X-551-69-493
PREPRINT

NASA TM X- 63773

AN ANALYTIC MONTE CARLO APPROACH TO PARAMETRIC STUDIES

P. ARGENTIERO

NOVEMBER 1969

GSFC GODDARD SPACE FLIGHT CENTER
GREENBELT, MARYLAND

X-551-69-493
Preprint

AN ANALYTIC MONTE CARLO APPROACH
TO
PARAMETRIC STUDIES

P. Argentiero

November 1969

GODDARD SPACE FLIGHT CENTER
Greenbelt, Maryland

AN ANALYTIC MONTE CARLO APPROACH TO PARAMETRIC STUDIES

P. Argentiero

SUMMARY

The following is a frequently occurring problem. A random variable Y is assumed to be a deterministic function of a one by N vector of random variables. From information on the multivariate statistical distribution of X, it is desired to answer questions concerning the statistical distribution of Y. If computational problems are surmountible, the most satisfactory solution to the problem is a Monte Carlo one. If this approach is too expensive, then another approach discussed in this paper should be considered. It involves obtaining an N-dimensional polynomial fit to data obtained from the simulation program connecting the input variables to the output variable, and obtaining Monte Carlo samples from the fitted polynomial instead of the simulation program. A drastic reduction in computer time is the usual result. This procedure also makes it quite easy to arrange the input variables in a hierarchy with regard to their impact on the distribution of the output variable.

This method was applied to obtain a histogram of the distribution of a minimum fuel midcourse velocity correction of a Venus-72 mission (a mission with a goal to orbit a probe around Venus). It is also shown that the down-track and radial-track velocity injection errors are the most significant injection errors with regard to their influence on the size of the midcourse correction of such a mission.

PRECEDING PAGE BLANK NOT FILMED.

CONTENTS

ILLUSTRATIONS

AN ANALYTIC MONTE CARLO APPROACH TO PARAMETRIC STUDIES

INTRODUCTION

The following is a generic problem which appears in many diverse areas of technology. A random variable Y is assumed to be a deterministic function of a one by n vector of random variables X. From information on the multivariate statistical distribution of X, it is desired to answer questions concerning the statistical distribution of Y. It is also frequently useful to be able to arrange the components of X into a hierarchy with regard to their impact on the distribution of Y or to perform a parametric study by obtaining estimates of the probability of Y exceeding a given critical value when certain components of X are fixed at certain values. If the functional relationship between Y and X is given in closed form, then in principle a closed form representation of the distribution of Y can be constructed from the given distribution of X. In practice, however, this fact is usually not relevant since in most cases either the closed functional form in question is too complicated to permit an analytic representation of the distribution of Y or the functional relationship is not given in closed form at all but in terms of a computing algorithm (i. e., a computer program). If computational problems are surmountible, the most satisfactory solution to the problem is a Monte Carlo one. The technique consists of repeatedly sampling from the given multivariate distribution for X and computing for each sampled value of X, the corresponding value for Y. The Y values are arranged in a histogram. The histogram obtained is then considered as an approximation to the true distribution of Y. The error produced in substituting the histogram for the true distribution for Y varies inversely as the square root of the sample size. Thus the demand for high confidence levels can lead to the necessity of very large sample sizes. In many cases, this becomes quite expensive in terms of computer time. The expense is even higher when parametric studies of the sort discussed above are required since this involves a repeated application of the entire Monte Carlo process.

When the Monte Carlo approach appears too expensive the root sum square procedure is sometimes utilized. In root sum square studies, individual one-sigma perturbations on the components of X are introduced into the functional relation, and the other components of X are set at zero. The resultant effects on Y are root-sum-squared to obtain the deviation on Y. If the investigator is willing to assume that Y is normally distributed, then, of course, the distribution of Y would be completely determined. This procedure is quite convenient but it relies on assumptions of linearity and statistical independance which frequently are not satisfied.

Recent authors have directed attention toward obtaining techniques for error analysis and parametric studies which avoid both the expense of standard Monte Carlo methods and the often times vitiating assumptions of normality, linearity, and statistical independence which accompany the root-sum-square technique.

Merel and Mullin (1) for example utilize covarience matrix analysis to separate linear and non-linear error sources and, after the effect of linear terms is computed, the effect of non-linear errors is obtained in a Monte Carlo simulation. Logsdon and Africano (2) make a series of perturbations on each individual input variable of a simulation. The effects on the output variable are then fit with a sequence of polynomial curves. Monte Carlo samples are then taken from the polynomials instead of the simulation program. Very accurate histograms and a drastic reduction in machine time are the reported results. Another advantage to this approach which should attract attention is that individual effects of input variables can be isolated and studied thus making parametric studies both cheap and convenient. For these reasons, further elaboration and improvement of the technique are offered in the present paper.

In order to exhibit the power and convenience of the present version of this polynomial fit-Monte Carlo approach, the technique will be utilized to obtain a parametric study of the effect of injection errors on the important first midcourse correction of a Venus-72 mission, a mission with a goal to orbit a probe around Venus during 1972. This example will also permit an opportunity to discuss some of the practical issues that occur in the application of this technique.

AN ANALYTIC MONTE CARLO PROCEDURE

Monte Carlo methods are frequently employed when it is desired to find the simultaneous effect of several input statistical variables on an output variable. The result of the application of the Monte Carlo method is usually a histogram which approximates the distribution of the output variable. But since the accuracy of such a histogram varies directly as the square root of the sample size, Monte Carlo procedures can prove quite expensive. In space work, for instance, high confidence levels are frequently demanded. Such demands can typically lead to Monte Carlo samples in excess of 10,000. If the simulation program is a time consuming one, then such sample sizes cause practical difficulties. In (2), Logsdon and Africano suggest a compromise procedure which in many cases produce histograms very similar to those produced by an honest Monte Carlo procedure but at a small fraction of the cost. This compromise procedure rests on the following analysis.

Let the output variable Y be a function of n random variables $x_1, x_2, \ldots x_n$. Thus $Y = f(x_1, x_2, \ldots x_n)$. A first order Taylor series expansion gives $Y = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} x_i$ where the expansion takes place about the origin of the n dimensional domain space. If the partial derivatives are analytically independent, the equation can be written $Y = \sum_{i=1}^{n} \Delta Y x_i$ where each term is a function only of its subscripted variable and hence can be expanded as a polynomial. Thus $\Delta Y x_i = a_0 + a_1 x_i + a_2 x_i^2 + \ldots a_k x_i^k$. The coefficients can be determined by making separate perturbations on the individual x_i 's while maintaining the other x_j 's,

$j \neq i$ at their zero values. Least square polynomial curve fits are then performed on the results.

Once the curve fits have been performed, the Monte Carlo samples of the x_i 's are substituted into the appropriate polynomial rather than the simulation program. The resultant functional values are then summed to obtain a sampled value of Y. These values are then arranged in a histogram which hopefully approximates the histogram which would have been obtained had a true Monte Carlo process been performed. And since the evaluation of polynomials is usually orders of magnitude faster than, say, the numerical integration of a long trajectory, in many situations a considerable saving in computer time is realized.

An important feature of this technique is that it conveniently permits an isolation of the effect of each input random variable on the statistics of the output variable. In fact, if one assumes that the input variables are normally distributed and statistically uncorrelated then the individual contributions of the input variables to the statistical mean and deviation of the output variable can be calculated directly from the coefficients of the least square polynomials. How these calculations are carried out is demonstrated in (2). If the input variables are significantly correlated or not normally distributed, then parametric studies can still be performed quite cheaply. One such method for the performance of parametric studies in the presence of significant correlations will be exhibited in a later section.

THE EFFECT OF ANALYTIC CORRELATION

The above analysis is contingent on the assumption that the partial derivatives in the Taylor series expansion of the simulation function are each functions of just one variable. This is a strong assumption and it is easy to see how its acceptance in some situations could lead to bad results. The essence of this assumption is that the output variable represents the sum of the effects of the individual input variables. But frequently pairs of input variables have a coupled or covarient effect on the output variable which can not be represented as a sum of their individual effects. This coupling, when it exists, will manifest itself in a mutual dependence of the associated partial derivatives. (It is analytic correlation which is under discussion here. This should not be confused with a possible statistical correlation of the variables in question. The two types of correlation are quite unrelated.) Allowances can be made for this coupling effect in the following way. Suppose input variables x_i and x_j are suspected to have a coupled or correlated effect on the output variable. Their contribution can be represented as $\Delta Y x_i, x_j = \sum_{k=1}^{m-1} a_k x_i^{m-k} x_j^k$ The coefficients a_k are then determined by simultaneously perturbing x_i and x_j and performing the usual least squares fit on the perturbations of the output variable. For each Monte Carlo sample that is then produced, $\Delta Y x_i, x_j$ is summed with the other contributors to obtain the values of Y which are arranged in a histogram.

In principle, there is no reason for not also considering the effects of correlated triples of input variables and so forth. Such contributions could be obtained in much the same way as the effect of correlated couples is obtained. This is not recommended either in (2) or in the present study. The reason is that the inclusion of such correlations increases the number of parameters to be estimated thus possibly causing numerical difficulties. It also increases the number of simulations necessary in order to obtain the polynomial fits. Thus the possible gain does not appear to be worth its price.

THE QUESTION OF FEASIBILITY

Why should it be that under many circumstances the analytic Monte Carlo procedure outlined above can give essentially the same results as a standard Monte Carlo procedure but at a small fraction of the cost in terms of computer time? To understand the reason, it is useful to recognize that every Monte Carlo procedure that leads to a quantative result (say the evaluation of the 99% critical value of an output random variable) may be regarded as an estimation procedure for the value of a multiple integral. Suppose that n random samples are deemed sufficient to estimate a given parameter. The results will be a function $\theta(\epsilon_1, \epsilon_2, \epsilon_3, \dots \epsilon_n)$ of the random numbers $\epsilon_1, \epsilon_2, \dots \epsilon_n$ which were chosen for the Monte Carlo process. This is an unbiased estimate of the correct answer which can be represented as $\int_0^1 \cdots \int_0^1 \theta(x_1, x_2, \dots x_n) dx_1 \dots dx_n$

So any Monte Carlo calculation is the result of an integrating or averaging process and is hence dependent on the global rather than the local properties of the simulation function. Hence, it is not surprising that in some cases a polynomial approximating surface can be utilized in the Monte Carlo process to produce almost the same histogram as would have been obtained had the simulation function been used. All that is necessary is that the polynomial surface fit the general contours of the multi-dimensional surface which defines the simulation function. If such a polynomial surface can be obtained, then it is quite sensible to utilize it since, as was mentioned before, it is frequently orders of magnitude faster to evaluate a polynomial than to evaluate the simulation function.

The saving involved in the analytic Monte Carlo process is strongly dependent on the nature of the simulation program. If it is a poorly behaved function with many sharp contours, then many simulations will be necessary in order to obtain sufficient data points for an adequate polynomial fit. But the purpose of the analytic Monte Carlo procedure is to keep the number of necessary simulations down. Thus, for a poorly behaved function, little may be purchased by an analytic Monte Carlo approach. Also the more poorly behaved a function the higher the orders of the polynomials must be in order to approximate the function. In this case the time involved in evaluating the high order polynomials may approach a significant fraction of the time necessary to evaluate the simulation function. In this way, also, a point of diminishing returns can be reached.

To see another difficulty with the version of the analytic Monte Carlo procedure presented in (2), it is necessary to look deeper into the analysis on which this version rests. Notice that the output variable in this analysis is represented as the sum of the uncoupled contributions of the input variables with provisions made for the coupling effect of certain pairs of the input variables. These contributions are then expanded as polynomials and the coefficients are determined by perturbing the appropriate input variable, keeping the others at zero, and fitting the concommitant perturbations of the output variable in a least-squares sense. This is equivalent to finding the least-squares multidimensional polynomial fit to the simulation function, with the data points lying along the axes of the domain space and in certain two-dimensional subspaces generated by pairs of basis vectors of the domain space. This particular distribution of data points offers significant numerical advantages since, in effect, it insures that the matrix of the normal equations contains a large number of zeroes which are in a pattern that can be exploited in the inversion process. Hence the problem of inverting large matrices is avoided. Inherent in this approach, however, is a very serious restriction. A least-squares polynomial approximation to an n-dimensional function is obtained while utilizing only information on the behavior of the function in certain selected one and two-dimensional subspaces of the domain space. This may lead to a good fit on the subspaces in question but a rather bad fit elsewhere. In general, the Monte Carlo samples used in the construction of the histogram will not be drawn exclusively from any particular subspace of the domain space. Hence, the analytic Monte Carlo procedure given in (2) while offering numerical advantages could in many situations also offer a warped histogram.

It should be clear that a modified version of the analytic Monte Carlo procedure given in (2), in which the data points could be chosen arbitrarily, would be desirable. With such a modification, for instance, one could choose the data points randomly from the same multivariate distribution from which the Monte Carlo samples are to be chosen. This would insure the most accurate histogram since the density of data points, and hence, the quality of the polynomial fit in a given region would be proportional to the number of Monte Carlo samples to be chosen from that region. Also, if one were particularly interested in a certain section of the histogram, typically a tail section, then one could choose the data points in a disproportionate manner from the region in the phase space which produces the values for that section of the histogram.

The possibilities of such a modified procedure depend on one's abilities to accurately invert large and not necessarily well conditioned matrices. The author has developed his own analytic Monte Carlo procedure which permits arbitrary selection of data points for the least-squares polynomial fit. The inversion problem is dealt with in this modified procedure by closely following the advice offered by Muhonen (3) and Lefferts (4) concerning proper numerical procedures in the inverting of large matrices.

NUMERICAL TECHNIQUES FOR LEAST SQUARES FITTING

In any least-squares fitting procedure, the inversion of a matrix or its numerical equivalent is required. The most straightforward approach with regard to least-square fitting is to form the normal equations, obtain the associated matrix and invert it. But this is not always the best approach. In fact, it can sometimes lead to useless results. Many alternative procedures exist. The proper procedure in a given situation depends on the particular trade off with regard to accuracy, computer time, computer storage, etc. That one desires. Recent research as reported in (3) and (4) has focused attention on five solutions to the least-squares fitting problem.

1. The Gauss-Jordan Algorithm on the normal equations,

2. An Andree Algorithm for the Penrose pseudoinverse on the normal equations,

3. A Gram-Schmidt orthogonalization pseudoinversion scheme on the normal equations,

4. The Gram-Schmidt procedure applied to the original rectangular data matrix, and

5. Householder's algorithm for direct triangulation of the original data matrix.

Some of the results obtained are listed below:

1. Procedures (4) and (5) above, which avoid the formation of the normal equations, give considerably better results than do the other schemes.

2. Procedure (4) should be used in double precision when there are no storage or timing restrictions. It will handle situations of reduced rank.

3. Procedure (5) should be used in double precision when timing but not storage restrictions exist. It does not handle situations of reduced rank.

An IBM 360/95, which is quite fast and has great storage capacity, was available during the course of this research. Hence, procedure (4) was utilized. It was found to be exceedingly accurate, reasonably fast, and demanding of considerable core storage when used for n-dimensional least-squares polynomial fits. Hence, it must be admitted that the version of the analytic Monte Carlo procedure presented in this paper is not practical unless an electronic computer with substantial storage capacity is available.

In Appendix A is a complete listing of a double precision Fortran program which utilizes the Gram-Schmidt procedure to obtain n-dimensional polynomial fits to arbitrary data. The program is a very flexible one as can be seen from the description of its input.

PROCEDURAL QUESTIONS

There are several practical questions to be considered in the application of the analytic Monte Carlo technique. The number of simulations which are necessary in a given application is an important issue. The particular type of n-dimensional polynomial to be used in the fitting procedure must be decided. These questions can only be answered within the context of a particular application. There are, however, general principles which ought to be followed. Obviously, the number of simulations should be kept as small as possible. Also, the total number of terms in the polynomial fit should be kept as small as possible. The number of necessary simulations and the number of terms necessary in the polynomial fit are dependent on the number of input variables, the nature of the simulation function, and the amount of filtering that is desired.

The issue of how much filtering is desired is dependent on the amount of noise the simulation values are thought to contain. The greater the amount of noise (computer round off, mathematical approximations, etc.) included in the simulation values, the greater the ratio of simulations to parameters in the polynomial fit must be. But other factors influence the minimum number of parameters in the polynomial fit which is tolerable. The more poorly behaved the surface of the simulation function is, the greater is the number of parameters that is required. Another factor to be considered is the nature of the multivariate distribution on which the Monte Carlo sampling is to be performed. The greater the dispersion of this distribution the larger the region is over which the polynomial fit must be adequate. This, of course, influences the number of parameters to be included in the fit. Once the number of parameters has been tentatively decided, the quantity of noise which is thought to be present then determines the number of simulations to be requested. It is advisable to be quite optimistic in the estimate of the number of necessary simulations. If the estimate proves too optimistic, then more can always be obtained.

There is no fixed format for deciding how the parameters of the polynomial fit are to be distributed. If the output variable is thought to be strongly influenced by a certain input variable, then the power on that variable in the polynomial fit should be large. Conversely, if a certain input variable is suspected of having little influence on the output variable, then a smaller power should be used. Physical intuition should also be relied on in deciding which input variables are to be correlated in the polynomial fit. It is recommended that considerable experimentation be performed with a variety of possible types of fit using as a criteria for quality some goodness of fit statistic calculated on the data points.

Another and perhaps more valuable method for comparing polynomial fits is possible if the nature of the problem fixes practical bounds on the size of the output variable. The procedure in testing the usefulness of a particular polynomial fit consists simply of choosing some data points randomly from the multivariate population on which the Monte Carlo process is to be performed and calculating their respective polynomial values. The number of values which exceed the practical bounds is obviously related to the quality of the polynomial fit. If for instance, the ratio of simulations to parameters in the fit was not sufficiently large, then the filtering would be insufficient and the quality of the fit off the data points could in some cases be surprisingly bad. This would not be shown by a goodness of fit statistic. But the above mentioned test would reveal that a significant fraction of the polynomial values are either too large or too small to be realistic.

In summation, the problem of obtaining a useful polynomial fit for the analytic Monte Carlo process is essentially a heuristic and experimental one in which physical intuition and common sense play an important part.

AN EXAMPLE

INTRODUCTION

One of the missions considered for the Planetary Explorer project was a Venus-72 mission whose goal is to orbit a probe around Venus. A fixed midcourse, minimum fuel guidance correction is planned at approximately five days after injection. It is desired to obtain information on the distribution of the midcourse velocity correction caused by injection errors and also to obtain hints concerning which types of injection errors have the most impact on the midcourse correction. For this purpose, a simulation program with Monte Carlo capabilities was available (5). The program utilizes an iterative scheme on an impulsive approximation to obtain midcourse corrections. The Monte Carlo sampling is performed on a six by six covarience matrix which defines the multivariate distribution of the injection errors in a local tangent coordinate set. For this particular mission the position and velocity vectors of the probe at injection are orthogonal. Hence what is referred to here as the down-track direction is the direction of the probe velocity vector at injection. A down-track velocity error is just a speed error.

Approximately ten thousand of these Monte Carlo samplings can be performed per hour on an IBM 360/95 computer. Clearly a straightforward Monte Carlo analysis using this simulation program could be quite costly in terms of computer time. Thus it was decided to apply the analytic Monte Carlo procedure. The results are reported in some detail below.

HOW THE POLYNOMIAL FIT WAS OBTAINED

The first step in applying the analytic Monte Carlo procedure is to obtain a polynomial fit to the data from the simulation program. The model for the polynomial fit was obtained by choosing one hundred data points from the population defined by the covarience matrix of the injection errors and fitting their associated midcourse velocity corrections with a variety of types of six dimensional polynomials. In doing so, the ratios of orders on the six variables, and the number and types of correlations were systematically varied and the one associated with the smallest goodness of fit statistic was chosen as the model. The model giving the smallest goodness of fit statistic had the same order on all six variables with correlations between the three position errors and between radial, down, and cross track position and velocity errors included. Next the ratio of data points to parameters in the fit was decided. For this sort of decision, a goodness of fit statistic is of no aid. A sample of one thousand output values from the simulation program, indicated that the midcourse correction should very seldom exceed one hundred and twenty meters per second. Since the corrections are always positive, effective bounds on the output variable are known. The proper number of parameters and data points can be determined by taking Monte Carlo samples of various fits and counting the number of values that fall outside the bounds. It was decided that the bounds to be used were minus ten meters per second and one hundred and thirty meters per second. It was also decided that a fit would be considered adequate when no more than one Monte Carlo value in a thousand fell outside these bounds. The intention, of course, was to choose the smallest number of parameters and data points which fulfill this condition. The fit which was finally accepted as adequate utilized three hundred data points and seventy-two parameters. This involved a polynomial order of six for each of the six variables with six parameters associated with each of the six analytic correlations. Since interest was focused on the right hand tail region of the resultant distribution of midcourse corrections, half the data points were chosen from the multinomial population of injection errors defined by the given covarience matrix and half were chosen from a distribution with twice the dispersion of the original distribution. This was believed to give a tighter fit in the region of large injection errors in the six dimensional domain space. Since this is the region associated with large midcourse velocity corrections, this procedure was believed to provide greater accuracy in the right hand tail region of the distribution. Ten thousand Monte Carlo samples can be obtained from the least-squares polynomial in less than ninety seconds on an IBM 360 model 95 computer. As mentioned previously, the obtaining of ten thousand Monte Carlo samples from the simulation function would require, approximately, and hour on the same computer.

THE RESULTS

Figure 1 is a histogram of midcourse velocity corrections based on a Monte Carlo sample of ten thousand. The values were obtained by sampling six tuples from the population defined by the given covarience matrix and obtaining the corresponding polynomial value from the six dimensional, seventy-two parameter polynomial fit. The ninety-nine and ninety-five percent critical values were respectively eighty-eight meters per second and sixty-seven meters per second.

It is of considerable interest to discover which of the injection errors have the most serious impact on the midcourse velocity correction. The analytic Monte Carlo procedure is a convenient tool for answering such questions. Each Monte Carlo value calculated from the polynomial may be thought of as the sum of twelve contributors, one contribution from each of the six input variables and six contributions from the terms giving the six analytic correlations of the polynomial fit. If the covariance matrix used to obtain the Monte Carlo samples were diagonal, then the mean value of these contributions could be readily calculated from the values of the appropriate coefficients in the polynomial fit. The method is discussed in (2). In this application, the covarience matrix was distinctly non-diagonal and this method could not be applied. An alternative procedure which should prove adequate in any situation is to calculate the mean of a few hundred sample values of each contributor and to accept the results as decent estimates of the true means. If one assumes that the polynomial surface coheres reasonably well to the true surface, then the relative sizes of these estimated means provide valuable information on the relative importance of the injection errors with regard to the midcourse velocity correction. The results are:

percent of mean due to radial-track position error	=	10.0 %
percent of mean due to down-track position error	=	0.0 %
percent of mean due to cross-track position error	=	13.0 %
percent of mean due to radial-track velocity error	=	27.0 %
percent of mean due to down-track velocity error	=	42.0 %
percent of mean due to cross-track velocity error	=	6.5 %

The contribution of the six correlation terms is negligible, the sum of their means amounting to less than 2 % of the total mean. Clearly, the radial-track and down-track velocity errors are the most critical with regard to a minimum fuel midcourse velocity correction. Figures two, three and four tend to corroborate these results. They represent the histograms of the midcourse velocity correction with respectively the radial-track position and velocity errors deleted, the down-track position and velocity errors deleted, and the cross-track position and velocity errors deleted.

The histogram of figure two was obtained by sampling ten thousand values of the down-track and cross-track posiiion and velocity errors from a covarience matrix obtained from the original covarience matrix by deleting the rows and columns associated with the radial-track position and velocity errors. The radial-track position and velocity errors are set at zero. The resultant ten thousand six-tuples were used as input into the polynomial and the output values were arranged in the histogram seen in figure two. Figures three and four were, of course, similarly obtained. The ninety-nine percent and ninety-five percent critical values associated with the histogram of figure two are respectively 62.5 m/sec and 47 m/sec. For the histogram of figure three, the same critical values are 57.6 m/sec and 44 m/sec. For the histogram of figure four, the critical values are 75.5 m/sec and 56 m/sec. These may be viewed as values toward which the 99 percent and 95 percent critical values will tend as greater and greater reductions in radial-track or down-track or cross-track errors are realized. The indications are that it is most advantageous to reduce down-track errors and least advantageous to reduce cross-track errors.

It ought to be noticed that these results are derived from the interaction of a multinomial distribution given by a covarience matrix with a polynomial fit to an analytic surface. If either one is substantially changed, these results may no longer be relevant.

RESULTS FOR A FIXED TIME OF ARRIVAL GUIDANCE LAW

The results obtained in previous sections concern the statistical characteristics of the mid-course velocity correction under a minimum fuel guidance law. In this section, the statistical characteristics of the mid-course velocity correction under a fixed time of arrival guidance law are reported. The same mission with the same covariance matrix of injection errors was utilized. Interest here was focused on the relative importance of the individual injection errors under such a guidance law and not on critical values. Hence for this study, three hundred fifty Monte Carlo samples of injection errors were chosen from the multivariate normal distribution defined by the covariance matrix of injection errors. The same polynomial model as used in the case of minimum fuel guidance was again utilized and the midcourse correction was again simulated at five days after injection. The resultant histogram based on ten thousand Monte Carlo samples was obtained from the least-squares polynomial and is displayed in Figure 5.

The relative importance of the individual injection errors on the fixed time of arrival midcourse velocity correction is revealed in the table below:

percent of mean due to radial-track position error	=	9.8 %
percent of mean due to down-track position error	=	8.5 %
percent of mean due to cross-track position error	=	0.0 %
percent of mean due to radial-track velocity error	=	22.9 %
percent of mean due to down-track velocity error	=	41.0 %
percent of mean due to cross-track velocity error	=	17.6 %

Again the contribution of the six correlation terms is negligible. The major difference between the above results and results obtained previously with regard to a minimum fuel guidance law is the increased importance of cross-track velocity errors.

CONCLUSION

The problem of obtaining information on the distribution of a random variable which is a known function of several other random variables of a given multivariate distribution is a common one. If computational difficulties are surmountible, then the most satisfactory solution to this problem is a Monte Carlo one. In many situations where a Monte Carlo approach is not feasible because of computer time requirements, the technique discussed in the present paper can be of considerable value. The technique involves fitting data points from a simulation function with a multidimensional polynomial in the least-squares sense. Monte Carlo samples are then obtained from the multinomial distribution of the input variables and their functional values are calculated from the multidimensional polynomial instead of the simulation function. The functional values are then arranged in a histogram which, one hopes, approximates the true distribution of the output variable. And since, presumably, the polynomial can be evaluated much more quickly than can the simulation function, a considerable saving in computer time may be realized.

A major advantage of this technique is that it isolates the influence of each input variable on the statistics of the output variable thus making parametric studies quite convenient. If the input variables are normally distributed and statistically independent, then the contribution to the mean and deviation of the output variable from each input variable can be calculated from the coefficients of the polynomial fit. Even if these conditions are not satisfied, the individual contributions can still be quite conveniently estimated.

The analytic Monte Carlo technique was applied to obtain a study of the statistical distribution of a minimum fuel midcourse velocity correction for a Venus-72 mission. A histogram of the midcourse velocity correction was obtained and is displayed in Figure 1 of the text. It was also discovered that the largest contributor to the mean of the midcourse velocity correction is the down-track velocity injection error. The next largest contributor is the radial-track velocity injection error. The other four contributors are substantially less important than these two.

If a fixed time of arrival guidance law is used for the midcourse velocity correction, the cross-track velocity error becomes substantially more important. Under this guidance law, the down-track velocity error is again the most important. The radial-track and cross-track velocity errors are next in importance having an approximately equal impact on the midcourse velocity correction.

ACKNOWLEDGEMENTS

I want to thank Mr. Eugene Lefferts and Mr. Charles Newman of the Goddard Space Flight Center and Mr. Daniel Muhonen of the Bendix Corporation for their aid during the course of this research. I also want to thank Mr. Alan Ritter of the Bendix Corporation for his valuable programming support.

REFERENCES

1. Merel, M. H. and Mullin, F. J., "Analytic Monte Carlo Error Analysis", J. Spacecraft and Rockets; Vol 5, No. 10, Nov. 1968, pp. 1304-1308.

2. Logsdon, T. S. and Africano, R. C., "A Modified Monte Carlo Procedure", A.I.A.A. Journal, Vol. 6, No. 6, June 1968, pp. 1111-1117.

3. Muhonen, D. P., "An Experimental Comparison of Several Approaches to the Linear Least Squares Problem", TR-01-1004, Bendix Field Engineering Corporation, Nov. 1968.

4. Lefferts, E. J., "Least Squares and Pseudoinversion", N.A.S.A. X-document X-551-69-73, September, 1968.

5. "Programmer's Manual For the Advanced Mission Analysis Programs", TR-DA1622, Philco-Ford Corporation, January, 1968.

APPENDIX A

Description and listing of a multidimensional polynomial fit program.

"POLY" is a double precision Fortran IV subroutine which obtains a least-squares multidimensional polynomial fit to input data. A Gram-Schmidt procedure applied to the original rectangular data matrix is utilized in the least-squares process. The program can provide a multidimensional least-squares fit to any degree up to and including ten. There are no restrictions on the individual orders of the variables in the polynomial model. Any or all possible binary correlations can be included in the model. Correlated triples and higher correlations cannot be included in the model and the total number of terms cannot exceed one hundred and twenty. An input and output description of "POLY" and a listing of the program are provided below.

INPUT

N - Dimension of polynomial fit

M - Number of times the function to be fitted was evaluated. (This number must exceed the number of parameters in the polynomial fit.)

F - An M dimensional array. The I^{th} element in F is the value of the function at the I^{th} evaluation.

A - An N by M array of numbers. The number A (I, J) is the value of the I^{th} input variable at the J^{th} evaluation of the function.

NP - An N dimensional array of fixed point numbers. NP (I) is the order desired on the I^{th} variable in the polynomial model.

K - The number of binary correlations included in the polynomial model.

IB - A k dimensional array of fixed point numbers giving the correlation numbers of the k correlations to be included in the polynomial fit. The correlation number associated with the correlation of input variables I and J where $I < J$ is obtained by the formula

$$\sum_{k=1}^{I} (10-k) - (10-J)$$

The correlation numbers in the IB array must be stored in the increasing order of magnitude.

OUTPUT

NNI - The number of terms in the least-squares polynomial.

XX - An NNI dimensional array giving the coefficients of the least-squares polynomial. In order to properly interpret the elements of XX, a standard sequence for the terms of the least-squares N dimensional polynomial must be given. The powers of the first variable are written first in the order of increasing power. The powers of the second variable are written next and so forth. There is no provision for a constant term in the polynomial fitting process. The correlation terms are written next and in the order of increasing correlation number. The terms associated with a given correlation are written in the order of increasing order of the lower indexed variable and decreasing order of the higher indexed variable. For instance, if the first element in the IB array is 2, then the correlation of the first and third variables is the first one to be written. If L is the smaller member of the set [NP (2), NP (3)], then the first term in the sequence related to the correlation of variables one and three is $X_2 \; X_3^L$. The next term is $X_1^2 \; X_3^{L-1}$ and the last term is $X_1^L \; X_3$. The XX array can now be interpreted. The Ith element in the XX array is the coefficient of the Ith term in the least squares N-dimensional polynomial when the polynomial is written in the above defined standard fashion.

```
      SUBROUTINE POLY(N,M,F,A,NP,K,IB,XX,NNI,NRANK)
C------                          . . .NOTE  NO MONTE
C23456 89  COLUMN SEVEN MARKER
      IMPLICIT REAL * 8 (A-H,O-Z)
      REAL*4 EPS
      DIMENSION U(100,100),AFLAG(100),ATEMP(100),XX(100)
      DIMENSION   F(M)    , A(10,M)     ,  NP(N ),   IB(K ),
     C  LC1(45),LC2(45),LC(45),NH(56),X(350,100)
C-----CALL ERRSET (210,256,20,0)
      CALL ERRSET(210,256,-1,1)
      MR=M
      NR=M
      EPS=3.0
C---------SET UP LC1 AND LC2 ARRAYS
C
C. . . . . . . . LC1 ARRAY FIRST NINE ELEMENTS=1
C                          NEXT EIGHT ELEMENTS=2
C                          ETC
C
C. . . . . . . . LC2 ARRAY ELEMENTS  1-9 =2-10
C                          ELEMENTS 10-17=3-10
C                          ELEMENTS 18-24=4-10
C                          ETC
C
C-----------LC1 ARRAY CODING
      KX=1
      NX=9
      LX=8
      DO 100 I=1,45
      IF(NX.LE.I) GO TO 25
      LC1(I)=KX
      GO TO 100
 25   NX=NX + LX
      LX=LX - 1
      LC1(I)=KX
      KX=KX + 1
 100  C O N T I N U E
C-----------LC2 ARRAY CODING
      KX=0
      DO 200 I=1,9
      L=I + 1
      DO 300 J=L,10
      KX=KX + 1
      LC2(KX)=J
 300  C O N T I N U E
 200  C O N T I N U E
```

```
C----------SET UP LC ARRAY
      DO 50 I=1,K
      L=IB(I)
      L1=LC1(L)
      L2=LC2(L)
      IF(NP(L1).LT.NP(L2)) GO TO 45
      LC(I)=NP(L2)
      GO TO 50
 45   LC(I)=NP(L1)
 50   C O N T I N U E
C-----------SET UP NH ARRAY
      NH(1)=0
      DO 90 J=1,N
      I=J + 1
      NH(I)=NH(J) + NP(J)
 90   C O N T I N U E
      L1= N + 1
      L2=K + N
      DO 91 J=L1,L2
      I=J + 1
      I1= J - N
      NH(I)= NH(J) + LC(I1)
 91   C O N T I N U E
      NOUT=N+K+1
C----------COMPUTE X(J,I) ARRAY
      NI=   N+K+1
      NNI= NH(NI)
      DO 120 I=1,NNI
      DO 110 J=2,NI
      IF(NH(J).LT.I) GO TO 110
      L=J - 1
      GO TO 111
 110  C O N T I N U E
 111  LL= I - NH(L)
      IF(N.LT.L) GO TO 118
      DO 115 I1=1,M
 115  X(I1,I)=A(L,I1)**LL
      GO TO 120
 118  LDEX=L + 1
      LLL=NH(LDEX) + 1 - I
      LN=L - N
      NNB=IB(LN)
      L1=LC1(NNB)
      L2=LC2(NNB)
      DO 119 I1=1,M
 119  X(I1,I)=A(L1,I1)**LL*A(L2,I1)**LLL
 120  C O N T I N U E
      NC=NNI
```

```
      CALL GINV2(X,U,AFLAG,ATEMP,MR,NR,NC,NRANK,EPS)
      DO 250  I=1,NC
      XX(I)=0.0D0
      DO 240 L=1,NR
      XX(I)= XX(I) + X(L,I) * F(L)
240   C O N T I N U E
250   C O N T I N U E
530   C O N T I N U E
 318   RETURN
      END
      SUBROUTINE GINV2(A,U,AFLAG,ATEMP,MR,NR,NC,NR1,EPS)
      DOUBLE PRECISION A(MR,NC),U(NC,NC),AFLAG(NC),ATEMP(NC)
      DOUBLE PRECISION FAC,DOT,DOT1,DOT2,TOL,DSQRT
      DO 10 I = 1,NC
      DO 5 J = 1,NC
    5 U(I,J) = 0.
   10 U(I,I) = 1.
      FAC = DOT(MR,NR,A,1,1)
      FAC = 1./DSQRT(FAC)
      DO 15 I = 1,NR
   15 A(I,1) = A(I,1)*FAC
      DO 20 I = 1,NC
   20 U(I,1) = U(I,1)*FAC
      AFLAG(1) = 1.
      N = 56
      NR1 = NC
      TOL = (10.**EPS*.5**N)**2
      DO 100 J = 2,NC
      DOT1 = DOT(MR,NR,A,J,J)
      JM1 = J-1
      DO 50 L = 1,2
      DO 30 K = 1,JM1
   30 ATEMP(K) = DOT(MR,NR,A,J,K)
      DO 45 K = 1,JM1
      DO 35 I = 1,NR
   35 A(I,J) = A(I,J)-ATEMP(K)*A(I,K)*AFLAG(K)
      DO 40 I = 1,NC
   40 U(I,J) = U(I,J)-ATEMP(K)*U(I,K)
   45 CONTINUE
   50 CONTINUE
      DOT2 = DOT(MR,NR,A,J,J)
      IF ((DOT2/DOT1)-TOL) 55,55,70
   55 DO 60 I = 1,JM1
      ATEMP(I) = 0.
      DO 60 K = 1,I
   60 ATEMP(I) = ATEMP(I)+U(K,I)*U(K,J)
      DO 65 I = 1,NR
      A(I,J) = 0.
```

```
      DO 65 K = 1,JM1
   65 A(I,J) = A(I,J)-A(I,K)*ATEMP(K)*AFLAG(K)
      AFLAG(J) = 0.
      FAC = DOT(NC,NC,U,J,J)
      FAC = 1./DSQRT(FAC)
      NR1 = NR1-1
      GO TO 75
   70 AFLAG(J) = 1.
      FAC = 1./DSQRT(DOT2)
   75 DO 80 I = 1,NR
   80 A(I,J) = A(I,J)*FAC
      DO 85 I = 1,NC
   85 U(I,J) = U(I,J)*FAC
  100 CONTINUE
      DO 130 J = 1,NC
      DO 130 I = 1,NR
      FAC = 0.
      DO 120 K = J,NC
  120 FAC = FAC+A(I,K)*U(J,K)
  130 A(I,J) = FAC
      RETURN
      END
      DOUBLE PRECISION FUNCTION DOT(MR,NR,A,J,K)
      DOUBLE PRECISION A(MR,1),X
      X = 0.D0
      DO 50 I = 1,NR
      X = X+A(I,J)*A(I,K)
   50 CONTINUE
      DOT = X
      RETURN
      END
      REAL FUNCTION VNORM*8(Z,W)
      IMPLICIT REAL*8(A-H,O-Z)
      DIMENSION Z(3), W(3)
1     SCL=DSQRT(Z(1)**2+Z(2)**2+Z(3)**2)
      IF(SCL.EQ.0.0D0) GO TO 20
      DO 2 I=1,3
2     W(I)=Z(I)/SCL
      VNORM=SCL
      RETURN
20    VNORM=0.0D0
      DO 5 I=1,3
5     W(I)=0.0D0
      RETURN
      END
      SUBROUTINE CROSS(A,B,C)
      IMPLICIT REAL*8 (A-H,O-Z)
      DIMENSION A(3), B(3), C(3)
```

```
      C(1)=A(2)*B(3)-B(2)*A(3)
      C(2)=B(1)*A(3)-A(1)*B(3)
      C(3)=A(1)*B(2)-B(1)*A(2)
      RETURN
      END
      SUBROUTINE MVTRN(A,B,C,M,N)
      IMPLICIT REAL*8(A-H,O-Z)
      DIMENSION A(9), B(9), C(9), NN(4)
      DATA NN/3,1,8,0/
C
  1   I1=1
      J1=1
      K1=1
      DO 4 I=1,N
      DO 3 J=1,3
      C(I1)=0.0D0
      DO 2 K=1,3
      C(I1)=C(I1) + A(J1)*B(K1)
      J1=J1 + NN(M)
  2   K1=K1 + 1
      I1=I1+ 1
      J1=J1 - NN(M+2)
  3   K1=K1-3
      K1=K1 + 3
  4   J1=1
      RETURN
      END
```

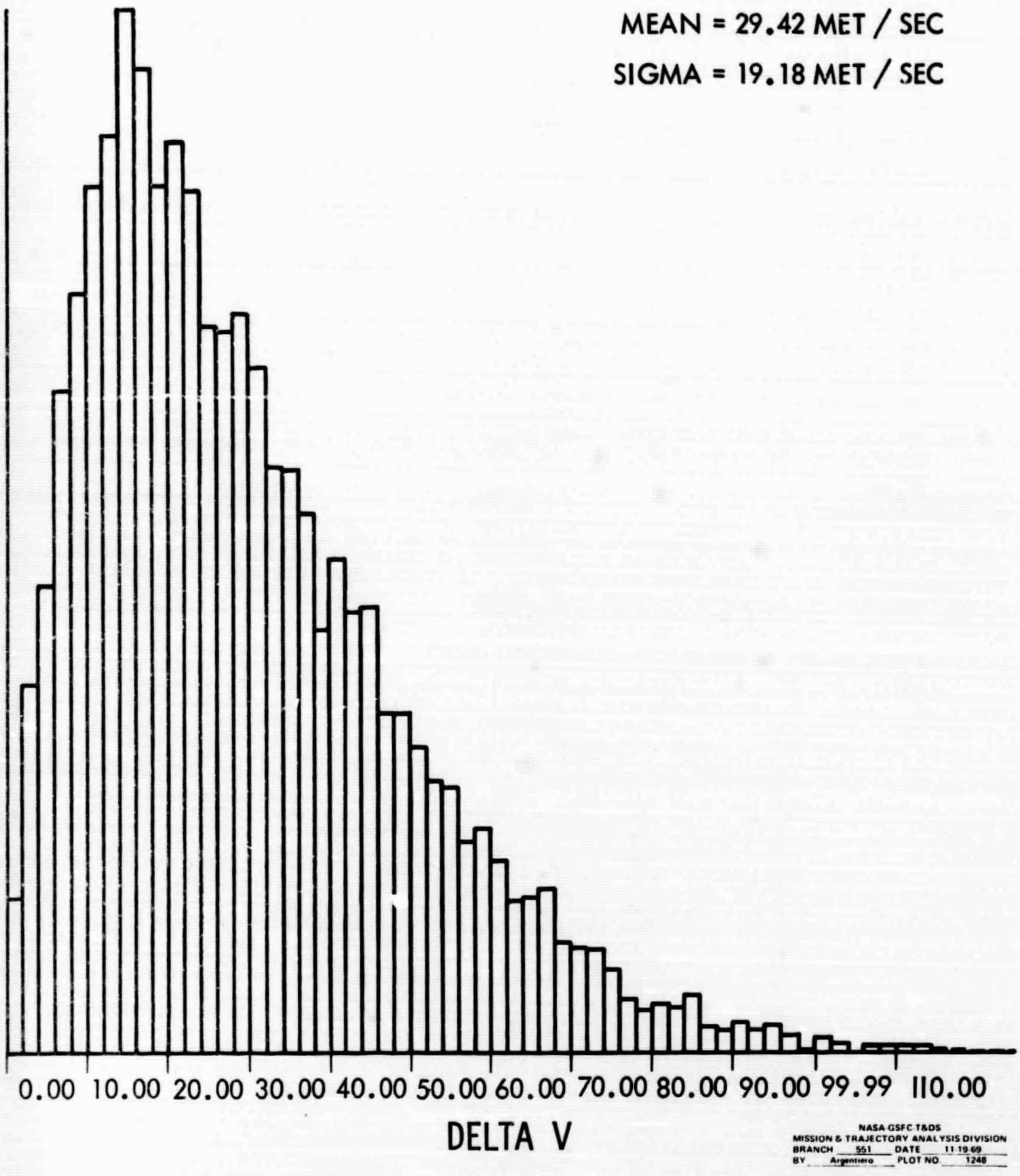

Figure 1. Histogram of Midcourse Velocity Corrections Under Minimum Fuel Guidance Law

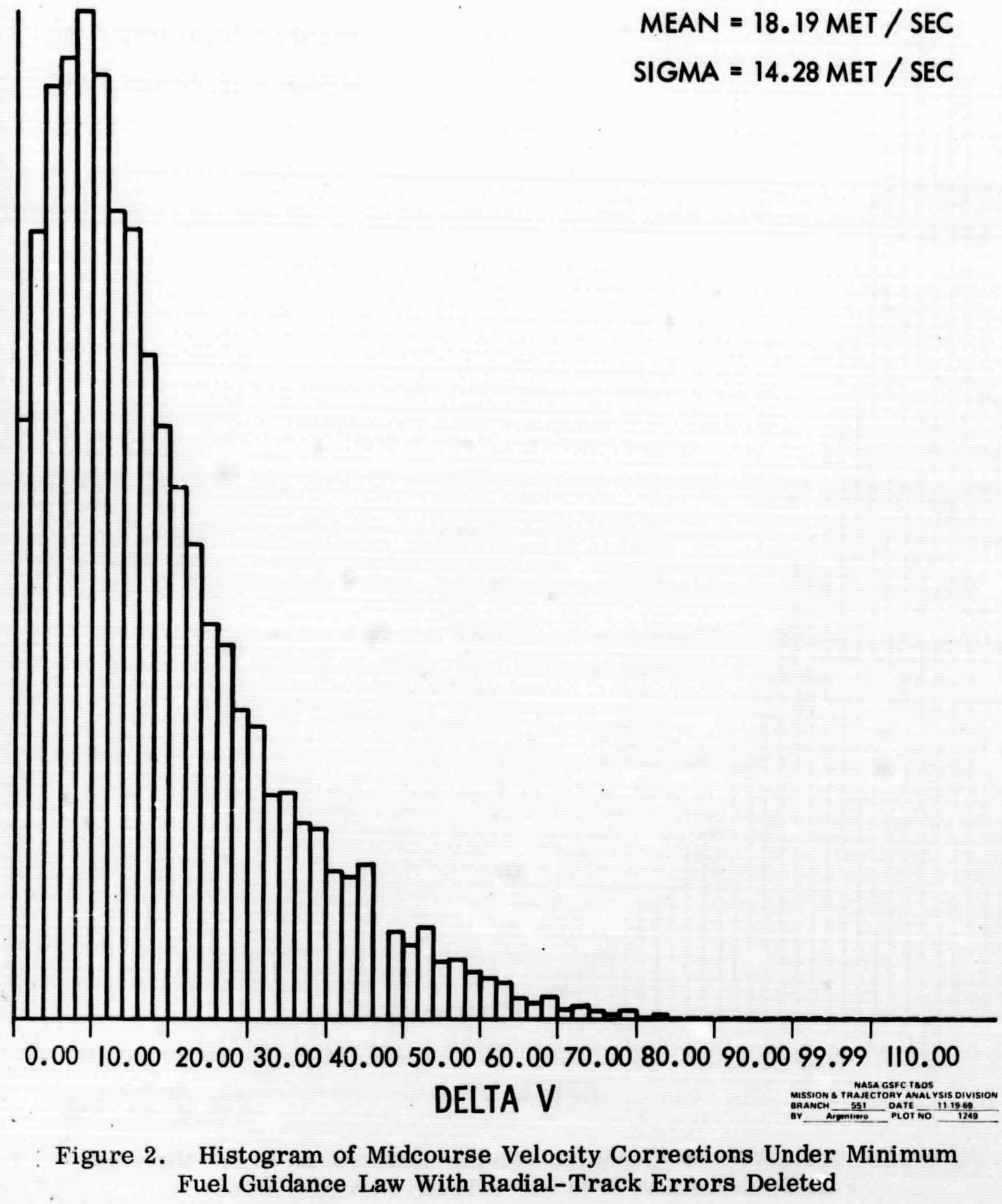

Figure 2. Histogram of Midcourse Velocity Corrections Under Minimum Fuel Guidance Law With Radial-Track Errors Deleted

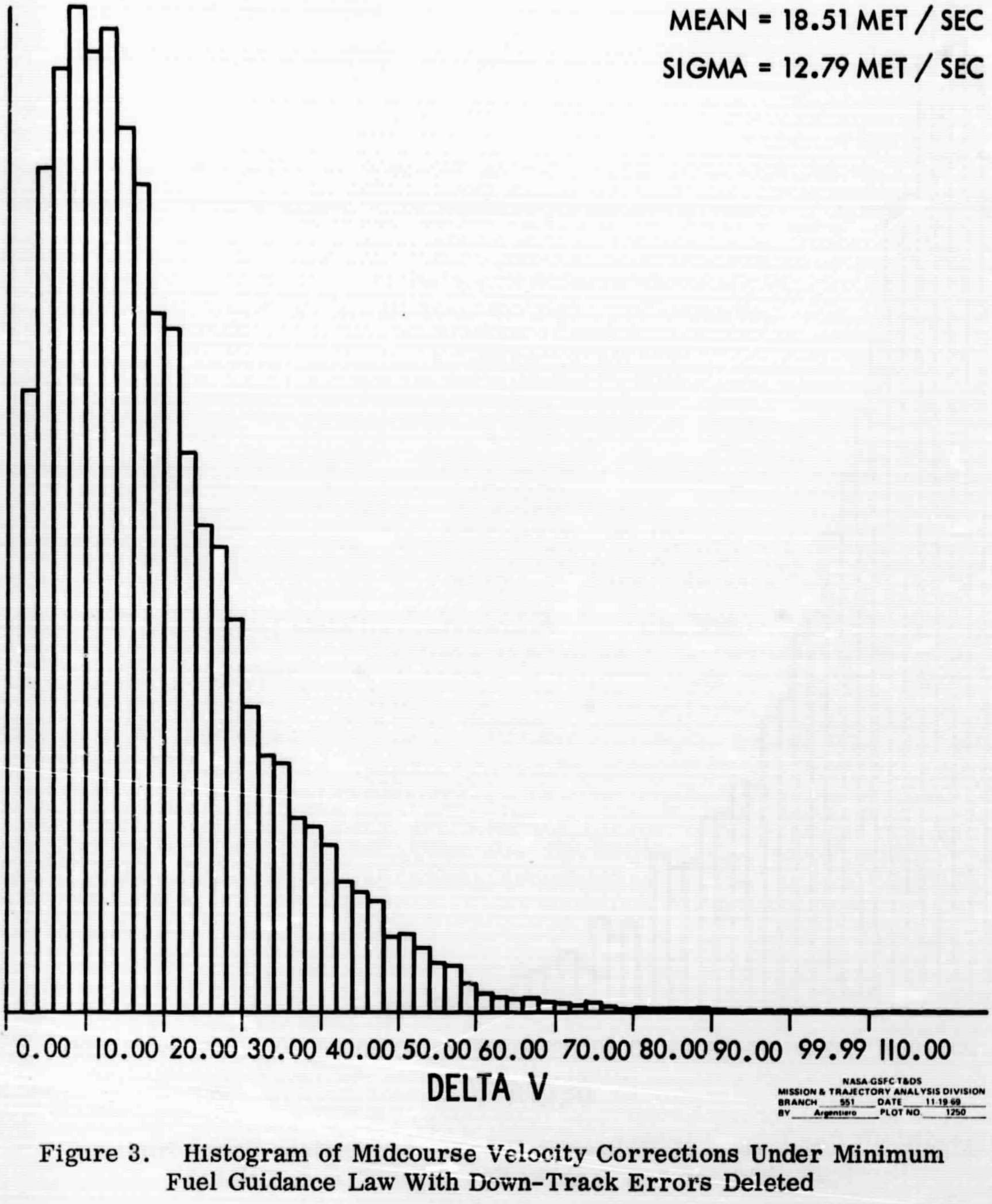

Figure 3. Histogram of Midcourse Velocity Corrections Under Minimum Fuel Guidance Law With Down-Track Errors Deleted

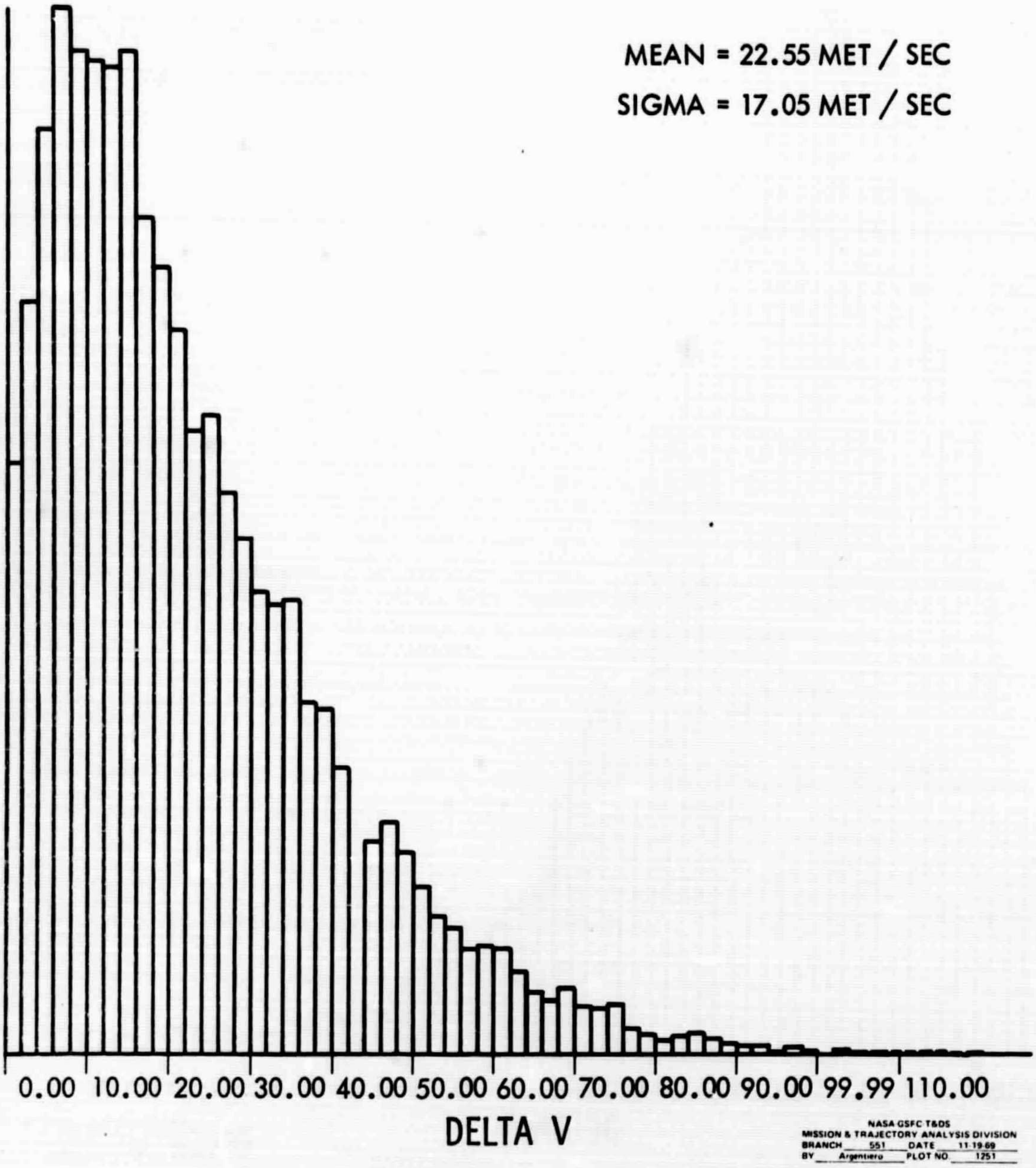

Figure 4. Histogram of Midcourse Velocity Corrections Under Minimum Fuel Guidance Law With Cross-Track Errors Deleted

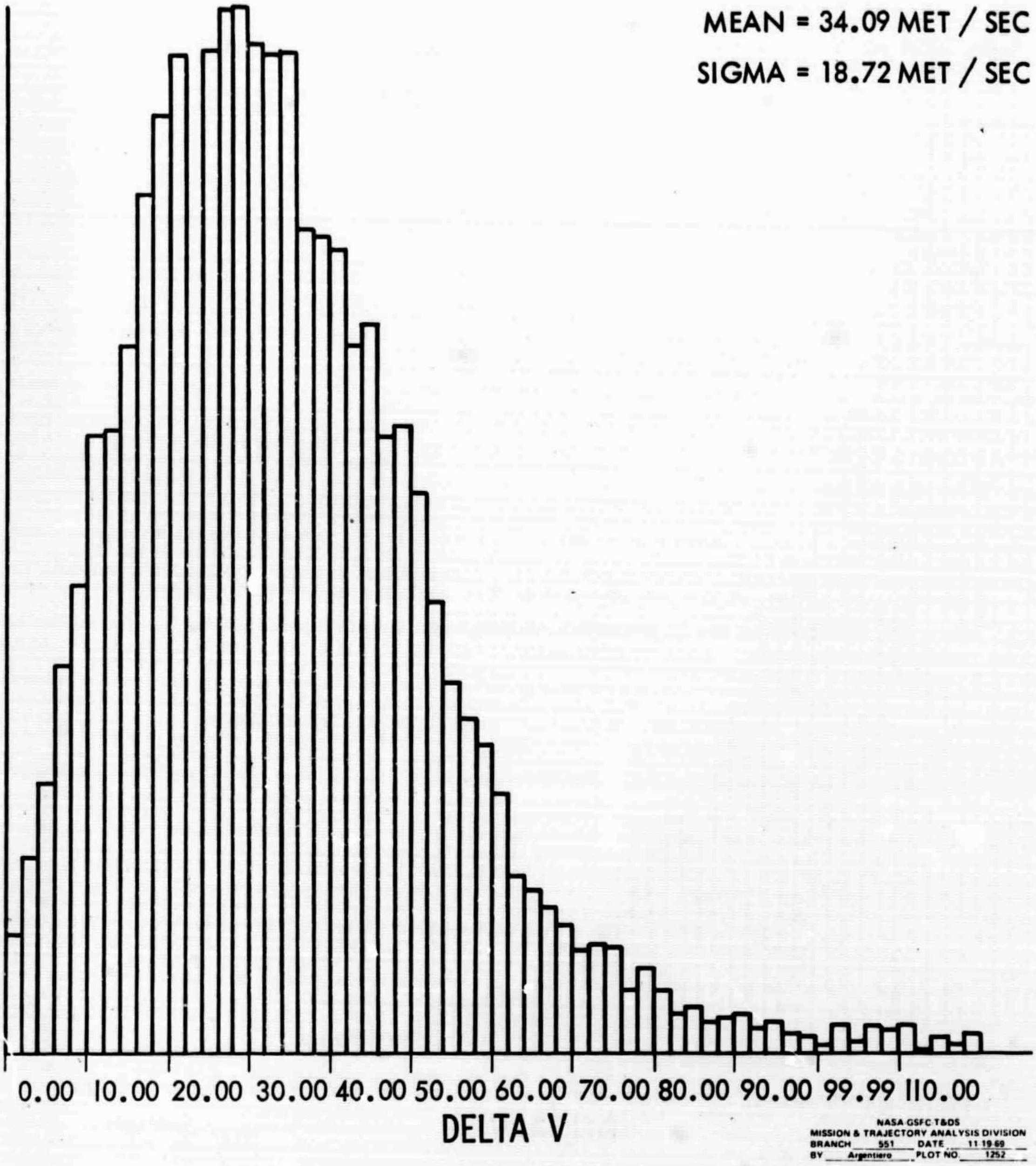

Figure 5. Histogram of Midcourse Velocity Corrections Under Fixed Time of Arrival Guidance Law