AMSAA

TECHNICAL REPORT NO. 346

MODEL VALIDATION TECHNIQUES

DWAYNE W. NUZMAN
JOSEPH K. WALD

NOVEMBER 1981

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

U. S. ARMY MATERIEL SYSTEMS ANALYSIS ACTIVITY
ABERDEEN PROVING GROUND, MARYLAND

DISPOSITION

Destroy this report when no longer needed. Do not return it to the originator.

DISCLAIMER

The findings in this report are not to be construed as an official Department of the Army position unless so specified by other official documentation.

WARNING

Information and data contained in this document are based on the input available at the time of preparation. The results may be subject to change and should not be construed as representing the DARCOM position unless so specified.

TRADE NAMES

The use of trade names in this report does not constitute an official endorsement or approval of the use of such commercial hardware or software. The report may not be cited for purposes of advertisement.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER Technical Report No. 346	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) MODEL VALIDATION TECHNIQUES		5. TYPE OF REPORT & PERIOD COVERED Technical 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Dr. Dwayne W. Nuzman Dr. Joseph K. Wald		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS US Army Materiel Systems Analysis Activity Aberdeen Proving Ground, MD 21005		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS DA Project No. 1R765706M541
11. CONTROLLING OFFICE NAME AND ADDRESS Commander US Army Materiel Development and Readiness Command 5001 Eisenhower Avenue Alexandria, VA 22333		12. REPORT DATE November 1981 13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release, distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

simulation model
validation
Monte Carlo bands
Monte Carlo median
Vulcan Air Defense system
Gun Low Altitude Air Defense
GLAAD
Division Air Defense gun
DIVAD

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report documents the development of some model validation techniques developed at AMSAA in the past several years. The methods of Monte Carlo band analysis and trend/spread plots are explained. Examples are provided and applications are discussed.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

CONTENTS

Page

1. INTRODUCTION 5

   1.1 Purpose 5

   1.2 Background. 5

   1.3 Organization. 5

2. VALIDATION TECHNIQUES. 7

   2.1 Monte Carlo Band Analysis 7

   2.2 Trend/Spread Plots. 7

3. CONCLUSIONS. 27

REFERENCES 29

DISTRIBUTION LIST. 31

Next page is blank

MODEL VALIDATION TECHNIQUES

1. INTRODUCTION

1.1 Purpose.

The purpose of this report is to describe some model validation procedures which were developed for and used in the validation of a number of air defense gun models over the past few years. The techniques are very general and are applicable to the validation of models of systems other than air defense guns.

The development of these procedures is an on-going process. They are being published at this time not because they have reached a final form, but rather because we feel that they will be useful to other model validators and, in addition, we expect that broader exposure of these procedures will lead to further refinements of them and, perhaps, to entirely new developments in this field.

1.2 Background.

The Air Warfare Division of AMSAA has been striving to seek out and develop simulation validation techniques for a number of years. With our study of the Vulcan Air Defense System (Reference 1) a detailed engagement simulation model was developed and the output was compared with test data. Work continued with the Gun Low Altitude Air Defense (GLAAD) Fire Control Test Bed, where another simulation model was developed (Reference 2) and the prototype test bed was tested using a test design developed primarily for model validation. As part of the GLAAD effort the Monte Carlo band technique described in this report was developed. It is a modification of the Crow band technique developed by Dr. Larry Crow from RAM Division of AMSAA during the GADES Model Validation Committee Study (Reference 3). We have continued the development of these validation techniques into the competitive Division Air Defense (DIVAD) Gun test program.

1.3 Organization.

The main body of the material in this report is contained in Section 2. The method of Monte Carlo band analysis is explained in Section 2.1 while the trend/spread plot technique is presented in Section 2.2. Also shown in Section 2.2 are two examples, each illustrating the use of both techniques.

Section 3 contains a description of a larger context in which these techniques can be profitably employed.

Next page is blank

2. VALIDATION TECHNIQUES

2.1 Monte Carlo Band Analysis.

The basic idea underlying our validation techniques is to generate enough data with the model to allow one to get an idea of the distribution of the output of the model, and then to compare the field test data with this distribution in order to attempt to determine whether the data can be distinguished from a member of the distribution.

The main tool in our procedure is the Monte Carlo band technique (Reference 2). The model is exercised N times (we have typically used N = 100) and N time histories of the model output variable X are collected. A criterion level, α, is chosen (we have typically used $\alpha = .9$) and a corresponding value $M = N\cdot(1-\alpha)/2$ is calculated (typically M = 5). At each timestep an upper band, X_U, which is greater than all but M of the N values of X at that timestep and a lower band, X_L, which is less than all but M of those N values are obtained. Plots of X_U, X_L, and the test data versus time can give a good qualitative feel for the validity of the model if scaling problems do not occur.

Section 2.2 contains some examples of Monte Carlo band plots as well as a discussion of (and a solution to) the above mentioned scaling problem. The plots in Section 2.2 contain not only the test data, X_U and X_L, but also a fourth curve, the Monte Carlo median, which is a curve (lying between X_U and X_L) obtained by computing the median of the N values of the variable X at each time step. The importance of the Monte Carlo median will become apparent in Section 2.2.

2.2 Trend/Spread Plots.

The process which results in the formation of the Monte Carlo bands, while having the virtue of presenting a very simple visual image against which one can compare test data, has the unfortunate drawback of allowing to be lost much information as to the shape of the curves representing individual Monte Carlo replications. In example 1 the Monte Carlo bands (dotted lines) appearing in Figure 2.1 were formed using a model which produced a population of high frequency, noisy, sine curves of varying phase while the Monte Carlo bands of Figure 2.2 were formed using a model which produced a population of curves which are noisy straight lines. One would be hard pressed to distinguish between these two models using only the Monte Carlo bands. In order to make just such distinctions, necessary for the in-depth analysis of the DIVAD gun test program, we developed an extension of the Monte Carlo band technique which we dubbed trend/spread plots.

Trend and spread are measures of differences between a time history of data, X(t), and the time history of median values, M(t). These measures are based on the deviation, $D(t) = X(t) - M(t)$, of

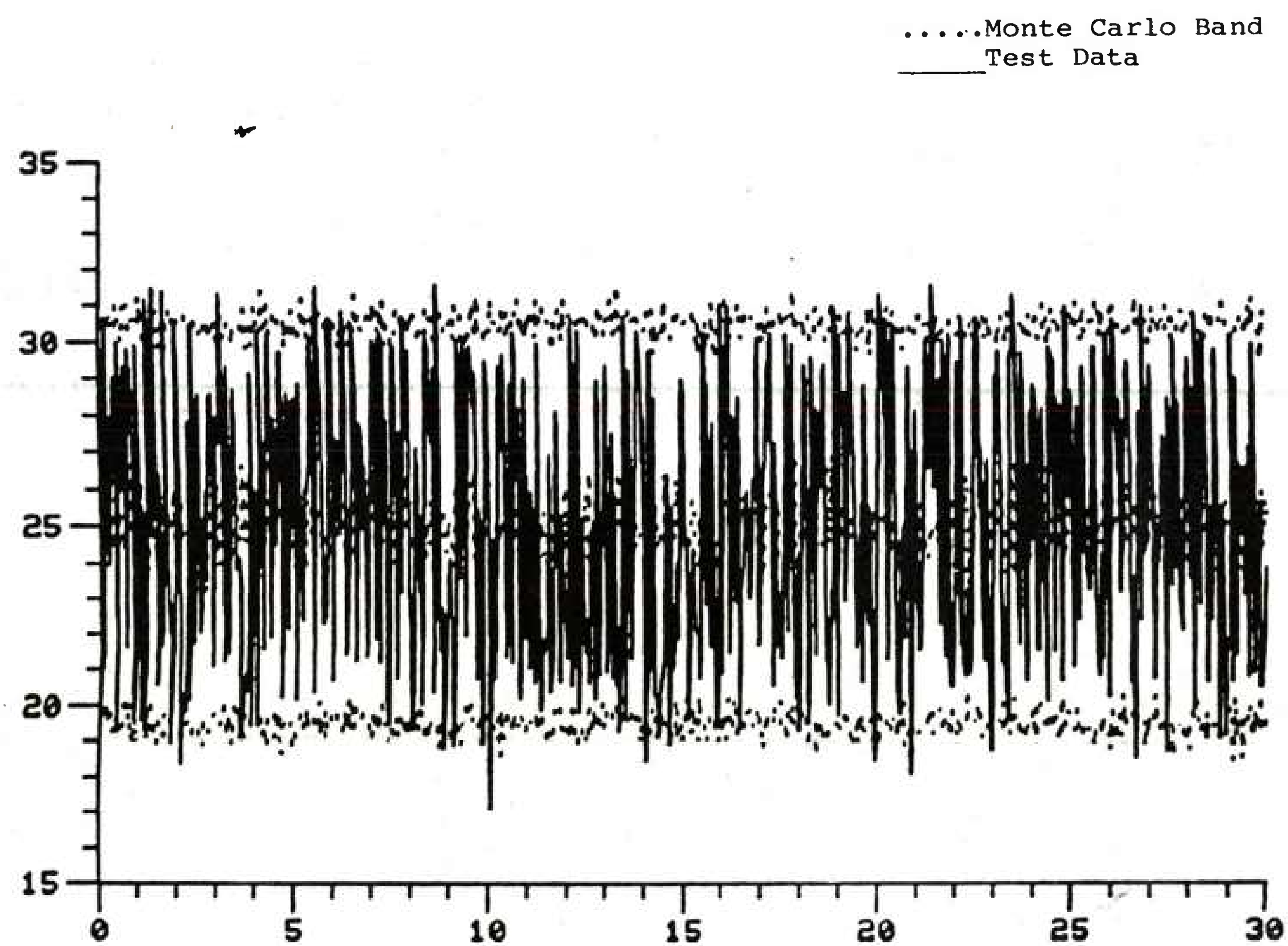


Figure 2.1

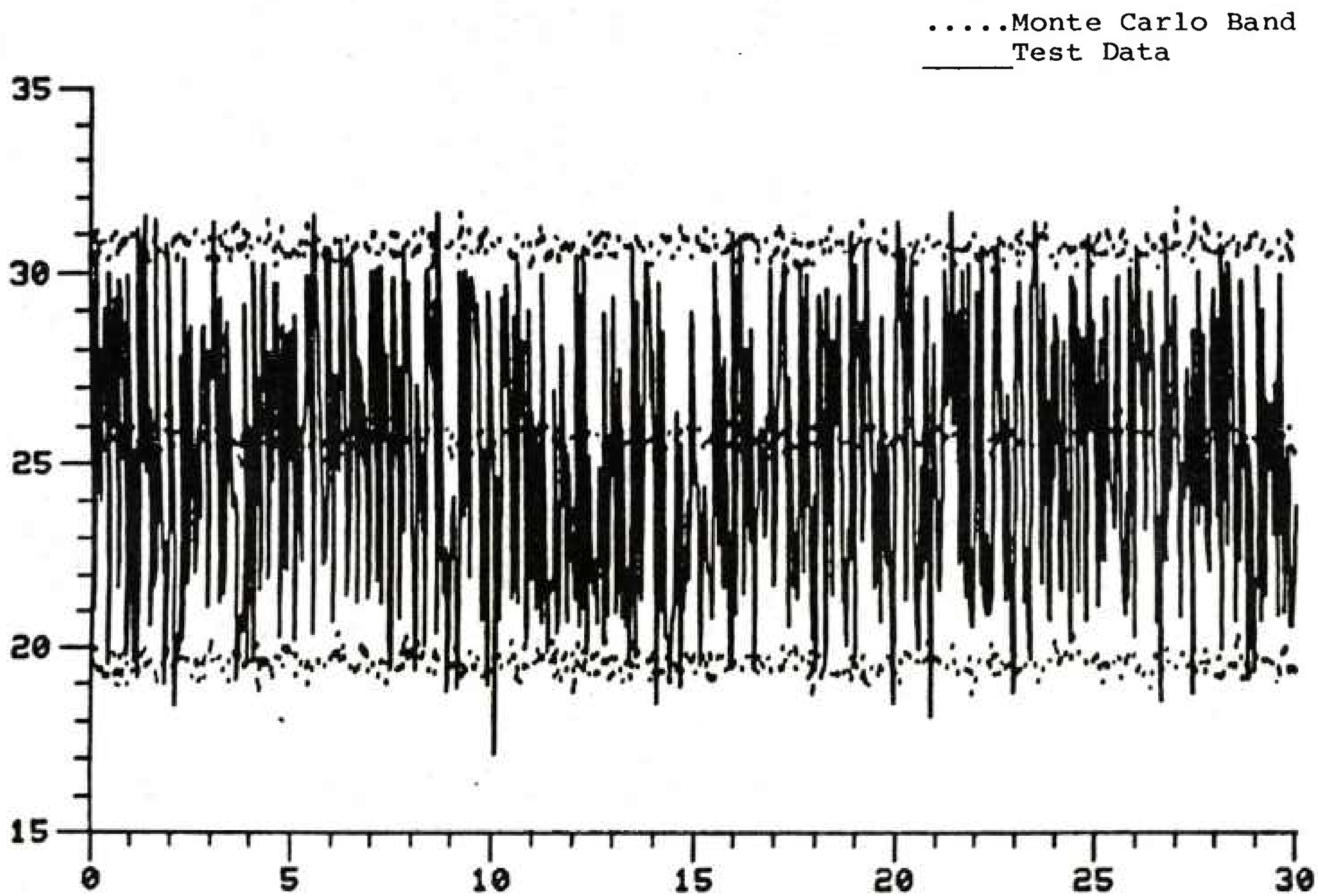


Figure 2.2

the data from the median. The trend, T, is defined as the average value of D(t) over the interval being considered. That is

$$T = \frac{1}{T_2 - T_1} \int_{T_1}^{T_2} D(t)\, dt.$$

Spread is a measure of the variability of D(t). It is defined by the equation

$$S = \frac{1}{T_2 - T_1} \int_{T_1}^{T_2} |D(t) - T|\, dt. \qquad (2.1)$$

If X(t) is the field test data plotted with the Monte Carlo bands, then the trend measures the average position of the data curve relative to the median curve. It is intended to detect systematic errors. The spread measures the variability of the data, relative to the median. It is intended to detect noise levels which are too high or too low.

In addition to their sensitivity to particular characteristics of the data, these particular parameters were chosen because they relate to the Monte Carlo band plots. Mean and variance as well as many other choices of parameters which also satisfy these criteria are available and could be used in their place. Similarly, curves other than the median could be used in these procedures. The median was chosen for this study for the following reasons:

- As the 50th percentile it is consistent with the use of percentile curves to bound the Monte Carlo bands.
- It is guaranteed to lie within the Monte Carlo bands.
- It is insensitive to outliers.
- The authors like it.

Assuming that the simulation exercise consisted of 100 Monte Carlo replications, we repeat the above procedure 100 times replacing the field test data, in turn, by the data generated by each of the 100 Monte Carlo replications. We now have a population of 100 ordered pairs (trend, spread) which we plot on a two-dimensional grid. On this plot we draw a pair of lines parallel to the vertical axis which forms a "strip" containing 95 percent of the points (2.5 to 97.5 percentile). A similar "95 percent strip" is formed from a pair of lines parallel to the horizontal axis. The intersection of these two strips forms a rectangular box. We now plot the ordered pair representing the test data on the same grid. If this point falls inside the box we conclude that the test data are indistinguishable from a member of the Monte Carlo population; i.e., the trend and spread of the model agree very closely with that of the test data. If the point falls outside the box we reject this conclusion. Note that it is possible for either the trend or the spread to be correctly modeled without the other being correct (see example 3).

The significance of the test for either trend or spread is simply $\beta = .05$. In the case of linear dependence the significance will again be $\beta = .05$. In the case of independence the significance will be $1-(1-\beta)^2 = .0975$. In any case other than these two extremes the significance level will lie somewhere between these two values. The question of significance level is not trivial; good estimates will require further research.

Figures 2.3 and 2.4 show the trend/spread plots corresponding to Figures 2.1 and 2.2, respectively. The diamond indicates the trend and spread of the data (which is the solid line in Figures 2.1 and 2.2 while the crosses correspond to the trend and spread of the 100 replications of the simulation. The difference in the position of the two diamonds relative to the rectangular boxes is quite apparent. This example is purely theoretical but it does illustrate the extra information available via the trend/spread plot, i.e. using Figures 2.3 and 2.4 one can immediately distinguish between the two models that produced the Monte Carlo bands which appear in Figures 2.1 and 2.2.

While not quite so striking, similar examples were found in the practical application of this method. Example 2 is presented in Figures 2.5 through 2.8. (Figure 2.6 is the trend/spread plot which corresponds to Figure 2.5 while Figure 2.8 is the trend/spread plot which corresponds to Figure 2.7). Each data curve appears to have approximately the same variation as the corresponding set of Monte Carlo bands, but the trend/spread plots show that such visual impressions can be misleading.

Like example 2, examples 3 and 4 were taken from actual field test data and corresponding simulation models. Example 3 consists of Figures 2.9 through 2.11. Figure 2.9 presents the data and the Monte Carlo bands (5 percent, median, 95 percent). In Figure 2.11 the data point lies in one strip but not in the other indicating that the trend is modeled correctly but that the spread is slightly low (i.e., the model is not noisy enough).

Figure 2.10 is obtained from Figure 2.9 by fitting a smooth curve through the Monte Carlo median and then subtracting this smooth curve from all four curves. (A smooth curve was chosen instead of the median itself in order to insure that any noise occurring in the median would not be transferred to the data.) The curves formed in the process of creating Figure 2.10 were not used in the trend/spread calculation, but rather were developed to obtain better insight into the situation depicted in Figure 2.9. The scale in Figure 2.9 is so large that it is difficult to compare the data with the bands. However, the much smaller scale of Figure 2.10 allows a more meaningful visual

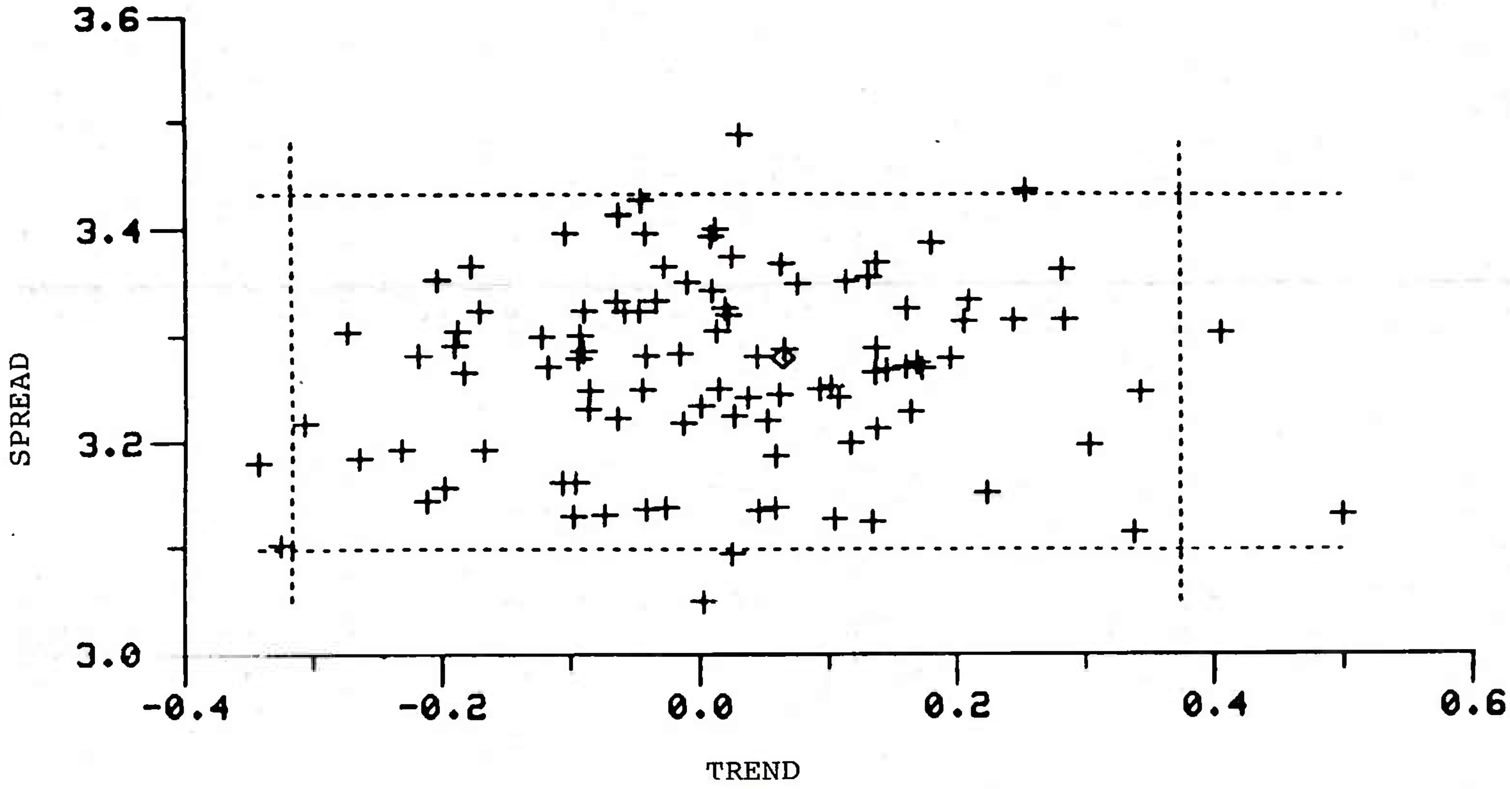


Figure 2.3

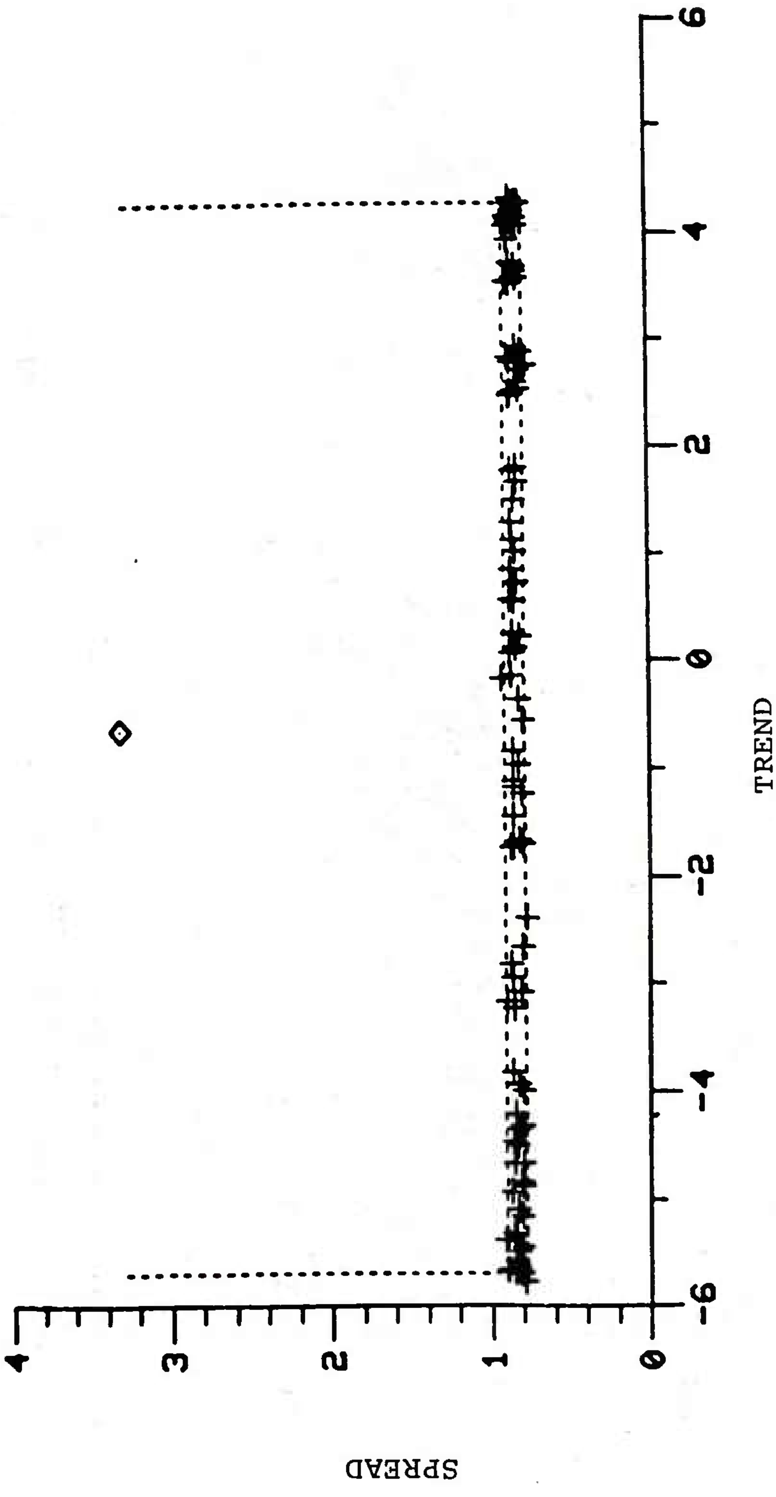


Figure 2.4

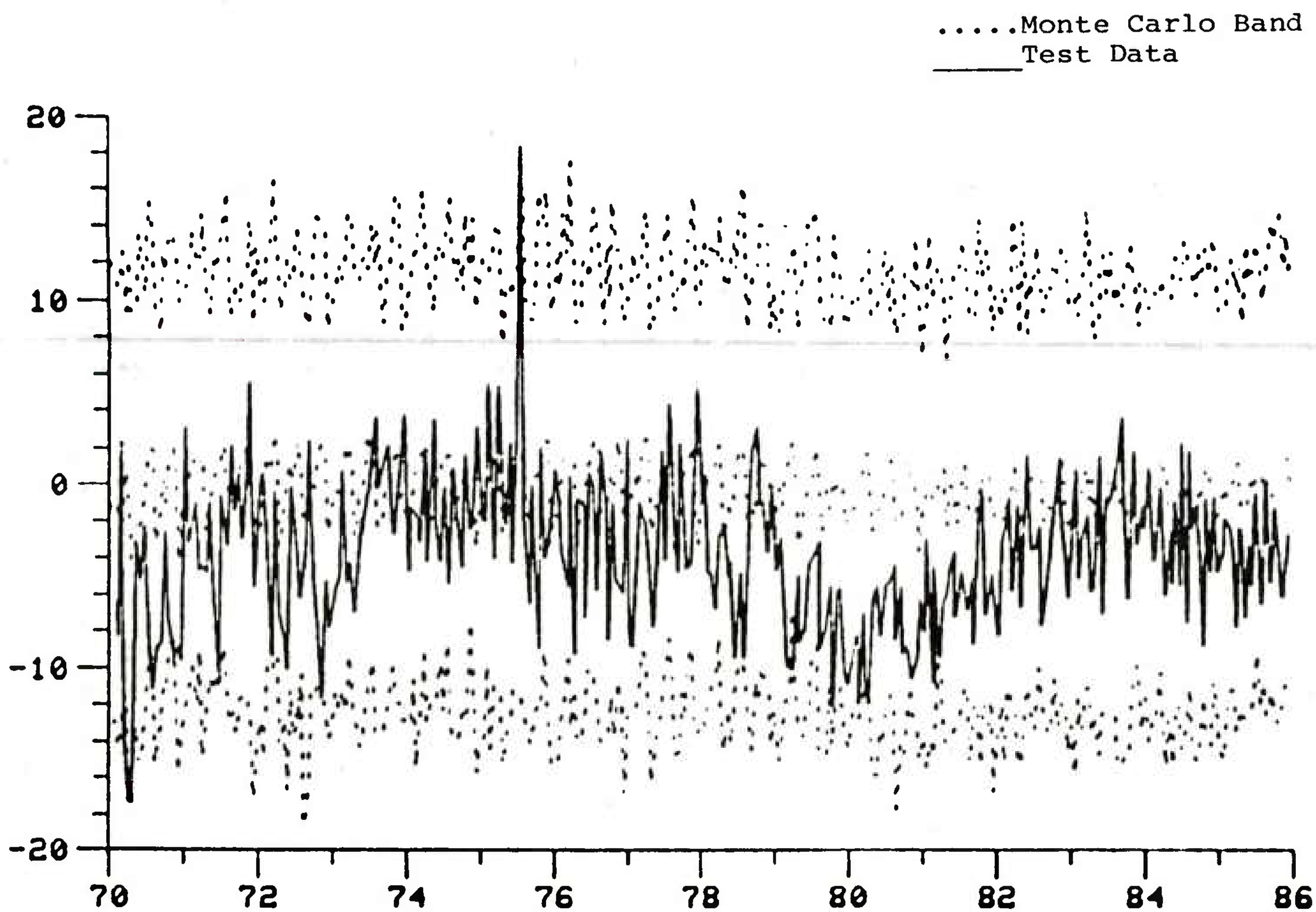


Figure 2.5

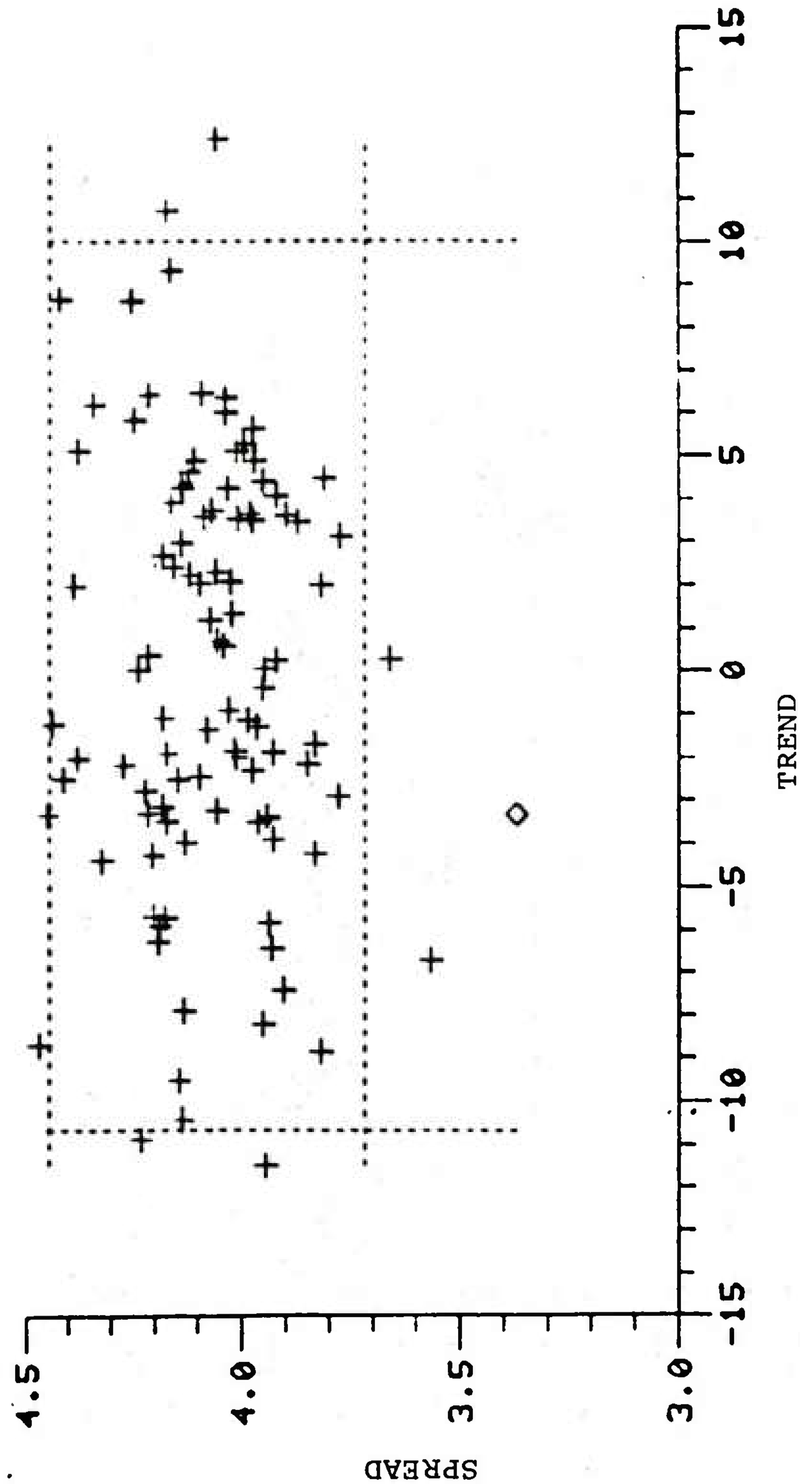


Figure 2.6

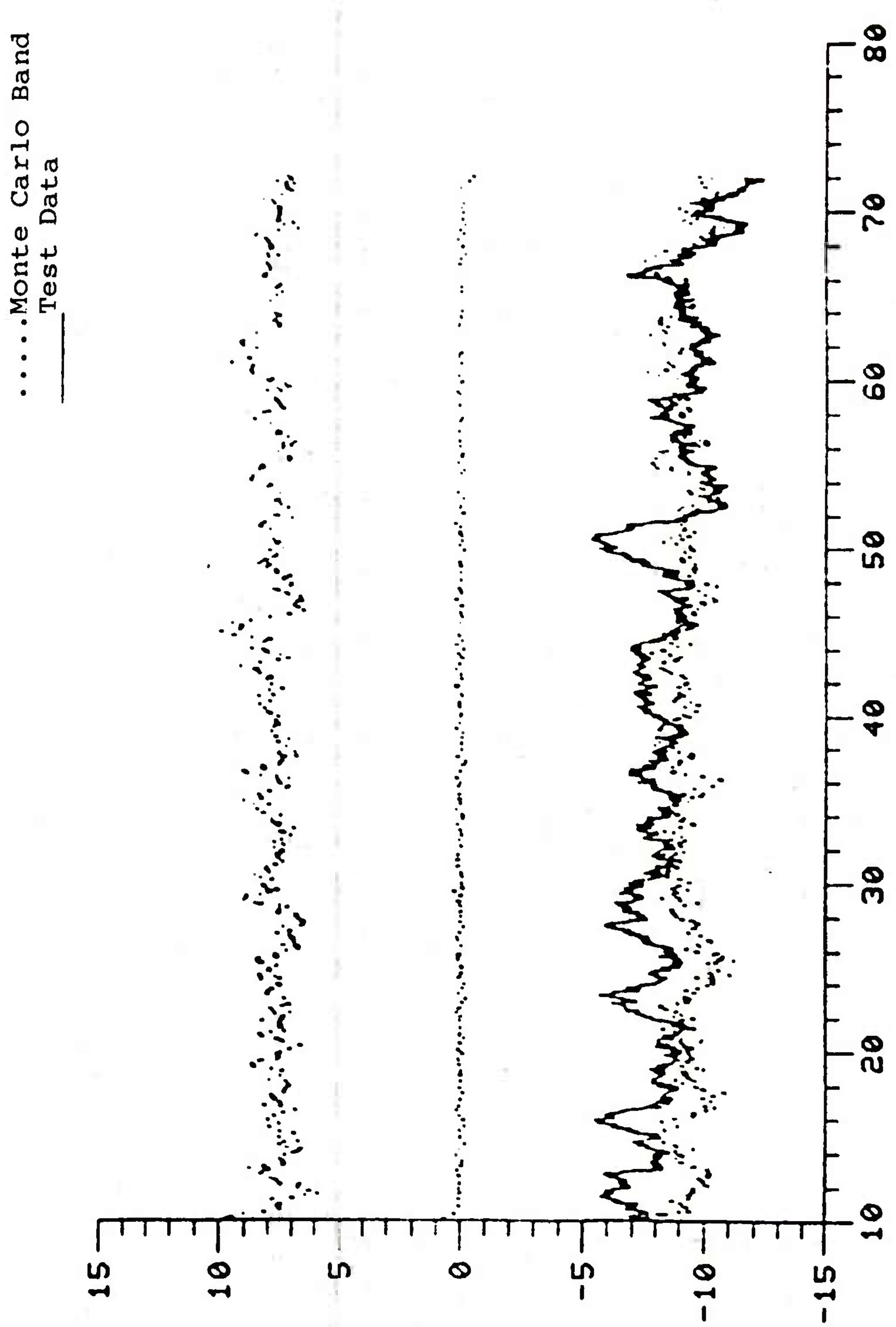


Figure 2.7

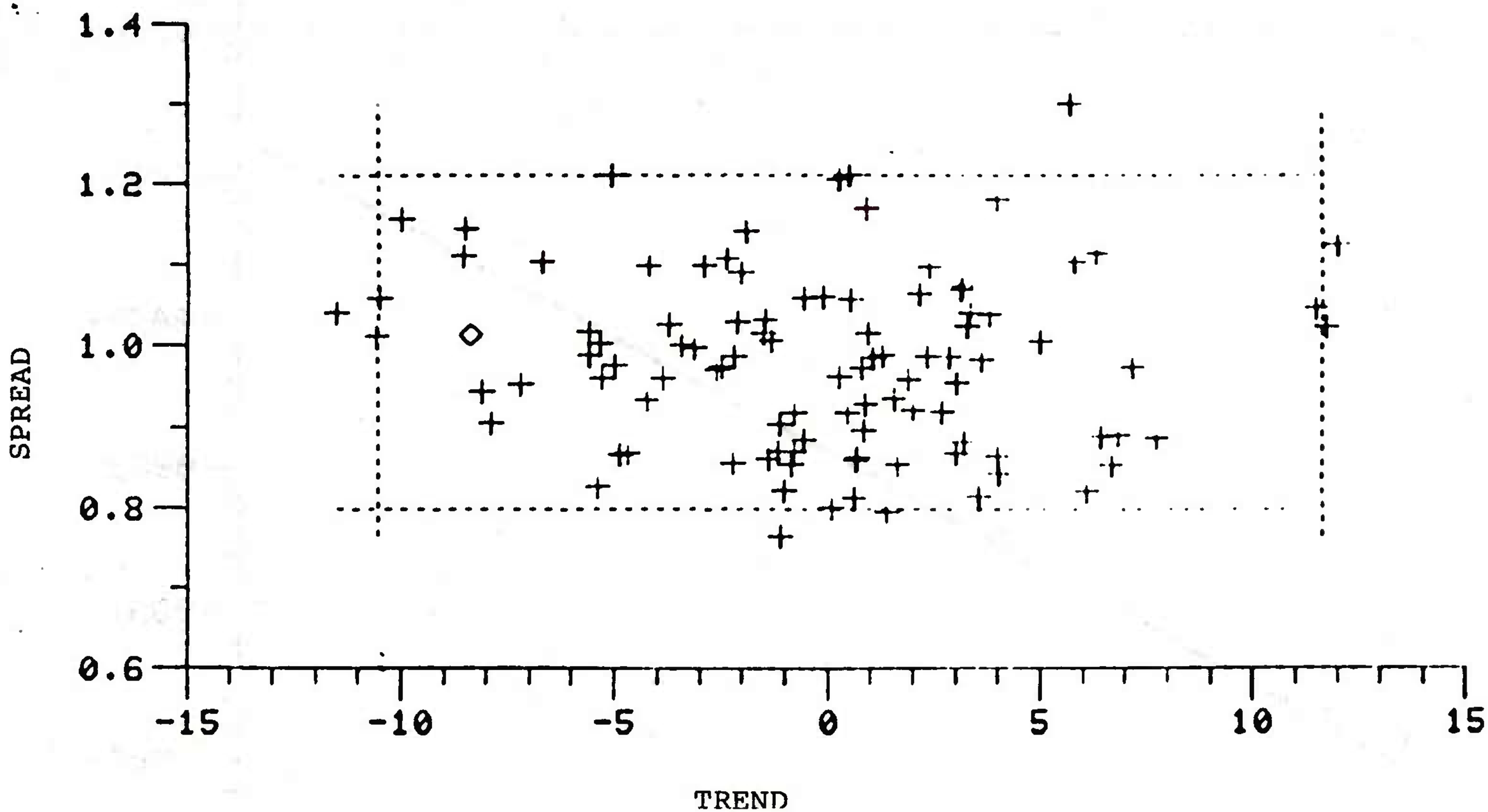


Figure 2.8

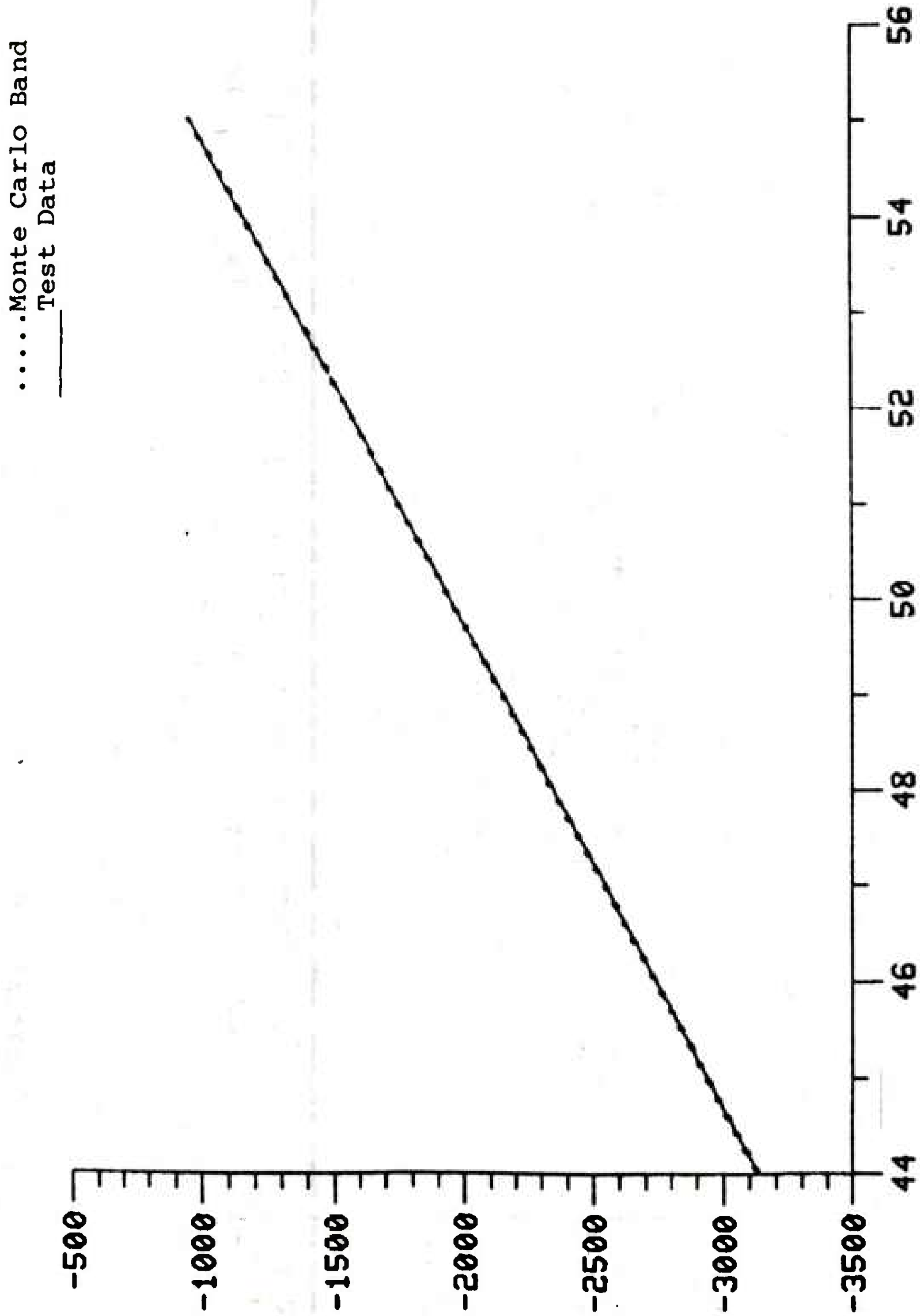

.....Monte Carlo Band
____Test Data
-500
-1000
-1500
-2000
-2500
-3000
-3500
44
46
48
50
52
54
56

Figure 2.9

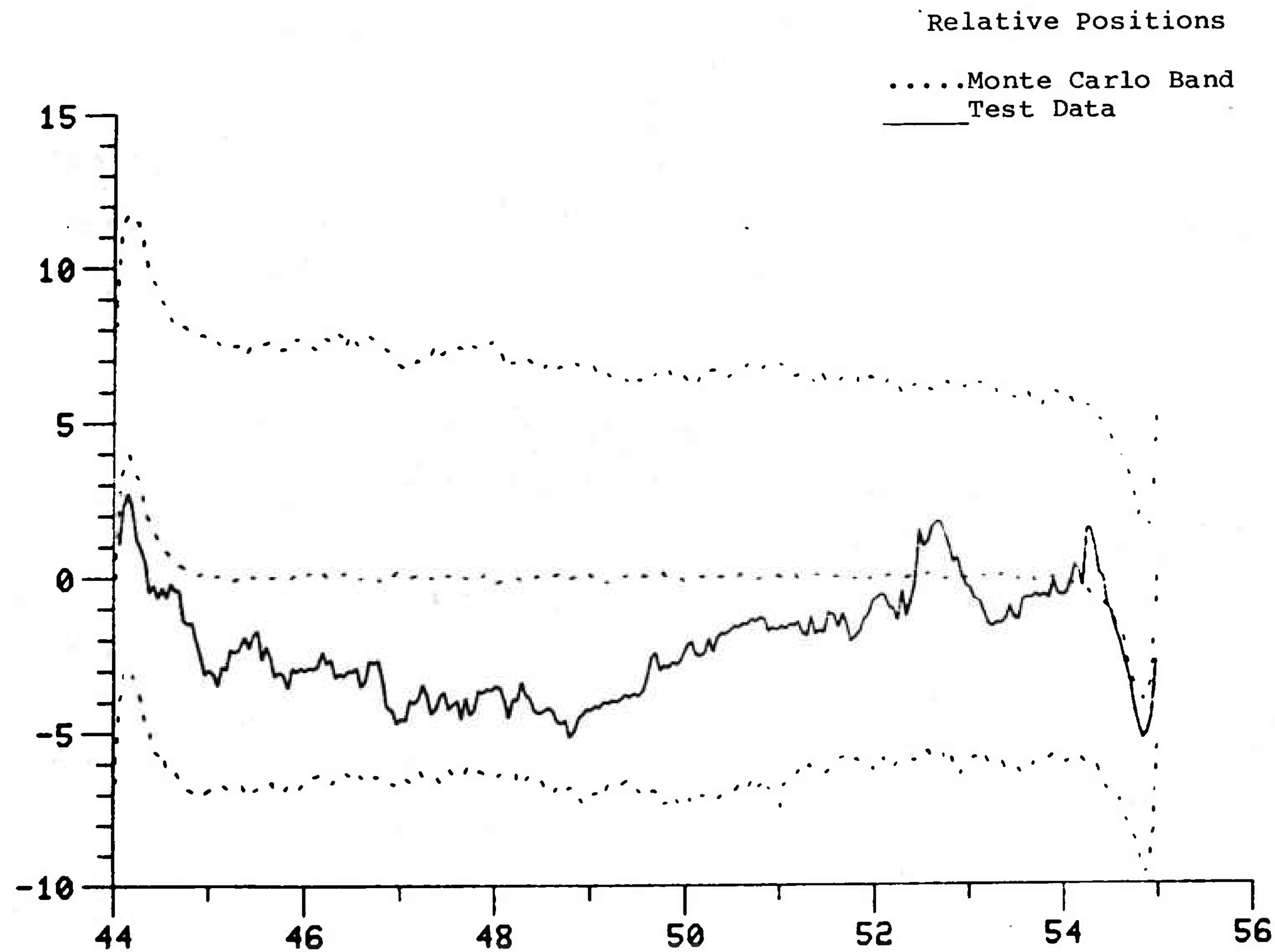

Relative Positions
.....Monte Carlo Band
_____Test Data
15
10
5
0
-5
-10
44
46
48
50
52
54
56

Figure 2.10

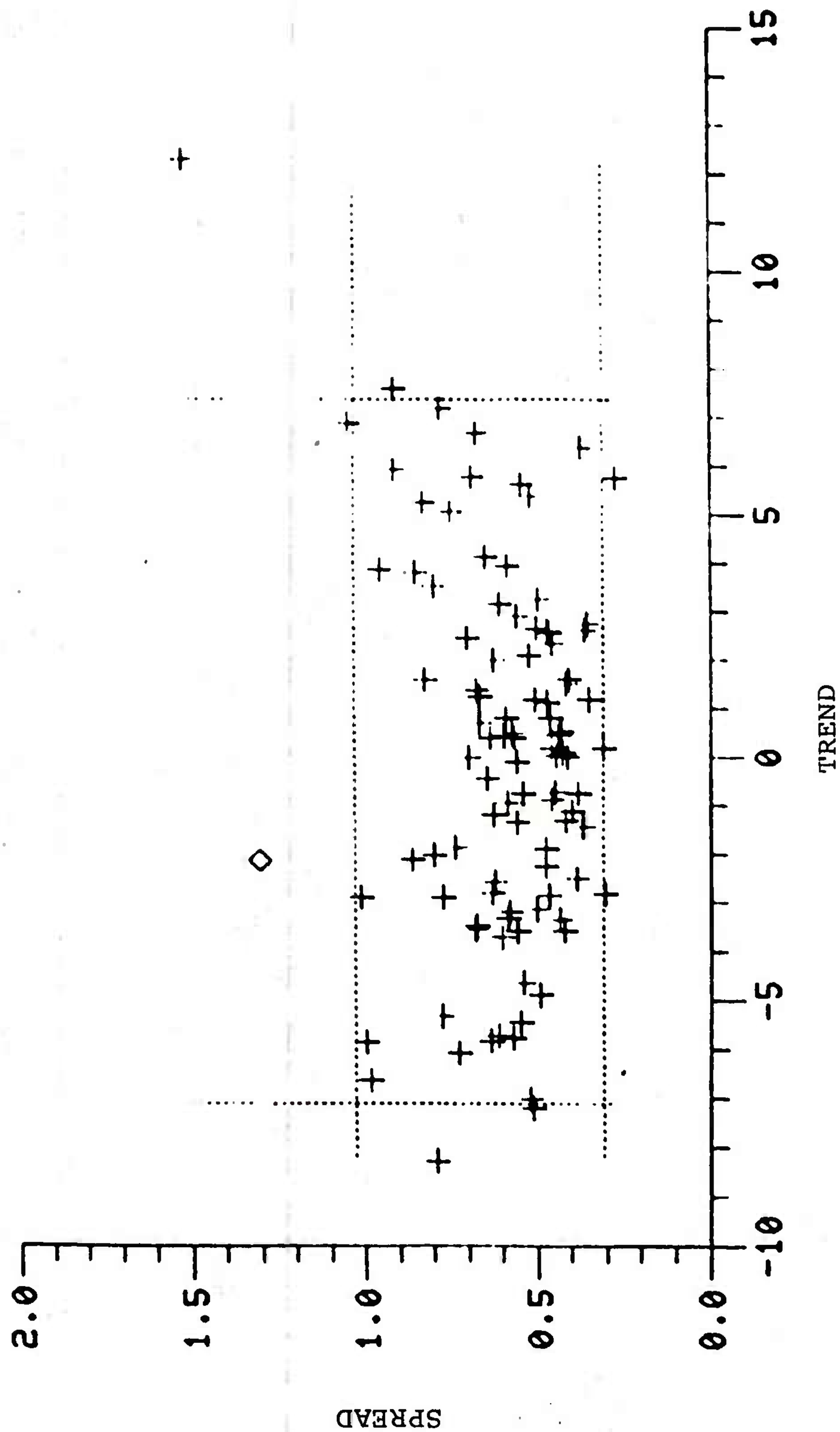


Figure 2.11

evaluation of the situation. Note that the solid line in Figure 2.10 represents not the data but rather the difference between the data and the Monte Carlo median. Also note that due to the fitting of the smooth curve through the Monte Carlo median the first few points on the plot and the last few points on the plot are not accurate and should be ignored.

The curves appearing in example 4 (Figures 2.12 through 2.14), which were obtained in a manner identical to that which produced the curves of example 3, illustrate the case in which one would conclude that both the trend and the spread are correctly modeled.

Clearly the number of Monte Carlo replications (100) and the breadth (95 percent) of the strips which form the rectangular box were used for illustrative purposes only and are subject to the demands of the particular problem under investigation.

One drawback of spread as a measure is that it is insensitive to frequency variation. For example, on the interval $[0,\Pi]$ both the functions $U(t) = \sin(t)$ and $V(t) = \sin(2t)$ have the same value for spread,

$$\frac{1}{\Pi} \int_0^{\Pi} |U(t)| dt = \frac{2}{\Pi} = \frac{1}{\Pi} \int_0^{\Pi} |V(t)| dt, \qquad (2.2)$$

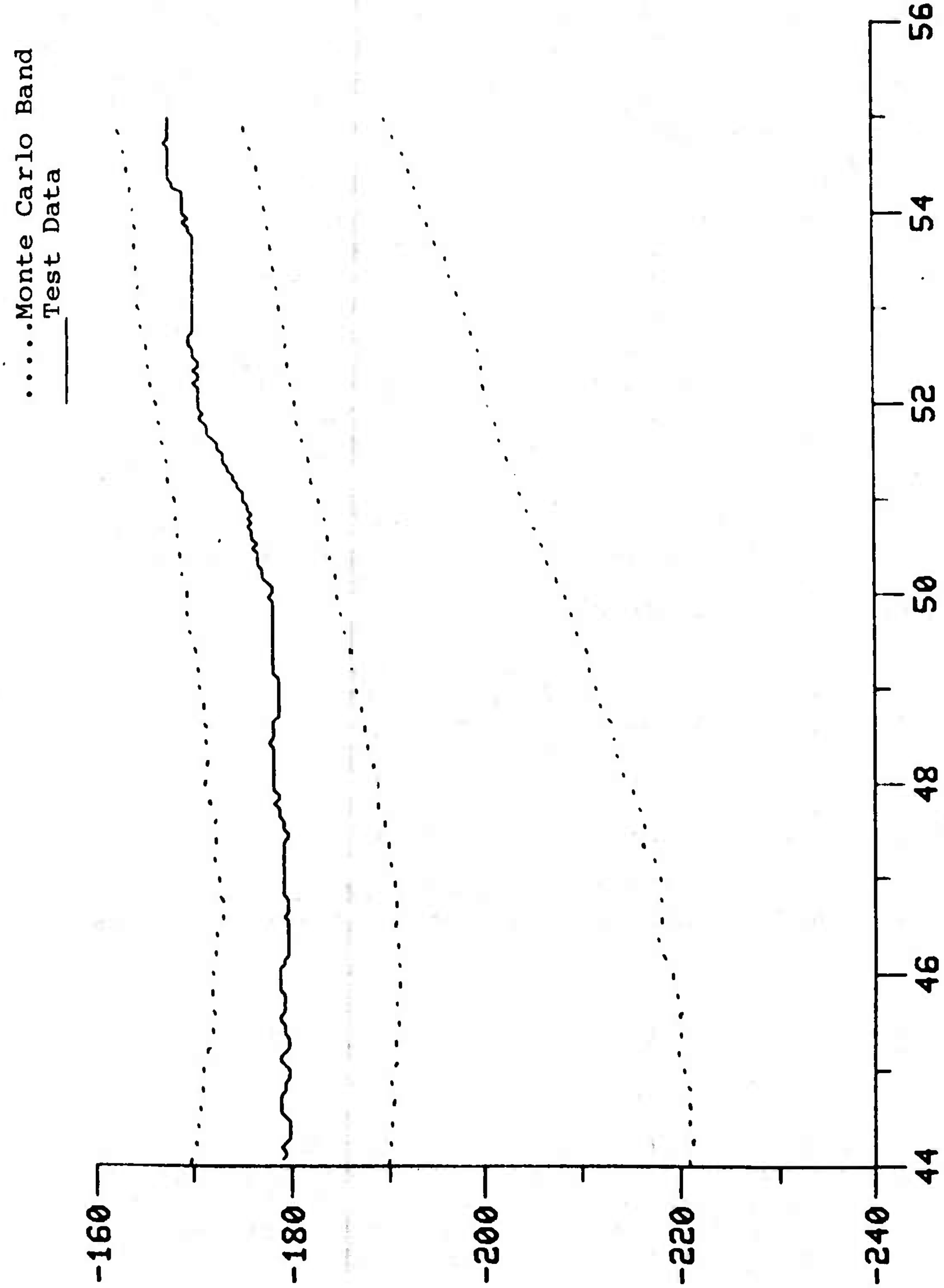
while their frequencies differ by a factor of 2.

A functional, similar in appearance to (2.1), which provides a means of distinguishing between these two cases is

$$\frac{1}{T_2-T_1} \sum_{i=1}^{n} |[f(t_i)-M(t_i)] - [f(t_{i-1})-M(t_{i-1})]|. \qquad (2.3)$$

Except for the factor $1/(T_2-T_1)$, (2.3) is nothing more than the total variation function of Lebesque measure theory (see for example Reference 4). Expression (2.3) can be thought of as the average rate of change or the average slope of $f(t)-M(t)$ over the interval $[T_1,T_2]$ since (2.3) corresponds to

$$\frac{1}{T_2-T_1} \int_{T_1}^{T_2} |f'(t) - M'(t)| \, dt$$



.....Monte Carlo Band
___ Test Data
-160
-180
-200
-220
-240
44
46
48
50
52
54
56

Figure 2.12

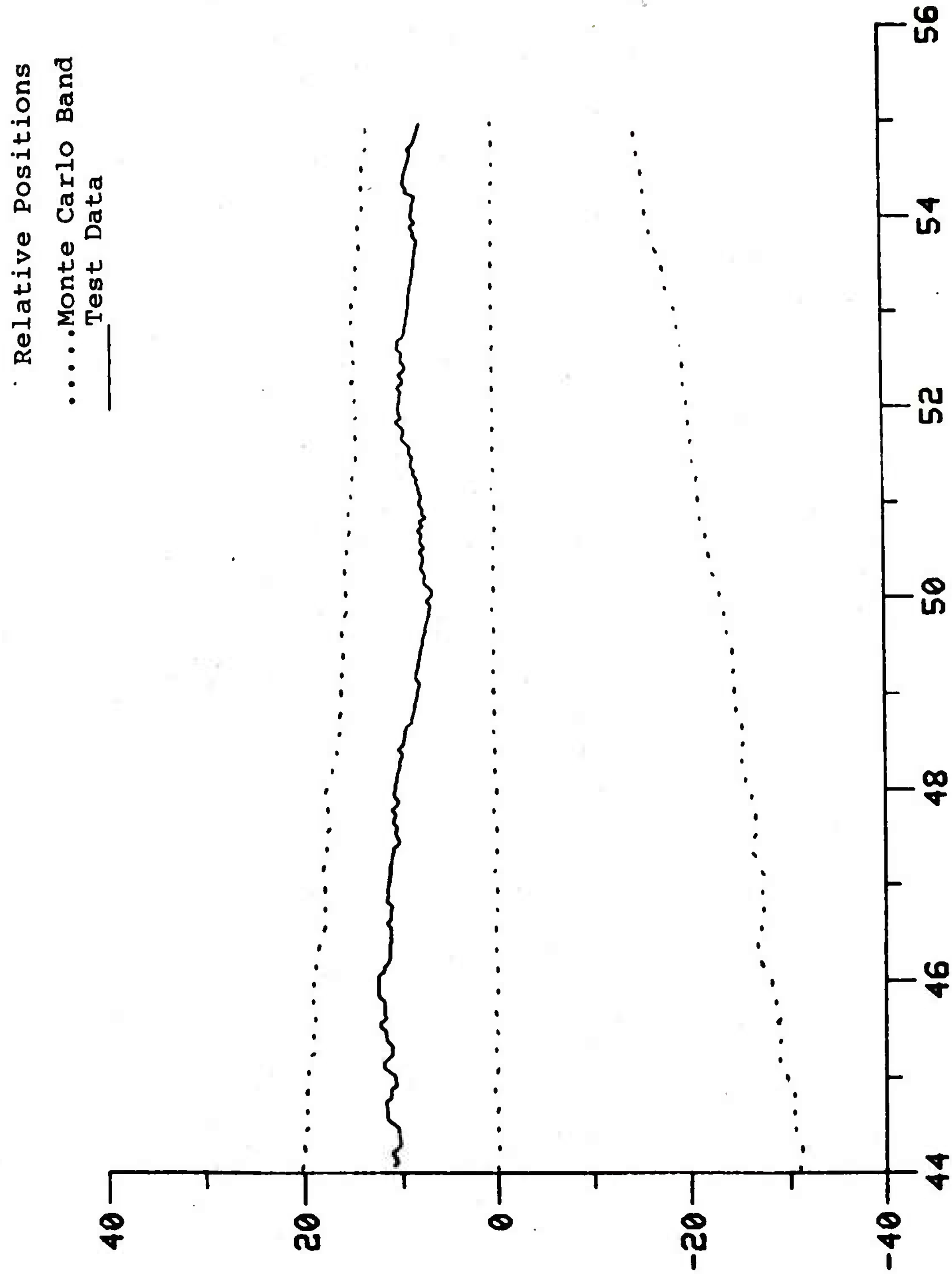

Relative Positions
.....Monte Carlo Band
___ Test Data
40
20
0
-20
-40
44
46
48
50
52
54
56

Figure 2.13

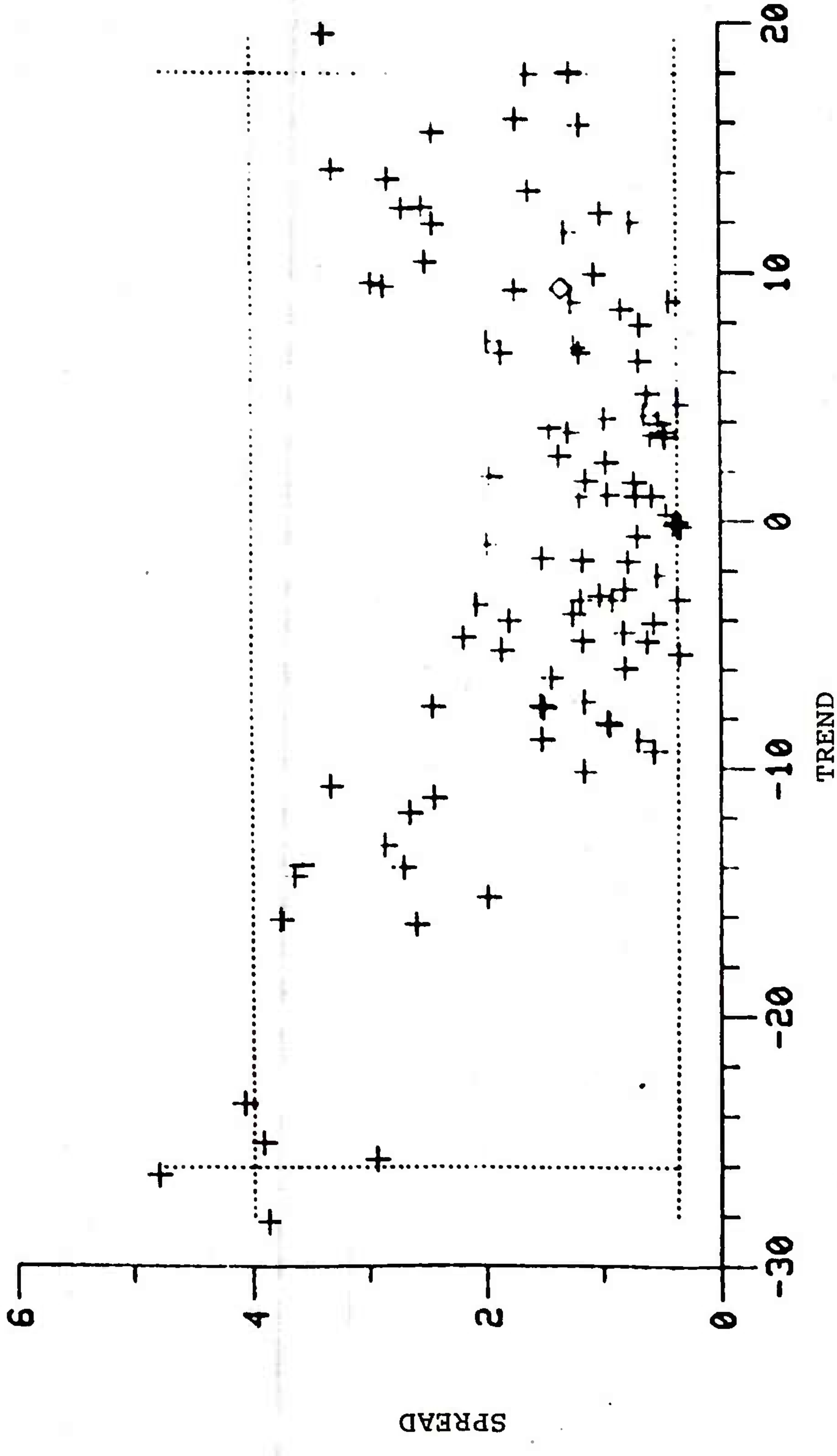


Figure 2.14

in the continuous case. Referring to (2.2) we see that this new measure of variation performs as claimed by noting that

$$\frac{1}{\pi} \int_0^{\pi} |U'(t)| \, dt = \frac{2}{\pi},$$

while

$$\frac{1}{\pi} \int_0^{\pi} |V'(t)| \, dt = \frac{4}{\pi}.$$

One can use this measure instead of or in addition to spread in the analysis procedure described in this section. The authors are investigating this possibility in ongoing work.

Next page is blank

3. CONCLUSIONS

The techniques of Monte Carlo band analysis and trend/ spread plots described in this report have been successful in identifying even fairly subtle flaws in simulation models (as in example 1 of Section 2.2). This confirms the fundamental value of these techniques.

From this point one can employ these methods as one part of a detailed iterative model design/validation procedure in which after a flaw has been discovered the model is suitably altered, exercised, and compared again to the test data. This process may be repeated as often as necessary in order to obtain a valid model.

We envision that as these techniques are used in conjunction with an increasing number of different types of models the scope of applications of the techniques will grow accordingly.

Next page is blank

REFERENCES

1. Meredith, J. L., and Ball, D., Validation of the Fire Unit Effectiveness Model, Report No. R-RR-S-5-96-73, December 1973, US Army Armament Command, Dover, NJ 07801, unclassified.

2. Meredith, J.L., Scheder, R.A., and Lufkin, B.M., Evaluation of the Gun Low Altitude Air Defense Fire Control Test Bed, Technical Report No. 149, April 1977, US Army Materiel Systems Analysis Activity, APG, MD 21005, unclassified.

3. Gades Group, Gun Air Defense Effectiveness Study, Volume II, Report Number R-TR-74-017, US Army Armament Command, Rock Island, IL 61201.

4. Royden, H.L., Real Analysis (2nd Edition), New York, Mcmillan, 1968.

Next page is blank

DISTRIBUTION LIST
TR-346

No. of Copies	Organization
1	Commander US Army Materiel Development and Readiness Command ATTN: DRCDM-P 5001 Eisenhower Avenue Alexandria, VA 22333
1	Commander US Army Materiel Development and Readiness Command ATTN: DRCDE-SD 5001 Eisenhower Avenue Alexandria, VA 22333
2	Commander Defense Technical Information Center ATTN: TCA Cameron Station Alexandria, VA 22314
2	Chief Defense Logistics Studies Information Exchange US Army Logistics Management Center ATTN: DRXMC-D FT Lee, VA 23801
1	HQDA (SAUS-OR) WASH DC 20310
1	HQDA (DAMO-RQA) WASH DC 20310
2	Commander US Army Concepts Analysis Agency ATTN: CSCA-MR CSCA-ADL 8120 Woodmont Avenue Bethesda, MD 20014
1	Commander US Army Operational Test and Evaluation Agency ATTN: CSTE-STS-L 5600 Columbia Pike Falls Church, VA 22041
5	Commander US Army Armament Research and Development Command ATTN: DRDAR-SE DRDAR-LCS DRDAR-SCF DRDAR-SCS DRCPM-ADG Dover, NJ 07801
2	Commander US Army Aviation Research and Development Command ATTN: DRDAV-B DRDAV-E PO Box 209 St. Louis, MO 63166
1	Commander US Army Electronics Research and Development Command ATTN: DRDEL-ST-SA (A. Sambuco) 2800 Powder Mill Road Adelphi, MD 20783
1	Commander US Army Troop Support and Aviation Materiel Readiness Command ATTN: DRSTS-BA 4300 Goodfellow Boulevard St. Louis, MO 63120

DISTRIBUTION LIST (Continued)
TR-346

No. of Copies	Organization
2	Commander US Army Combat Developments Experimentation Command ATTN: ATEC-PL-M ATEC-PL-TL FT Ord, CA 93941
2	Director US Army TRADOC Systems Analysis Activity ATTN: ATAA-PL ATAA-TG White Sands Missile Range, NM 88002
1	Commandant US Army Air Defense School ATTN: ATSA-CD-A FT Bliss, TX 79916
1	Commander US Army Aviation Center and Fort Rucker ATTN: ATZQ-CD FT Rucker, AL 36362
1	President US Army Air Defense Board ATTN: ATZC FT Bliss, TX 79916
2	Commander US Army White Sands Missile Range ATTN: STEWS-TE-ML STEWS-NR-PA White Sands Missile Range, NM 88002
1	Commander US Army Tank-Automotive Command ATTN: DRDTA-V Warren, MI 48090
1	Commander US Army Missile Command ATTN: DRSMI-DS Redstone Arsenal, AL 35809
1	Commander US Army Armament Materiel Readiness Command ATTN: DRSAR-IEP-L Rock Island, IL 62199
1	AFELM, The Rand Corporation ATTN: Library-D 1700 Main Stree Santa Monica, CA 90406

Aberdeen Proving Ground

No. of Copies	Organization
3	Cdr, USATECOM ATTN: DRSTE-CT DRSTE-AD DRSTE-CM
1	Dir, HEL ATTN: DRXHE-SP
1	Dir, BRL ATTN: DRDAR-BLB
7	Dir, AMSAA ATTN: DRXSY-A (D. O'Neill) DRXSY-AD (H. Peaker) DRXSY-AA (T. Coyle) DRXSY-ADG (J. Meredith) DRXSY-ADS (D. Nuzman) DRXSY-ADG (J. Wald) DRXSY-A (Editorial Ofc)