$

Toggle WRITEPRN if you really want to save results:

$WRITEPRN(\text{err22}) := M$

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	September 1995	Final—January - June 1995

4. TITLE AND SUBTITLE

Statistical and Modeling Uncertainties in the Thermal Target Acquisition Model Improvement Program (TAMIP) Predictions

5. FUNDING NUMBERS

DASW01 94 C 0054
A-162

6. AUTHOR(S)

James D. Silk

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

Institute for Defense Analyses
1801 N. Beauregard St.
Alexandria, VA 22311-1772

8. PERFORMING ORGANIZATION REPORT NUMBER

IDA Paper P-3078

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

Advanced Research Projects Agency
3701 N. Fairfax Drive
Arlington, VA 22203-1714

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 180 words)*

We analyze the uncertainties that are associated with the Thermal Target Acquisition Model Improvement Program (TAMIP) target detection predictions. The deviations of the measured probabilities from the predicted ones are much smaller than for the previous model, but still exceed those that would be expected from the finite statistical samples. Therefore there is still some residual error in the prediction. We determine confidence limits on the predictor variable which are unbiased in the sense that they are accurate at both high and low detection probabilities. We validate the confidence interval using a second data set that was not used in the development of this phase of the target detection model.

14. SUBJECT TERMS

target acquisition, model

15. NUMBER OF PAGES

22

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	SAR