N73-11811

FINAL REPORT

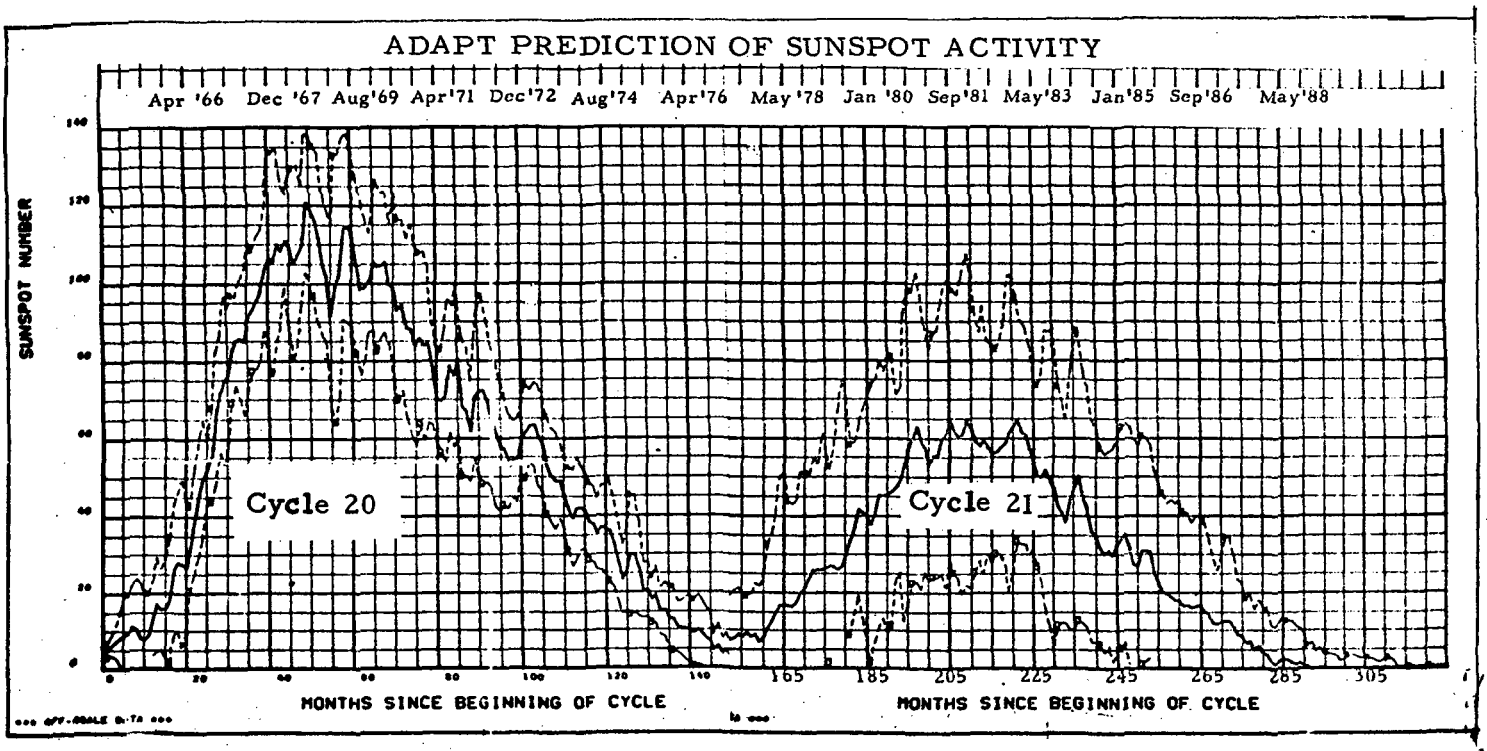
APPLICATION OF AVCO DATA ANALYSIS AND PREDICTION TECHNIQUES (ADAPT) TO PREDICTION OF SUNSPOT ACTIVITY

By:

Herbert E. Hunter and Richard A. Amato

AVSD-0287-72-CR August 1972

Contract: NAS 8-28087



Prepared For:

National Aeronautics & Space Administration
George C. Marshall Space Flight Center
Aerospace Environment Division
Huntsville, Alabama 35807

Prepared By:

Avco Government Products Group
Avco Systems Division
Wilmington, Massachusetts 01887

ACKNOWLEDGEMENT

This work was supported by NASA Marshall Space Flight Center under Contract NAS 8-28087. In addition, the authors want to thank Messrs. Harold Euler, R. E. Smith, and W. W. Vaughan of NASA/MSFC's Aerospace Environment Division for suggesting this application and their interest and support during the course of the work. Thanks are also due to the following Avco personnel: Dr. Nelson Kemp for many helpful discussions, Mr. Jay Gottesfeld for programming support and Mr. Jake Apsel for carrying out much of the computation required for this study.

TABLE OF CONTENTS

Use of ADAPT to Extrapolate Sunspot Data

ABSTRACT

This report presents the the results of the application of Avco Data Analysis and Prediction Techniques (ADAPT) to derivation of new algorithms for the prediction of future sunspot activity. The ADAPT derived algorithms show a factor of 2 to 3 reduction in the expected 2-sigma errors in the estimates of the 81-day running average of the Zurich sunspot numbers. The report presents: (1) The best estimates for sunspot cycles 20 and 21, (2) a comparison of the ADAPT performance with conventional techniques, and (3) specific approaches to further reduction in the errors of estimated sunspot activity and to recovery of earlier sunspot historical data.

The ADAPT programs are used both to derive regression algorithm for prediction of the entire 11-year sunspot cycle from the preceding two cycles and to derive extrapolation algorithms for extrapolating a given sunspot cycle based on any available portion of that cycle. It is suggested that further improvement in sunspot predictions is possible by including more data in the learning set, accounting for the present value of the sunspot number in the immediate future and for the extrapolation algorithm, using a three cycle base instead of a single cycle base.

The estimates obtained show that cycle 20 should last somewhat longer than previously anticipated, with the minimum of the 81-day running average occurring early in 1977. The estimates also show a lower peak activity for cycle 21 than previous estimates, with a maximum sunspot number of approximately 60 for cycle 21.

LIST OF FIGURES

LIST OF FIGURES (Cont'd.)

LIST OF FIGURES (Cont'd)

1.0 INTRODUCTION

This report presents the results of a study which has the objective of developing improved numerical techniques for predicting future sunspot numbers. The improved techniques are based on the use of the Avco Data Analysis and Prediction Techniques (ADAPT). ADAPT is a unique set of programs which first obtains the best representation for any given set of data. This best representation then allows one to characterize the data and to derive empirical prediction, classification, extrapolation and/or clustering laws in an extremely efficient manner. Previous applications of the ADAPT programs to reentry physics, sonar, engine diagnostics, medical, meteorological, and solar physics problems have demonstrated that the empirical prediction and extrapolation laws derived using the ADAPT programs have significant advantages relative to those developed by more classical techniques (see refs. 1 thru 7). The methods currently employed for estimating future sunspot activity are based primarily upon a classical empirical regression scheme developed by McNish and Lincoln (see ref. 8). Thus, the ADAPT techniques should provide significant improvement in the capability for estimating future sunspot activity. This report presents the results of a study which demonstrates this improvement.

The development of improved estimates of future sunspot activity is important to the study of solar physics in general and possibly to astrophysics. The estimates of solar activity have a very practical importance in that they are part of many geophysical models for predicting such quantities as satellite life times, cosmic ray intensity, atmospheric and climatic phenomena.

The linear regression techniques developed by McNish and Lincoln have been modified by different investigators by using different lengths of past data for deriving the predictions of future sunspot numbers. The current estimates for cycle 20 are based on the application of these techniques to the data obtained from cycles 1 thru 19. References 9 and 10 summarize these current predictions, which are reported monthly by solar activities indices memos such as References 11 and 12. Until the second quarters of 1972, this method was used by NASA/MSFC for the current estimates of sunspot cycle 21. Beginning in the second quarter of 1972 the NASA/MSFC estimates for cycle 21 have been based on a modification of this method introduced by Sleeper in Reference 13. This modification consists primarily of an analysis of the similarity of sunspot cycles which has produced classification of the sunspot cycles from cycles 1 thru 20 according to their polarity and their mode. By limiting the selection of the cycles used in the linear regression forecast to that class of cycle which is being predicted, Sleeper has been able to obtain improved predictions.

In order to evaluate the advantages of the ADAPT prediction techniques, it is necessary to compare them with the currently-used techniques. For this purpose we shall call the application of the McNish and Lincoln linear regression

techniques to the first 19 sunspot cycles as described in Ref. 10 simple regression, and their application to a limited set of 9 negative cycles based upon the criteria outlined in Ref. 13 selective regression. In addition to the current simple regression and selective regression techniques, two separate ADAPT techniques are evaluated. The first of these is designated the ADAPT prediction technique, which refers to the algorithms developed for predicting the current sunspot cycle from the preceding two sunspot cycles. This provides the capability to extend sunspot activity beyond the present cycle. For completing the present cycle, the ADAPT programs have been used in their extrapolation mode to develop extrapolations of the present cycle. This is referred to as the ADAPT extrapolation technique. Both of these ADAPT techniques have been used to predict the 81 day running average of the daily Zurich sunspot numbers. This introduces some minor difficulties in comparing the ADAPT results with the simple and selective regression methods, since these latter methods have been used to estimate the twelve month running average of the mean monthly Zurich sunspot numbers.

This report will present the results of the studies carried out and the recommendations for the best method for estimating future solar activity, improvements which may still be made in the methods, and the application of the ADAPT techniques to further understanding of sunspot activity. The report also contains a description of the ADAPT programs and a detailed description of the efforts carried out to develop both the ADAPT predictions and extrapolations. Additional applications of the ADAPT programs to analysis of sunspot data are also outlined. The performance of the ADAPT sunspot estimates are compared with the performance of the currently used techniques.

2.0 RESULTS AND RECOMMENDATIONS

The primary results of the application of the ADAPT techniques to the problem of predicting future sunspot activity is a factor of 2 to 3 reduction in the RMS error of the 1-sigma estimate of the Zurich sunspot numbers for the remainder of the current sunspot cycle and the first half of the next sunspot cycle. To achieve this reduction in the error one must use both the ADAPT extrapolative and predictive algorithms. The present study, as well as comparison of the present study with that of Reference 13, provides evidence that further significant improvement in the ADAPT derived algorithms is almost certain if the analysis recommended in this report is carried out.

The best ADAPT estimate of cycles 20 and 21 as compared to the latest available conventional estimate (Ref. 12) is presented in Figure 2.1. In comparing the ADAPT estimates with the conventional estimates the reader must realize that the conventional estimates are for a 12-month running average while the ADAPT estimates are for an 81-day running average. The effects of this difference are primarily that the 12-month running average reaches a given value approximately 3 to 4 months after the 81-day running average has reached that value. The 81-day running averages should have higher peaks and lower minimums than the 12-month running average. The ADAPT analysis indicates that the next minimum for the 81-day running average will occur in February of 1977 which translates to May or June 1977 for the 12-month running average. Considering the three to four month correction which should be incorporated in the 12-month running average, ADAPT predicts generally higher values of the sunspot activity for the remainder of cycle 20 but approximately a 25% lower peak activity for cycle 21. For a detailed comparison of these ADAPT predictions with both the prediction in Fig. 2.1 and the latest predictions of Ref. 13, the reader is referred to Section 4.4. It is important to note, however, that the mid-1977 date for the next minimum and the greater sunspot activity during the remaining portion of cycle 20 is in remarkable agreement with best estimate presented by Sleeper in Ref. 13. Although the maximum sunspot number of 60 predicted for cycle 21 by ADAPT is lower than the peak values of approximately 80 predicted by Sleeper in Ref. 13, it is in better agreement with the trends of peak magnitude for negative cycles and the maximum sunspot number versus period correlations in Ref. 13.

Figure 2.2 compares the performance of the two ADAPT methods developed in this study with the simple and selective regression techniques. This curve also compares all four of these techniques with the simple assumption that the sunspot cycle is equal to the mean of sunspot cycles 1 thru 19. Again the reader is cautioned that the simple and selective regression as well as the mean sunspot cycle shown in this figure are for 12-month averages whereas the errors for the ADAPT predictions are for the 81-day running average. Figure 2.2 plots the expected 2-sigma error (i.e. 95% confidence limit) as a function of position as defined by number of months since start of cycle. Both the simple and selective regression

techniques have errors presented for long and near term estimates. The curves indicated by the circled numeral 1 are for the long term estimates. Here we define long term estimates as those estimates which use the available portion of the present cycle to predict the next cycle. The curves indicated by the circled numerals 2 and 3 designate the short term estimates using the simple and selective regression. The short term estimates are defined as estimates using the available portion of the present cycle to predict the remainder of the present cycle. Thus, in their functional use, the short term estimates correspond essentially to the ADAPT extrapolations and the long term estimates correspond to the ADAPT prediction. The ADAPT prediction performance is indicated by the solid line interrupted by circles. The solid lines interrupted by plus signs, crosses, and squares indicate the performance of the ADAPT extrapolations using 38, 76, and 93 months of the current cycle to extrapolate the remainder of the cycle.

Although detailed discussion of the conclusions is presented in Section 4.4, we may summarize the conclusions reached from this study as follows:

1. The best method for estimating future 81-day running averages of the Zurich sunspot number is to use the ADAPT prediction algorithm for estimating sunspot activity in all cycles for which less than 70 to 80 months of the current cycle are available. For those cycles for which 70 to 80 months are available the ADAPT extrapolation should be used. A simple interpolation from the current value to the ADAPT extrapolated value for the period in the immediate future, i.e. the next approximately 3 to 6 months will provide further improvement to this ADAPT estimate for the very near term.

2. The ADAPT predictions are approximately a factor of 2 to 3 better than the long term predictions based on either the simple or selective regressions over the first half of the sunspot cycle. The ADAPT predictions are similar to both the simple and selective predictions over the third quarter of the sunspot cycle and again the ADAPT predictions show a significant advantage for the end of the sunspot cycle.

3. The ADAPT extrapolations based on the first quarter to the first half of the data in the sunspot cycle is similar to both the selective and simple regression methods except for the immediate future (i.e. 3 to 6 months) when the ADAPT techniques are inferior since they do not use the knowledge of the present value to correct the immediate future. The ADAPT extrapolations of the first quarter and first half also shows significant advantages near the end of the cycle. The ADAPT third quarter extrapolation is significantly better than any of the selective or simple regression techniques except in the immediate future which can be corrected as outlined in conclusion 1 above.

4. Approximately three quarters of the variation between sunspot cycles occurs in the first half of the cycle. ADAPT's ability to account for this is the major contribution to the improved accuracy of the ADAPT predictions over the first half of the cycle.

5. The ADAPT prediction based on the preceding two cycles is better than the ADAPT extrapolation based on the first quarter of the cycle.

6. Incorporation of the preceding 2 cycles in the ADAPT extrapolation base will significantly improve ADAPT extrapolations using the first quarter and the first half of the cycle and may improve even the third quarter extrapolations.

7. The ADAPT derived algorithms can be significantly improved by the addition of variables such as those outlined by Sleeper, including such items as the angular momentum of the solar system, the polarity, and position in the 180 year cycle of the cycles being predicted.

8. For purposes of the prediction techniques currently available the 81-day running average may be assumed to be identical to the three-month running average.

9. After a period of three to five years, the simple regression techniques are equivalent to assuming that the predicted value is equal to the mean cycle of the cycles which are used to carry out the predictions.

10. Sunspot cycle 19 is an anomalous cycle.

The above conclusions are further confirmed by the summary of the RMS errors of various quantities which are presented in Table 2.1. The first column of Table 2.1 presents the RMS error of the 1-sigma error in the estimate. The second column presents the RMS error of the estimates of all of the sunspot cycles in the learning data. The third column designated by $23.4\,\sigma_{RAT}$ is an estimate of the quantity in the second column which can be performed relatively simply from the standard outputs of the ADAPT algorithms. The fourth and fifth columns present the RMS error for the predictions of cycles 19 and 20. There is considerable evidence presented in this report of the anomalous nature of cycle 19, and therefore, it is not a good basis for evaluating the ability of the learning data to project the performance of an algorithm. The anomalous nature of this cycle was indicated by some of the ADAPT validity criteria.

Analysis carried out in this report has shown that the ADAPT techniques should also be extremely useful for recovering earlier sunspot data, for predicting sunspot cycle properties and for performing cluster analysis. The ADAPT scatter plots clearly show the separation into the mode 1 and mode 2 cycles as introduced by Sleeper in Reference 13.

The most important recommendations resulting from this study are as follows:

1. Future estimates of the 81-day running average of the Zurich sunspot number should be based on the techniques outlined in conclusion 1 above and updated on approximately a quarterly basis.

2. The present ADAPT algorithms should be immediately upgraded using the existing data and technology which has been suggested elsewhere in this report.

3. Studies should be performed to use the ADAPT techniques in conjunction with conventional approaches to recover additional sunspot cycles.

4. After completion of the studies to recover additional sunspot cycles this information should be used to develop additional ADAPT algorithms for the long range forecasting of solar activity.

5. ADAPT should be used for clustering studies of sunspot cycles.

TABLE 2.1 - Comparison of RMS Errors for Various Prediction Methods

Method	RMS Error of 1-σEst	Learning	23.4σ_{RAT}	Cycle-19	Cycle-20*
Estimate is Mean of Learning Data	25.5	23.4	23.4	73.7	
Simple Regression (9/71 9/83)	25	---	---	---	
Selective Regression (3/72 3/84)	20.4	---	---	---	
ADAPT Prediction	13.8	17.4	19.7	44.3	13.7
ADAPT Extrapolation 1st Quarter Cycle	15.2	---	14	13.3**	8.8**
ADAPT Extrapolation 2nd Quarter Cycle	10.9	---	7	18.4**	7.8**
ADAPT Extrapolation 3rd Quarter Cycle	8.1	---	5	18.3	9.6**

* First 93 Months Only

** For Entire Available Cycle

FIGURE 2.1 COMPARISON OF ADAPT AND CONVENTIONAL PREDICTION FOR CYCLES 20 AND 21

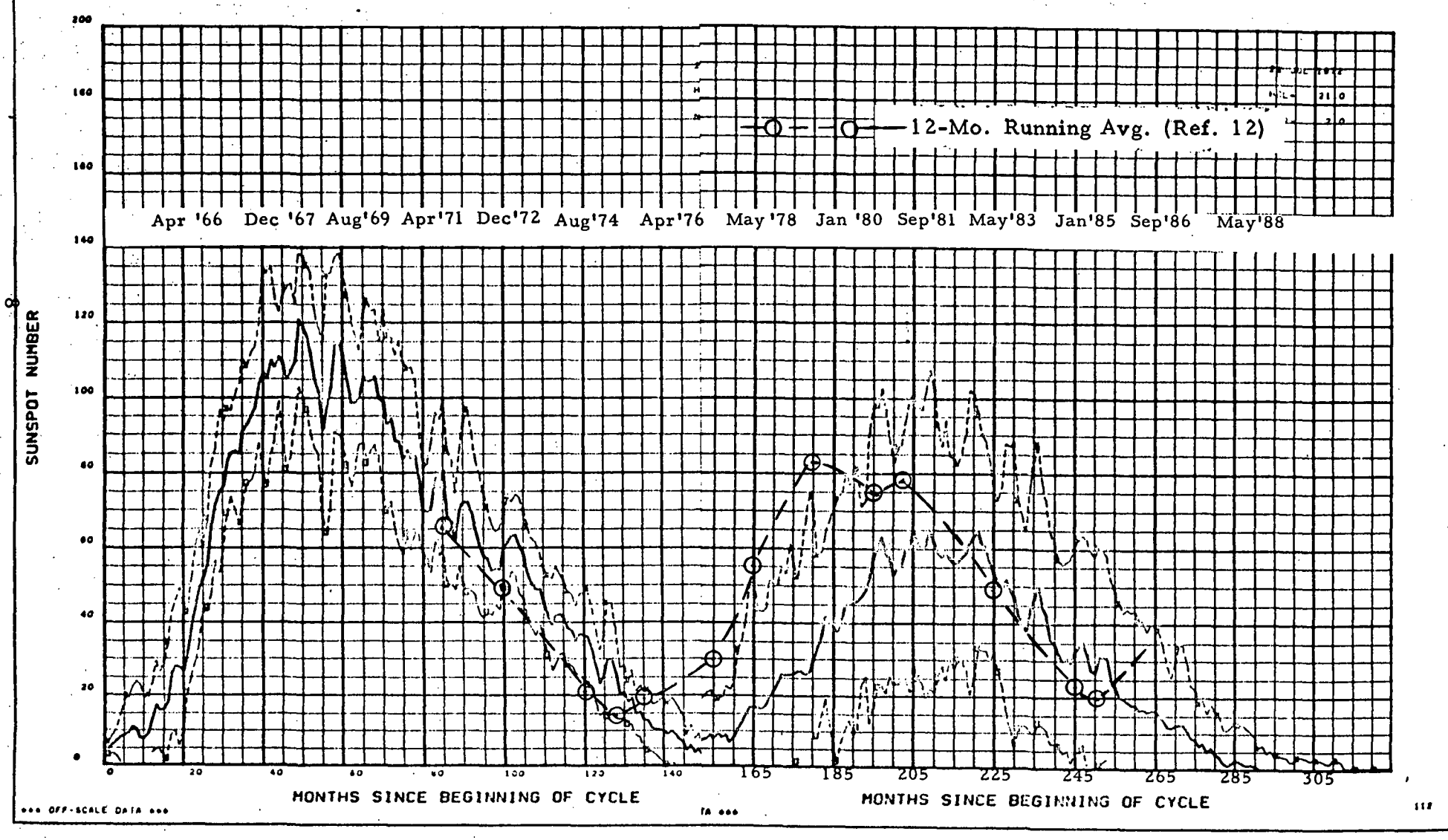
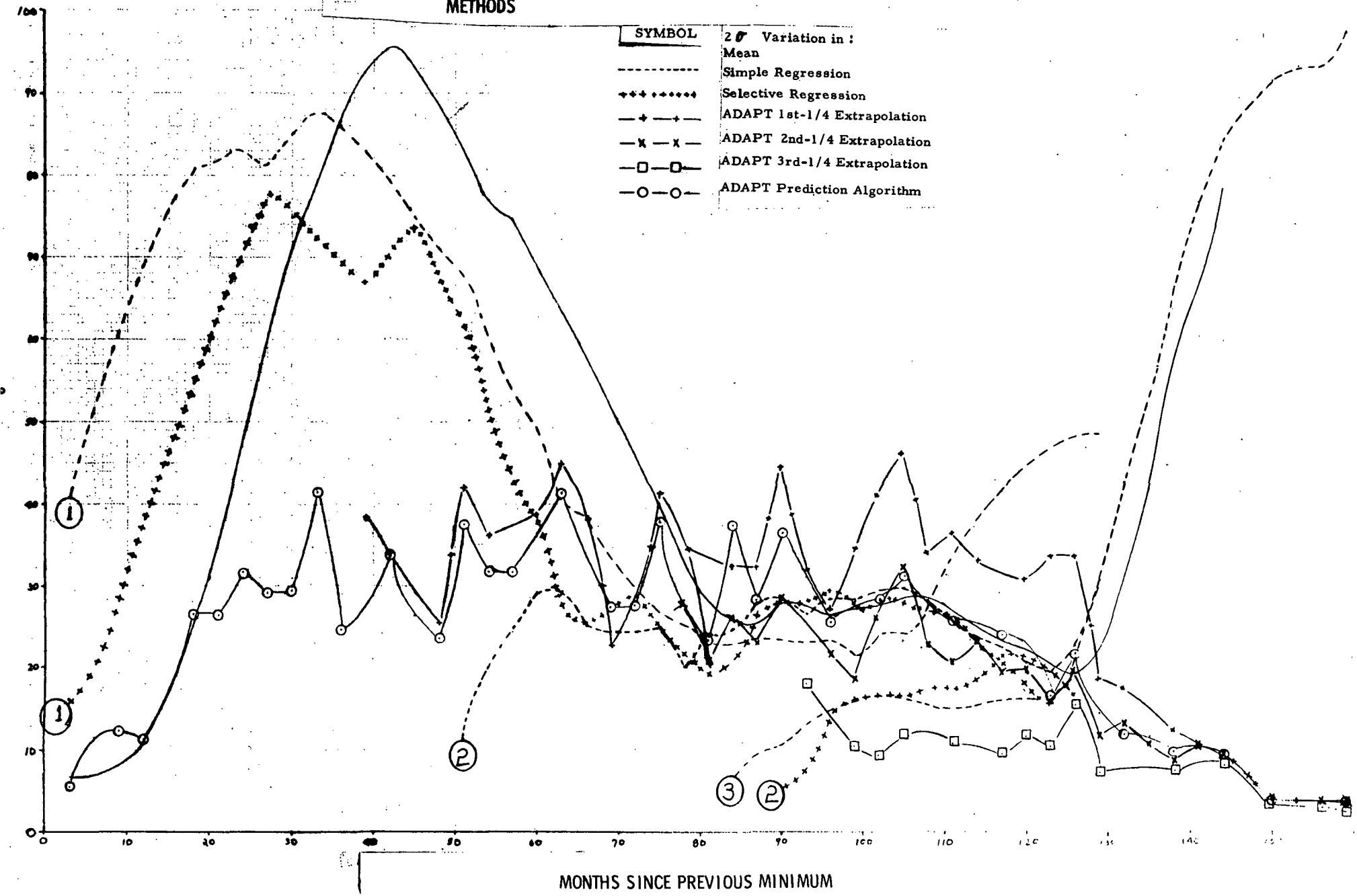


FIGURE 2.2 COMPARISON OF 2 SIGMA ERROR FOR VARIOUS PREDICTION METHODS



3.0 DESCRIPTION OF ADAPT

3.1 Definition of Data Histories

The ADAPT techniques address themselves to the representation and empirical analysis of data which appear as data histories, i.e., an indexed series of numbers. The features of ADAPT which make it advantageous for empirical analysis are reviewed in Appendix A. In the present case the indexing variable is time, in months. The histories may consist of numbers with different physical meaning; for example, quantities such as cycle type, mode, and/or position in the 180 year Jose cycle may be adjoined to the sunspot numbers. This was not done in the present study but offers an interesting method of incorporating additional information for the predictions.

The histories may be given in continuous (analog) form or in discrete form; since the ADAPT programs operate in digital computers, analog histories are each digitized into a finite set of N numbers, so a history is treated as an N-dimensional vector in Euclidean space. If there are M histories, the result is an N x M matrix of numbers.

The choice of the N numbers to represent each history is to some extent arbitrary, the chief criterion being that the desired physical phenomena are properly contained in the N numbers. It may be desirable to perform some pre-processing on the given data to bring out these features before entering the ADAPT programs. Such pre-processing could include Fourier transforms, normalization, taking logarithms, etc. From a theoretical viewpoint, one could even use continuous data at this stage, since the first step in ADAPT (discussed below) produces a discrete output even when the input is continuous functions instead of vectors. However, the realities of numerical analysis on digital computers require that the input be in vector (digitized) form rather than functional (analog) form.

3.2 Optimal Representation of Data Histories

With the M input history vectors defined, the first step in ADAPT is to construct from them an orthonormal set of base vectors by the classical Gram-Schmidt procedure. This ignores any history vectors linearly dependent on others, and results in a set of NC orthonormal N-component vectors where NC is less than or equal to the smaller of N and M. (The maximum number of linearly independent N-component vectors is N, so if $M > N$, some of the histories are surely linearly dependent on others. If $M < N$, then there are a maximum of M orthogonal base vectors.) The data history vectors are now expressed in the Gram-Schmidt base by their components along the NC Gram-Schmidt vectors, so each history is given by NC components, and there are M x NC components altogether.

This step has accomplished the task of discretizing the data, regardless of the form of the input or of its dimensionality. The Gram-Schmidt base vectors

have N components (or would be continuous functions if the input histories were functions instead of vectors) but the representation of the histories by their components in this base is independent of N, depending only on NC. Therefore, the Gram-Schmidt representation is largely independent of the particular way the data histories were digitized, assuming only that the numbers chosen to represent the histories properly contain their important features. In addition, there is usually a reduction in the number of numbers at this stage, since the M x N original components have been reduced to M x NC.

However, there is no reason to believe that the Gram-Schmidt base is the best one for representing the data. It is really an arbitrary orthonormal set of base vectors determined solely by the order in which the histories are arranged. The next step is to find another orthonormal base which is in some sense the best for the given data as a whole. *

To achieve this, a new set of NC N-dimensional orthonormal vectors, rotated from the Gram-Schmidt set, is postulated. This set is to be chosen in an ordered fashion, so that the first vector is the best, and so on. Only a limited number, NR $\leq$ NC, of these vectors will be used as new base vectors for representing the histories. They are chosen as follows: Each history is represented by its coefficients in the Gram-Schmidt base, and is projected onto the NR new vectors, giving M x NR components in the new base. If there were as many new vectors as Gram-Schmidt vectors, NR = NC, this would be an exact representation of the history vectors, but since NR$\leq$ NC, it is only approximate, leaving an error vector as the difference between the history vector and its representation in the new vector base. The square magnitude of this error vector is a measure of the error for each history, and the average of these square magnitudes for all histories is the mean square error incurred by representing the history vectors in only NR new base vectors.

The new orthonormal set of vectors is chosen by minimizing this mean square error, thus defining the meaning of a "best" set of vectors. If only one vector is used, NR = 1, it is that vector which makes the one-vector representation error the smallest. If a second vector is used also, it is chosen so that together

*The approach taken is analogous to the expansion of functions in a set of orthonormal functions, of which Fourier series is the most common example. When one of the classical boundary value problems of mathematical physics is solved, the appropriate differential equation defines a set of orthonormal functions. To satisfy a given function on the boundary, this boundary function is expanded in this set of orthonormal functions. In the present case, there is no differential equation to define a particular set of orthonormal functions. However, it is possible to make this data define its own best set of such functions, or vectors.

with the first vector, it minimizes the two-vector representation error. This is continued for as many vectors, i.e., as large a value of NR $\leq$ NC, as is necessary or desirable.

When formulated mathematically, this criterion requires the maximization of a quadratic form whose unknowns are the Gram-Schmidt components of one of the "best" base vectors, and whose coefficient matrix is the covariance matrix of the Gram-Schmidt components of the input histories. This problem is a classical one in linear algebra, which often appears under the name Karhunen-Loeve Expansion or principal components analysis of a matrix. * The solutions for the unknown vector components are the normalized eigenvectors of the covariance matrix, and the resulting values of the quadratic form are the eigenvalues of this matrix. Once they are obtained, they are simply arranged in order of decreasing size of the eigenvalues. The largest eigenvalue gives the most reduction in mean square error that can be achieved with only one new base vector and the corresponding eigenvector is this new base vector. The next largest eigenvalue gives the most reduction in the error that can be achieved by using a second new base vector in addition to the first one found above, and this second vector is the eigenvector of this second largest eigenvalue. This process can be continued until the desired accuracy is achieved. The sum of the NR largest eigenvalues gives the maximum mean square error reduction which can be achieved with NR new base vectors; when adding additional eigenvalues does not significantly increase this sum, the use of the corresponding eigenvectors as additional base vectors does not significantly improve the representation.

A convenient measure of the degree of representation achieved with a given number of base vectors is the sum of the eigenvalues of the vectors used, divided by the average square magnitude of the original data history vectors. This represents the reduction in mean square error achieved divided by the total error reduction possible; in statistical terms this is the percent of the variation of the data explained by the representation used. Since information is only conveyed by the variation in the data and the variation has the form of an energy, the percent variation explained is also known as the information energy. A similar measure of representation which is applied to the individual data vectors is the ratio of the square magnitude of the data vector in the NR base vector system to the original square magnitude of the data vector. This provides a measure of the adequacy of the empirically derived base for representing each history, and when applied to a test history serves as the basis for

*For a detailed discussion of the Kahunen-Loeve Expansion and its advantages in empirical data analysis see: S. Watanabe, "Karhunen-Loeve Expansion and Factor Analysis Theoretical Remarks and Application", Transactions of the Fourth Prague Conference Information Theory, Statistical Decision Functions and Random Processes, 1965, pp. 635-660.

the apriori test of the validity of applying the empirical data analysis to the test case.

For each history the NR components in the optimal system are the optimal representation of the data in the sense described above. Alternatively, these components may be interpreted as coefficients of the Fourier series of optimal orthonormal functions representing the history.

The optimal components are used in all further empirical analysis. Thus, the original M x N numbers representing M histories have been reduced to M x NR components, plus N x NR numbers to define the optimal vector base. Since the base system is optimal, the number of terms, NR, necessary to give a useful representation of history is small, often of the order of 10 or less, and the reduction in the number of numbers is usually large.

In the process described so far, the optimal vectors are represented by their NC components in the Gram-Schmidt base, but this means they are a linear combination of the NC Gram-Schmidt vectors, the coefficients being these NC components. Since the Gram-Schmidt vectors are N-dimensional vectors, the optimal vectors can also be represented in the original N-dimensional space of the data history vectors by performing the linear combination.

The ADAPT representation process just outlined can be clarified with the simple example of two input histories, which has been carried through analytically in Appendix B. For this special case the first optimal function is proportional to the average of the two history functions, the second to their difference, a result in accord with simple intuition. The relative sizes of the two eigenvalues is found to depend on the degree of correlation of the two histories, which has implications discussed later.

3.3 Sunspot Estimates and Analysis Using Optimal Representation

Having arrived at the optimal (Karhunen-Loeve) representation, attention is now turned to use of the optimal components for generating empirical algorithms to perform the parameter prediction and extrapolation required for this study. It should be noted that this optimal representation is also well suited for empirical clustering analysis, classification and clutter subtraction. For clustering analysis, one represents each history by a point in optimal coordinates, and the degree of similarity of two histories can be defined as the distance between the two points. If the optimal representations are normalized, this distance is simply related to the correlation of the two histories. Thus, the application of visual, nearest neighbor, or other cluster identification schemes to points (i.e. data histories) of the optimal space will lead to identification of natural clusters and algorithms to identify their members.

For classification (including the special classification problem of detection) the same representation of a history as a point in optimal coordinates is used. A number of parametric schemes and linear non-parametric schemes which can be applied are included in the ADAPT programs. They may be extended to multi-class problems by repetitive application, separating a different class with each application. If the statistics of the learning data are Gaussian the maximum likelihood technique, which is included as an option in ADAPT, may be used for multi-class classification problems.

Clutter subtraction is the unique capability of the ADAPT programs to subtract out characteristics associated with identifiable phenomena in the process of constructing the optimal space. The phenomena to be omitted from the optimal space is first characterized and then the directions associated with this phenomena are given a low or zero weighting in constructing the optimal space. The resulting optimal space or functions can then be used to reconstruct histories which will not contain characterizeable portions of the signature due to this phenomena.

The two types of empirical analysis which will be used in conjunction with the ADAPT representation to perform the prediction of future sunspot cycles and to extrapolate the present sunspot cycles are the use of parameter estimation and extrapolation techniques respectively. The mathematical basis for these operations will now be discussed.

The ADAPT technique for constructing an algorithm to predict a physical parameter associated with each history again makes use of the components of each history in the optimal system. For every history in the learning data, the known value of the parameter is written as a linear combination of the optimal components. The unknowns are the coefficients in this linear combination, which are taken to be the same for every history. The sum, over all histories, of the square error of this linear representation is then minimized to determine the coefficients. This amounts to a regression of the parameter on the optimal components. When the coefficients are found, they can then be used with optimal components of any new history to obtain an estimate of the value of the parameter for that history.

Parameter estimation may also be used to predict data histories rather than single parameters. The approach is to utilize the ADAPT parameter estimation programs to predict the components of the ADAPT representation of the history to be predicted. Thus, the ADAPT representation plays two roles in this type of analysis. The first role is to define the optimal coordinate system in which to represent the history to predict, so that the number of components which must be predicted is minimized. The second role of the representation is the usual ADAPT role of representing the data histories used as predictors,

to minimize the dimensionality of the space in which the regression is to be carried out. For the case of predicting future sunspot cycles both of these representations could be the same if a single sunspot cycle were used to predict the next sunspot cycle. However, for the present study it was decided, after consultation with NASA, that the preceding two sunspot cycles should be used to predict any given sunspot cycle. Thus, the components to be predicted are the components of the optimal representation for a single sunspot cycle. The data histories to be used to make this prediction, which must first be represented in the optimal coordinate system prior to carrying out the regression analysis, are a set of two adjacent sunspot cycles. Regression is then performed to relate the components of a single sunspot cycle to the components of the preceding two sunspot cycles. Once the components of the sunspot cycle have been predicted from this regression equation, they may be used in the Fourier series representation of the sunspot cycle, which provides the prediction of the sunspot number as a function of data.

It is not necessary to actually find the optimal coefficients of a new history which is being investigated to apply an ADAPT derived algorithm. The transformation from the N-dimensional data vector space to the NR-dimensional optimal vector space can be inverted and incorporated into the algorithm vectors. Then the process of applying this algorithm to a new data vector involves primarily the dot product or combination of dot products of this N-dimensional data vector with an N-dimensional algorithm vector or vectors, a rather simple procedure. Thus, the algorithm for predicting the coefficients of the next cycle can be expressed as a dot product of the end of the month values of the 81-day running averages of the daily Zurich sunspot number for the preceding two cycles with the ADAPT-derived relative importance vector.

ADAPT offers a unique approach to extrapolating data histories. The entire learning data history, including the region over which one hopes to eventually extrapolate is used to find the optimal representation for the histories. One then determines the best components for the history to be extrapolated by making a least square fit of the available portion of the history to a generalized Fourier series using the part of the optimal orthogonal functions which cover the available portion of the history. These components are then used to reconstruct the entire history from the complete optimal orthogonal functions. Clearly, the number of points in the known portion of the history must be greater than or equal to the number of components which are being estimated. In the special case where the number of points to be matched equals the number of components to be determined, an exact fit rather than a least square fit can be found. A detail description of the ADAPT history extrapolation program which performs this extrapolation is given in Appendix C.

3.4 Evaluation of Performance and Validity of Estimates

An objective of the ADAPT approach to empirical data analysis is to provide the analyst with information regarding both the performance and the validity of the

algorithms which he develops. Performance tells the analyst how good his algorithm is when it is applied to test data belonging to the same population as the learning data used to derive the algorithm. The validity criteria is a measure of how well the test data belongs to the population of the learning data. Thus, the availability of performance data allows the analyst to select the best algorithm and to verify that the performance of the algorithm is sufficient to insure that it is based on physics and not merely a fortuitous mathematical manipulation of the data. The validity criteria provides the user a measure of the applicability of the algorithm to the particular case being tested.

In the ADAPT programs the performance of regression algorithms is defined by the classical correlation coefficient and by the ratio, (σ_{RAT}) of the standard deviation of the error to the standard deviation about the means of the learning data. In the sunspot study these latter measures for the regression analysis are used to decide which algorithms should be used in the prediction of the coefficients of the sunspot cycle.

Equations 1 thru 3 define σ_{RAT} for the estimation of each coefficient of a data history.

$$\sigma_{RAT} = \frac{\sigma_{ERR}}{\sigma_{\bar{\sigma}}} \tag{1}$$

$$\sigma_{ERR} = \sqrt{\frac{1}{M}\sum_{j=1}^{M} (Y_j - \hat{Y}_j)^2} \tag{2}$$

$$\sigma_{\bar{\sigma}} \sqrt{\frac{1}{M}\sum_{j=1}^{M} (Y_j - \bar{Y})^2} \tag{3}$$

where $\hat{Y}_j$ is the estimated coefficient, $\bar{Y}$ is the mean of the coefficients of of the learning data, Y_j is the actual value of the coefficient of the ADAPT representation, and M is the number of cases used in the learning data. Note that this performance estimate, like all of the other performance estimates provided as part of the derivation of ADAPT algorithms, is based on the learning data.

The evaluation of the prediction of entire histories, such as the sunspot cycle, requires other methods of evaluating the performance of the prediction, rather than the correlation coefficients or σ_{RAT}. One classical method of evaluating the performance of the prediction of a data history is to consider the two-sigma error bounds about the history. This measure of performance is particularly useful for understanding the performance of any particular estimate as a function of the indexing variable. However, for comparing a large number of algorithms or methods of predicting histories, the large number of numbers involved makes the use of this measure rather awkward. To overcome this, a single number which summarizes the performance over the entire history has been used in this study to compare various methods and algorithms in the initial phases. This measure is the RMS error between the estimated and actual data history.

The calculation of the two-sigma band for a data history requires that one estimate the expected standard deviation between the estimated and actual data histories. This is an example of a problem for which the ADAPT formulation offers considerable simplification in the amount of caluclations required. Consider the task of evaluating the standard deviation between the estimated sunspot number, $\hat{R}jt$, and the actual sunspot number Rjt. Since ADAPT considers both the actual and estimated values as represented by Fourier series of optimal orthogonal functions the actual and estimated values can be expressed as:

$$R_{jt} = \overline{R_t} + \sum_{\ell=1}^{L} Y_{j\ell} H_{\ell t} \tag{4}$$

$$\hat{R}_{jt} = \overline{R_t} + \sum_{\ell=1}^{Q} \hat{Y}_{j\ell} H_{\ell t} \tag{5}$$

where L is the number of terms required to achieve 100% representation, Q is the number of terms utilized in the ADAPT representation used to estimate the sunspot number, the bar indicates the average value, the hat indicates the estimated value, the Y's are the coefficients of the generalized Fourier series utilizing the optimal orthogonal functions designated by $H_{\ell t}$. Note that the coefficients are independent of the indexing parameter t and that the optimal functions are the same for all histories (i.e., all j's).

Noting that the standard deviation, σ_t, is given by $\sqrt{\frac{1}{M}\sum_j (R_{jt} - \hat{R}_{jt})^2}$ at each value of time (in months) one may utilize equations 4 and 5 to write the standard

deviation of the estimate at time t as:

$$\sigma_t = \sqrt{\frac{1}{M} \sum_j \sum_{\ell=1}^{L} \left(\tilde{Y}_{j\ell} \; H_{\ell t}\right)^2} \tag{6}$$

where:

$$\tilde{Y}_{j\ell} = \begin{cases} Y_{j\ell} - \hat{Y}_{j\ell} & \ell \leq Q \\ Y_{j\ell} & L \geq \ell > Q \end{cases} \tag{7}$$

Thus, $\tilde{Y}_{j\ell}$ is the difference between the actual and the estimated coefficient (i.e. the error in estimating the coefficient) and simply equal to the actual coefficients for ℓ greater than Q. This follows from the fact that since only Q terms are being used in this analysis, this is equivalent to estimating zero for the coefficients of all of the terms beyond the Qth term in the series.

The standard deviation is found at each point in the data history and thus one has one value of the standard deviation for each indexing value applicable to all histories. The RMS error between the estimated and actual history will be defined in such a way that one has only a single value for the entire history. However, one will now have an RMS error for every history. The RMS error for history j is defined by:

$$E_{RMSj} = \sqrt{\frac{1}{T} \sum_{t=1}^{T} (R_{jt} - \hat{R}_{jt})^2} \tag{8}$$

Since there exists an RMS error for each history, one may define the expected average RMS error based on all of the histories in the learning data. This is the average of the RMS error when the algorithm is applied to the learning data,

and may simply be expressed as:

$$\overline{E}_{RMS} = \frac{1}{M} \sum_{j=1}^{M} E_{RMSj} \tag{9}$$

The calculation of the RMS error for the special case where the estimate is the mean of the learning data is another example when the ADAPT formulation significantly simplifies the computation involved. For this special case equations 4 and 8 combined with the orthogonal property of the optimal functions, $H_{\ell t}$ gives the following expression for the RMS error of the jth history:

$$E_{RMSj} = \sqrt{\frac{1}{T} \sum_{\ell=1}^{L} Y_{j\ell}^2} \tag{10}$$

in terms of the coefficients of the optimal series, $Y_{j\ell}$.

The RMS error between the nominal estimate and the one sigma error, $E_{RMS\sigma}$ can also be calculated by substituting σ_t for $R_{jt} - \hat{R}_{jt}$ in equation 8. Although this is not identical to the average RMS error, $\overline{E}_{RMS}$, both $E_{RMS\sigma}$ and $\overline{E}_{RMS}$ are estimates of the expected value of E_{RMSj} for any given algorithms.

The ADAPT programs also provide validity criteria which are based on the ability of the optimal functions derived from the learning data to represent the test data. These validity criteria are identical for and applicable to all ADAPT classification prediction and clustering algorithms. The validity criteria essentially makes use of the data vector's geometric property of length. The length of the learning data vectors may be calculated in the original data space and then compared with the new length when the learning data is represented in the optimal ADAPT space. A similar comparison can be made between the length of the test data vector as it is represented in the original data space and the optimal ADAPT space. If the test data vector's length is reduced significantly more than that of the learning data vectors when it is represented in the optimal space, this is indication that the test data is from a different population than the learning data used to develop that algorithm. Thus, it is not valid to apply the algorithm to that particular case.

The validity criteria for the case of extrapolated data histories must be modified since the learning data is now identical to the first portion of the data histories or sunspot cycles and was not used to make the data base. However, the data which was used to make the base also contains the portion covering the identical

range of the indexing variable as the learning portion of the data history to be extrapolated. One may then compute the RMS error, E_{RMS_j}, for the first portion of all the learning data histories. One may then take the average of this, finding the average RMS error, $E_{RMS\ L}$, for all the learning data histories and also the standard deviation $\sigma_{\bar{E}}$ of these RMS errors. One may then compare the RMS error of the test case with the average and standard deviation of the RMS error for the corresponding region of the learning data and calculate the confidence in the validity of the extrapolation. For example, if the RMS error of the test cases falls outside of the range of the average RMS error for the learning data plus or minus its two-sigma value one has only 5% confidence that the extrapolation will be accurate to the degree indicated by the performance estimate based on the learning data.

The next section of this report will present the detailed results of the representation, prediction and extrapolation of the sunspot cycles using the methods which have been outlined above.

4.0 ESTIMATE OF FUTURE SUNSPOT NUMBERS

4.1 Representation of Sunspot Data

The objective of the predictions which are being investigated in this study is to provide an estimate of the 81-day running average of the daily sunspot number. The daily sunspot numbers are available from 1890 or from cycle 13* to cycle 20. Since NASA desired to reserve cycles 19 and 20 as test cases, this would leave a total of 6 cycles for learning data. This is believed to be insufficient learning data to provide good algorithms. On the other hand, monthly sunspot numbers are available from 1750 or cycle 1 and if one could use the monthly data as learning data instead of the 81-day running average one would have 18 learning cases instead of 6.

If one formulates the 81-day running average of the sunspot numbers and compares this with the 3-month running average of the sunspot numbers, one discovers that 81 components of the two averages are identical and usually only 9 additional components are added to the monthly average. Thus, the total error made in using the three-month running average instead of the 81-day running average is the order of 10% of the difference in the average of these 9 components and the average of the 81-day running average. Since the 9 components of the average are adjacent to the 81-day running average, one would expect that their average would be quite similar to the 81-day running average. Since the total expected error is only 10% of the small difference between these averages, one would expect a very small error in using the three-month running average as an approximation to the 81-day running average. This is verified by Figure 4.1 which shows the difference between the 81-day running average and the three-month running average evaluated every five months for cycle 19. The maximum difference observed is less than 3 sunspots. This difference is considerably smaller than the expected error in any of the estimates which will be discussed in this report. Thus, we conclude that the three-month running average is a completely adequate approximation to the 81-day running average for the studies in this report and should be used since the factor of 3 gain in learning data significantly enhances the probability of success in this study.

The first step in an ADAPT analysis is to find the optimum representation of the data histories which will be used to predict any given quantity. For the case of the sunspot estimates there are two data bases of interest. The first data base is made up of single eleven-year cycles. This data base is required for both the extrapolation and the prediction approaches to estimating future sunspot cycles. In both of these cases this base will be used in an unconventional manner when compared to the normal ADAPT procedures. In particular, for the prediction

*The starting date for the cycles used in this study are given in Table 4.1

of future sunspot cycles based on the preceding two sunspot cycles this base will be used to reduce the approximately 180 numbers required to define a sunspot cycle to two numbers. Thus, we will only have to make two prediction algorithms rather than 180 prediction algorithms to predict the sunspot cycle.

The extrapolation of sunspot cycles only requires one base to carry out extrapolation as defined in Section 3. For this study the single cycle base will be used for the extrapolation. The double cycle base is the classical ADAPT base which will be used to predict the two numbers required to define the next sunspot cycle.

Figures 4.2 through 4.9 define the significant characteristics of the single cycle base. Figure 4.2 is the average of sunspot cycles 1 through 18 which were used to make the single cycle base. This average cycle was first subtracted from each of the 18 cycles and then these data histories were processed through the ADAPT programs to find the optimum empirical orthogonal functions to represent the sunspot cycles.

Figure 4.3 shows the amount of information energy or the amount of variation from the average input vector which is explained by each of the terms in the optimum generalized Fourier series expansion of the sunspot cycles. The first term in the optimal series explains slightly more that 60% of the variation from the average cycle. The second term explains approximately 18% of the variation, and thus the first two terms explain nearly 80% (as illustrated by the upper curve in Figure 4.3) of the variation in the sunspot cycles. Approximately another 10% of the variation is explained by the third through sixth terms of this optimal series. Examination of Figure 4.3 indicates that remaining eleven terms in the expansion provide very little additional physical information. This follows from the term-by-term or lower curve in Figure 4.3 which shows that from the seventh through the seventeenth term in the series the change in information energy as one goes from term to term is approximately equal.

The first optimal function, representing about 60% of the variation from the average is shown in Figure 4.4. The most striking feature of this first optimal function is that the great majority of the variation explained by this function occurs in the first half of the cycle (i.e., before month 76). This is extremely significant since more than half of the variation of the cycles from the average is explained by this function. This implies that a minimum of slightly more than a half of the entire variation from sunspot cycle to sunspot cycle occurs in the first half of the cycle. Moreover, this variation is the most highly correlated of all the variations occurring and is thus the most easily predicted.

Examination of the second optimal function shown in Figure 4.5 shows that this function, which explains the next 18% of the variation, provides approximately equal correction over the entire span of the sunspot cycle. Combining this with the conclusions from the first optimal function, one can estimate that between 70 and 80% of the variation between sunspot cycles occurs during the first 76

month of the sunspot cycles. This implies that extrapolations of the sunspot cycles should be quite good if one uses the first 76 month as a basis for the extrapolation. It also implies that if one is going to predict the sunspot cycle from the preceding sunspot cycles, the prediction of only the first coefficient will result in a very accurate prediction for the first 76 month of the sunspot cycle but will yield a prediction which is little better than using the mean of all of the cycles for the last half of the sunspot cycle. Thus, one believes that the minimum number of terms which must be predicted to provide a significant improvement in the second half of the sunspot cycle is two terms of the optimal series. This is consistent with the results of Figure 4.3, which shows that 80% of the variation will be explained by using these two terms.

Examination of Figure 4.3 indicates that the third through sixth terms, although making a relatively small contribution to the variation, might still contain some significant information. These four optimal functions are presented in Figures 4.6 through 4.9 and have the general characteristic that they apply approximately equally throughout the entire sunspot cycle. They also have the characteristic that they define specific spikes in the sunspot cycle. This follows from the spikey nature of these optimal functions. The one exception to this is in the rear-most portion of the sunspot cycle from approximately month 100 through month 160, where optimal functions three, four and six each appear to make a uniform contribution to the last portion of the sunspot cycle. This indicates that these three optimal functions may be providing a correction to the length of the sunspot cycle.

Figures 4.10 through 4.15 present similar information for the double cycle base. The Double cycle base is constructed using two adjacent eleven year cycles or approximately 22 year cycles of sunspot numbers. Just as in the case of the single cycle base one still desires to keep the nineteenth and twentieth cycles as test cases and thus the learning data for predictions of future cycles based on the preceding two cycles was limited to that data required to predict cycles three through eighteen. Since the two preceding cycles are not available for predicting cycles one and two, the number of learning cases available to make this base is reduced by two from the number of learning cycles available for producing the single cycle base. This base was constructed using cycles one-two; two-three, three four, . . . up to cycles sixteen-seventeen. Note that the double cycle seventeen-eighteen, although available for use in the representation would not be available as a predictor in the learning data since this double base would be used to predict cycle nineteen which is being withheld as proof data.

The average of these double cycles for cycles constructed from cycles 1 through 17 in the method outlined above is presented in Figure 4.10 and is the average input vector which was subtracted from each of the double cycle learning cases prior to constructing the optimal orthogonal functions for representing these histories. Since there are only 16 cycles in the learning data, a total of 15 optimal functions are sufficient to completely explain the variation in the data. The amount

of explained variation as a function of the number of optimal functions (i.e., number of dimensions in optimal space) used is presented in Figure 4.11. Here we see that for the double cycle the first optimal function only explains about 37% of the variation. However, the second, third and fourth optimal functions explain 25, 12 and 9% of the variation, respectively, which is considerably greater than the variation explained by the corresponding functions in the single cycle base.

Examination of Figure 4.11 shows that there are breaks or changes in the slopes of the explained variation occurring after the fourth, sixth, and twelfth optimal functions. Thus, the most likely groupings of interesting information consists of either the first through fourth optimal functions, the first through sixth optimal functions, or the first through twelfth optimal functions. Actually, it is quite likely that the last group, the seventh through twelfth optimal functions, explain peculiar characteristics of this set of learning data and will not be useful for analysis of sunspots. Thus, one could guess that the first six optimal functions would contain information which might be useful for predicting future sunspot cycles. We shall see in Section 4.2 that this is indeed the case. Thus, let us consider the first six optimal functions in more detail.

Comparison of the first optimal function for the double cycle base presented in Figure 4.12 with the single cycle first optimal function presented in Figure 4.4 shows that the most highly correlated portion of the variation is still from the first portions of the two eleven-year cycles which combine to make up the double cycle. The second optimal function for the double cycle base shown in Figure 4.13 has the same characteristic and thus it is now taking two optimal functions to explain the variation occurring in the first 76 month of the eleven year cycles.

The third and fourth optimal functions presented in Figures 4.14 through 4.15 provide relatively uniform corrections over the entire cycle and thus play a role similar to the second and third optimal functions of the one cycle base. Since the first two optimal functions explaining a total of 62% of the information deal with essentially the first 76 month of the eleven year cycle, we have a further confirmation of the result obtained from examination of the single cycle optimal functions that approximately 70 to 80% of the variation occurring over the eleven year sunspot cycle occurs in the first 76 month of the cycle. The fifth and sixth optimal function presented in Figures 4.16 and 4.17 appear to make detail corrections to the oscillations of the sunspot cycle and possibly minor adjustments to the length of the sunspot cycle.

Thus, we conclude that one should estimate at least two coefficients and no more than six coefficients of the optimal Fourier series representation of each of the cycles to be reconstructed. If these coefficients are to be estimated from the preceding two cycles the preceding two cycles should be represented by at least three

and not more than six terms in the optimum generalized Fourier series representation. We have also seen that the three-month running average of the monthly averages are a reasonable approximation to the 81-day running average. Thus, the remainder of Section 4 will explore the use of these two representations of the three-month running average to estimate future sunspot numbers either by extrapolation of the current cycle to its end or by predicting the future cycle from the preceding two cycles.

4.2 Predictions from the Preceding Two Cycles

Prior to constructing the reference algorithm both the dimensionality and the type of regression must be selected. For the prediction of sunspot cycles there are two dimensionalities which must be considered. The first is the number of terms which will be used in reconstructing the new sunspot cycle. The analysis of the ADAPT single cycle representation presented in the preceding section has already indicated that the reconstruction should be based upon between two and six terms in the optimal series. The second dimensionality which must be considered is that of the space in which the regression algorithm is derived. Again the analysis in the preceding section indicated that the dimensionality of this base should lie between 3 and 6. Thus, the remainder task is to select the best dimensionality within these ranges.

There are two general types of regression algorithms available in the ADAPT programs. The first is a canonical regression which amounts to a simultaneous regression between the independent variables and all of the dependent variables which are to be predicted. In the present case the number of dependent variables to be predicted is equal to the number of terms which will be used to reconstruct the sunspot cycle. A classical multiple regression which fits each dependent variable separately to all of the independent variables individually is also available in the ADAPT programs. The advantages of the canonical regression are twofold. First a single processing derives the algorithms for all of the dependent variables, thus saving computer time and manpower in deriving the algorithms. The more important consideration is that the simultaneous fitting of all of the dependent parameters to the independent variables makes it more difficult for the mathematics to make a fortuitous fit to the data which will not be applicable to the test cases. The disadvantage of the canonical regression (i.e., the advantage of the classical multiple regression) is that the canonical regression results in a slightly larger least square error between the estimated and actual values and thus does not provide as small a value of σ_{RAT} as is provided by the classical multiple regression technique.

Since the canonical regression is less expensive to apply to a large number of dependent variables, the first step in further refining the estimate of the dimensionality to be used was to apply the canonical regression for several different dimensionalities. The results of this are included in Table 4.2. Algorithms were derived using 4, 6 and 8 dimensions. In the case of the 4 dimensional

algorithm, 4 coefficients were predicted. In 6 dimensions 2 algorithms were derived, one for predicting 4 coefficients simultaneously and the other for predicting 6 coefficients simultaneously, and 6 coefficients were predicted in the 8 dimensional space. In each of these cases the performance of the algorithm for predicting each coefficient as measured by the standard deviation of the error in the estimate relative to the standard deviation of the input data about its mean, σ_{RAT}, and the correlation coefficient, ρ_{ZVK} are summarized in Table 4.2 The expected reduction in the standard deviation about the mean for the estimate of the entire sunspot cycle using the first few terms of the series is also presented in Table 4.2 and is designated by the quantity $\hat{\sigma}_{RAT}$. This quantity is simply calculated by summing the reduction in explained variation, E_ℓ, times the value of $\sigma_{RAT-\ell}$ for each term used in the estimate and adding to this the amount of the explained variation which is not included in the estimate. Thus $\hat{\sigma}_{RAT}$ is given by:

$$\hat{\sigma}_{RAT} = \sum_{\ell=1}^{Q} E_\ell \; \sigma_{RAT-\ell} + \sum_{\ell=Q+1}^{L} E_\ell \qquad (11)$$

This quantity, $\hat{\sigma}_{RAT}$, multiplied by the RMS error between the mean cycle and the learning cycles should approximate the RMS error for the prediction using Q dimensions. Table 4.2 also gives the performance of these canonical algorithms in predicting cycles 19 and 20. This performance is summarized as the root mean square error (see Section 3) between the estimated and actual values for these two cycles.

The canonical results presented in Table 4.2 lead to two conclusions. The first conclusion is that the prediction algorithms should be derived in the ADAPT optimum sixth dimensional space. Secondly, that algorithms derived from higher dimensional spaces will tend to be overdetermined; that is, a significant portion of the performance of the algorithm on the learning data is due to a fortuitous fit to the data and not the physics of the problem. The first of these conclusions is reached by noting that the best predictions of the entire sunspot cycle as indicated by either $\hat{\sigma}_{RAT}$ or the RMS error for cycles 19 and 20 which have been circled in Table 4.2 all occur for algorithms derived in a 6 dimensional space.

This conclusion is further enhanced by examination of the relative importance spectrum for predicting the 6 coefficients which are presented in Figures 4.18 through 4.23. The relative importance spectrum is related to the spectrum in

classical Fourier analysis except that the trigonometric functions have now been replaced by the optimum empirical orthogonal functions and frequency no longer has a physical interpretation but merely is a number identifying the term in the generalized Fourier series. The relative importance spectrum tells the importance of each of the optimal dimensions, in this case 6, to the particular algorithm in question. Thus, examination of the relative importance spectrum presented in Figure 4.18 indicates that the most important direction for calculating the first coefficient is the fifth optimal direction and that the second and sixth optimal directions make significant contributions to this prediction. Thus, it is clear if one were to use less than 5 dimensions, there would be a significant increase in the error associated with the prediction of the first coefficient of the next sunspot cycle. Similarly, Figure 4.19 shows that the sixth dimension is dominant in predicting the second coefficient and that the error in the prediction of the second coefficient would be significantly increased if all six coefficients were not used. The same conclusion applies to the prediction of the third coefficient as can be seen by examination of Figure 4.20. In general, by examining Figures 4.21 through 4.23 we see that the fifth and sixth optimal directions make significant contributions to all the predictions and thus should be retained.

The on-set of the "overdetermined" condition as one moves from six optimal dimensions to eight optimal dimensions can be seen by noting that although the performance of the learning data as indicated by σ_{RAT} and ρ_{ZVK} improve significantly as one increases the dimensionality of the space in which the algorithm is derived from 6 to 8, the performance of the algorithms on the independent test cases (i.e., cycles 19 and 20) decreases; that is, the RMS error is larger for the algorithm derived in the 8 dimensional space than in the 6 dimensional space. This is the characteristic of an overdetermined algorithm; namely, it has a significantly better performance on the learning than on independent test cases.

Based on these results 6 dimensions of the optimal space were used to derive the prediction algorithm. For this reason the classical multiple regression algorithms were only applied in 6 dimensions. The results of the application of these algorithms are also shown on Table 4.2 for the prediction of 4 coefficients in 6 dimensions. Note that since each coefficient is predicted independently, sets B and C are identical for the multiple regression (i.e., M.R.) algorithms. This is indicated by the X's in the performance regions of set C. Examination of the performance of the 6 dimensional algorithms for predicting the entire sunspot history reveals the interesting fact that the best performance for cycle 20 occurs when only 2 coefficients are predicted. The σ_{RAT} also indicates that the greatest gain in prediction accuracy is achieved in the first two coefficients since the decrease in this parameter as one goes from the second to third or third to fourth coefficients is quite small. Thus, we have an indication that one should use two coefficients for reconstructing the predicted sunspot cycle. The discussion of Table 4.2 should have made clear the advantages of a single performance criteria

for measuring both the performance of the prediction algorithm and the degree of matching between the predicted and actual sunspot cycles which were claimed in Section 3.

Table 4.2 also provides a basis for selecting the type of regression to be used. Since the major advantage of the canonical regression is in reduction of the likelihood of the "overdetermined" condition and since the standard multiple regression algorithm always performs better on the learning data, the only justification for using the canonical algorithm is its performance on the test data. If it performs better, it is an indication that a significant portion, i.e., sufficient to account for the difference between the multiple and canonical regression of the performance observed on the learning data is due to the "overdetermined" nature of the algorithm. Examination of the performance of the multiple and canonical regressions shows that for this case this is not true. In fact for cycle 20 the multiple regression algorithm has significantly better performance than the canonical algorithm. For cycle 19 the canonical algorithm has slightly better performance than the multiple regression; however, as will be discussed in Section 4.3, cycle 19 is an anomolous cycle and is probably not a valid cycle for making decisions as to the best way to construct the prediction algorithms.

Thus, the prediction of the future sunspot cycle will be based on the use of the classical least square multiple regression applied in the first six dimensions of the ADAPT optimal space to predict the first two coefficients of the single cycle generalized Fourier series representation of the sunspot cycle. The accomplishment of this prediction may be divided into two parts: 1) the prediction of the first two coefficients of the sunspot cycle and 2) the reconstruction of the sunspot cycle using these first two coefficients.

The relative importance spectrum for the algorithms recommended for predicting the first two coefficients of the next sunspot cycle is presented in Figures 4.24 and 4.25 for the first and second coefficients respectively. These may be compared with the relative importance spectrum obtained for the corresponding coefficients using the canonical regression which were presented in Figures 4.18 and 4.19. Comparison of Figures 4.18 and 4.24 show that both types of regression give very similar relative importance spectra and therefore similar algorithms for predicting the first coefficient. Comparison of Figure 4.19 and 4.25 show that the sixth optimal direction is dominant for both the canonical and least square multiple regression algorithms for predicting the second coefficient. However, the canonical prediction made considerable more use of the first, third and fourth coefficients than the least square multiple regression algorithms.

It is interesting to note that the prediction of the first coefficient is primarily based upon a term containing about 30 percent of the variation in the sunspot

cycles and has significant contribution from the first portion of each of the preceding two cycles. On the other hand, the prediction of the second coefficient is based on only approximately 3 percent of the variation of the data and has a relatively uniform contribution from both the first and second halves of both the preceding cycles. These conclusions are reached by comparing the information energy and relative importance spectra with the corresponding optimal functions presented in Figures 4.13 and 4.17.

Figures 4.26 and 4.27 present the relative importance vectors for these two algorithms. These relative importance vectors represent the vectors which when multiplied (dot product) by the sunspot numbers associated with the preceding two cycles will yield a number equal to the coefficient for the next sunspot cycle. Thus, the relative importance vector is the algorithm for predicting the coefficients of the next sunspot cycle, and as such also defines the importance of each portion of the sunspot cycle for predicting the next sunspot cycle. These same relative importance vectors are included in the tabulation of the algorithms which are presented in Table 4.3. Table 4.3 has been constructed so that it may be used independent of this report to calculate the coefficients of the next sunspot cycle.

The second step in constructing the sunspot cycle consists of utilizing the predicted coefficients in conjunction with the first and second optimal functions to reconstruct the sunspot cycles. The detailed procedure for this is outlined in Table 4.4. Briefly, this procedure consists of taking the average sunspot cycle presented in Figure 4.2 and adding to it, for each month the product of the first coefficient times the corresponding value of the first optimal function for that month plus the product of the second coefficient times the corresponding value of the second optimal function for that month. This procedure is carried out for each month in the cycle and the result will produce the predicted sunspot cycle history. This has been accomplished for the predictions of the learning data (cycles 3 through 18), the predictions of the proof test cycles (cycles 19 and 20), and for cycle 21. The resulting reconstructions for the learning data are presented in Appendix D. We will now discuss the reconstruction of the proof test cases.

Examination of Figure 4.26 shows that the decision to use the two preceding cycles rather than just a single preceding cycle to predict the future sunspot cycle was a wise one. We see that the second preceding sunspot cycle has slightly more influence on the prediction of the first coefficient of the sunspot cycle than the immediately preceding sunspot cycle. In particular, the second half of the first of the two preceding sunspot cycles makes a significantly greater contribution to the prediction than the corresponding second half of the sunspot cycle immediately preceding that being predicted. One also can see that if the preceding two cycles decrease in amplitude the first coefficient will tend to be larger than if the preceding two cycles have increasing amplitude. Examination of Figure 4.4 shows that if the first coefficient is larger, the first portion of

the sunspot cycle will tend to have lower sunspot numbers than the mean cycle. Thus, one may make the general observation that if the first fifty months of the preceding sunspot cycle have lower sunspot numbers than the corresponding fifty months of a sunspot cycle, the next sunspot cycle will tend to have a relatively slow rise in sunspot numbers as compared to the mean cycle.

Figures 4.28 and 4.29 measure the performance of each of these algorithms for predicting the coefficients of the learning data. The ordinate in these figures is the estimated value of the coefficient whereas the abscissa is the actual value of the coefficient; thus the solid line drawn on these figures represents a perfect prediction. The dash lines have been placed on these figures to indicate the approximate bounds of the error in the coefficient which would yield an error in sunspot number of $\pm$ 20.

Since the sunspot cycles are being represented by two numbers, namely, the coefficients of the first and second terms in the generalized Fourier series representation of the sunspot cycle, it is possible to display the cycles on a two dimensional graph. Figure 4.30 is such a display. This display is known as a scatter plot display and is simply a plot of the second coefficient of the optimal generalized Fourier series representation versus the first coefficient of this representation for each history. Thus, each sunspot cycle appears as a single point on this plot. The scatter plot presented in Figure 4.30 is constructed on the single cycle base and thus represents 80 percent of the variation or information contained in the sunspot cycles. We shall see later that this plot is very useful for studying groupings of sunspot cycles but it is also useful for comparing estimated and actual sunspot cycles. Figure 4.30 shows all of the actual sunspot cycle locations for cycles 1 through 20. These actual locations are indicated by the circles with the sunspot number shown inside the circle. The estimated position of the sunspot cycles is indicated at a sunspot cycle number enclosed in a square. If the estimated and actual sunspot cycle were to fall on the same place in this scatter plot that would indicate that the two term reconstructions would be identical. Since two terms of the optimal representation account for 80 percent of the variation in the sunspot cycles, it is a very good prediction. This scatter plot shows that the prediction of cycle 20 is rather typical of the predictions in the learning data. The prediction of cycle 19 is not very typical of the accuracy of the predictions in the learning data. However, since the actual value of cycle 19 is far removed from any other cycle on this plot there is a strong indication that cycle 19 is anomalous.

Table 4.5 compares the RMS error of the learning data and the proof test cases with the average and standard deviation of this RMS error. This figure also shows the value of the ADAPT validity parameter, Q for the two cycles used for each of the predictions. This verifies that cycle 20 is an extremely good prediction. Examination of the representation criteria (Q) indicates that cycle 19 has a relatively poor representation, namely .74 as compared to an average

representation of .83 with a standard deviation about this of .13. This representation can be taken as an indication that one should exercise some caution in utilizing the prediction for cycle 19.

The value of .74 for the validity criteria is sufficiently low that it would not pass the more severe validity test of requiring that the representation be greater than the mean representation of the learning data. However, this severe criteria will limit the applicability of the predicted algorithm to a maximum of approximately half of the cases to which it would be applied. More reasonable validity criteria for situations such as this where there is only one estimate upon which to make a decision is the mean minus 1 or 2 standard deviations. The validity criteria value of .74 would pass either of these two less severe but more realistic representation requirements. It appears from this that in terms of the predictive algorithms one can probably have high confidence in those predictions which have a representation test or Q value greater than approximately 80 percent. For Q values less than 80 percent one must still use the predictions even if the confidence is lower since approximately half of the valid cases will have such a value but it is possible that invalid cases would also be in this region. Thus, some caution must be exercised when one observes a validity criteria below .8.

Figure 4.31 shows the predicted sunspot number (solid lines) and two sigma bounds (dashed lines) on this prediction for cycle 19. This prediction is compared with the actual sunspot numbers (solid line) for cycle 19 in Figure 4.32. We see that there is a great discrepancy between the actual and predicted sunspot values for cycle 19, especially over the first 76 months of the sunspot cycle. This is entirely consistent with the scatter plot positions shown in Figure 4.30 for this cycle. The first optimal function presented in Figure 4.4 was completely dominated by the first 76 months of the cycle. Thus, a large error in the first coefficient of cycle 19 would result in an extremely large error in estimating the sunspot numbers over the first 76 months of the sunspot cycle. In particular, since the estimated value of the first coefficient is considerably larger than the actual value one would expect the prediction to significantly under predict the sunspot numbers for the first 76 months of the sunspot cycle. On the other hand, the prediction of the second coefficient is considerably better than the first coefficient and one would expect that the second half of the sunspot cycle as well as the length of the period might be predicted considerably more reasonably. Examination of Figure 4.32 shows this to be the case. In fact, since the minimum of the mean cycle is approximately five sunspots one should discontinue the predicted or dash curve in Figure 4.32 when it crosses the five sunspot value. This occurs at approximately 136 months which compares with the actual period of 125 months or just slightly less than a year in error. Furthermore, the actual sunspot numbers from approximately the 80th through the 120th month are in good agreement with the prediction. Figure 4.33 compares the actual sunspot number (solid line) for sunspot cycle 19 with the two sigma bounds on the prediction. Again we see that cycle 19 is extremely anomalous and as we compare the ADAPT results with other results we shall see that cycle 19 is indeed an anomalous cycle which should be included in the base but which is not likely to reoccur for at least 50 and possibly 150 or more years.

Figure 4.34 presents the predicted (solid line) and two sigma bounds (dashed lines) the sunspot numbers for sunspot cycle 20. This prediction was made using cycles 18 and 19, and is more typical of the performance observed on the learning data as can be seen from Figure 4.30. Examining Figure 4.35 which compares the actuals to date (solid line) with the prediction (dashed lines), one sees that the prediction for cycle 20 is indeed quite good. Figure 4.36 compares the actual values (solid lines) of cycle 20 with the two sigma bounds (dashed lines) on this prediction. We see that the two sigma bounds have been exceeded once at about 38 months after the beginning of the cycle. In evaluating the meaning of these two sigma bounds one must remember that the present predictions are made on a monthly basis and therefore for the typical cycle there are approximately 132 opportunities to exceed these bounds. Since the two sigma bound is the 95 percent confidence bound one would expect five to ten months during each typical cycle in which the actual values would exceed these two sigma bounds. Thus, the performance of cycle 20 tends to verify the validity of the two sigma bounds.

The same algorithm was used to predict cycle 21 and the predicted values of the coefficients for cycle 21 are presented on the scatter plot in Figure 4.30. The actual predictions for this cycle as well as the two sigma bounds about this prediction have been included in Figure 2.1 and represent the best estimate for cycle 21. This prediction will be discussed in more detail in Section 4.4.

4.3 Extrapolation of Sunspot Cycles

The extrapolation of sunspot cycles will be carried out using the single cycle base in the manner outlined in Section 3.3. As discussed in Section 4.1, examination of the single cycle base showed that one should use at least two and no more than six dimensions for the extrapolation of the history. To evaluate the effect of dimensionality on the performance of the extrapolations we shall use the parameter $\hat{\sigma}_{RAT}$ as defined in equation 11. Figure 4.37 presents this quantity for each of the three extrapolations which will be carried out in this section. The dash line in Figure 4.37 is the result that would be obtained if the first term on the right hand side of equation 11 were zero. In other words, this is the result that one would obtain as a function of the number of terms used if the estimates of all of the coefficients obtained by the extrapolation were perfect. Actually, the estimates of the coefficients will have some error and as the number of terms increased one would expect that prediction to improve. Thus, the value of $\hat{\sigma}_{RAT}$ should decrease until the point is reached where the extrapolation procedure no longer reduces the error in the estimated coefficient. At this point the performance of the algorithm will degrade until the overdetermined characteristic sets in. When this occurs the curve will approach the dash line. This behavior is illustrated by the solid or third quarter extrapolation in Figure 4.37. The first and second quarter extrapolations have only been carried through to their first minimum since one should use the number of terms at which this minimum occurs for the extrapolation.

This study will consider three different extrapolations. The first will use the first 38 months of the cycle to extrapolate the entire cycle, which is designated as the first quarter extrapolation. The second quarter extrapolation uses approximately half of a typical cycle or the first 76 months. The third quarter extrapolation uses 93 months of the cycle to extrapolate the entire cycle. This time for the third quarter extrapolation was picked so that there would be sufficient data to extrapolate the 20th cycle using this approach. In order to withhold cycles 19 and 20 as proof test cases these cycles were not included in the single cycle base for the extrapolation. The locations of the first minima on the curves in Figure 4.37 suggests one should use two terms for the second quarter extrapolation and four terms for the third quarter extrapolation.

Extrapolations were formed for both the learning data, cycles 1 through 18, the proof test data, cycle 19 and the test data, cycle 20. The extrapolation for all of these cases is carried out in the same manner. The portion of the cycle to be used as the basis for extrapolation, i.e., the first 39 months the first 76 months or the first 93 months, is substituted into the linear relationship between the coefficients to be predicted, the optimal orthogonal functions and the values of the sunspot number. One equation is obtained for each of the months for which data is available for extrapolation. Thus, we have 38, 76 and 93 equations for determining the two, two, and four unknowns for the first quarter, second quarter and third quarter extrapolations, respectively. This overdetermined problem is solved by a standard least square fit procedure to determine the best coefficients to satisfy the entire set of equations. These coefficients are then assumed to be the correct coefficients for the entire cycle. As in the case of the predictions the performance of the prediction can be evaluated to a great extent by simply examining these coefficients. Again, the scatter plot is a convenient way to examine them. For the case of the first and second quarter extrapolations which are performed in two dimensions the scatter plot shown in Figure 4.38 is a complete comparison of the estimated and actual values of the coefficients which will be used to predict this sunspot cycle. In the case of the third quarter extrapolation it is a comparison of the dominant information; however, two additional coefficients which are not shown on this figure will also be used in the prediction and may result in slightly different performance than would be obtained from the examination of Figure 4.38. Examination of this figure shows that the 3rd quarter extrapolation is significantly better than either the first or second quarter extrapolation in agreement with Figure 2.2. It is also interesting to note that extrapolation is the first prediction technique to give reasonably good performance for cycle 19.

The performance of each of these extrapolations for each of the dimensions considered on each of the learning and test cycles is summarized in terms of the RMS errors between the estimated and actual cycles in Table 4.6. The mean and the standard deviation of the RMS errors for the learning data are

also presented on this table. As discussed in Section 3.4 these values of the mean and standard deviation provide a validity criteria for the extrapolation for data histories. For example, if the RMS error of the extrapolated portion of a test history exceeds the mean of the RMS of the learning data plus twice the standard deviation of this RMS error of the learning data one knows that only 5 percent of the cases belonging to the population of the learning data could have values of the RMS error which were this large. Thus, it is quite reasonable to assume that this case is significantly different from the learning data and caution should be exercised when using this extrapolation. Examination of the RMS error for cycle 19 as compared to the means of the standard deviations shows that this 95 percent confidence level is exceeded. This then is a strong indication that cycle 19 is indeed an anomalous cycle. Thus, the ADAPT validity criteria does appear to work for the extrapolation.

Figure 4.39 through 4.56 compare the extrapolated data histories with the actual histories, with the anticipated two sigma variation in the prediction. Figure 4.39 presents the extrapolated sunspot cycle 19 and its two sigma bounds based on the extrapolation of the first 38 months of the cycle. Figure 4.40 compares the extrapolated history with the actual data history. Here we see the surprising result that despite the indication from the validity criteria that cycle 19 is an odd cycle, we have an unusually good estimate of this cycle when compared to other techniques. It must be pointed out that although this estimate is quite good compared to other techniques it is not nearly as good as the estimates which can be expected by this extrapolation for normal sunspot cycles. But even the first 38 months of cycle 19 provided sufficient information to allow significantly better prediction of this cycle than any other technique has been able to do by utilizing only data from preceding cycles. This appears to be an important attribute of the extrapolation technique, namely, that it has a better chance of accounting for the anomalous cycle than the prediction techniques. Figure 4.41 compares cycle 19 actual values with the estimated two sigma errors that would be expected from the 38 month extrapolation. One would expect only five to ten months during the sunspot cycle in which the two sigma bounds should be exceeded. Examination of Figure 4.41 shows that the two sigma bounds are exceeded for approximately 20 months of sunspot cycle 19. Thus, the validity criteria indication that cycle 19 is anomalous and its extrapolation would be poorer than expected is verified.

The first quarter extrapolation of the 20th cycle and its expected two sigma errors are presented in Figure 4.42. This cycle is compared with the actual values in Figure 4.43 and the predicted values are in good agreement with the actual values. When one compares the actual values with the expected two sigma errors in Figure 4.44 one finds that the two sigma error is only exceeded two times during this history. Thus, cycle 20 appears to be a reasonable extrapolation based on just the first 38 months of the cycle. Figures 4.45 through 4.47 present

the same data for cycle 19 based on the extrapolation using the first 76 months of the cycle. Here we see, as would be expected from examination of the optimal functions, very good agreement between the estimated and the actual although we still see an unexpectedly large number of cases for which the actual values exceed the expected two sigma error bounds. Figures 4.48 through 4.50 provide the same information for the second quarter extrapolation of cycle 20 and the conclusions are similar to the first quarter extrapolation with the exception that the accuracy of the extrapolation has been somewhat improved.

Figures 4.51 through 4.53 present the results of the third quarter extrapolation of cycle 19. The results are very similar to the first and the second quarter extrapolations of this cycle with the exception that the error bands have been significantly reduced. Figures 4.54 through 4.56 present the third quarter extrapolation of cycle 20 and again the only significant difference between the third quarter extrapolation and the extrapolation using 76 or 38 months of the cycle is the reduction in the two sigma error.

4.4 Comparison of Predictions

The preceding two sections have developed and presented the results of the two ADAPT approaches to predicting future sunspot numbers. Section 4.2 presented the ADAPT predictive approach which provides a capability to perform long term predictions. Section 4.3 presented the ADAPT extrapolation approach to completing the present cycle. The detailed results of the predictions for cycles 19, 20 and 21 for these two methods have been given in those sections. In this section we shall compare the results of the ADAPT predictive and ADAPT extrapolative predictions with the simple and selective regression models which have been used for predicting the sunspot numbers.

Comparison of Predicted Values

Figure 2.1 presents the comparison of the latest available estimate (June 1972) with the ADAPT estimate for sunspot cycles 20 and 21. In examining this figure it must be realized that the conventional prediction is for a 12-month running average and evaluated quarterly, whereas the ADAPT predictions are for 81-day running averages and evaluated monthly. This difference has three major effects on the predictions: the first is that the ADAPT predictions, being based on shorter running averages and evaluated more often, tend to have more of the detailed oscillations retained than the longer 12-month running average. The second effect is that since the 12-month running average contains data from earlier times the 12-month running average will reach a given sunspot number later than the 81-day running average. Third, the 12-month running average will tend to lower the peak values and raise the minimum values associated with the sunspot cycle.

Thus, we see there is considerable disagreement between the ADAPT methods and the current predictions of the sunspot cycles. The most significant of these is the disagreement in the time of the next minimum and therefore also the prediction of the time of the next maximum. Figure 2.1 shows a minimum on the current predictions of the 12-month running average of June 1975 as compared to a minimum of February 1977 for the ADAPT predicted curve. However, realizing that the ADAPT curve based on an 81-day running average will reach the minimum approximately 3 to 4 months earlier than the 12-month running average of the same data, February of 1977 is equivalent to April to June of 1977 for the 12-month running average. Thus, we see that the ADAPT predictions indicate that the end of cycle 20 will occur approximately 1 1/2 to 2 years later than the current predictions.

The ADAPT predictions presented here are a composite of the extrapolation for cycle 20 and the prediction for cycle 21. The extrapolation for cycle 20 is the best available extrapolation based on extrapolation the first 93 months of cycle 20 to the end of cycle 20. Since extrapolation techniques are only suitable for completing the present cycle, the prediction of cycle 21 was based on the predictive approach. The two predictions are attached together at the point where they each reach a value of approximately 5 for the sunspot number. This is based on the result that the means of the minimum of the 81-day running average sunspot number for the first 18 cycles is approximately 5.

The expected time difference between the 12-month running average and the 81-day running average implies that for the remainder of the present cycle one would expect the estimated of the 12-month running average to remain higher than the estimate of the 81-day running average. Thus, the fact that the 12-month running average lies slightly below the 81-day running average in Figure 2.1 is an indication that the difference between the present method of estimating sunspot numbers and the ADAPT extrapolation of cycle 20 is somewhat greater than would be indicated by Figure 2.1. Similarly, for the beginning of cycle 21 one would expect the 12-month running average to lie underneath the 81-day running average and therefore Figure 2.1 indicates considerable difference in the estimates for cycle 21. However, the major portion of this difference is due to the difference in the predicted time of the next minimum, i.e. the start of cycle 21. Clearly, the approximately year and half to two years later start of cycle 21 predicted by ADAPT accounts for the major difference between the estimates of cycle 21 based on ADAPT predictions and the conventional estimates. One other major difference in the estimates for cycle 21 is that the June projection for cycle 21 indicates a maximum sunspot number for the 12-month running average of slightly over 80 sunspots; whereas, the ADAPT prediction for cycle 21 indicates a maximum of the order of 65 sunspots. This is particularly significant if one realizes and recalls that the 12-month running average should tend to have lower peaks than the 81-day running average.

Since the June projection of the remainder of cycle 20 and cycle 21 sunspot numbers is a combination of the simple regression for the remainder of cycle 20 and the results of Reference 13 for cycle 21, it does not represent a fair comparison between the results of Reference 13 and the ADAPT predictions for the sunspot numbers through cycle 21. Figure 4.57 presents a figure similar to Figure 2.1 which compares the best estimate presented in Reference 13 (see Figure 4.67 of Reference 13) with the ADAPT predictions. This figure shows remarkable agreement between these two methods for the remainder of cycle 20. The estimate provided by Sleeper remains just slightly above the ADAPT 81-day running average for the remainder of cycle 20 which is exactly what would be expected based on the fact that the Sleeper prediction is for the 12-month running average whereas the ADAPT prediction is for the 81-day running average. Cycle 21 comparison between these two methods is identical to that in Figure 2.1 corrected for the approximately year and half to two-year difference in start time for cycle 21. That is, the Sleeper prediction for cycle 20 yields the same start time for cycle 21 as does the ADAPT extrapolation on cycle 20. Thus, this presents a better comparison of the cycle 21 predictions based on the selective regression proposed by Sleeper which utilizes only the 9 negative cycles to predict negative cycle 21. The only significant difference between these two predictions is that the ADAPT predicts a lower peak activity for cycle 21 than does the selective regression method.

Comparison of Expected Accuracy

The most significant way to compare the expected accuracy of the various methods of predicting the sunspot cycles is to compare plots of their 95% confidence bounds. These plots comparing the four ADAPT predictions, simple and selective regression techniques based on 18 sunspot cycles are shown in Figure 2.2. The mean value of the first 18 cycles is presented as a solid curve on this figure. The 3 dashed curves No. 1, 2, 3 represent the results of the simple regression. The dash curve 1 represents the results of the simple regression for predicting the next cycle based on a portion of the current cycle. Dash (-) curves 2 and 3 represent the results of predicting the remainder of the present cycle starting at 51 and 84 months respectively. The curves consisting only of plus (+) signs, known hereafter as the plus curves, present similar results for the selective regression technique. Again, the plus curve identified by Number 1 is the prediction of the next cycle from a portion of the current cycle. The plus curve Number 2 shows the results for predicting the remainder of the current cycle starting at month 90. The results for the simple and selective regression have been taken from References 10 and 13.

The solid lines interrupted by plus signs (+), crosses (x), squares (□), and circles (O) present the 95% confidence bounds for the ADAPT predictions. The solid line interrupted with plusses represents the ADAPT extrapolation based on using only the first quarter of the sunspot cycle. The solid line interrupted

by crosses represents the ADAPT extrapolation utilizing the first half of the sunspot cycle and the solid line interrupted by squares indicates the ADAPT extrapolation based on the first three quarters of the sunspot cycle. These extrapolations are essentially based on the same information as the dash curves 2 and 3 and plus curve 2. Comparison of these six curves shows that in general, with approximately half of the sunspot cycle available, the simple and selective regressions can be expected to give better estimates than the ADAPT extrapolations for periods of approximately 30 to 40 months. After that time the results of the two approaches become quite similar until near the end of the cycle, where the simple and selective regression approaches have difficulty associated with the large variation about the mean being introduced by the following cycle.

Comparison of the solid line interrupted with squares with the dash line 3 and the plus line 2 indicates that when approximately 80 months of the cycle are available the selective and simple regressions only hold their advantage for a period somewhat under a year after which the ADAPT extrapolation proves considerably better for the remainder of the cycle.

If one wishes to project from the present cycle to the next cycle the comparison of the solid line interrupted by circles with the dash line 1 and plus line 2 indicates that the ADAPT prediction is significantly better for the first half of the cycle and approximately equal to the other methods during the greatest portion of the second half of the cycle with the exception of the very back portion of the cycle when the simple regression has difficulties associated with the large variation expected around the beginning of the next cycle. It is interesting that ADAPT prediction from the preceding two cycles performs as well or slightly better than either the first or second quarter ADAPT extrapolations.

In the first approximately 70 to 80 months of the cycle both the selective and simple regression techniques have 95% error bounds significantly larger than the ADAPT methods. The reason for this is apparent if one recalls the shape of the first optimal function for representing the sunspot cycles. Figure 4.4 showed that the first optimal function is almost entirely composed of information in the first 70 to 80 months of the sunspot cycle. Recalling that this explained approximately 75% of the variation from the mean, it is clear that any technique which tends to compensate equally throughout the cycle will make much larger errors in this first 70 to 80 months due to the fact that this is where the greatest variation lies.

The simple and selective regression techniques are essentially methods of utilizing the average or mean of the preceding cycles as a basis for extrapolating the present cycles. This can be seen by considering the regression curves started at later dates in the cycle as indicated for the simple regression by the dash curves starting with the numbers 2 and 3. The number 2 dash curve starts

at approximately month 51 and by approximately month 80 has reached the dash curve 1 which represents the results of extrapolating forward an entire cycle. Dash curve No. 3 starts at approximately month 85 and reaches this extrapolation of an entire cycle forward approximately at month 120. Similarly, a selective regression prediction indicated by the plus curve 2 starting at month 90 also reaches the full cycle forward prediction based on the selective regression indicated by the plus curve 1 at approximately 120 month. From this we conclude that the effect of the regression portion of the simple and selective regressions is to buy 30 to 40 months of improvement over simply assuming that the next sunspot number is the mean of all the preceding sunspot cycles. In other words, the simple and selective regressions amount to an extrapolation procedure to account for the additional information that the knowledge of the present position of the sunspot cycle provides the user. We conclude that we would achieve similar results to both the simple and selective regression by applying these regressions over no more than a three or four year period and at the end of this three to four year period simply assuming the remainder of the cycle is the mean cycle.

Thus, it appears that the best methods currently available to predict future sunspot numbers are as follows: The prediction over the next three to six month period from any given time is presently best made using an interpolation between the predicted value and an extension of the present value using the predicted variation. If one has more than approximately half of the present cycle available the prediction to the end of the present cycle can best be accomplished by using the ADAPT extrapolation. Projection of the next cycle regardless of the position in the present cycle is best made by using the ADAPT prediction algorithms.

The preceding results also provide strong indications of how the prediction of future sunspot numbers can be further significantly approved. The first major improvement which can be made to the ADAPT techniques is to incorporate some extrapolative capability to make use of the knowledge that the sunspot number in the immediate future is very strongly influenced by the present value. This is particularly true because of the fact that one is using running averages and it is not possible for radical changes in the sunspot number to occur in very short times. Thus, it is recommended that an interpolative procedure be added to the ADAPT extrapolations to better account for the current value of the sunspot number.

The preceding 2 cycles contain extremely important information for the prediction of the sunspot cycle. On the other hand, the ADAPT extrapolations have shown that considerable advantage can be gained by utilizing information in the present sunspot cycle. Thus, it is recommended that the best procedure for estimating future sunspot numbers is to develop an ADAPT extrapolative procedure utilizing

the preceding 2 cycles plus the available portion of the present cycle. For long term estimates, the ADAPT prediction algorithms based on the preceding two cycles will give the best results. The fact that one has reasonably good estimates based on predicting the current cycle from the preceding two cycles suggests that this procedure should be good for a period of at least 22 years and probably significantly more. The next section of this report will summarize these recommendations as well as outline a program to implement it.

Table 2.1 presented a more compact summary of the errors of the various prediction techniques. The first column in this table presents the RMS error defined in Section 3 of the one sigma error band relative to the estimated value of the sunspot number. The second column presents the RMS error of the learning data when one assumes the estimate to be the mean of the learning data and for the ADAPT predictive techniques. An approximation to the RMS error is equal to the reduction in RMS error expected for particular algorithm times the RMS error achieved by using the mean of the sunspot cycles. This estimate of the RMS error is presented in column 3 which is headed $\hat{\sigma}_{RAT}$. This table also presents the RMS error observed for predictions of cycles 19 and 20 utilizing each of the methods.

The methods considered are simply taking the sunspot number as the mean of the corresponding point in the learning data histories, the simple regression over the period of September 1971 to September 1983, and the selective regression over the period of March 1972 to March 1984. These are compared with the four ADAPT estimates: 1) ADAPT predictions, 2) the ADAPT extrapolation over the first quarter cycle, 3) the ADAPT extrapolation using first and second quarters, and 4) the ADAPT extrapolation using the first three quarters.

Table 2.1 provides further verification of Figure 2.2, i.e., that considering the entire cycle one finds that the best prediction is made by the ADAPT extrapolation using the third quarter data, the next best is the second quarter extrapolation and the third best is either the ADAPT prediction or the ADAPT extrapolation of the first quarter. It is clear from the preceding discussion of Figure 2.2 that these gross summaries do not give the entire story because there are regions in which some of the techniques which show up relatively poorly as a predictor of an entire cycle show certain significant advantages for a portion of the cycle.

A more detailed examination of the performance of the ADAPT derived prediction on the learning data can be made by comparing the actual and predicted learning sunspot cycles. The information required to perform this comparison for the ADAPT prediction and third quarter extrapolation algorithm is presented in Appendix D.

4.5 Recommendations

The application of ADAPT to estimating future sunspot cycles which has been described in Section 4 leads to recommendations in two general areas. The first is that of defining how one should best make estimates of future sunspot numbers using the available algorithms. The second is the definition of analysis which should lead to significant improvement in the available algorithms for estimating future sunspot numbers.

Sunspot Estimates Using Available Algorithms

Based on the preceding analysis it is recommended that the prediction algorithm presented in Tables 4.3 and 4.4 be used to predict all future sunspot cycles (i.e., cycles beyond the current cycle) and for the first 75 months of the current cycle. For months 75 through the end of the current cycle the ADAPT extrapolation as described in Section 4.3 and Appendix C should be used. In both cases, the immediate future, that is the next 3 to 6 months, can be improved by interpolating between the predicted value and the value which would have been obtained by extending the current sunspot number utilizing the predicted variation for the next six months.

The short term (i.e. six months) correction to the ADAPT predictions has been recommended to overcome the disadvantage which ADAPT has as a result of predicting the entire sunspot cycle without insuring that the prediction actually goes through the most recent known values of the sunspots. It is believed that if the above recommendations are followed, predictions having approximately a factor of three improvement over any currently available can be achieved. It is also been shown that these prediction techniques offer an opportunity to provide reasonable predictions as much as two or more cycles in advance of the current sunspot cycle.

Improvement in Sunspot Prediction Techniques

Although the ADAPT analysis to date has produced significant improvement in the ability to estimate future sunspot numbers, considerably greater improvement can still be achieved by making use of what has been learned from this study and from the studies outlined in Reference 13. The improvements in prediction capability can be expected in the two areas of methodology and an improved data base.

The present study has shown that the two cycles preceding any given cycle contains significant information for predicting that cycle. Furthermore, the results from cycle 19 showed that this information is not completely redundant with the information contained in the first part of a sunspot cycle. Thus, it follows that one could significantly improve the extrapolations which have already

been quite successful using the ADAPT approach by using a 3 cycle rather than a 1 cycle base for the extrapolation. Both the extrapolation and the predictive approaches can be improved over the present results by including a procedure to account for the fact that the present value for the sunspot cycle is known and in general different from the present value estimated by the ADAPT approach. It is also possible to make use of the fact that negative sunspot numbers are inadmissible. This is perhaps more important for the studies which will be disucssed in Section 5 but could make some additional contribution to the accuracy of the estimates of future sunspot cycles. To incorporate this in the extrapolation algorithm requires the use of a nonlinear programming analysis in place of the least square fit for the extrapolation.

In addition to improving the methods as outlined above, it is clear from the results from the present study and of the work in Reference 13, considerable additional information is available which can be used to improve sunspot predictions. The first thing that should be included in the predictions is all of the available sunspot data. Avco believes that the present study has adequately demonstrated the advantages of the ADAPT approach to predict future sunspot cycles and any future applications should include all of the available data. Thus, it is proposed that as a minimum cycles 19 and 20 be included in the base for developing any further algorithms. In addition, Reference 13 has shown that there may be a high correlation between sunspot cycles and such quantities as the angular momentum of the solar system (dP/dt), the position of the sunspot cycle in the 180 year period, the polarity of the expected sunspot cycle and the mode classification of sunspot cycle. With the exception of the mode of the sunspot cycles, all of these quantities are known prior to the beginning of the prediction task. They may therefore be included in the data vector used to predict the sunspot cycle. It should be noted that the angular momentum is a data history in itself and may be included in the same way as the preceding sunspot cycles. It is suggested that an annual measure of angular momentum of the solar system for the preceding two cycles as well as for the period over which the sunspot history is to be predicted should be included in the data history. The position in the 180 year sunspot cycle should be included in two different ways to account for possible nonlinear effects. The first is simply to assign a value to this variable equal to the number of months from the start of the most recent 180 year period to the start of the present sunspot cycle. In addition to this, 16 binary variables should be introduced which have the value of zero except for that variable corresponding to the position (i.e. number of cycles since beginning of the 180-year history) for the sunspot cycle being predicted. Inclusion of these additional variables in the derivation will probably lead to improved predictions, and at the very least provide a conclusive determination of the importance of these variables to the estimation of sunspot number.

The present study has indicated that there is a reasonably high probability that application of the ADAPT techniques to the available data plus an analysis similar

to that carried out in Reference 14 can result in the addition of a significant number of sunspot cycles to the learning data. Justification and approach to accomplishing this will be presented in more detail in Section 5; however, to the degree to which this is successful more learning data will be available at the conclusion of such an effort. Clearly, this additional learning data should be included in the development of algorithms for predictions of future sunspot cycles. It may be possible to recover sufficient data to make the estimate of future 180 year cycles feasible. If this is the case, the prediction of the next 180 year cycle provides additional information which should be incorporated in the data history as learning data.

Based on the preceding discussion, it is recommended that a two-phase program be implemented to upgrade the prediction of future sunspot cycles. The first phase would be aimed at immediately improving techniques for estimating sunspot cycles. The second phase would be aimed at a long term upgrading of the techniques for estimating sunspot cycles, by making use of additional information and techniques such as the recovery of additional sunspot cycles prior to 1750 which would take a considerable length of time to achieve. This two-phase program is recommended since it is believed that significant improvements, even relative to the new ADAPT derived algorithms, are possible in a matter of months. This can be accomploshed by using all of the currently available data, the three cycle base for extrapolation, and the available auxiliary information such as polarity, angular momentum and position in 180 year history. On the other hand, it is also believed that after the use of the ADAPT techniques to carry out the recovery of additional sunspot cycles further significant improvements especially in the long range prediction (greater than 15 to 20 years) of the sunspot cycles are likely.

Immediate improvements are recommended for both the extrapolation and prediction techniques using the ADAPT technology. For both the extrapolation and prediction techniques it is recommended that the data base consist of all available data from cycles 0 through 20, as well as the polarity and position in 180 year cycle of the cycle being predicted. In addition, the angular momentum of the solar system over the period of the preceding two cycles and for the period of the cycle being predicted should be included in the data history.

The extrapolation should be based on the use of the data from the preceding two cycles as well as the available portion of the cycle being predicted. In addition a short term correction algorithm should be developed to account for the fact that the present values of sunspot number is known but slightly different from the extrapolated value for the present sunspot number. This correction algorithm would take as input the predictions over the next six to twelve months and the actual value and provide as output corrections to the predicted values to account for the present actual value of the sunspot numbers. The ADAPT prediction algorithms should be developed exactly as they were in the present study

with the improved data base described above. Analysis should also be carried out to determine the feasibility and complexity associated with introducing the constraint of positive values of the sunspot numbers into the extrapolation algorithm.

The long term improvement in sunspot prediction accuracy should rest primarily on the addition of the data developed by the studies recommended in Section 5. When these studies are completed it is recommended that the new data be used to develop improved algorithms in essentially the same manner as outlined above. It may also prove desirable to develop algorithms for predicting the 180 year cycle if sufficient sunspot data can be recovered. It also may prove possible to recover annual data considerably further back in time than monthly data in which case algorithms should be developed to predict annual sunspot averages for long periods in the future. Clearly, the detail definition of the phase two tasks must await the completion of the studies for recovering additional sunspot cycles which will be discussed in Section 5.

TABLE 4.1

SUNSPOT CYCLE START AND END DATES AS USED IN ADAPT ANALYSIS

CYCLE NO.	BEGIN DATE	END DATE
1	June, 1755	Aug., 1766
2	Aug., 1766	Sept., 1775
3	Sept., 1775	July, 1784
4	July, 1784	July, 1798
5	July, 1798	Sept., 1810
6	Sept., 1810	June, 1823
7	June, 1823	May 1834
8	May, 1834	Oct., 1843
9	Oct., 1843	Sept., 1855
10	Sept., 1855	Feb., 1867
11	Feb., 1867	March, 1879
12	March, 1879	April, 1890
13	April, 1890	June, 1902
14	June, 1902	June, 1913
15	June, 1913	March, 1924
16	March, 1924	Dec., 1933
17	Dec., 1933	June, 1944
18	June, 1944	June, 1954
19	June, 1954	Sept., 1964

TABLE 4.2

EFFECT OF DIMENSIONALITY ON PREDICTED ALGORITHM PERFORMANCE

SET	No. of Dimensions Used/No. Predicted Simultaneously	Coef. Predicted	σ_{RAT}		ρ_{ZVK}		$\hat{\sigma}_{RAT}$		RMS Error For Cycle 19		RMS Error For Cycle 20	
			MR	C	MR	C	C	MR	Canonical	Mult. Reg.	Canonical	Mult. Reg.
	4/4	1	--	1.05	--	.45	1.03	--	31.5	--	30.1	--
A		2	--	1.15	--	.33	1.06	--	30.9	--	30.9	--
		3	--	1.18	--	.30	1.07	--	30.8	--	34.6	--
		4	--	1.12	--	.38	1.07	--	30.8	--	34.3	--
	6/4	1	.77	.90	.54	.59	.94	.86	40.2	44.7	25.6	18.2
B		2	.90	1.07	.44	.43	.95	.84	39.6	44.3	22.7	13.7
		3	.75	.91	.67	.59	.95	.83	39.6	44.3	24.0	14.8
		4	.85	1.07	.55	.43	.95	.82	39.6	44.3	23.8	14.5
	6/6	1	X	.91	X	.59	.94	X	44.6	X	19.7	X
C		2	X	1.12	X	.37	.97	X	44.1	X	16.4	X
		3	X	.95	X	.55	.96	X	44.2	X	17.1	X
		4	X	1.12	X	.38	.97	X	44.3	X	17.2	X
	8/6	1	--	.88	--	.76	--	--	45.0	--	20.5	--
D		2	--	.71	--	.94	--	--	46.0	--	21.0	--
		3	--	.79	--	.8	--	--	45.9	--	20.8	--
		4	--	1.09	--	.76	--	--	46.1	--	20.9	--

TABLE 4.3

TO PREDICT THE FIRST COEFFICIENT OF SUNSPOT CYCLE i, CONSTRUCT A DATA HISTORY BY COMBINING CYCLE i-2, (FILL OUT CYCLE WITH ZERO VALUES UNTIL 180 POINTS IS OBTAINED) AND CYCLE i-1 (AGAIN FILL WITH ZEROS - 180 POINTS) AND APPLY THE FOLLOWING ALGORITHM:

$$Y^1 = A^1 \cdot SS - 28.49865$$

Mo. After Start of Cycle	A^1	Mo. After Start of Cycle	A^1	Mo. After Start of Cycle	A^1	Mo. After Start of Cycle	A^1	Mo. After Start of Cycle	A^1	Mo. After Start of Cycle	A^1	Mo. After Start of Cycle	A^1	Mo. After Start of Cycle	A^1
1	1.8110339D-03	46	4.8147024D-03	91	-2.4201762D-02	136	1.7476314D-02	181	-5.5013395D-03	226	4.1101308D-02	271	5.6663634D-02	315	-3.5859359D-03
2	5.2690946D-03	47	-5.7442852D-03	92	-6.9384030D-03	137	1.0702179D-02	182	-9.5701526D-03	227	5.6863433D-03	272	7.0006737D-02	316	-2.1449446D-03
3	7.4610355D-03	48	-2.5201867D-02	93	9.6009898D-03	138	8.8252418D-03	183	-1.4738586D-02	228	-4.0145129D-02	273	6.3031799D-02	317	2.2031929D-03
4	1.3319841D-02	49	-4.0228362D-02	94	2.6355914D-02	139	1.3424560D-02	184	-2.1369570D-02	229	-1.0251833D-01	274	5.4561352D-02	318	5.0351919D-03
5	1.5434493D-02	50	-4.9986403D-02	95	3.4109162D-02	140	1.6610122D-02	185	-2.2204501D-02	230	-1.0288475D-01	275	5.0658004D-02	319	8.1723122D-03
6	8.3521646D-03	51	-4.8803919D-02	96	2.8920344D-02	141	1.7097173D-02	186	-2.0693128D-02	231	-7.6341102D-02	276	3.6043440D-02	320	9.4119914D-03
7	4.0548408D-03	52	-3.6200348D-02	97	1.3265467D-02	142	1.6703510D-02	187	-2.3287912D-02	232	-6.6977573D-02	277	1.4565127D-02	321	1.2711272D-02
8	-1.3805064D-02	53	-1.8342074D-02	98	7.7183470D-03	143	1.5327003D-02	188	-2.1232328D-02	233	-6.7172416D-02	278	1.2308567D-02	322	1.5999036D-02
9	-1.3332246D-02	54	5.6104340D-03	99	2.7834631D-03	144	1.4793259D-02	189	-1.2739712D-02	234	-6.4059888D-02	279	1.9223502D-02	323	1.5837211D-02
10	-7.3142834D-03	55	-1.0428739D-02	100	8.4560445D-03	145	1.2732790D-02	190	-1.0081252D-02	235	-5.0879949D-02	280	2.2910905D-02	324	1.4242190D-02
11	-1.8987134D-02	56	-2.5914067D-02	101	1.2182780D-02	146	7.6401838D-03	191	-4.1870201D-03	236	-2.0996729D-02	281	2.5934543D-02	325	1.3291135D-02
12	-2.8292837D-02	57	-5.5356337D-02	102	1.2278028D-02	147	8.8149991D-03	192	-1.4581590D-02	237	-1.9410989D-02	282	1.6736508D-02	326	7.5843015D-03
13	-3.5170668D-02	58	-3.3662718D-02	103	1.1544937D-02	148	5.7086514D-03	193	-1.3046836D-02	238	-1.0794520D-02	283	7.2916614D-04	327	1.0723099D-02
14	-2.0826111D-02	59	-1.8696685D-02	104	1.8223659D-02	149	6.9032063D-03	194	-2.1314016D-02	239	-3.1262701D-02	284	-1.3623270D-02	328	7.3189709D-03
15	-1.8168357D-02	60	1.2272305D-02	105	1.7625547D-02	150	4.4906106D-03	195	-2.8213365D-02	240	-1.5941122D-02	285	-6.3186248D-03	329	8.9178736D-03
16	-1.1226647D-02	61	4.6878573D-03	106	1.9650790D-02	151	5.2290228D-03	196	-3.4199964D-02	241	1.4072229D-02	286	8.3987011D-03	330	5.8036956D-03
17	-2.9794611D-02	62	-8.1878232D-03	107	2.4423152D-02	152	4.4623334D-03	197	-4.4594349D-02	242	1.0174724D-02	287	1.5451021D-02	331	6.6401666D-03
18	-5.0265882D-02			108	3.5930122D-02	153	4.2432794D-03	198	-5.7411647D-02	243	-2.1479041D-02	288	1.7467037D-02	332	5.6492920D-03
19	-7.7439457D-02	63	-3.7550276D-02	109	3.7070150D-02	154	2.2453050D-03	199	-7.7702786D-02	244	-7.1548224D-02	289	1.2527582D-02	333	5.3651851D-03
20	-5.5807241D-02	64	-5.2503720D-02	110	2.9200934D-02	155	2.7929405D-03	200	-8.7195075D-02	245	-6.8896808D-02	290	7.1118168D-03	334	2.9018475D-03
21	-3.5239783D-02	65	-6.2292718D-02	111	2.6737550D-02	156	4.1072658D-03	201	-8.0152049D-02	246	-3.2200001D-02	291	-8.0248706D-03	335	3.6096153D-03
22	-1.0801435D-02	66	-4.1582210D-02	112	3.5775873D-02	157	4.1620291D-03	202	-7.1783860D-02	247	1.0405685D-02	292	-2.5916562D-03	336	5.3082583D-03
23	-1.2706208D-02	67	-5.3458015D-03	113	4.3336705D-02	158	3.8882114D-03	203	-6.0093466D-02	248	4.9381410D-02	293	8.8571478D-04	337	5.3790347D-03
24	-1.7310449D-02	68	4.0700340D-02	114	4.2358956D-02	159	2.9024674D-03	204	-5.7580499D-02	249	3.4284109D-02	294	1.2402186D-02	338	5.0251508D-03
25	-2.0372385D-02	69	4.9695517D-02	115	3.7033328D-02	160	3.3953397D-03	205	-5.5904322D-02	250	2.4300903D-02	295	6.3559962D-03	339	3.7511687D-03
26	-4.2040752D-02	70	5.8823104D-02	116	3.6032962D-02	161	3.3405759D-03	206	-5.4732735D-02	251	1.5467533D-02	296	7.5595675D-03	340	4.3881601D-03
27	-3.4672028D-02	71	3.3143486D-02	117	3.6368666D-02	162	2.8477042D-03	207	-6.1057129D-02	252	8.3740713D-03	297	7.3748845D-03	341	4.3173830D-03
28	-1.0485025D-02	72	1.1066178D-02	118	3.9591641D-02	163	1.9714873D-03	208	-4.1291728D-02	253	2.0222612D-03	298	1.0131261D-02	342	3.6803923D-03
29	-2.2408129D-02	73	-3.3440595D-02	119	4.0507882D-02	164	1.6429063D-03	209	-2.8199307D-02	254	4.8833530D-04	299	7.3807196D-03	343	2.5479636D-03
30	2.4404691D-02	74	-4.5667467D-02	120	4.1060002D-02	165	3.3405759D-03	210	-2.1147368D-02	255	1.3887980D-02	300	1.2147765D-02	344	2.1233033D-03
31	1.7706654D-02	75	-5.9259456D-02	121	3.9364965D-02	166	3.1762852D-03	211	-2.1753993D-02	256	2.8683041D-02	301	1.2341454D-02	345	4.3173830D-03
32	-1.2563384D-02	76	-4.9890998D-02	122	3.3026586D-02	167	2.4643595D-03	212	-4.7528055D-02	257	1.7872484D-02	302	1.4631959D-02	346	4.1050526D-03
33	-1.9721641D-02	77	-4.7322349D-02	123	2.5003302D-02	168	2.1905417D-04	213	-6.1054098D-02	258	2.7798125D-02	303	4.6521966D-03	347	3.1849550D-03
34	-3.5311465D-02	78	-2.1918664D-02	124	1.8532466D-02	169	0.0	214	-7.2391757D-02	259	5.4038410D-02	304	1.4028447D-02	348	2.8310709D-04
35	-3.9991259D-03	79	7.0530530D-03	125	2.6231189D-02	170	0.0	215	-5.1320389D-02	260	5.5759720D-02	305	2.2466520D-02	349	0.0
36	-4.4862687D-03	80	1.7475614D-03			171	0.0	216	-5.5276416D-02	261	6.0159465D-02	306	2.6898309D-02	350	0.0
37	1.9521671D-02	81	5.6879528D-03	126	3.1337260D-02	172	0.0	217	-3.1200081D-02	262	4.7518832D-02	307	1.8449397D-02	351	0.0
38	1.1352854D-02	82	-1.2334460D-02	127	3.6176468D-02	173	0.0	218	-7.8834888D-03	263	7.9239280D-02	308	1.1995358D-02	352	0.0
39	2.4245949D-02	83	1.2410716D-02	128	2.9738741D-02	174	0.0	219	5.2392707D-02	264	6.5849500D-02	309	7.9048300D-03	353	0.0
40	8.3526149D-03	84	1.5573906D-02	129	1.5652268D-02	175	0.0	220	6.9081695D-02	265	5.3455719D-02	310	1.0172918D-02	354	0.0
41	1.6298055D-02	85	1.7609058D-02	130	1.4723446D-02	176	0.0	221	8.4529904D-02	266	2.0682277D-02	311	7.0159677D-03	355	0.0
42	1.8855757D-02	86	-5.4324672D-03	131	1.2837791D-02	177	0.0	222	8.1951168D-02	267	2.8100507D-02	312	5.8673755D-03	356	0.0
43	2.3939086D-02	87	-1.8744188D-03	132	1.7420275D-02	178	0.0	223	9.6792431D-02	268	2.6768065D-02	313	1.8677422D-03	357	0.0
44	2.0673048D-02	88	-8.7378648D-03	133	1.7088664D-02	179	0.0	224	7.6704828D-02	269	2.5510894D-02	314	-1.3570695D-03	358	0.0
45	1.8754732D-02	89	-1.6483192D-02	134	2.0541873D-02	180	0.0	225	6.0721918D-02	270	3.5238711D-02			359	0.0
		90	-3.1354563D-02	135	1.8673066D-02									360	0.0

TABLE 4.3 (cont'd)

IN ORDER TO PREDICT THE SECOND COEFFICIENT OF SUNSPOT CYCLE 1 , APPLY THE ALGORITHM BELOW TO THE DATA HISTORY CONSTRUCTED ABOVE:

$$Y^2 = A^2 \bullet SS + 66.01652$$

Mo. After Start of Cycle	A^2	Mo. After Start of Cycle	A^2	Mo. After Start of Cycle	A^2	Mo. After Start of Cycle	A^2	Mo. After Start of Cycle	A^2	Mo. After Start of Cycle	A^2	Mo. After Start of Cycle	A^2	Mo. After Start of Cycle	A^2
1	3.72412800-03	46	1.50368530-01	91	5.65232850-02	136	1.92261300-03	181	1.12398680-02	226	-2.14224530-01	271	1.08035590-01	315	1.76301350-02
2	-2.06259490-03	47	8.60520620-02	92	8.80928620-02	137	8.87141000-03	182	1.68870240-03	227	-2.07519940-01	272	8.26299570-02	316	2.28794840-02
3	-1.03342120-02	48	-3.84379970-02	93	5.07069590-02	138	1.37815270-02	183	-9.46735250-03	228	-1.24232990-01	273	5.79240460-02	317	2.51227960-02
4	-1.54816300-02	49	-1.42248620-01	94	6.39640760-02	139	2.45003020-02	184	-2.87390820-02	229	-6.89066440-02	274	5.01731180-02	318	3.10814970-02
5	-8.46238040-03	50	-1.98932470-01	95	7.41444210-02	140	2.85109440-02	185	-1.98283170-02	230	-5.99960130-02	275	3.13028570-02	319	3.61273540-02
6	-4.97371910-03	51	-1.66984750-01	96	8.02486740-02	141	3.45970270-02	186	-3.99813700-05	231	-1.45806010-01	276	3.55538540-02	320	3.78133410-02
7	-4.02317590-03	52	-1.54838270-01	97	5.38052330-02	142	4.21029550-02	187	2.40376250-02	232	-1.73132240-01	277	3.85801880-02	321	3.41596240-02
8	-2.48139680-03	53	-1.27481940-01	98	4.60733790-02	143	4.17442950-02	188	3.95184160-02	233	-1.85763120-01	278	2.87612640-02	322	3.40744900-02
9	-5.95810070-03	54	-1.13395180-01	99	3.40986000-02	144	4.34366100-02	189	4.89573220-02	234	-1.92154380-01	279	8.90046990-03	323	3.19693790-02
10	-9.76737630-03	55	-8.85711430-02	100	4.28839630-02	145	3.97408550-02	190	5.61369590-02	235	-1.97439060-01	280	5.62456220-03	324	2.66009740-02
11	-2.42762690-02	56	-5.53745980-02	101	3.13433880-02	146	2.71540470-02	191	6.93313060-02	236	-1.98306390-01	281	3.27592340-03	325	2.34196330-02
12	-8.51379360-04	57	-6.68271240-02	102	7.02111170-02	147	3.05058290-02	192	7.27959870-02	237	-1.33959640-01	282	2.53828150-02	326	1.35414680-02
13	9.00510040-03	58	-5.58002790-02	103	8.39084090-02	148	1.92908900-02	193	7.99673890-02	238	-7.24015830-02	283	1.40286790-02	327	1.94520050-02
14	1.79901540-02	59	-8.89168370-02	104	1.11227190-01	149	2.35320770-02	194	4.88125780-02	239	-2.89453660-02	284	3.27112320-02	328	1.31223350-02
15	-2.84407260-04	60	-5.32190070-02	105	9.98089290-02	150	1.53145270-02	195	2.94826800-02	240	-6.03581750-02	285	2.25537400-02	329	1.59794740-02
16	1.62977430-02	61	8.42634610-03	106	7.54511530-02	151	1.74776160-02	196	1.28691450-02	241	-5.02743560-02	286	8.34220920-03	330	1.03993400-02
17	2.75085450-02	62	-3.73574300-04	107	4.80671680-02	152	1.48629410-02	197	5.68339610-02	242	-6.01193570-02	287	-8.06577400-03	331	1.19138730-02
18	1.32581240-02	63	-6.94587560-02	108	3.56494850-02	153	1.41158920-02	198	8.52071820-02	243	-7.68836460-03	288	-3.27283480-02	332	1.01383760-02
19	-4.74484240-02	64	-1.85131160-01	109	4.35517640-02	154	7.65726250-03	199	9.26237910-02	244	-5.67433520-03	289	-2.97682280-04	333	9.63109190-03
20	-5.88091250-02	65	-1.52734120-01	110	3.26685530-02	155	9.52488780-03	200	2.74310710-02	245	3.29249880-02	290	1.65881410-03	334	5.19966960-03
21	-4.08359270-02	66	-9.43766550-02	111	6.61660660-03	156	1.40071890-02	201	-3.18777940-02	246	3.26727880-02	291	-9.67313800-03	335	6.46788180-03
22	1.52271720-02	67	-1.16857470-02	112	1.59898990-02	157	1.41939510-02	202	-6.87557050-02	247	-1.37157960-03	292	-1.89818490-02	336	9.51159200-03
23	-4.81736400-02	68	-5.35745100-02	113	2.68069430-02	158	1.32601380-02	203	-6.43387850-02	248	-5.71621710-02	293	-3.18759080-02	337	9.63841250-03
24	9.28711020-02	69	7.71978340-02	114	4.56385870-02	159	9.89841250-03	204	-4.52686820-02	249	-9.71561220-02	294	-8.04814630-03	338	9.00430640-03
25	8.71661820-02	70	-9.07281490-02	115	2.37371210-02	160	1.15792760-02	205	-5.55641790-02	250	-1.96651010-01	295	-2.38513160-02	339	6.72152400-03
26	9.12805210-02	71	7.72864440-02	116	3.21397880-02	161	1.13925130-02	206	-3.52035880-02	251	-1.83704980-01	296	-9.41021540-03	340	7.86291580-03
27	8.87842790-02	72	5.07290750-02	117	3.33856780-02	162	9.71165100-03	207	-3.56504790-02	252	-1.63969370-01	297	-1.01492800-02	341	7.73609410-03
28	1.50744720-01	73	2.59355750-02	118	3.43273360-02	163	6.72344990-03	208	-4.49789770-02	253	-9.00070630-02	298	1.27190250-03	342	6.59470350-03
29	1.82781500-01	74	1.80544010-03	119	3.14570980-02	164	5.60287580-03	209	-6.67519740-02	254	-4.05195730-02	299	2.19306500-02	343	4.56556340-03
30	2.03142150-01	75	-3.28441000-03	120	3.22088280-02	165	1.13925130-02	210	-2.25137860-02	255	2.86424980-02	300	3.10548080-02	344	3.80463680-03
31	1.97157210-01	76	1.03133090-02	121	3.47759070-02	166	1.08322250-02	211	1.78846880-02	256	8.34877510-02	301	2.84184630-02	345	7.73609410-03
32	1.99444020-01	77	3.11122010-02	122	4.46083670-02	167	8.40431370-03	212	6.30666760-02	257	1.14569830-01	302	2.33353920-02	346	7.35563020-03
33	2.11168850-01	78	7.83643180-02	123	3.90426980-02	168	7.47050060-04	213	3.49984920-02	258	1.09939940-01	303	1.44735330-02	347	5.70695520-03
34	2.32736740-01	79	1.24509930-01	124	4.53529610-02	169	0.0	214	2.77361320-02	259	1.00798600-01	304	1.24561860-02	348	5.07284880-04
35	2.40049500-01	80	8.36170470-02	125	5.05287860-02	170	0.0	215	-3.10683490-02	260	6.01529470-02	305	1.71245390-02	349	0.0
36	1.41967560-01	81	6.09135520-02	126	5.46574860-02	171	0.0	216	-6.86087490-02	261	-1.59219160-02	306	2.29246910-02	350	0.0
37	1.16569610-01	82	1.63696520-02	127	4.76813650-02	172	0.0	217	-1.14135700-01	262	-5.52460990-02	307	2.52897430-02	351	0.0
38	9.27287610-02	83	9.41237490-02	128	3.66735900-02	173	0.0	218	-1.12525390-01	263	-4.18132580-02	308	2.71517230-02	352	0.0
39	2.01333820-01	84	7.34583360-02	129	2.01122250-02	174	0.0	219	-6.20655990-02	264	-1.36196100-02	309	2.36830590-02	353	0.0
40	2.16104400-01	85	1.08217640-01	130	3.35417380-02	175	0.0	220	1.81611230-02	265	-7.60588720-03	310	1.90889880-02	354	0.0
41	2.41061200-01	86	3.58525460-02	131	3.24922930-02	176	0.0	221	3.12557220-02	266	-1.22371340-02	311	1.88220000-02	355	0.0
42	2.13538420-01	87	7.44894210-02	132	4.36863710-02	177	0.0	222	1.61969220-02	267	1.67274220-02	312	2.39886120-02	356	0.0
43	2.00783730-01	88	5.31280910-02	133	2.31686390-02	178	0.0	223	-3.86904020-02	268	6.13072630-02	313	2.40501010-02	357	0.0
44	1.75998620-01	89	4.75231500-02	134	1.56804680-02	179	0.0	224	-8.97236330-02	269	1.01936340-01	314	2.11086020-02	358	0.0
45	1.83608140-01	90	5.67930970-02	135	2.16896380-03	180	0.0	225	-1.75395610-01	270	1.18878460-01			359	0.0
														360	0.0

TABLE 4.4

No. Month After Start	Mean Cycle	H_1	H_2
1	0.2294D 01	-0.61157D-02	-0.422990-02
2	0.3617D 01	-0.74578D-02	-0.478000-02
3	0.4944D 01	-0.11099D-01	-0.434460-02
4	0.6611D 01	-0.10261D-01	-0.3C838D-02
5	0.8528D 01	-0.12008D-01	0.770740-02
6	0.1016D 02	-0.12121D-01	0.145700-01
7	0.1193D 02	-0.16170D-01	0.151640-01
8	0.1294D 02	-0.18012D-01	0.221850-01
9	0.1286D 02	-0.13242D-01	0.246610-01
10	0.1274D 02	-0.15882D-01	0.285580-01
11	0.1342D 02	-0.14282D-01	0.134360-01
12	0.1538D 02	-0.26348D-01	0.202780-01
13	0.1752D 02	-0.32471D-01	0.194130-01
14	0.1860D 02	-0.42414D-01	0.419380-01
15	0.2110D 02	-0.46694D-01	0.483110-01
16	0.2459D 02	-0.5C235D-01	0.626200-01
17	0.2901D 02	-0.60342D-01	0.6C5850-01
18	0.3179D 02	-0.62896D-01	0.652050-01
19	0.3467D 02	-0.65395D-01	0.766780-01
20	0.3632D 02	-0.69587C-01	0.856260-01
21	0.3897D 02	-0.81661D-01	0.819110-01
22	0.4310D 02	-0.98780C-01	0.847200-01
23	0.4734C 02	-0.11090D 00	0.837770-01
24	0.5151D 02	-0.11643D 00	0.108580 00
25	0.5344D 02	-0.11723D 00	0.107370 00
26	0.5636C 02	-0.11981D 00	0.101800 00
27	0.6029C 02	-0.12892C 00	0.803970-01
28	0.6371D 02	-0.14273D 00	0.815150-01
29	0.6729D 02	-0.15572C 00	0.930660-01
30	0.6703D 02	-0.15447D 00	0.853320-01
31	0.6744D 02	-0.15531D 00	0.605810-01
32	0.6739D 02	-0.15486D 00	0.708000-01
33	0.7256D 02	-0.17458D 00	0.120540 00
34	0.7563D 02	-0.17981C 00	0.151630 00
35	0.7856C 02	-0.18208D 00	0.120960 00
36	0.8010D 02	-0.17909D 00	0.639620-01
37	0.8322D 02	-0.18153D 00	0.400110-01
38	0.8947D 02	-0.18536D 00	0.505450-01
39	0.9237D 02	-0.19999D 00	0.631530-01
40	0.9201D 02	-0.20563D 00	0.460060-01
41	0.8729D 02	-0.19566D 00	0.221C40-01
42	0.8594D 02	-0.17E35D 00	-0.212590-02
43	0.8586D 02	-0.16187C 00	0.113000-02
44	0.8578D 02	-0.16108D 00	-0.206370-02
45	0.8404C 02	-0.15343C 00	-0.357630-02

No. Month After Start	Mean Cycle	H_1	H_2
46	0.8031D 02	-0.13976C 00	-0.134C3D-01
47	0.8403D 02	-0.14892D 00	-0.30431D-01
48	0.8575D 02	-0.15205D 00	-0.707550-01
49	0.8287D 02	-0.17193D 00	-0.10370D 00
50	0.8421D 02	-0.15505D 00	-0.10938C 00
51	0.8421D 02	-0.15198D 00	-0.96913D-01
52	0.8399D 02	-0.14023D 00	-0.91696C-01
53	0.8907D 02	-0.12935C 00	-0.95099D-01
54	0.8299D 02	-0.11047C 00	-0.950850-01
55	0.7678D 02	-0.10140C 00	-0.720210-01
56	0.7842D 02	-0.98325D-01	-0.767850-01
57	0.8155D 02	-0.11061D 00	-0.807930-01
58	0.8583D 02	-0.11325C 00	-0.10609D 00
59	0.8751D 02	-0.11693D 00	-0.11453D 00
60	0.8499D 02	-0.91659D-01	-0.11901D 00
61	0.7968D 02	-0.82706D-01	-0.113590 00
62	0.7644D 02	-0.75117D-01	-0.110510 00
63	0.7784D 02	-0.79304D-01	-0.125460 00
64	0.8124D 02	-0.77372D-01	-0.125390 00
65	0.8372D 02	-0.98442D-01	-0.108050 00
66	0.8336D 02	-0.10610D 00	-0.825840-01
67	0.8038D 02	-0.98388D-01	-0.905910-01
68	0.7536D 02	-0.74202D-01	-0.115220 00
69	0.7257D 02	-0.68446D-01	-0.938860-01
70	0.7286D 02	-0.74424D-01	-0.760680-01
71	0.7280D 02	-0.70735D-01	-0.572530-01
72	0.7136D 02	-0.62217D-01	-0.850130-01
73	0.6303D 02	-0.45656D-01	-0.941800-01
74	0.6612D 02	-0.39558D-01	-0.123180 00
75	0.6541D 02	-0.19075D-01	-0.107750 00
76	0.6460D 02	-0.95462D-02	-0.104330 00
77	0.6289D 02	-0.12953D-01	-0.795790-01
78	0.5934D 02	-0.38321D-02	-0.888440-01
79	0.5559D 02	-0.23053D-02	-0.761370-01
80	0.5275D 02	-0.39934D-02	-0.624940-01
81	0.5119D 02	-0.10209D-01	-0.701770-01
82	0.5115D 02	-0.13487D-01	-0.850070-01
83	0.5106D 02	-0.79570D-02	-0.116230 00
84	0.5022D 02	-0.15359D-01	-0.115470 00
85	0.5031D 02	-0.26604D-01	-0.124110 00
86	0.4762D 02	-0.30344D-01	-0.908890-01
87	0.4376D 02	-0.21603D-01	-0.884320-01
88	0.4096D 02	-0.13560D-01	-0.868460-01
89	0.4365D 02	-0.65207D-02	-0.119220 00
90	0.4600D 02	-0.80067D-02	-0.119670 00

$= \text{Mean Cycle} + Y^1 H_1 + Y^2 H_2$

TABLE 4.4 (CONT'D)

No. Month After Start	Mean Cycle	H_1	H_2
91	0.4658D 02	0.27033D-02	-0.11586D 00
92	0.4328D 02	-0.61714D-02	-0.10685D 00
93	0.4018D 02	-0.23651D-02	-0.10490D 00
94	0.3615D 02	-0.93553D-02	-0.88440D-01
95	0.3432D 02	-0.74548D-02	-0.84183D-01
96	0.3226D 02	-0.10076D-01	-0.84202D-01
97	0.2972D 02	-0.80408D-02	-0.93163D-01
98	0.2922D 02	-0.35327D-02	-0.10308D 00
99	0.2931D 02	-0.10035D-01	-0.11377D 00
100	0.3107D 02	-0.61577D-02	-0.12285D 00
101	0.3199D 02	-0.30548D-02	-0.12227D 00
102	0.3110D 02	0.54397D-02	-0.11640D 00
103	0.2945D 02	-0.18458D-02	-0.11910D 00
104	0.2558D 02	-0.56185D-02	-0.11465D 00
105	0.2429D 02	-0.13934D-02	-0.11784D 00
106	0.2437D 02	0.14877D-01	-0.11226D 00
107	0.2397D 02	0.16857D-01	-0.10971D 00
108	0.2334D 02	0.16054D-01	-0.10920D 00
109	0.2122D 02	0.32458D-02	-0.10748D 00
110	0.1952D 02	0.31105D-02	-0.10530D 00
111	0.1752D 02	-0.93158D-04	-0.10286D 00
112	0.1635D 02	0.24659D-02	-0.52386D-01
113	0.1584D 02	0.16146D-02	-0.98927D-01
114	0.1487D 02	0.15602D-02	-0.55254D-01
115	0.1406D 02	0.41686D-02	-0.10838D 00
116	0.1309D 02	0.35616D-02	-0.98835D-01
117	0.1303D 02	0.47010D-02	-0.96435D-01
118	0.1244D 02	0.53902D-02	-0.90191D-01
119	0.1290D 02	0.43184D-02	-0.54667D-01
120	0.1302D 02	0.42561D-02	-0.92226D-01
121	0.1261D 02	0.10660D-02	-0.85511D-01
122	0.1157D 02	0.15867D-02	-0.78277D-01
123	0.9739D 01	0.20447D-02	-0.66695D-01
124	0.9128D 01	0.45455D-02	-0.54644D-01
125	0.9835D 01	0.39314D-02	-0.58888D-01
126	0.1076D 02	0.17471D-03	-0.65515D-01
127	0.1038D 02	0.13497D-02	-0.67559D-01
128	0.8878D 01	0.70247D-03	-0.60763D-01
129	0.7033D 01	0.22496D-02	-0.52718D-01
130	0.6650D 01	0.14103D-02	-0.53489D-01
131	0.5322D 01	0.55049D-03	-0.46014D-01
132	0.5700D 01	0.16514D-02	-0.47256D-01
133	0.4250D 01	-0.11580D-02	-0.36713D-01
134	0.4850D 01	0.67881D-03	-0.35452D-01
135	0.3961D 01	-0.14556D-02	-0.35906D-01

$= + Y^1 + Y^2 =$

No. Month After Start	Mean Cycle	H_1	H_2
135	0.3694D 01	-0.38124D-02	-0.34866D-01
137	0.2556D 01	-0.57756D-02	-0.28449D-01
138	0.2244D 01	-0.80815D-02	-0.27381D-01
133	0.2361D 01	-0.53162D-02	-0.23698D-01
141	0.2572D 01	-0.29094D-02	-0.21704D-01
141	0.2322D 01	-0.24199D-02	-0.17378D-01
142	0.2339D 01	-0.53256D-02	-0.18469D-01
143	0.1911D 01	-0.55198D-02	-0.16431D-01
144	0.1394D 01	-0.42724D-02	-0.11221D-01
145	0.1461D 01	-0.43801D-02	-0.98035D-02
146	0.1076D 01	-0.23961D-02	-0.55335D-02
147	0.9500D 00	-0.49662D-02	-0.94558D-02
148	0.5778D 00	-0.32839D-02	-0.65912D-02
149	0.7000D 00	-0.40751D-02	-0.80781D-02
150	0.4556D 00	-0.26521D-02	-0.52572D-02
151	0.5278D 00	-0.29131D-02	-0.59378D-02
152	0.4500D 00	-0.24604D-02	-0.50402D-02
153	0.4278D 00	-0.23310D-02	-0.47838D-02
154	0.2278D 00	-0.13260D-02	-0.26286D-02
155	0.2333D 00	-0.16495D-02	-0.32697D-02
156	0.4167D 00	-0.24257D-02	-0.48084D-02
157	0.4222D 00	-0.24580D-02	-0.48725D-02
158	0.3944D 00	-0.22963D-02	-0.45519D-02
159	0.2944D 00	-0.17141D-02	-0.33979D-02
160	0.3444D 00	-0.20052D-02	-0.39749D-02
161	0.3389D 00	-0.19729D-02	-0.39108D-02
162	0.2889D 00	-0.16818D-02	-0.33338D-02
163	0.2000D 00	-0.11643D-02	-0.23080D-02
164	0.1667D 00	-0.97027D-03	-0.19234D-02
165	0.3389D 00	-0.19729D-02	-0.39108D-02
166	0.3222D 00	-0.18759D-02	-0.37185D-02
167	0.2500D 00	-0.14554D-02	-0.28850D-02
168	0.2222D-01	-0.12937D-03	-0.25645D-03
169	0.0	0.0	0.0
170	0.0	0.0	0.0
171	0.0	0.0	0.0
172	0.0	0.0	0.0
173	0.0	0.0	0.0
174	0.0	0.0	0.0
175	0.0	0.0	0.0
176	0.0	0.0	0.0
177	0.0	0.0	0.0
178	0.0	0.0	0.0
179	0.0	0.0	0.0
180	0.0	0.0	0.0

TABLE 4.5

COMPARISON OF RMS ERROR AND REPRESENTATION USING 6 DIMENSIONS

CYCLE NO. (n)	RMS-ERROR E_{RMS}	$Q_{n-1,\ n-2}$
3	26.13	.781
4	18.51	.890
5	15.79	.950
6	8.85	.963
7	21.28	.950
8	20.56	.951
9	19.68	.923
10	11.88	.978
11	14.83	.904
12	23.51	.815
13	14.81	.780
14	19.81	.552
15	12.47	.698
16	20.78	.627
17	11.23	.680
18	19.17	.801
Avg.	17.5	.83
Std. Dev.	4.7	.13
19	44.3	.74
20	13.7	.91
21	-	.92

TABLE 4.6

SUMMARY OF RMS ERROR, E_{RMS}*, FOR EXTRAPOLATION OF SUNSPOT CYCLES

No. Dimensions	1		2			3		4	5	5
Cycle #	E_2*	E_1	E_3	E_2	E_1	E_3	E_2	E_3	E_3	E_3
1	11.19	---	11.87	10.84	5.54	9.45	9.69	8.94	9.11	8.53
2	15.51	11.53	14.5	14.1	11.34	13.84	13.36	8.54	13.77	6.11
3	21.50	12.67	11.55	12.11	12.67	10.88	11.86	7.63	9.88	6.88
4 II	10.21	7.45	8.26	7.74	6.61	8.10	7.72	5.49	5.50	4.29
5	8.25	5.46	6.81	7.22	4.94	6.80	7.04	6.25	6.31	6.24
6	8.91	3.76	8.39	7.99	3.51	8.38	7.98	7.76	7.87	7.76
7	9.09	4.63	9.93	7.69	4.34	9.48	7.69	8.29	8.65	7.15
8	22.78	16.7	12.67	12.52	10.11	8.03	8.60	7.38	7.58	7.34
9 II	26.36	7.41	14.38	15.3	7.41	13.86	13.69	5.21	5.37	4.09
10	14.57	7.56	8.89	9.39	4.06	8.15	9.28	8.11	8.12	8.11
11 II	17.31	7.11	14.05	14.71	6.68	10.27	10.88	8.56	8.56	8.07
12	11.09	7.21	9.87	9.90	6.69	10.66	9.74	9.93	9.95	9.87
13	13.15	8.96	8.32	8.71	7.19	7.27	7.68	7.12	7.25	5.41
14	12.13	6.71	9.69	9.77	4.18	9.44	9.66	8.73	9.42	8.66
15	16.23	8.49	14.47	13.50	8.23	13.87	13.49	6.4	11.37	6.37
16	15.16	12.32	11.72	11.73	10.26	11.20	11.70	8.59	9.28	8.52
17	15.53	8.46	13.94	14.19	6.42	13.39	13.11	8.92	10.02	8.6
18 II	14.12									
Mean	14.62	8.53	11.13	11.25	7.07	10.21	10.19	7.76	8.71	7.18
σ	4.83	3.21	2.49	2.7	2.6	2.28	2.23	1.26	2.03	1.55
19 II		13.27	18.4		13.07	18.38		17.77	18.29	17.36
20 II		8.8	11.3		7.79	10.5		9.51	9.61	9.51

* E_X = Error in X-Quarter Extrapolation

FIGURE 4.1

COMPARISON - END OF MONTH VALUE OF 81-DAY RUNNING AVERAGE WITH 3-MONTH RUNNING AVERAGE

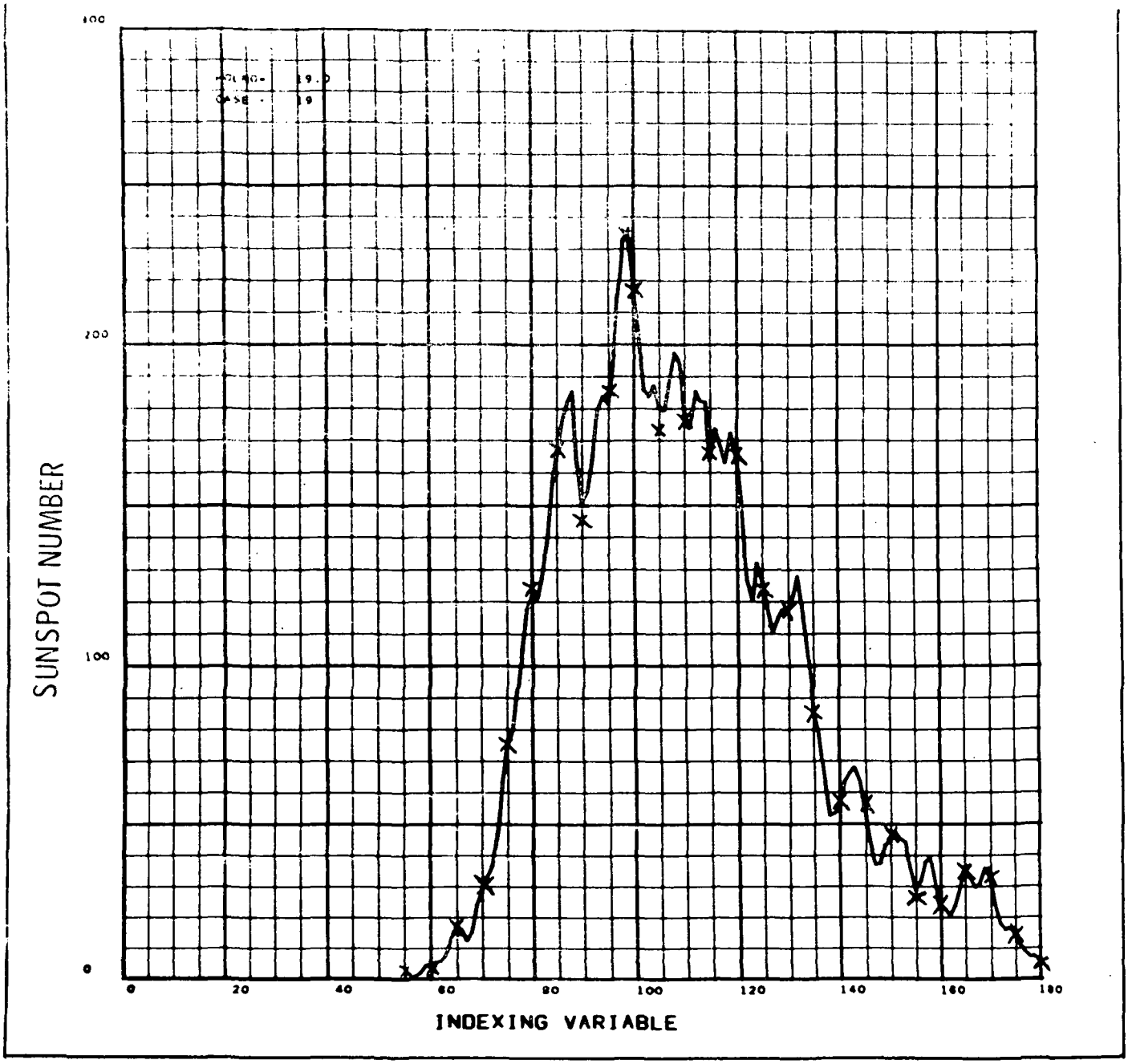


FIGURE 4.2 SINGLE CYCLE AVERAGE INPUT VECTOR

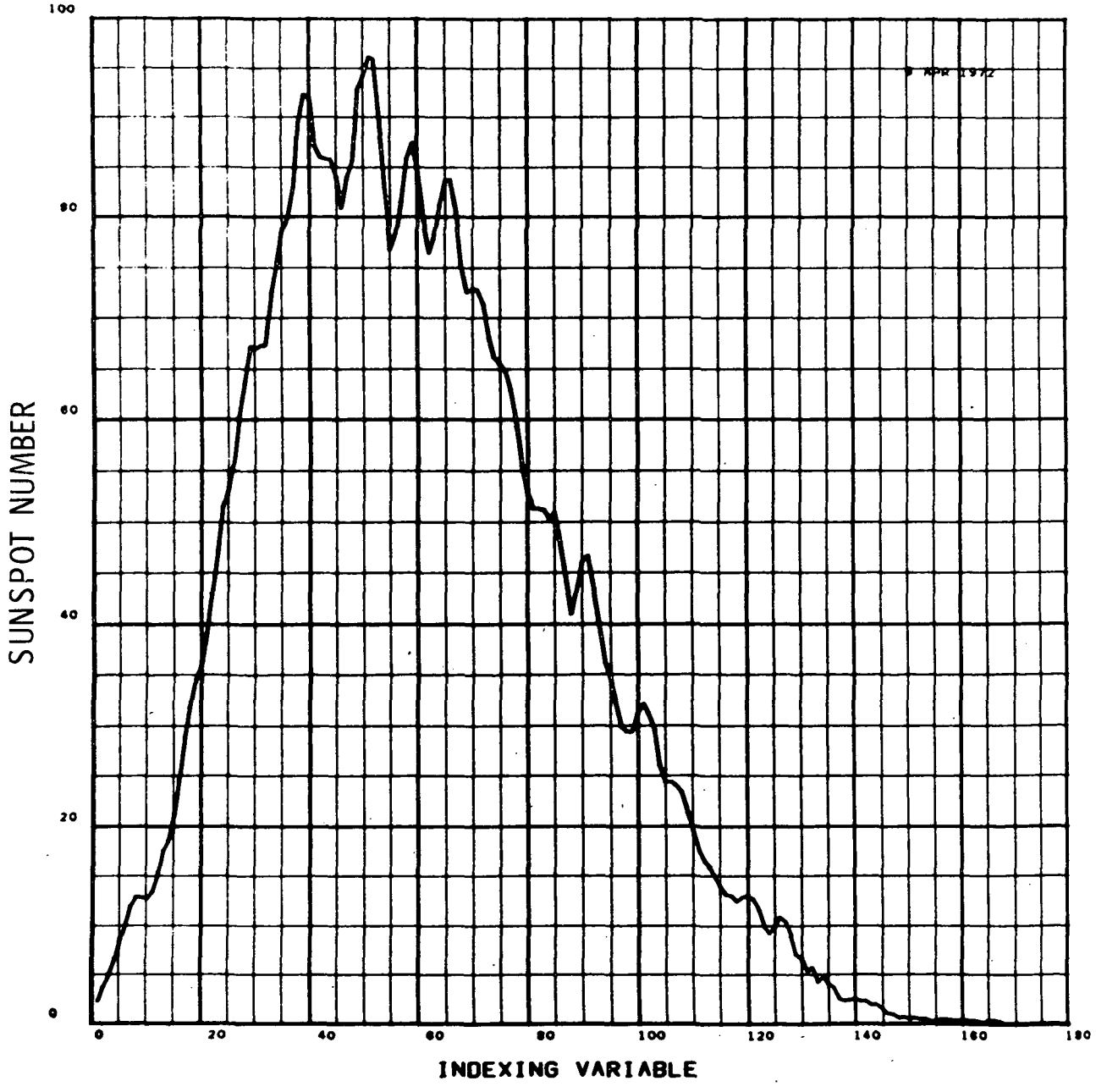


FIGURE 4.3 SINGLE CYCLE INFORMATION ENERGY

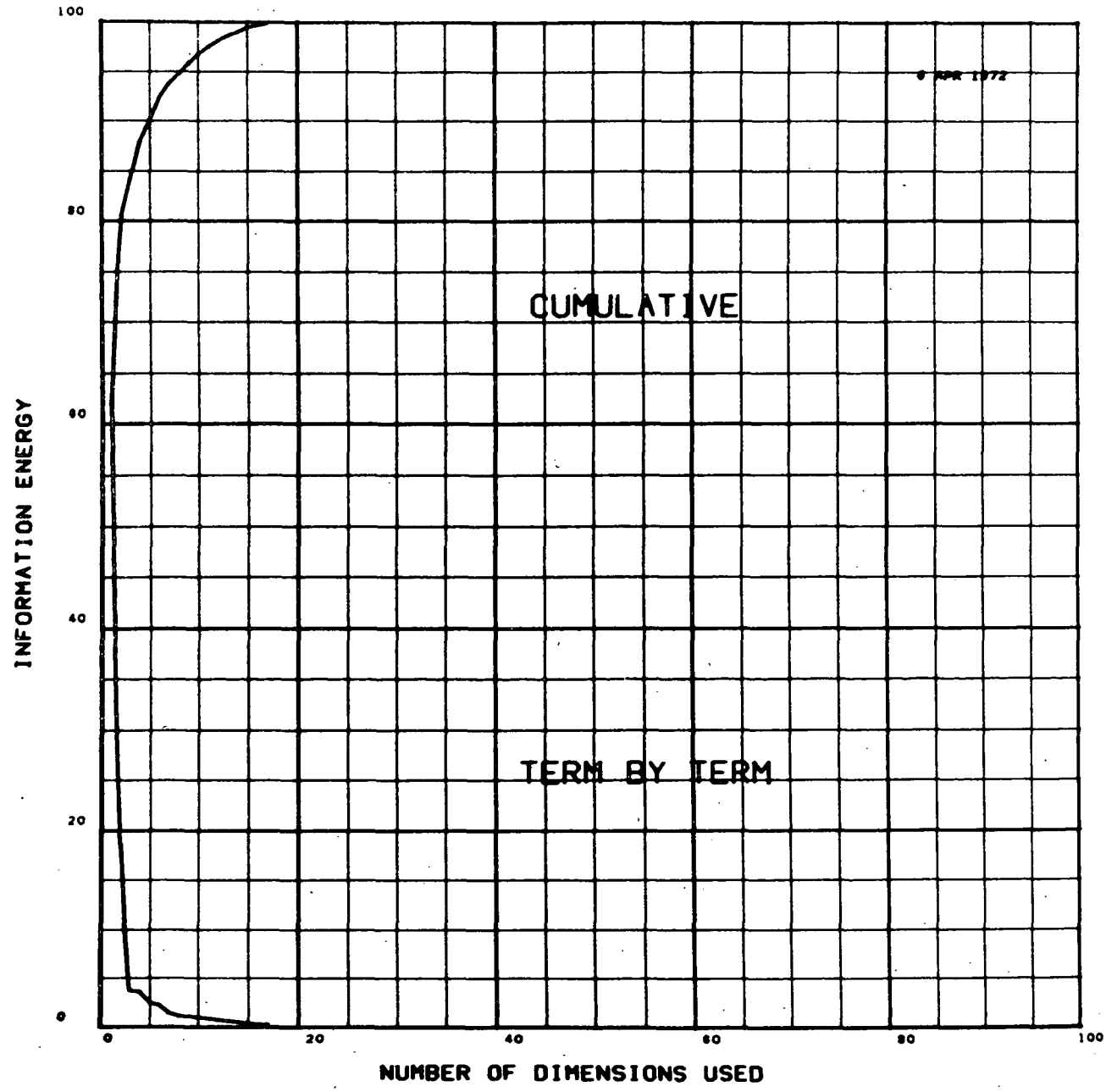


FIGURE 4.4 SINGLE CYCLE FIRST OPTIMUM FUNCTION

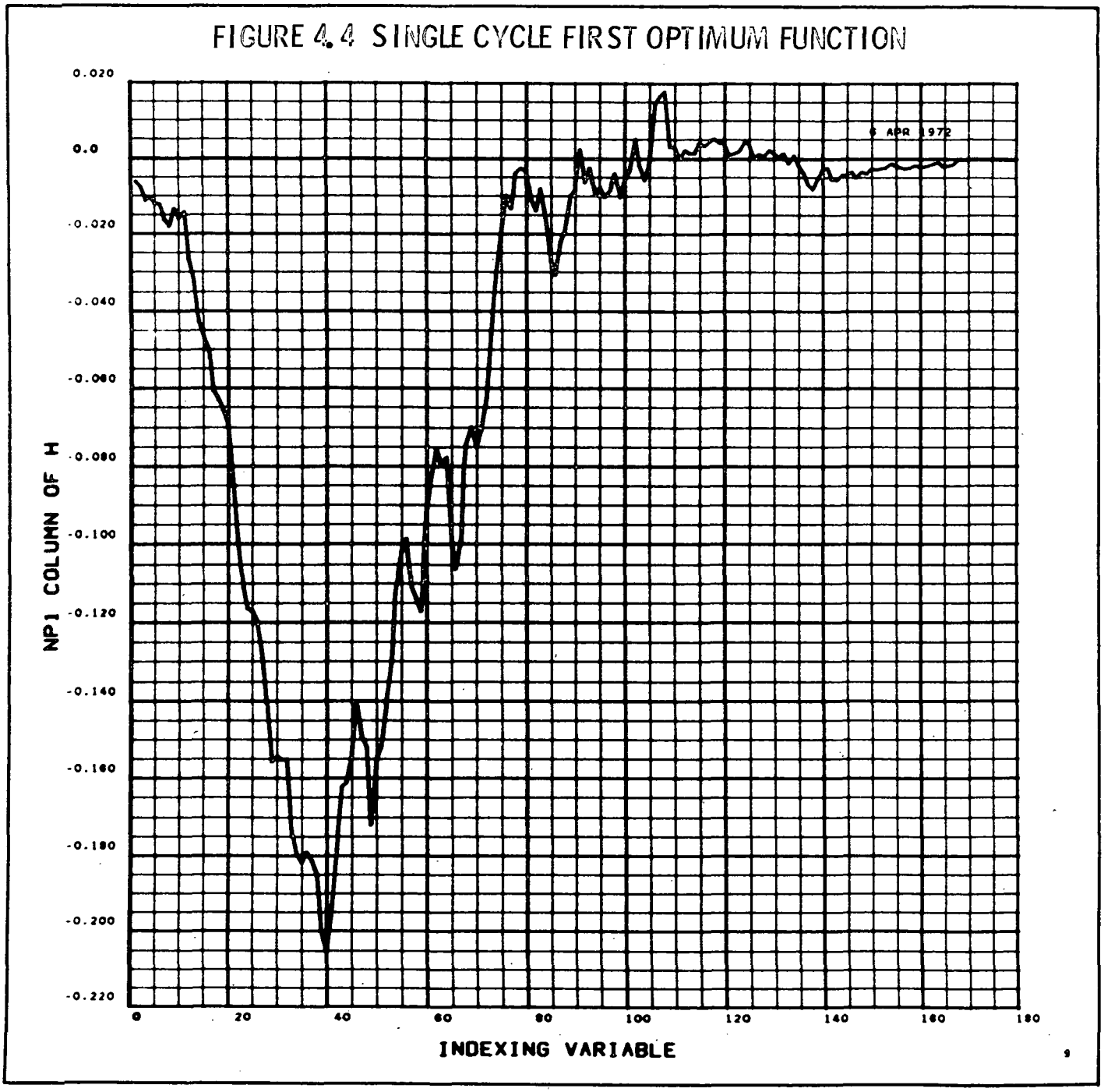


FIGURE 4.5 SINGLE CYCLE SECOND OPTIMUM FUNCTION

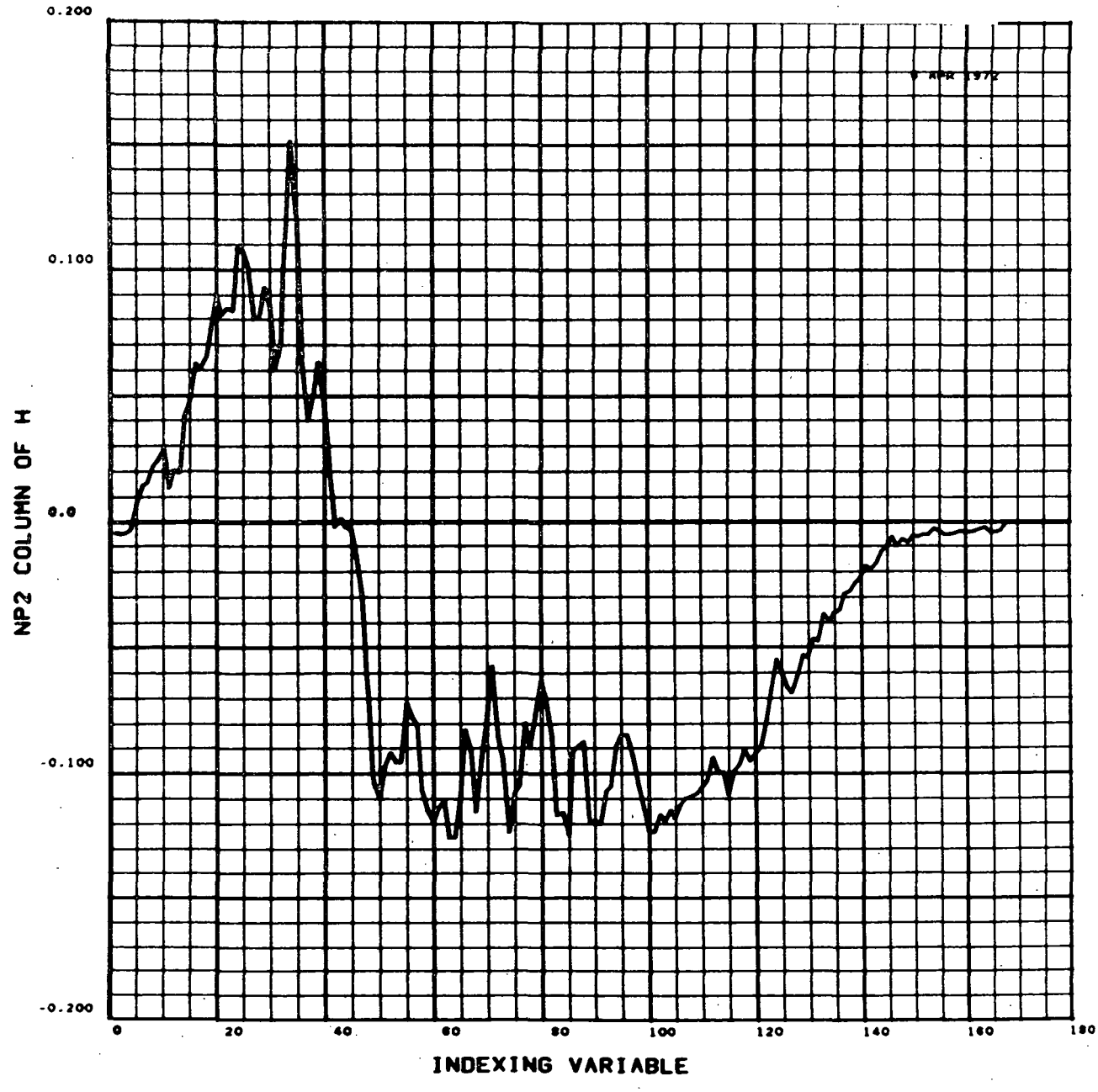


FIGURE 4.6 SINGLE CYCLE THIRD OPTIMUM FUNCTION

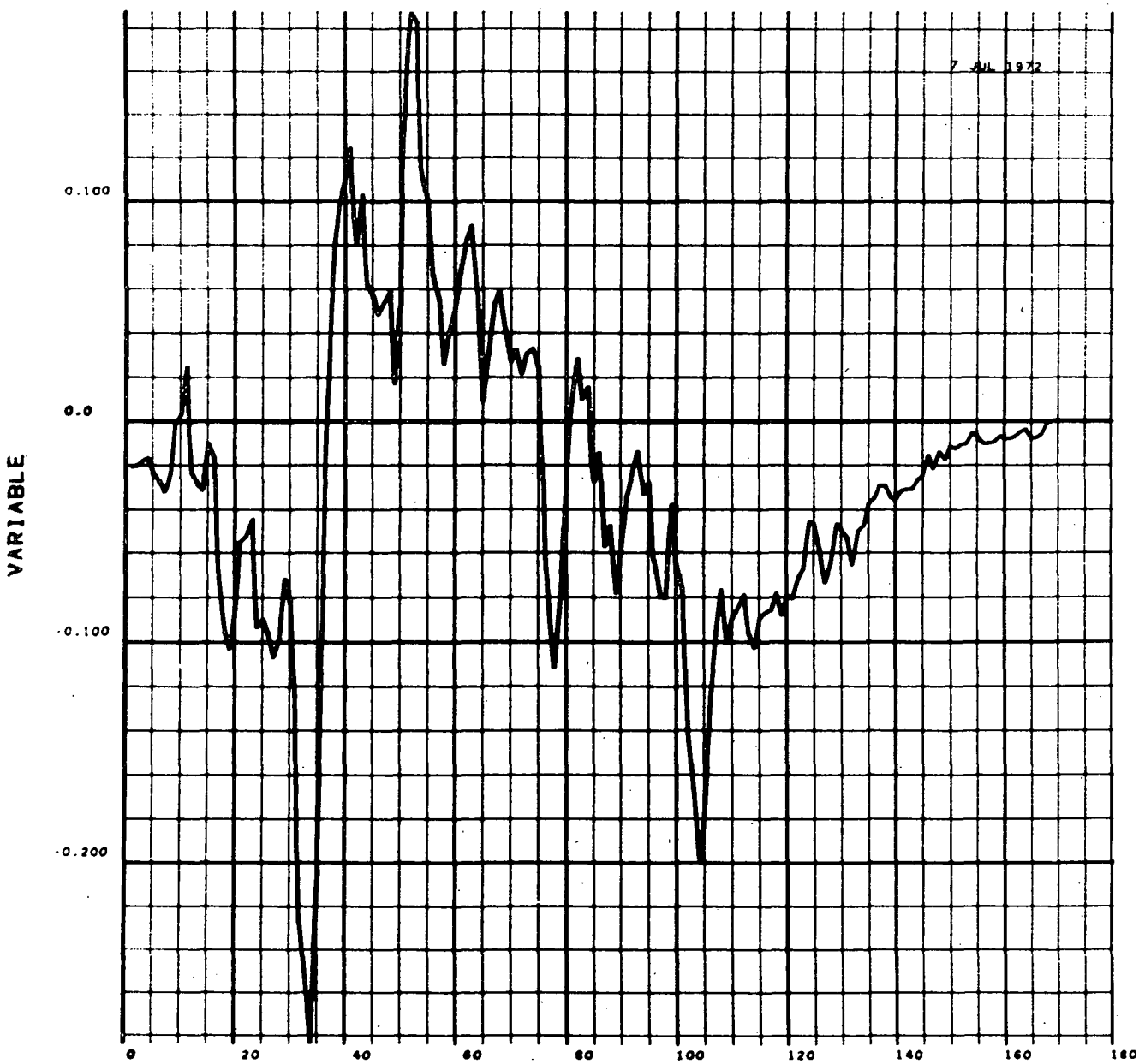


FIGURE 4.7 SINGLE CYCLE FOURTH OPTIMUM FUNCTION

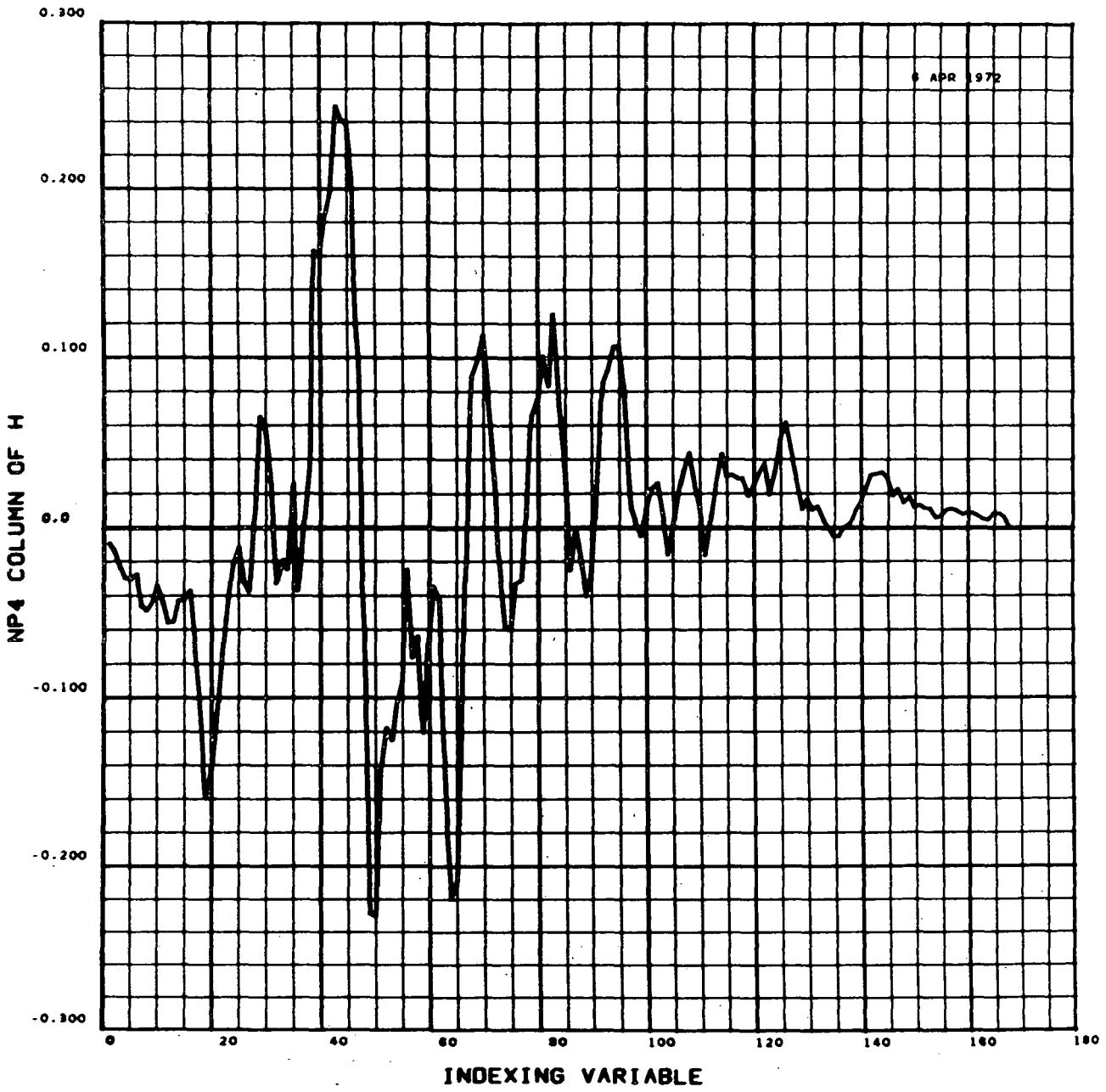


FIGURE 4.8 SINGLE CYCLE FIFTH OPTIMUM FUNCTION

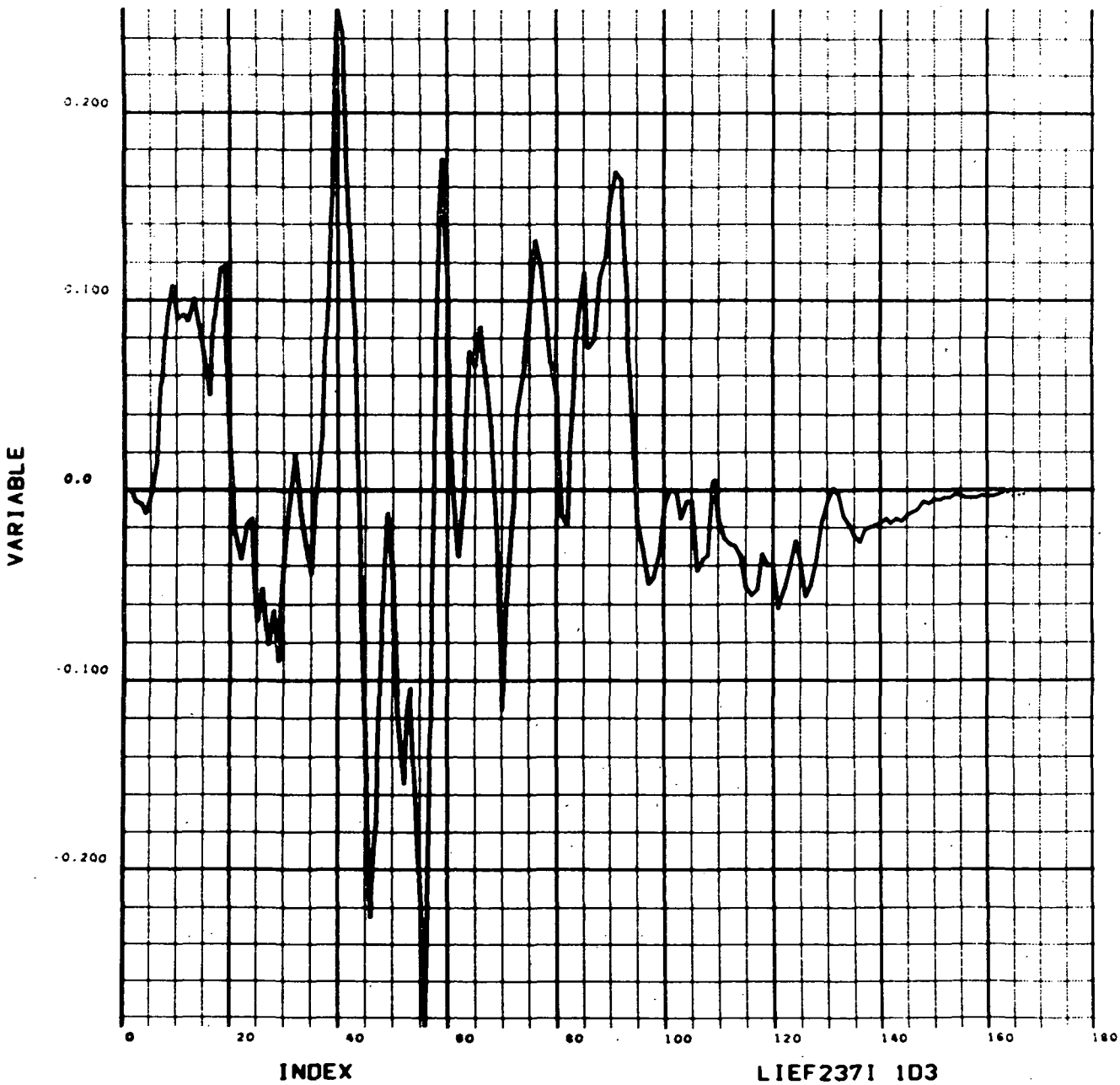


FIGURE 4.9 SINGLE CYCLE SIXTH OPTIMUM FUNCTION

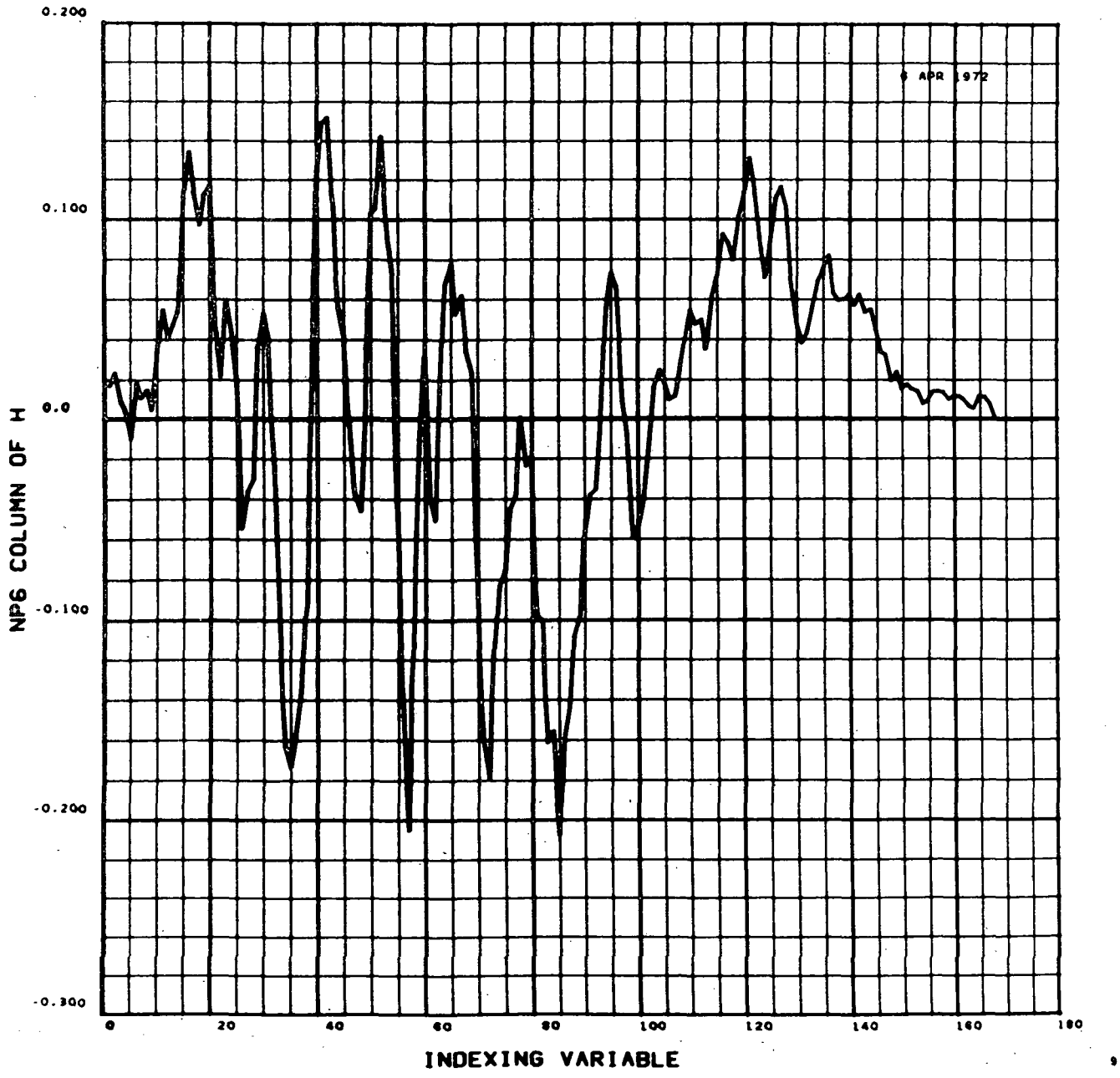


FIGURE 4.10 DOUBLE CYCLE AVERAGE INPUT VECTOR

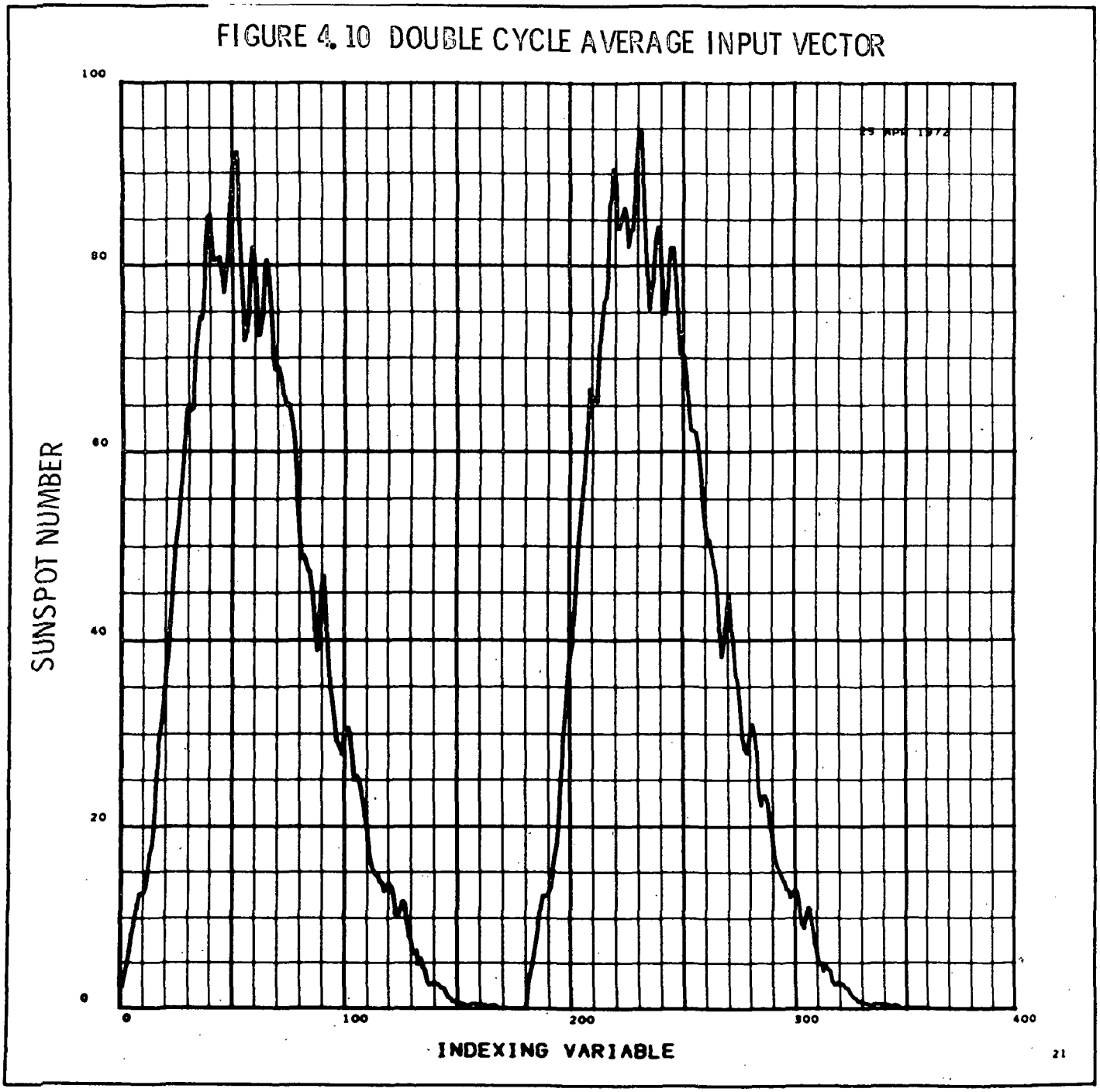


FIGURE 4.11 DOUBLE CYCLE INFORMATION ENERGY

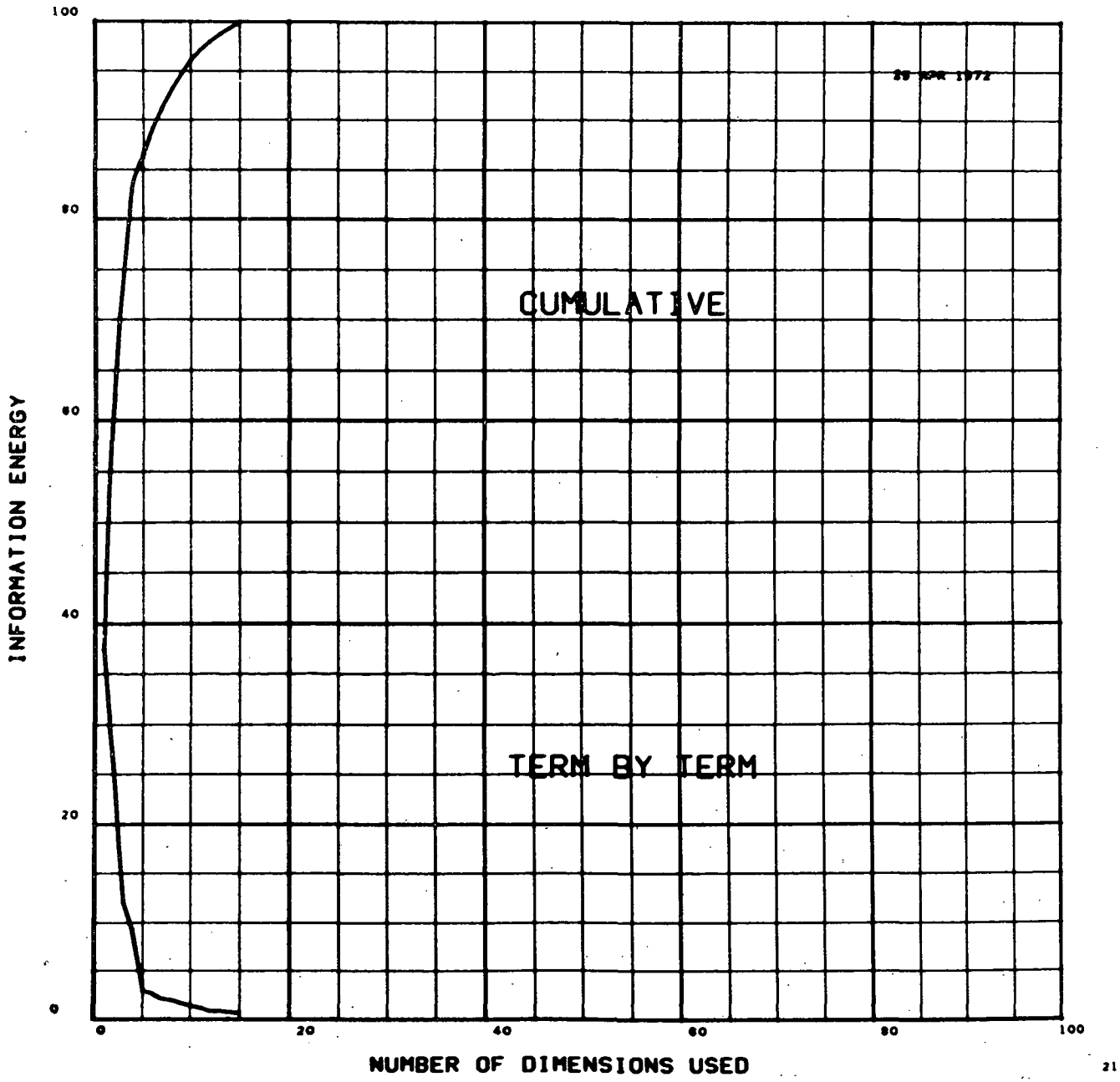


FIGURE 4.12 DOUBLE CYCLE FIRST OPTIMUM FUNCTION

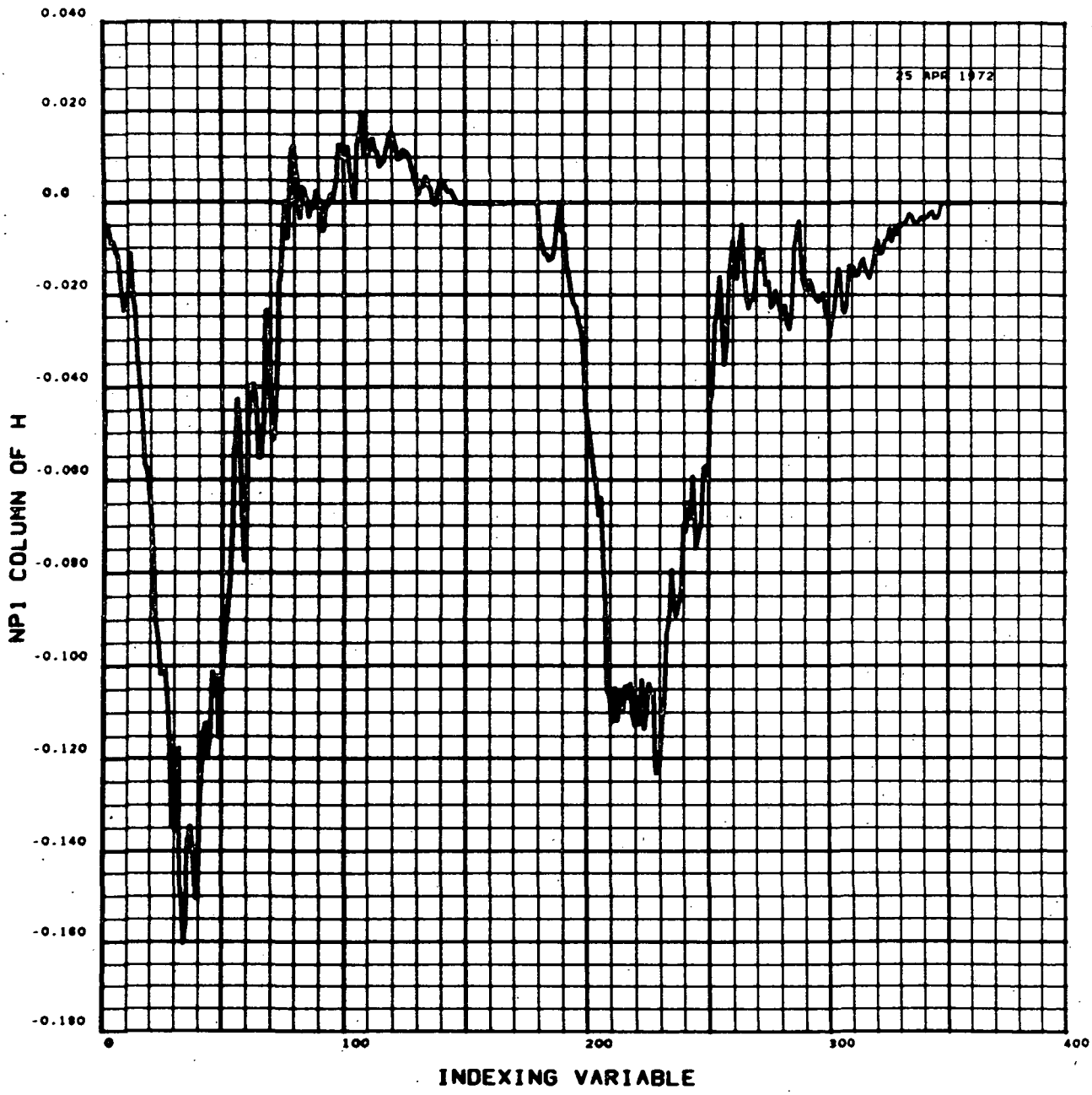


FIGURE 4.13 DOUBLE CYCLE SECOND OPTIMUM FUNCTION

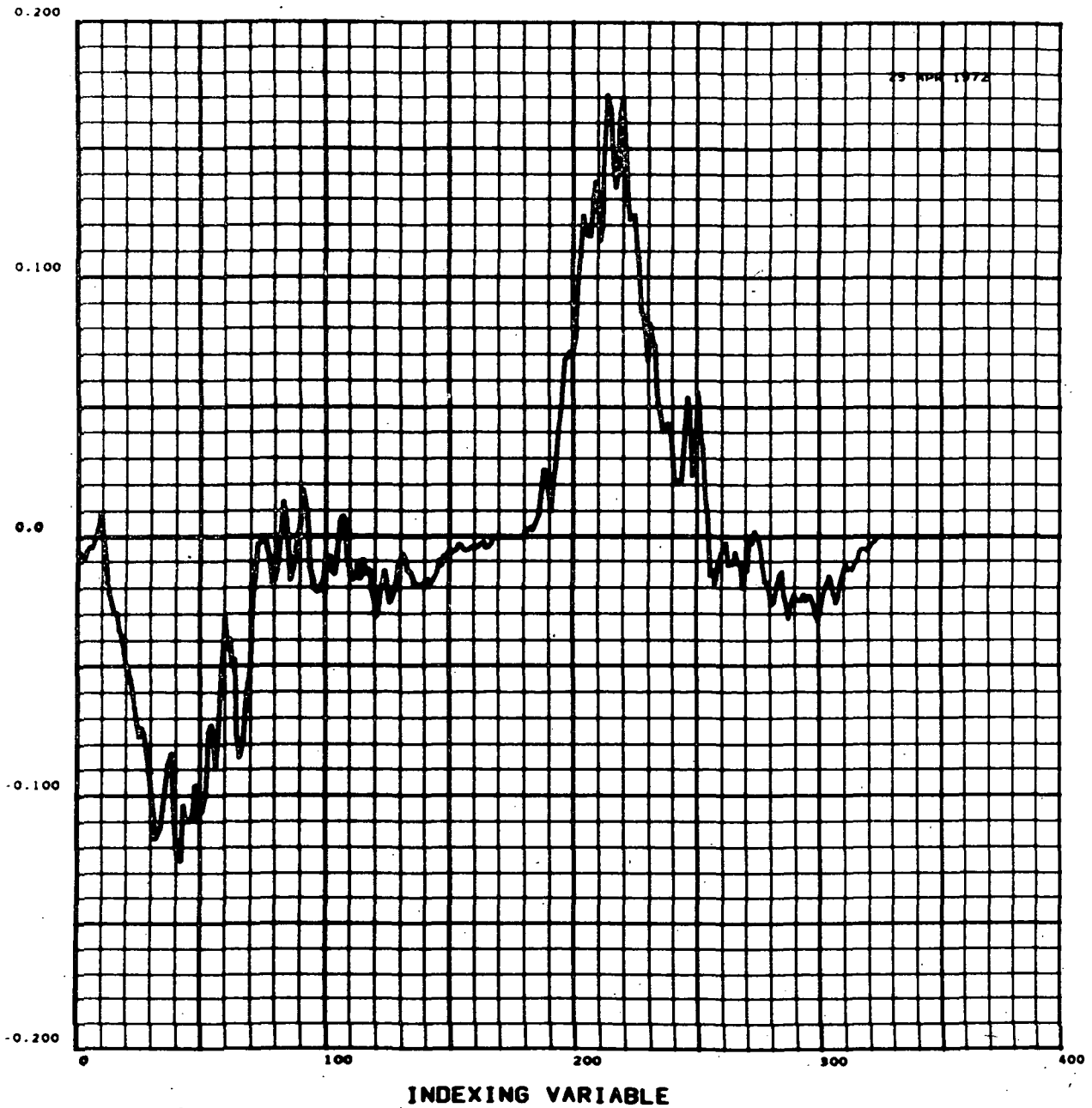


FIGURE 4.14 DOUBLE CYCLE THIRD OPTIMUM FUNCTION

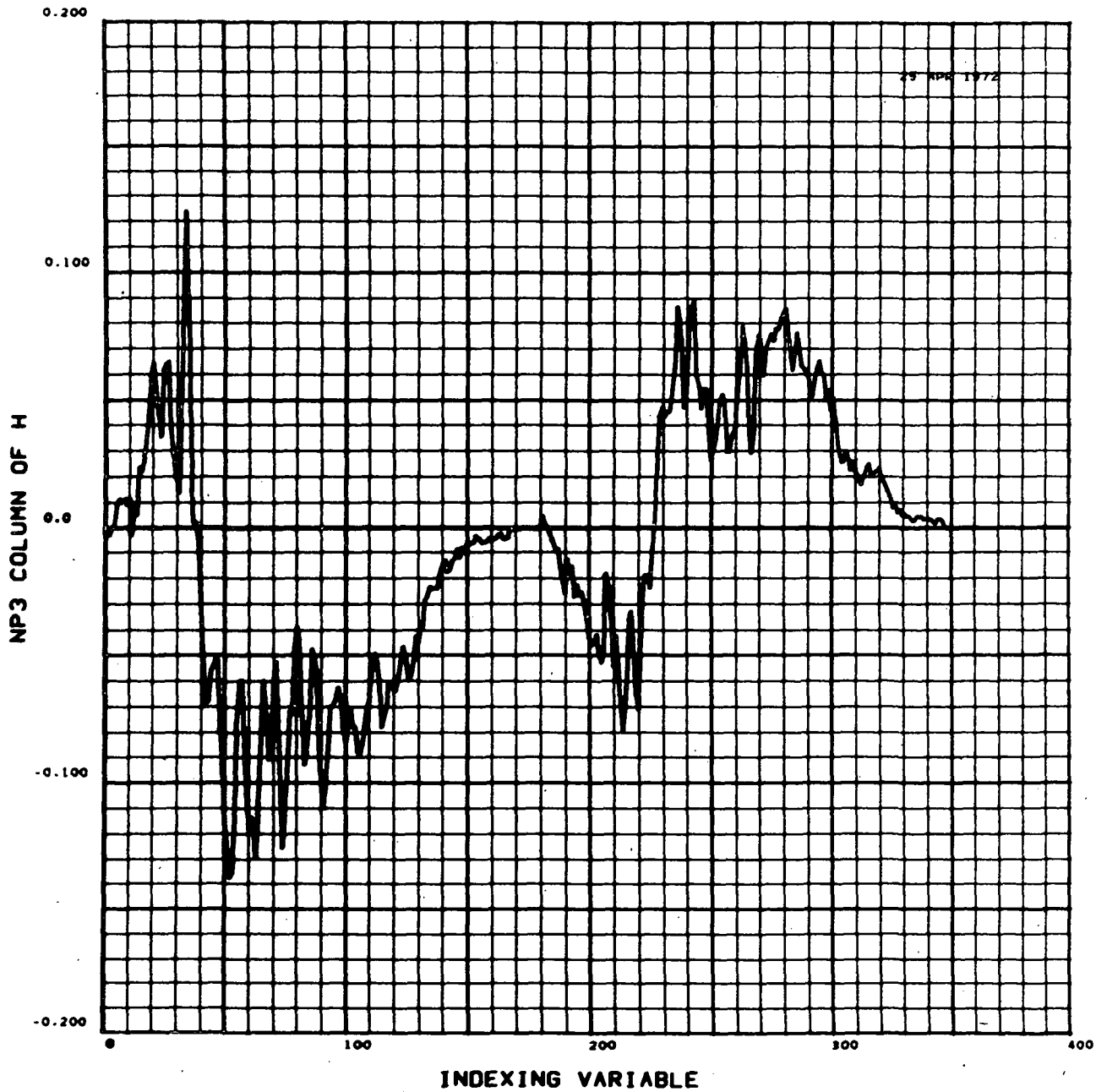


FIGURE 4.15 DOUBLE CYCLE FOURTH OPTIMUM FUNCTION

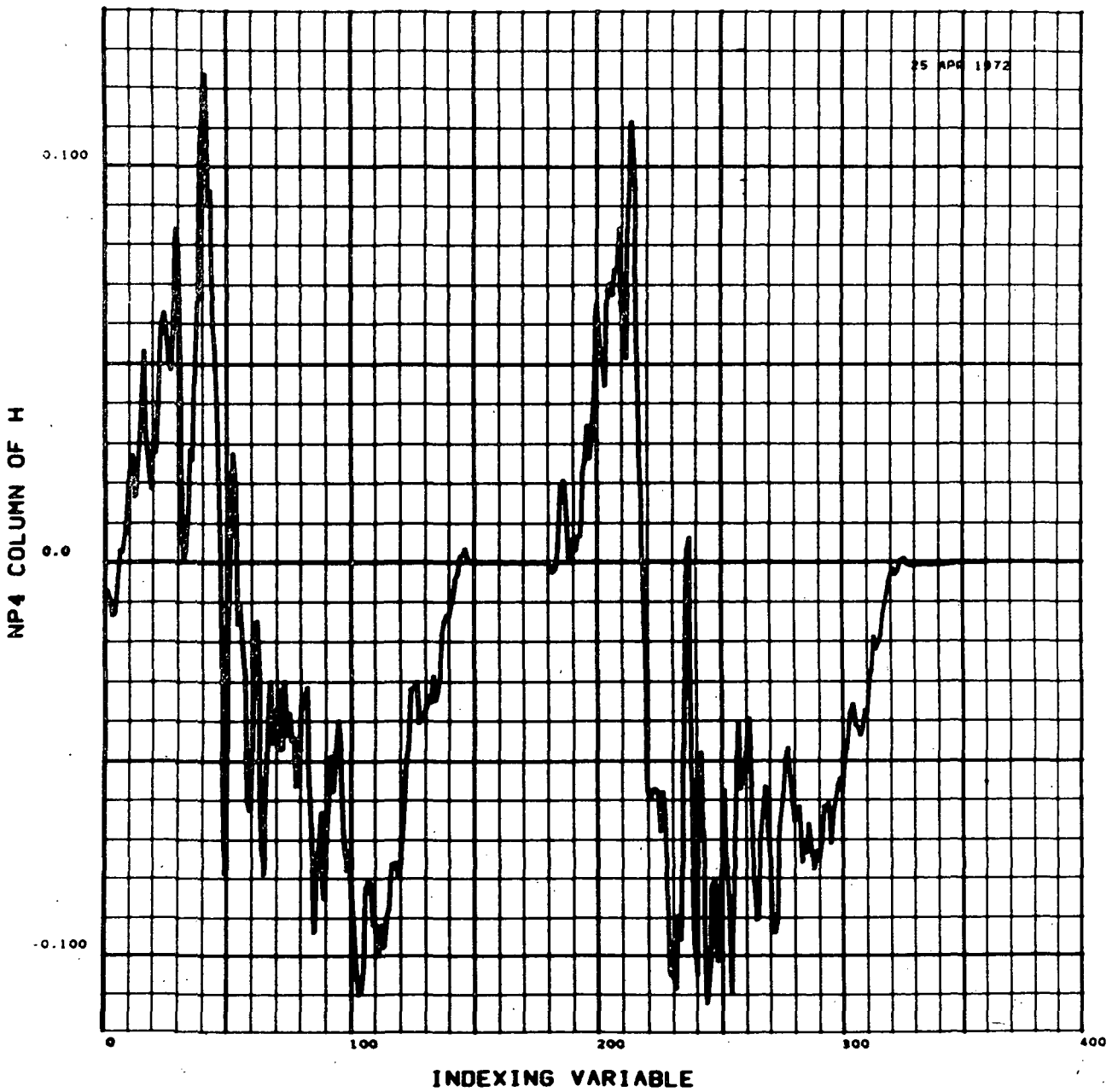


FIGURE 4.16 DOUBLE CYCLE FIFTH OPTIMUM FUNCTION

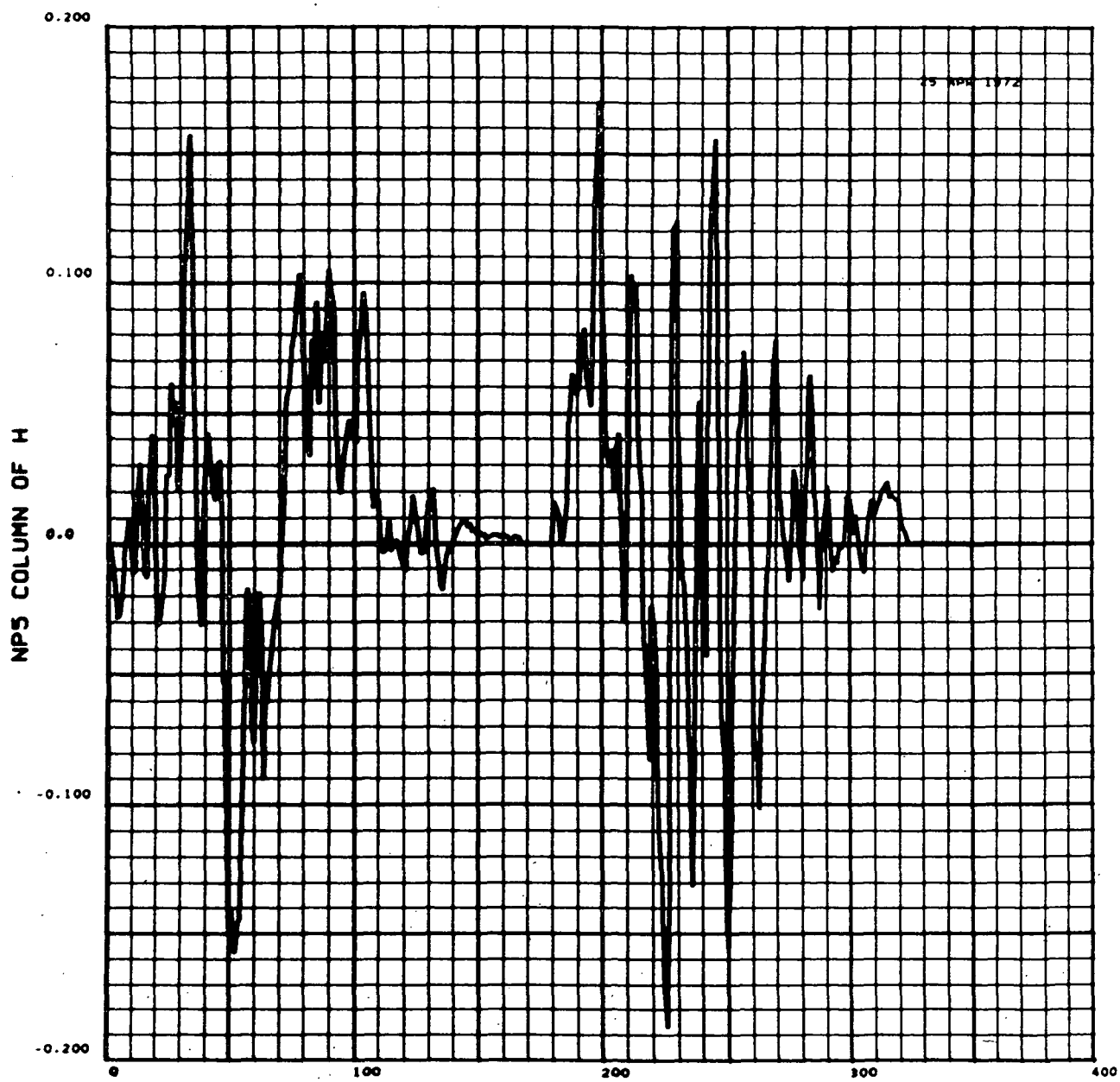


FIGURE 4.17 DOUBLE CYCLE SIXTH OPTIMUM FUNCTION

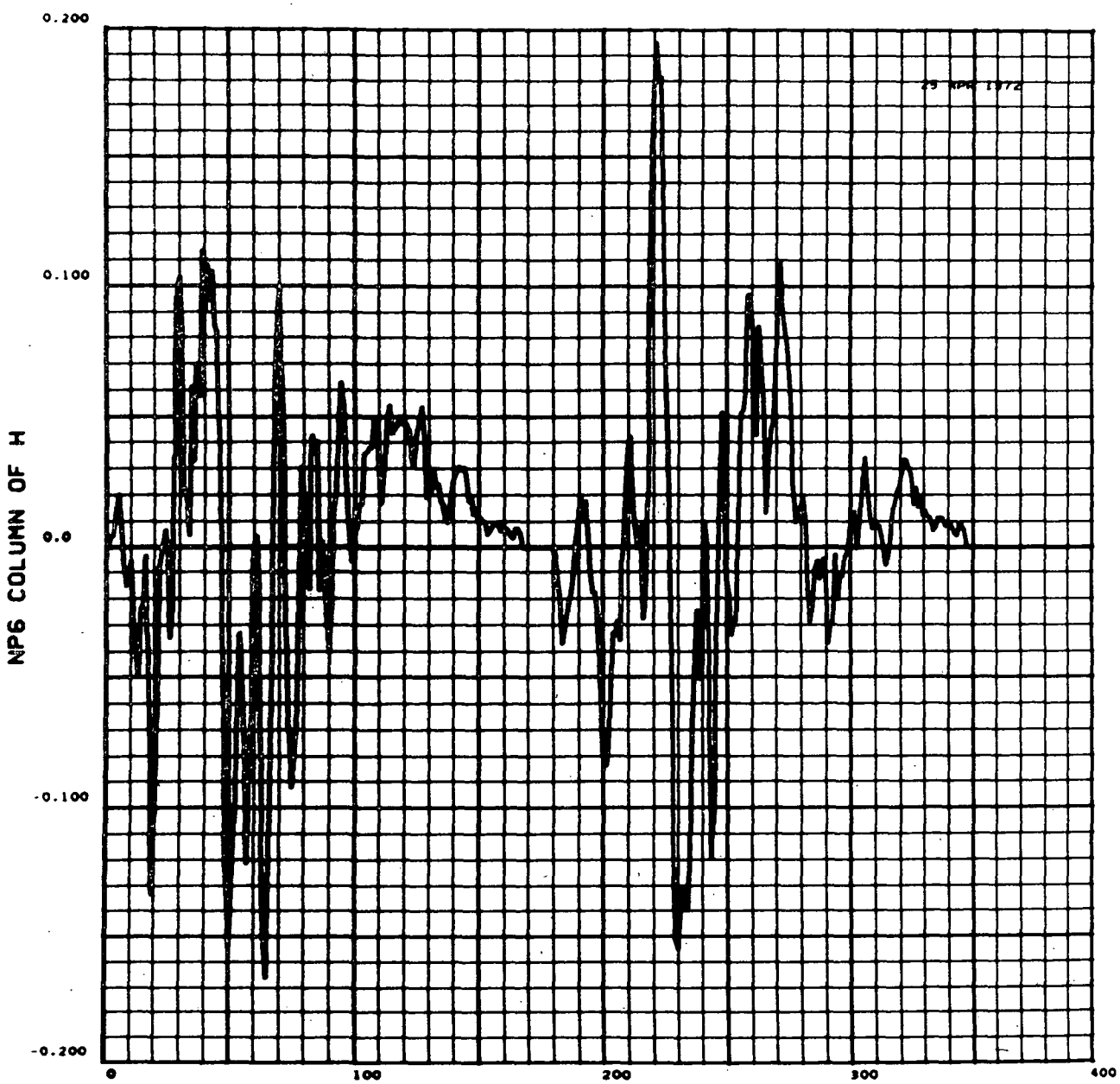


FIGURE 4.18 RELATIVE IMPORTANCE SPECTRUM FOR CANONICAL PREDICTION OF FIRST COEFFICIENT

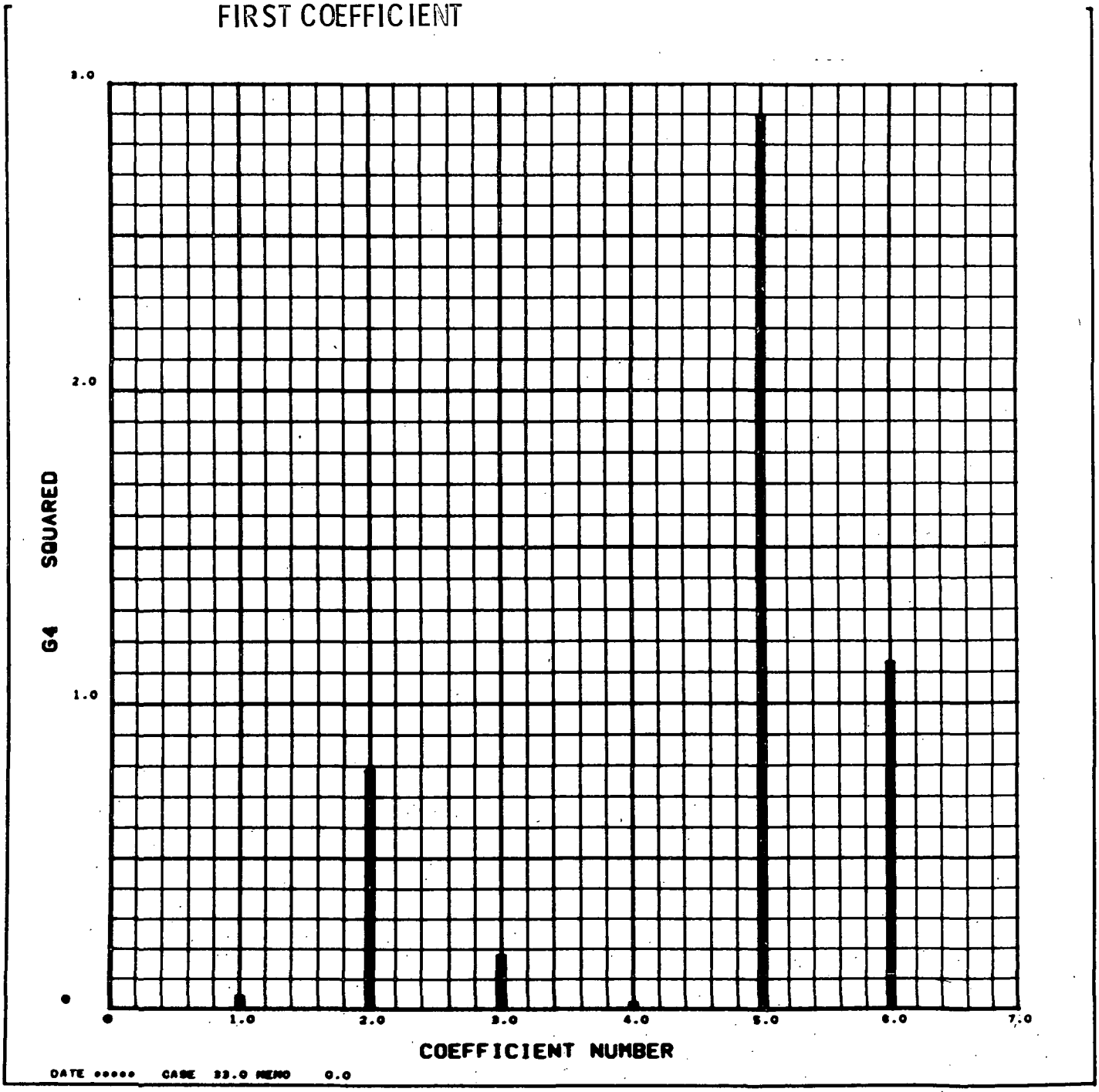


FIGURE 4.19 RELATIVE IMPORTANCE SPECTRUM FOR CANONICAL PREDICTION OF SECOND COEFFICIENT

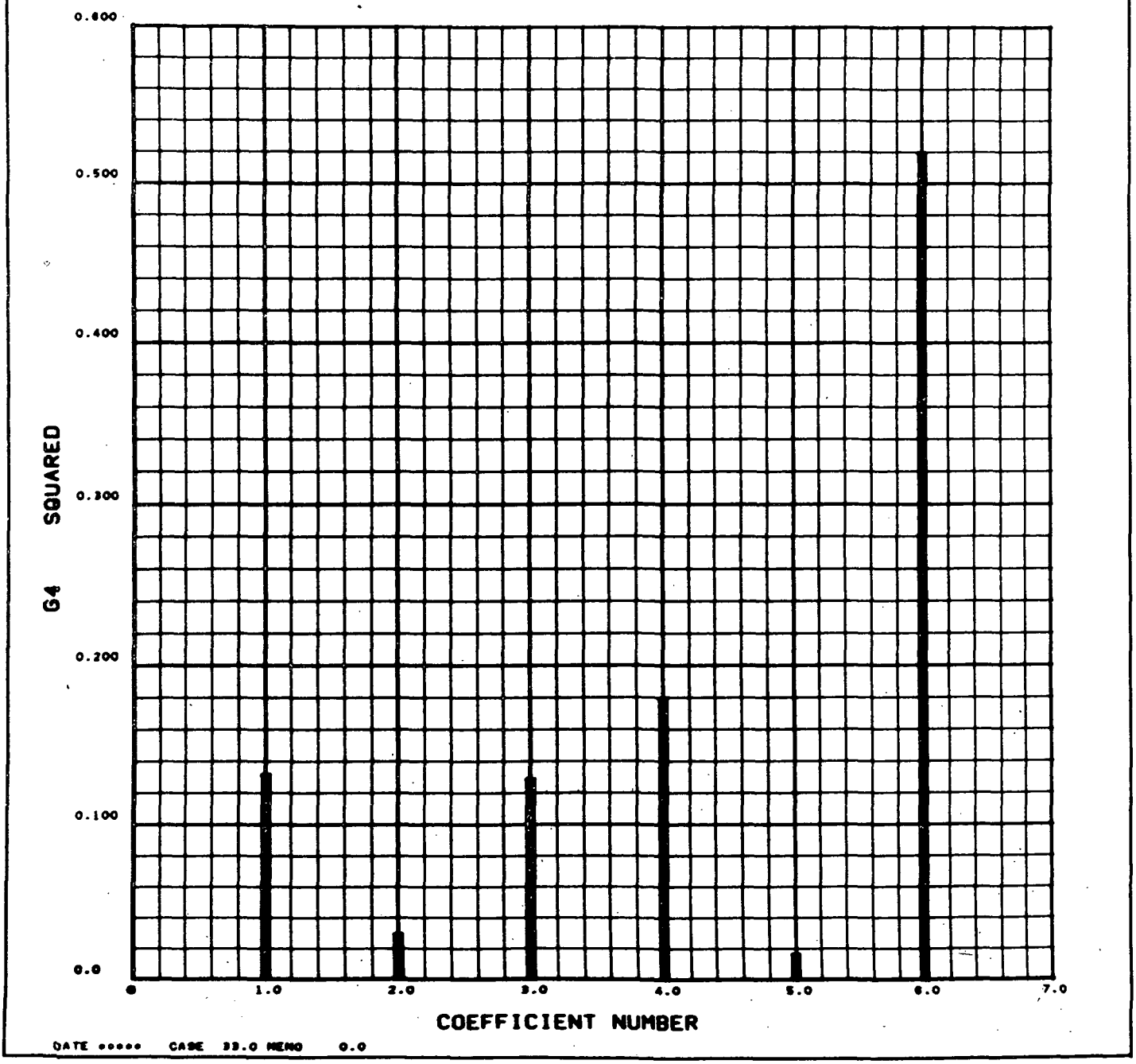


FIGURE 4.20 RELATIVE IMPORTANCE SPECTRUM FOR CANONICAL PREDICTION OF THIRD COEFFICIENT

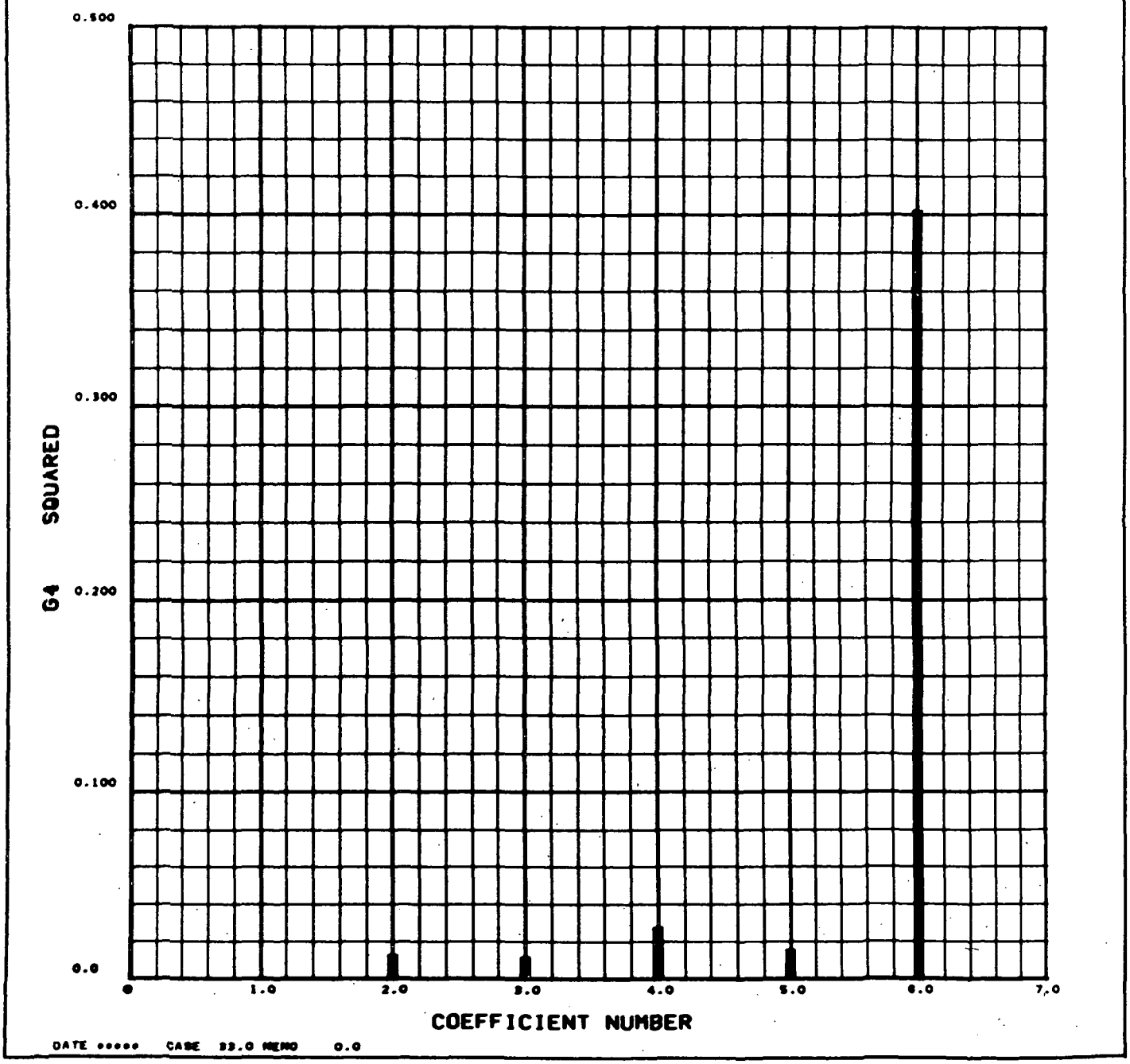


FIGURE 4.21 RELATIVE IMPORTANCE SPECTRUM FOR CANONICAL PREDICTION OF FOURTH COEFFICIENT

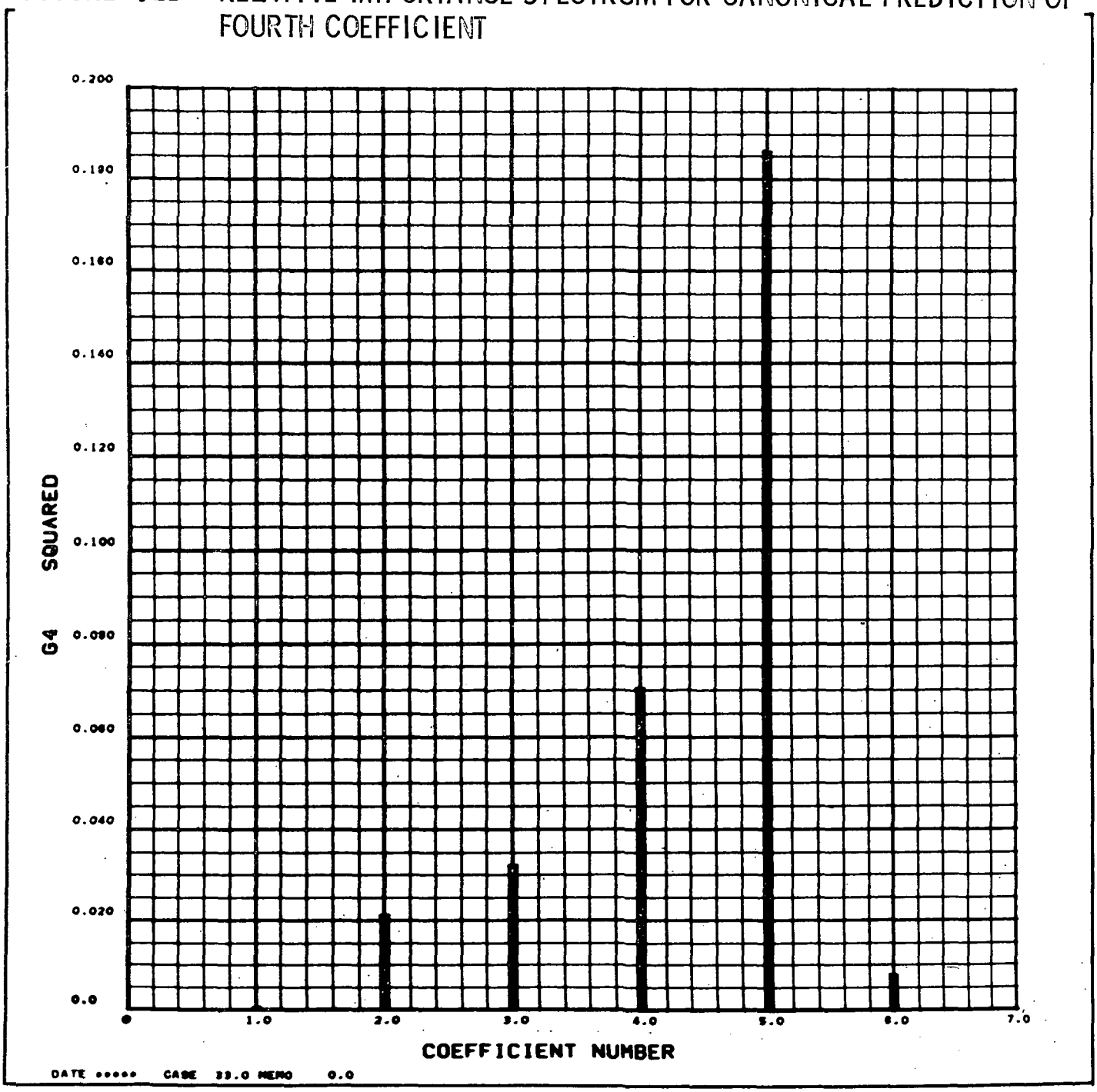


FIGURE 4.22 RELATIVE IMPORTANCE SPECTRUM FOR CANONICAL PREDICTION OF FIFTH COEFFICIENT

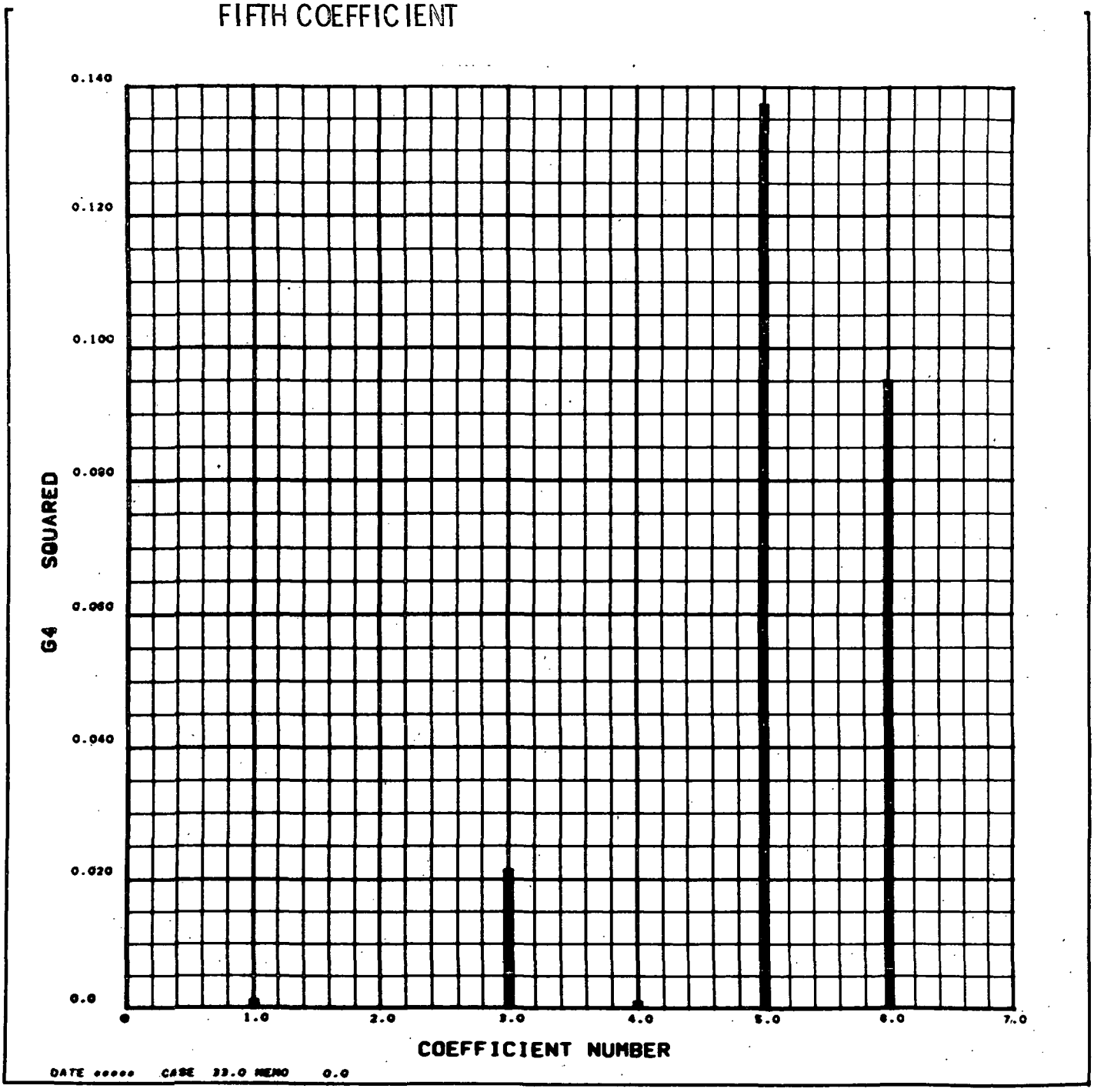


FIGURE 4.23 RELATIVE IMPORTANCE SPECTRUM FOR CANONICAL PREDICTION OF SIXTH COEFFICIENT

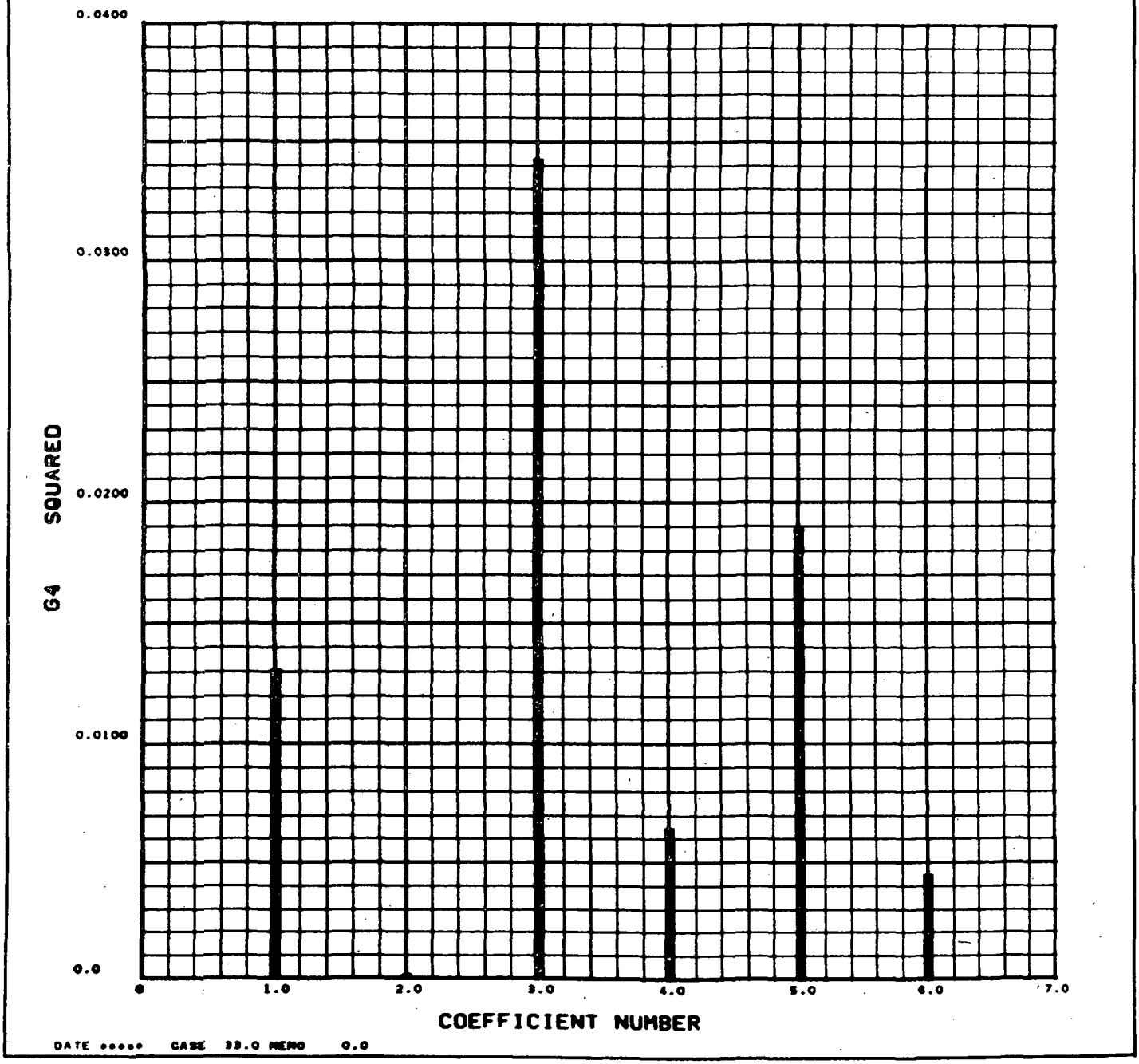


FIGURE 4.24 RELATIVE IMPORTANCE SPECTRUM FOR PREDICTION OF FIRST COEFFICIENT

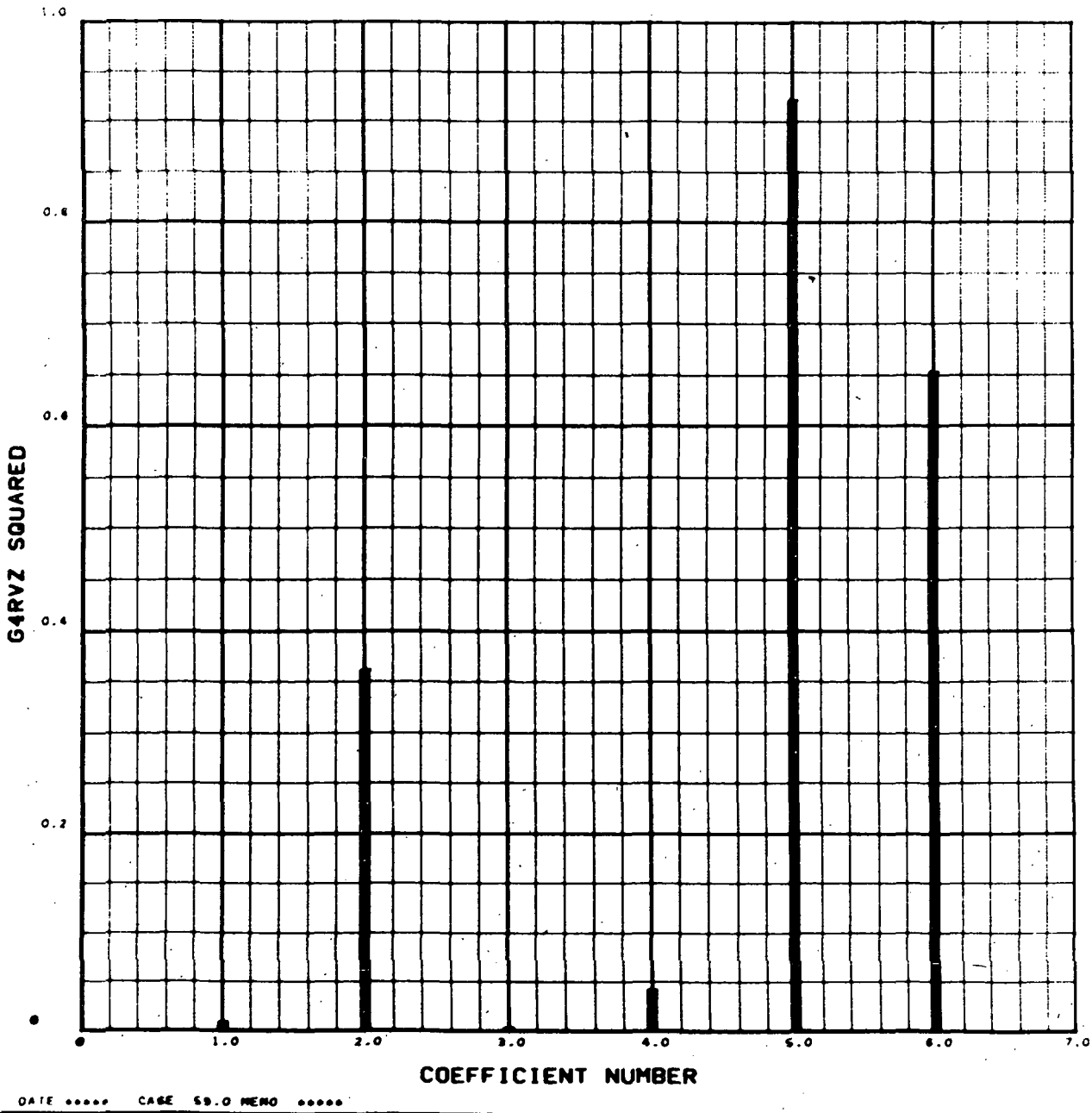


FIGURE 4.25 RELATIVE IMPORTANCE SPECTRUM FOR PREDICTION OF SECOND COEFFICIENT

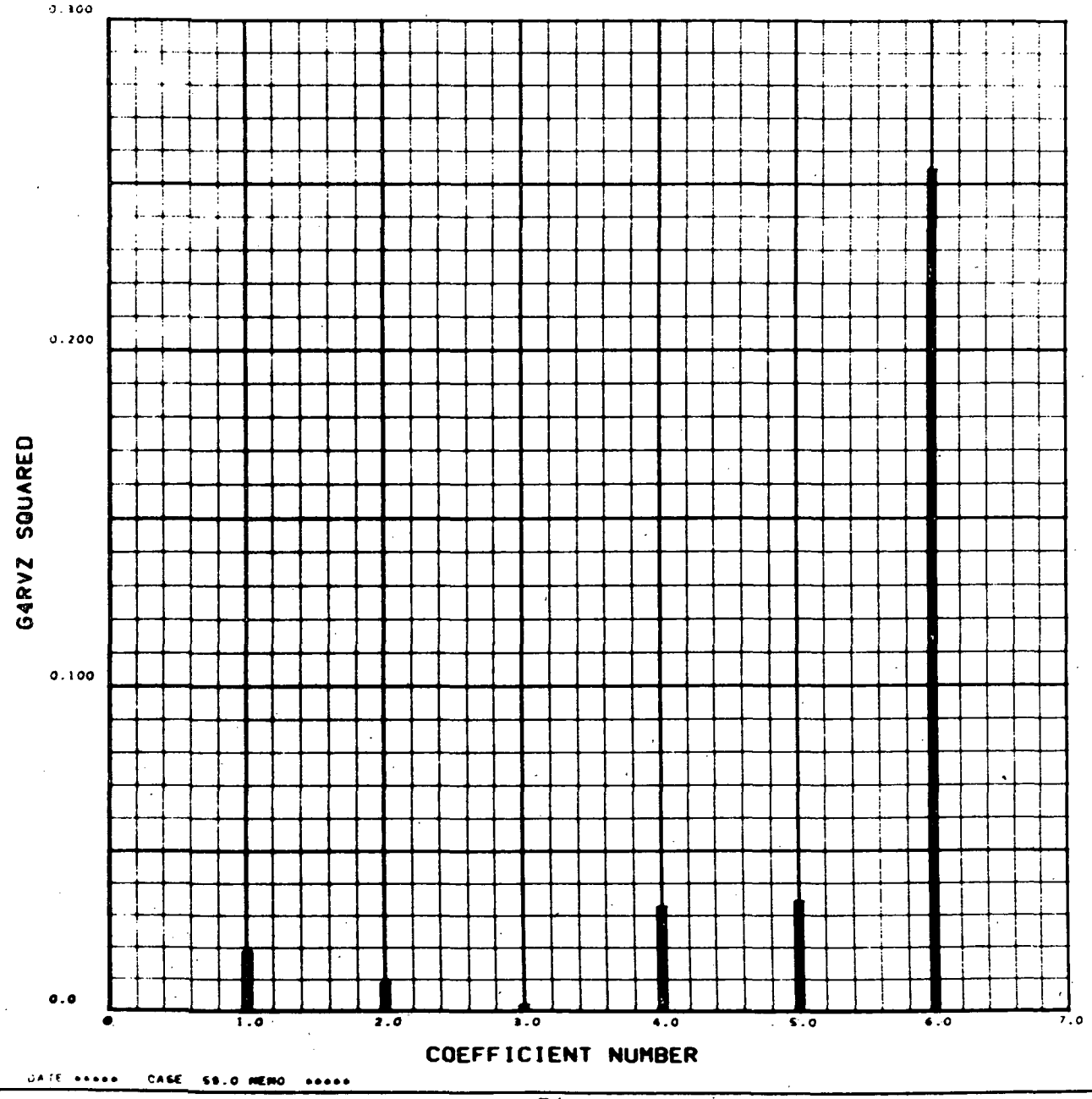


FIGURE 4.26 RELATIVE IMPORTANCE VECTOR FOR PREDICTION OF FIRST COEFFICIENT

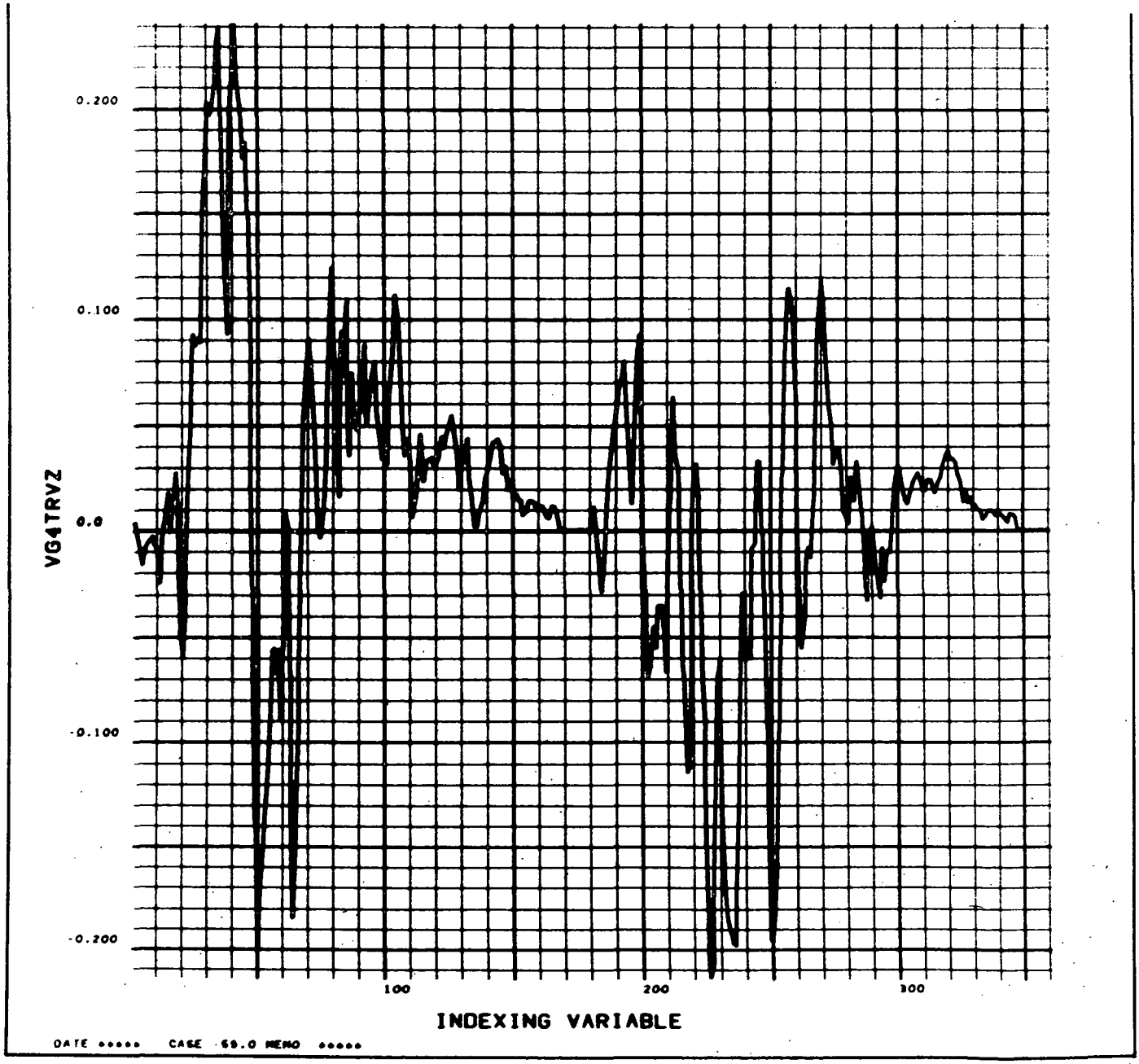


FIGURE 4.27 RELATIVE IMPORTANCE VECTOR FOR PREDICTION OF SECOND COEFFICIENT

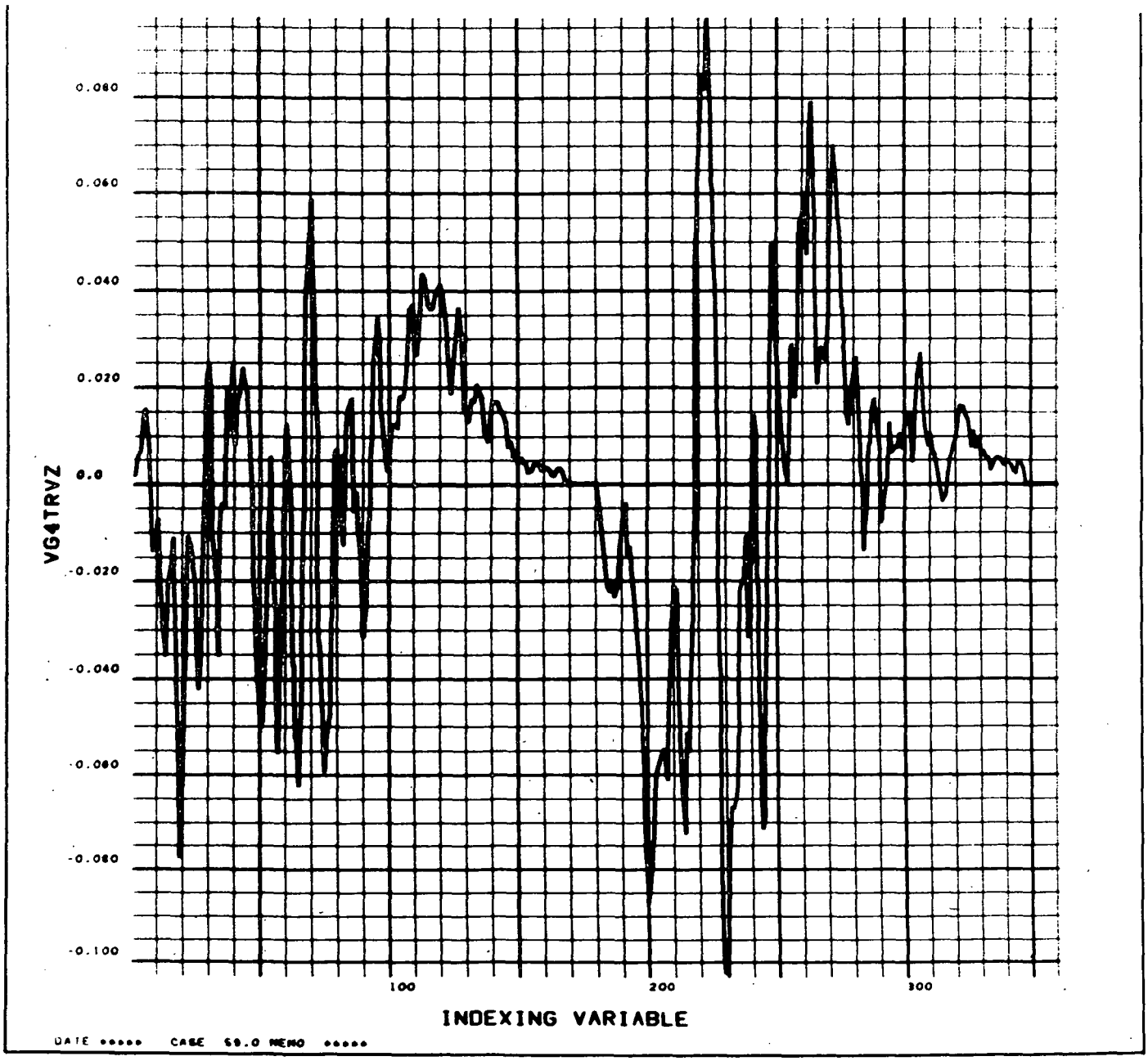


FIGURE 4.28 COMPARISON OF PREDICTED AND ACTUAL FIRST COEFFICIENT

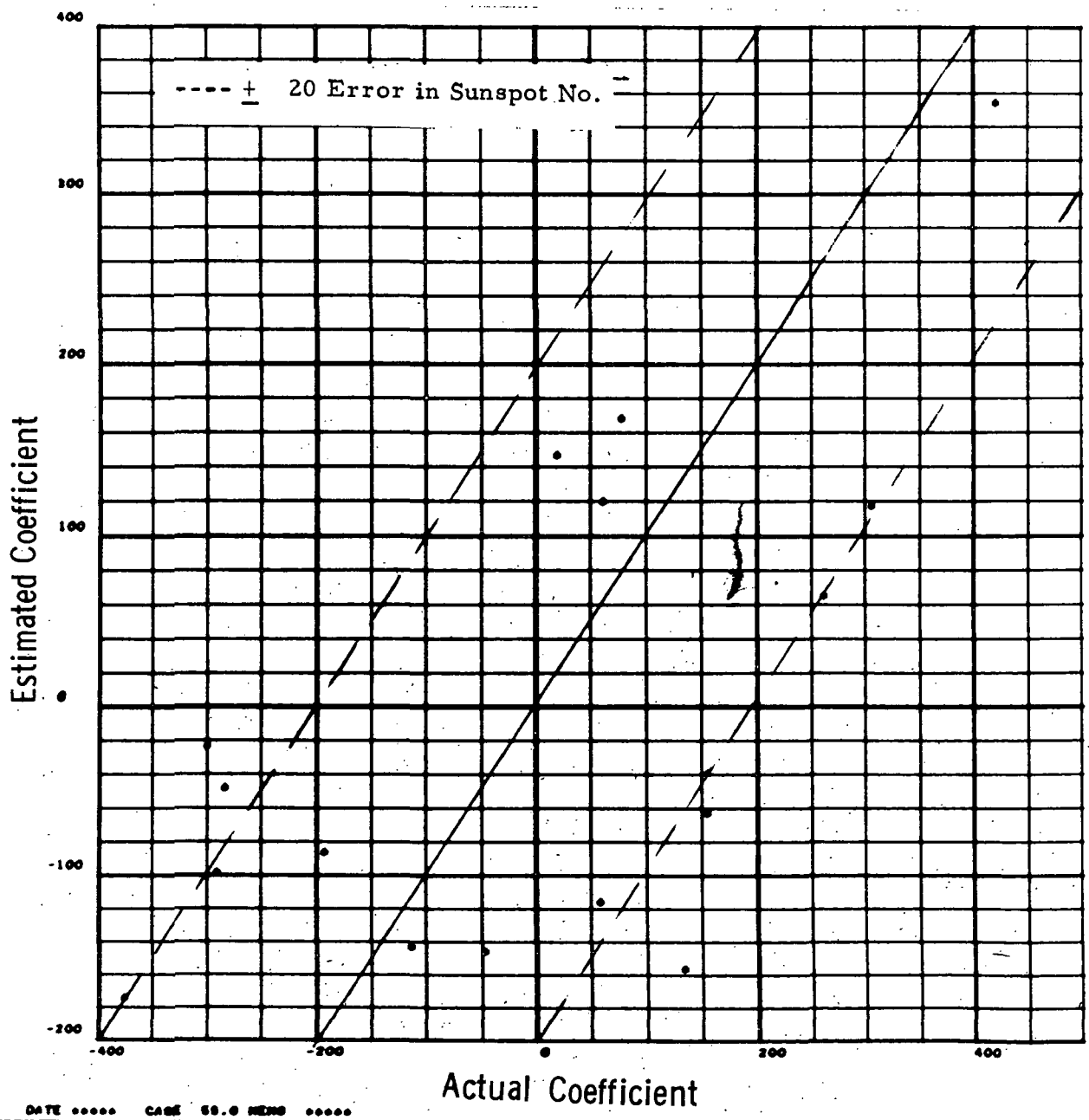


FIGURE 4.29 COMPARISON OF PREDICTED AND ACTUAL SECOND COEFFICIENT

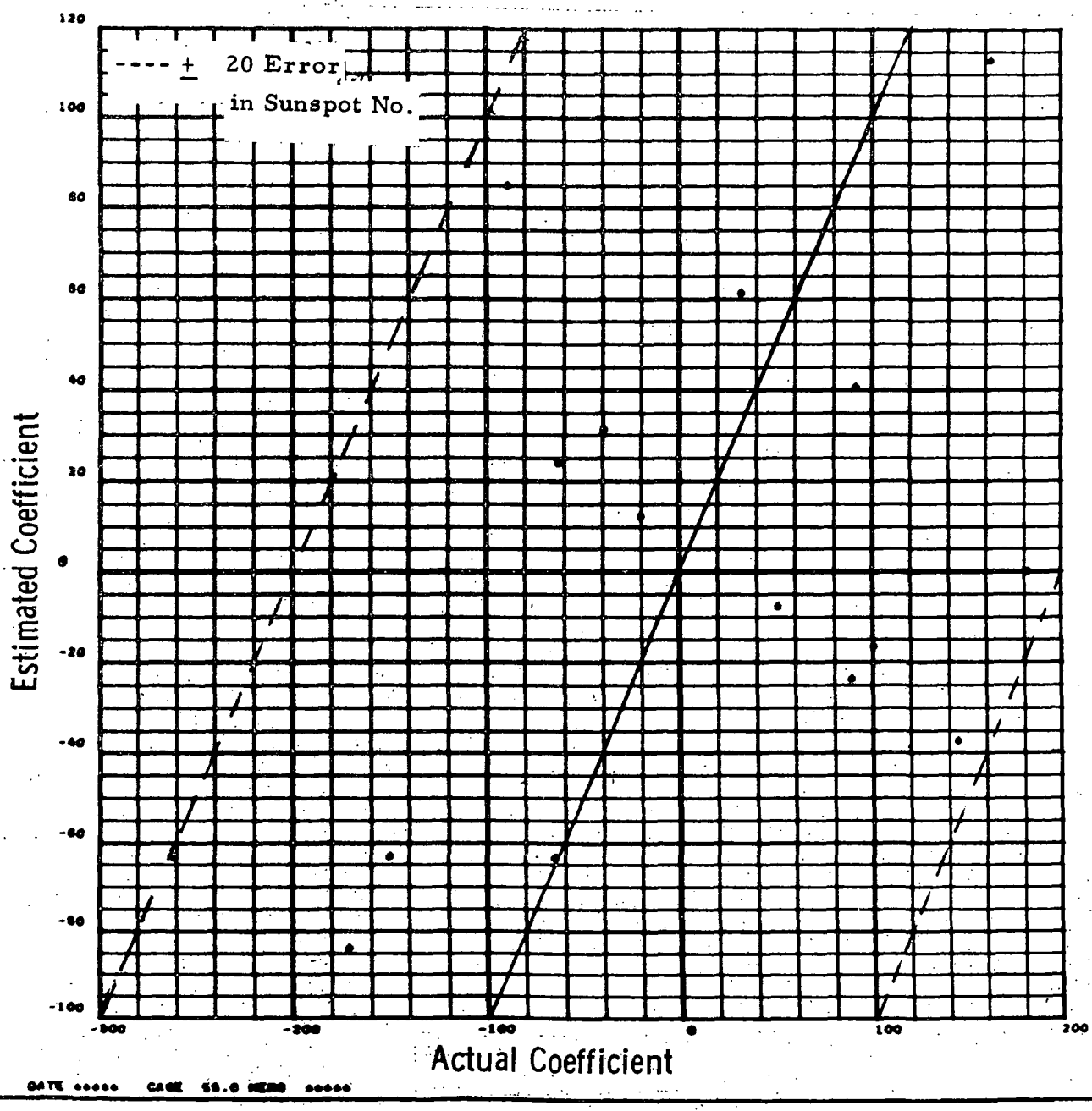


FIGURE 4.30 COMPARISON OF ACTUAL AND PREDICTED SCATTER PLOT

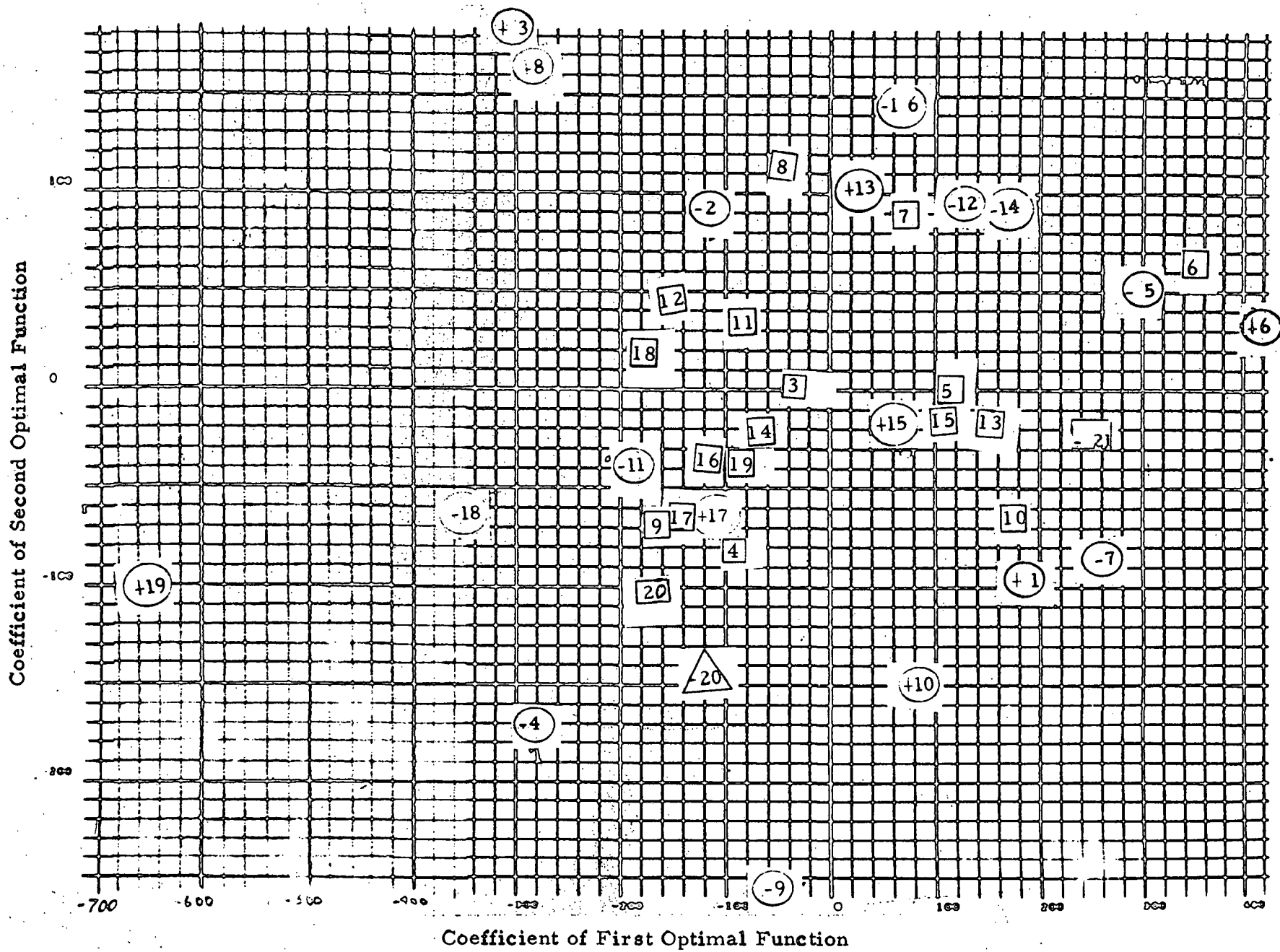


FIGURE 4.31 PREDICTED (USING CYCLES 17 AND 18) SUNSPOT NUMBER AND 2-SIGMA ERROR BOUNDS FOR CYCLE 19

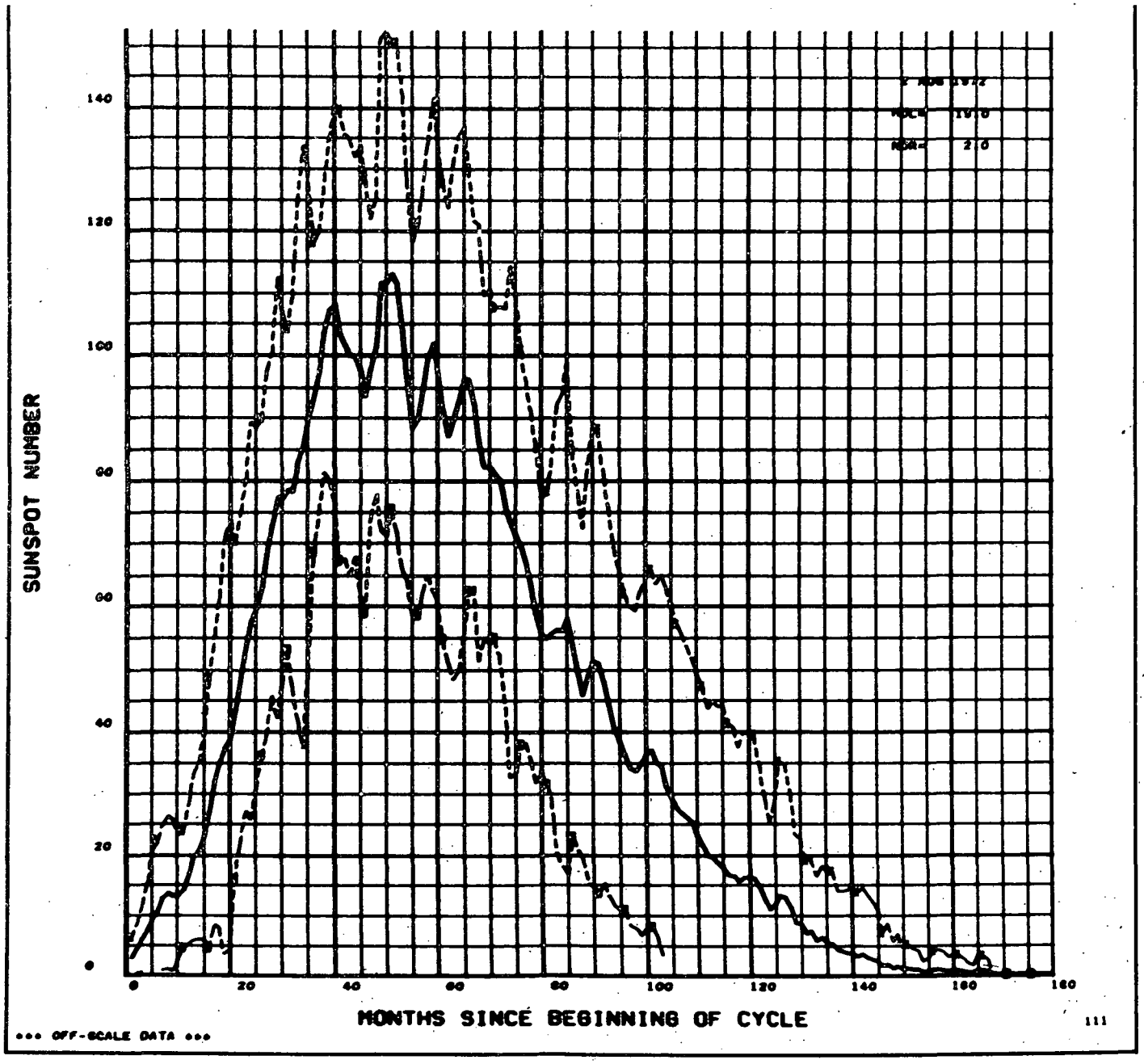


FIGURE 4.32 COMPARISON OF PREDICTED AND ACTUAL SUNSPOT NUMBER FOR CYCLE 19

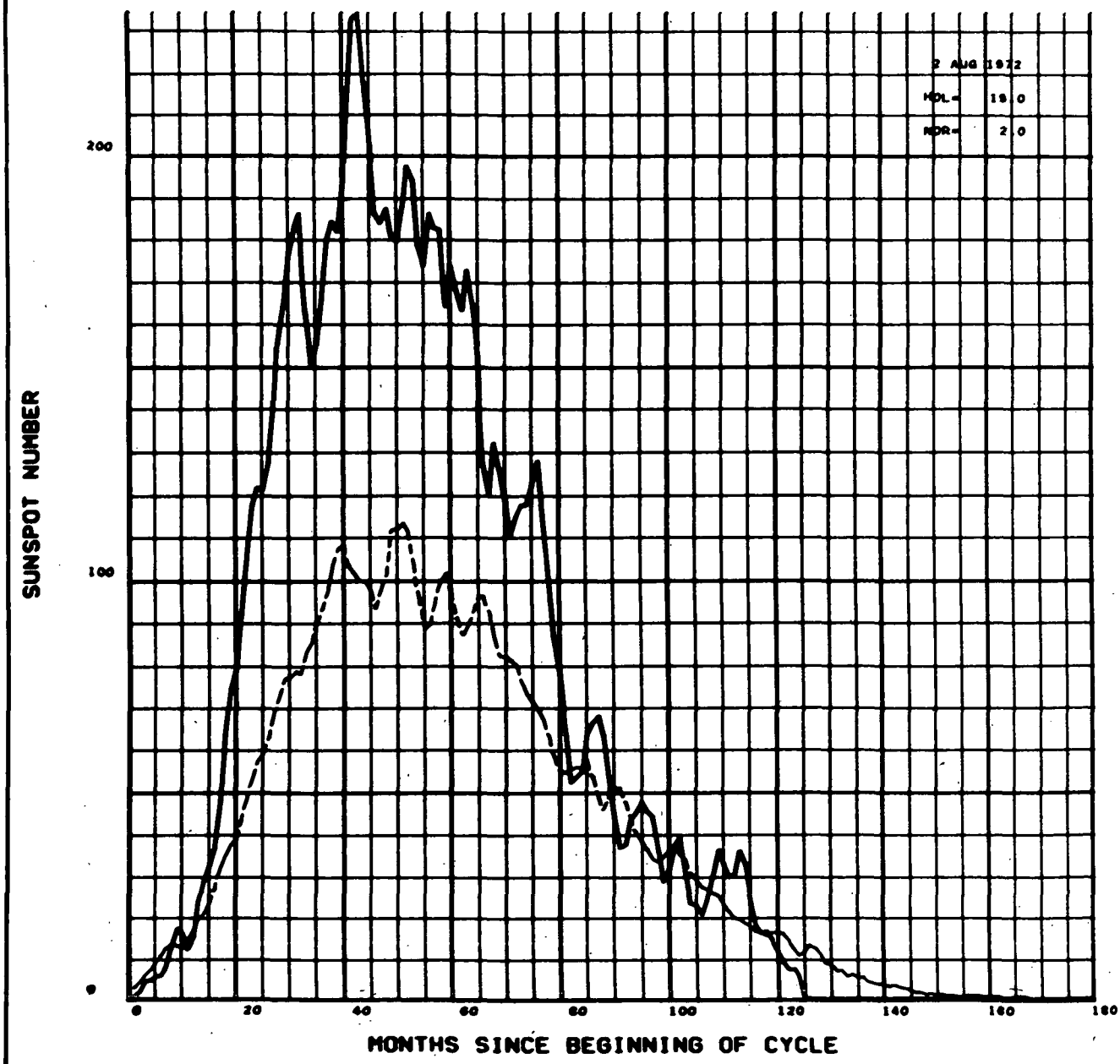


FIGURE 4.33 COMPARISON OF ACTUAL SUNSPOT NUMBER AND 2-SIGMA BOUNDS ON PREDICTIONS FOR CYCLE 19

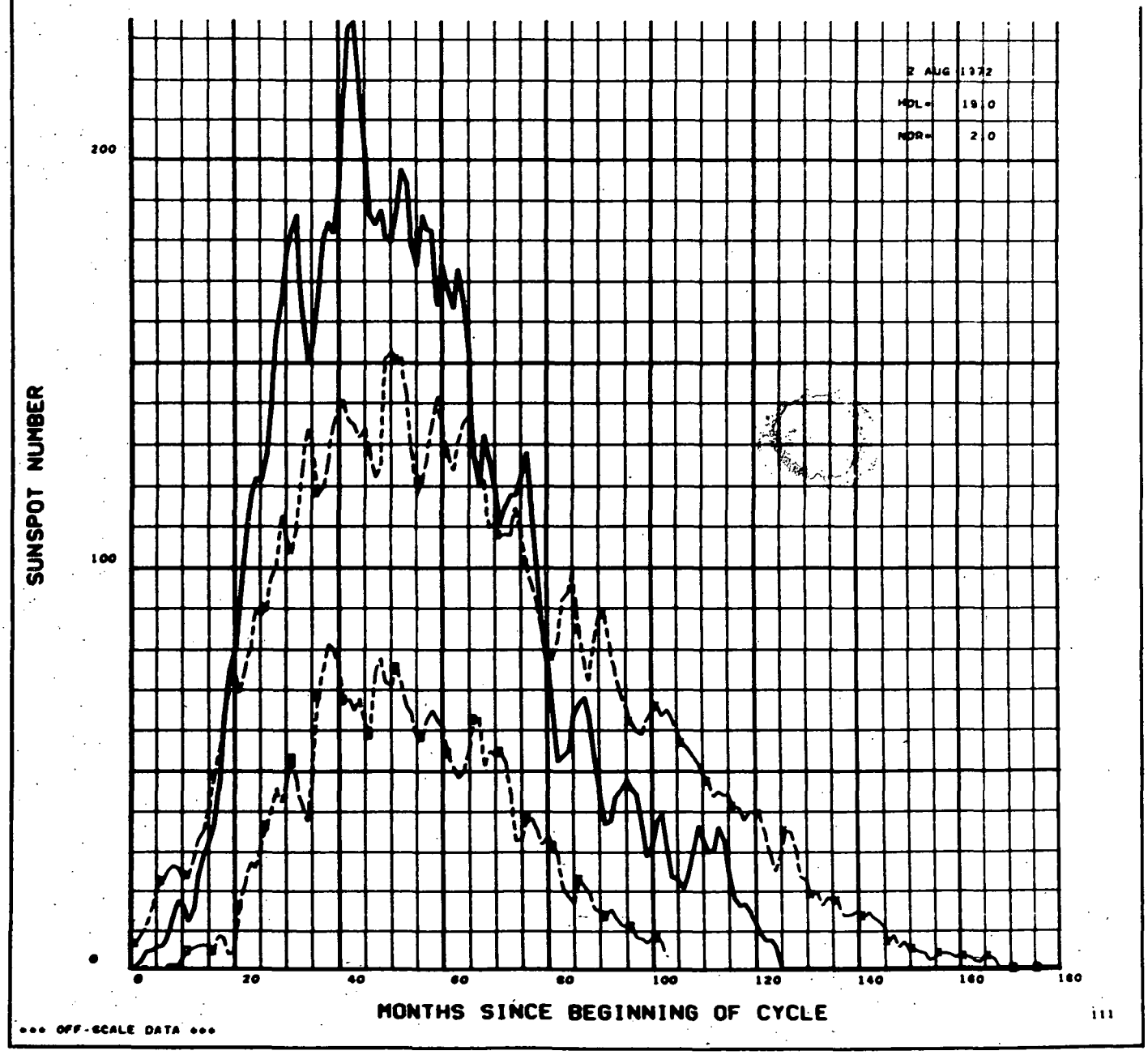


FIGURE 4.34 PREDICTED (USING CYCLES 18 AND 19) SUNSPOT NUMBERS AND 2-SIGMA ERROR BOUNDS FOR CYCLE 20

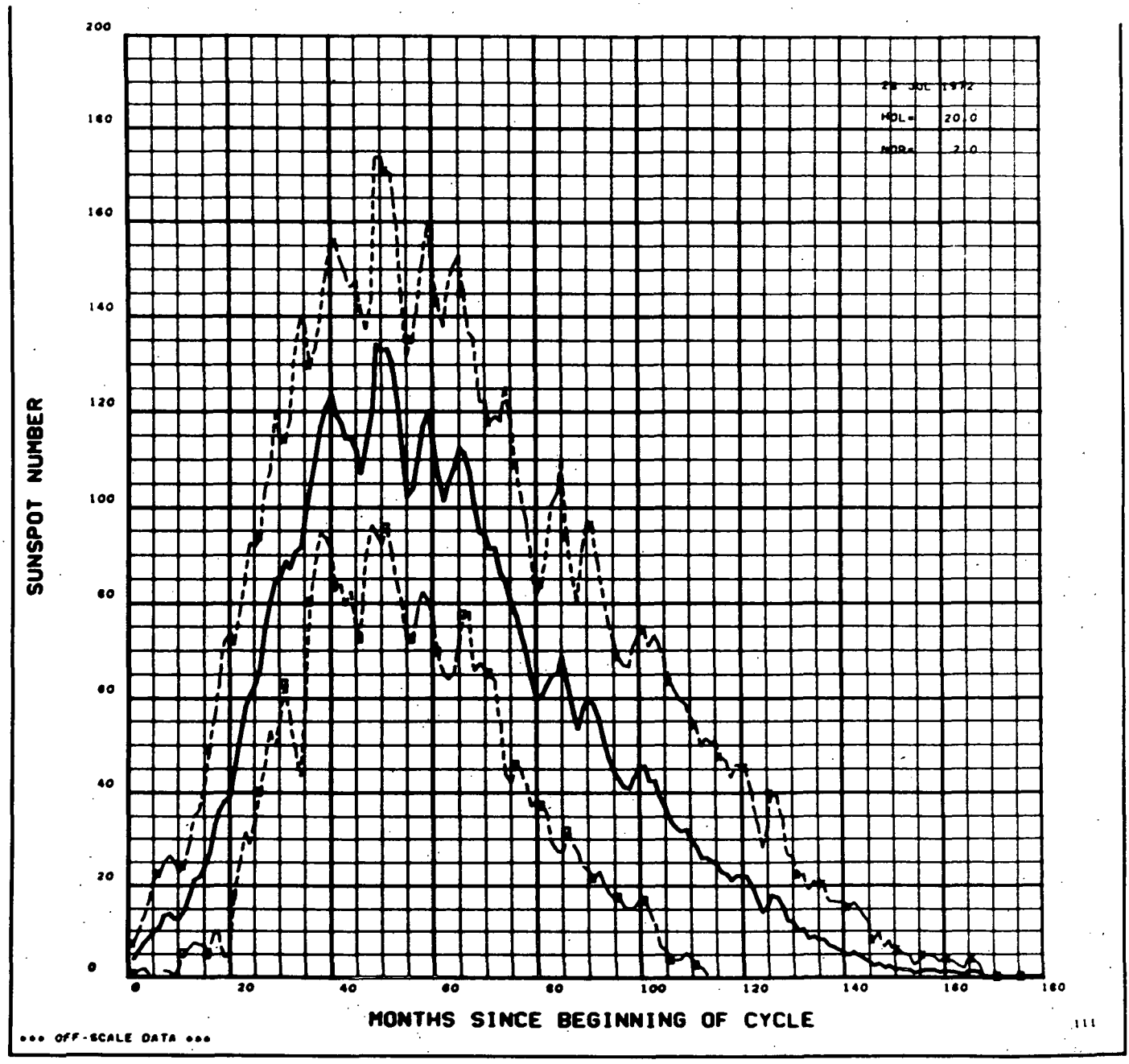


FIGURE 4.35 COMPARISON OF PREDICTED AND ACTUAL SUNSPOT NUMBERS FOR CYCLE 20

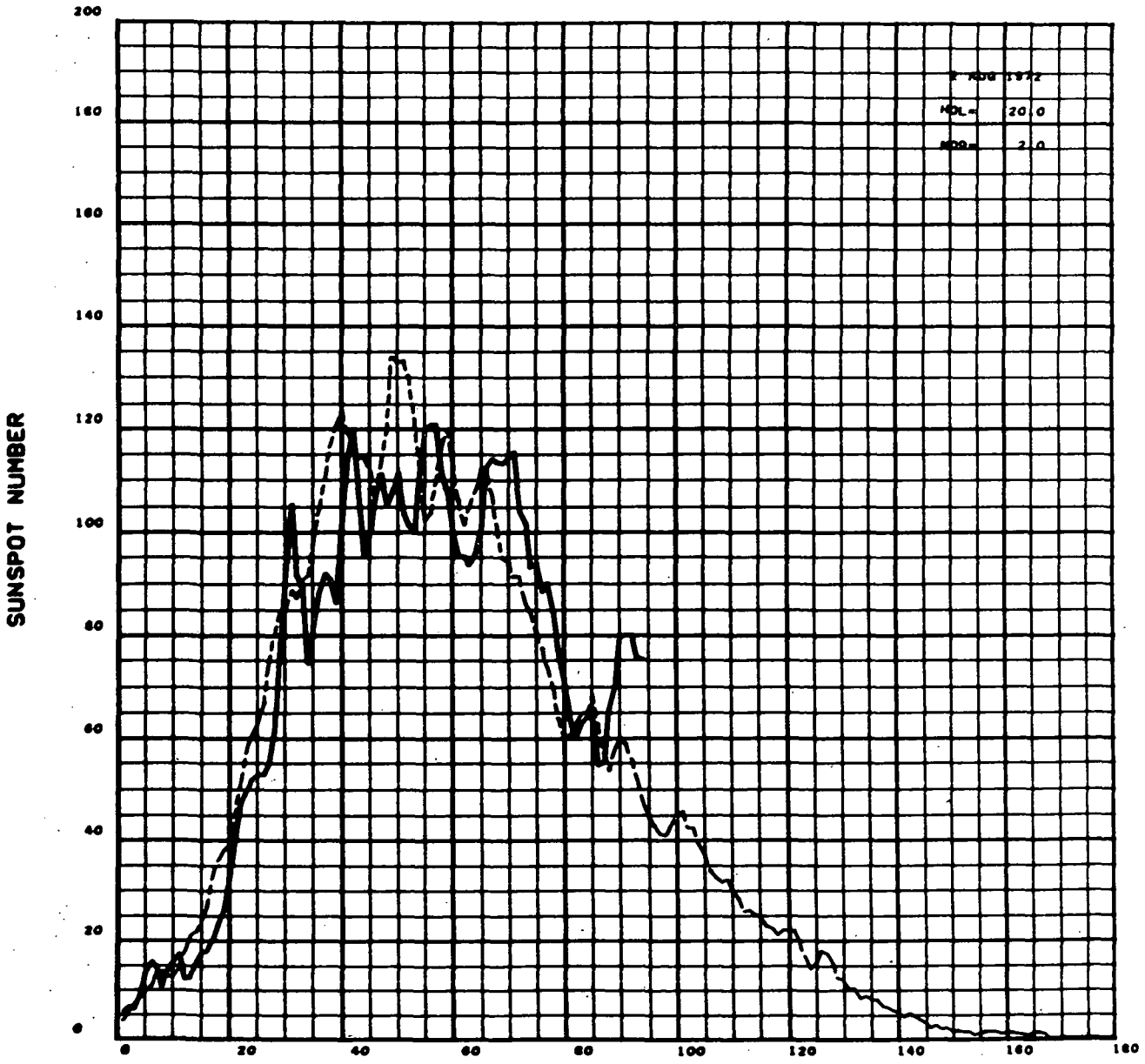


FIGURE 4.36 COMPARISON OF ACTUAL SUNSPOT NUMBER AND 2-SIGMA BOUNDS ON PREDICTION FOR CYCLE 20

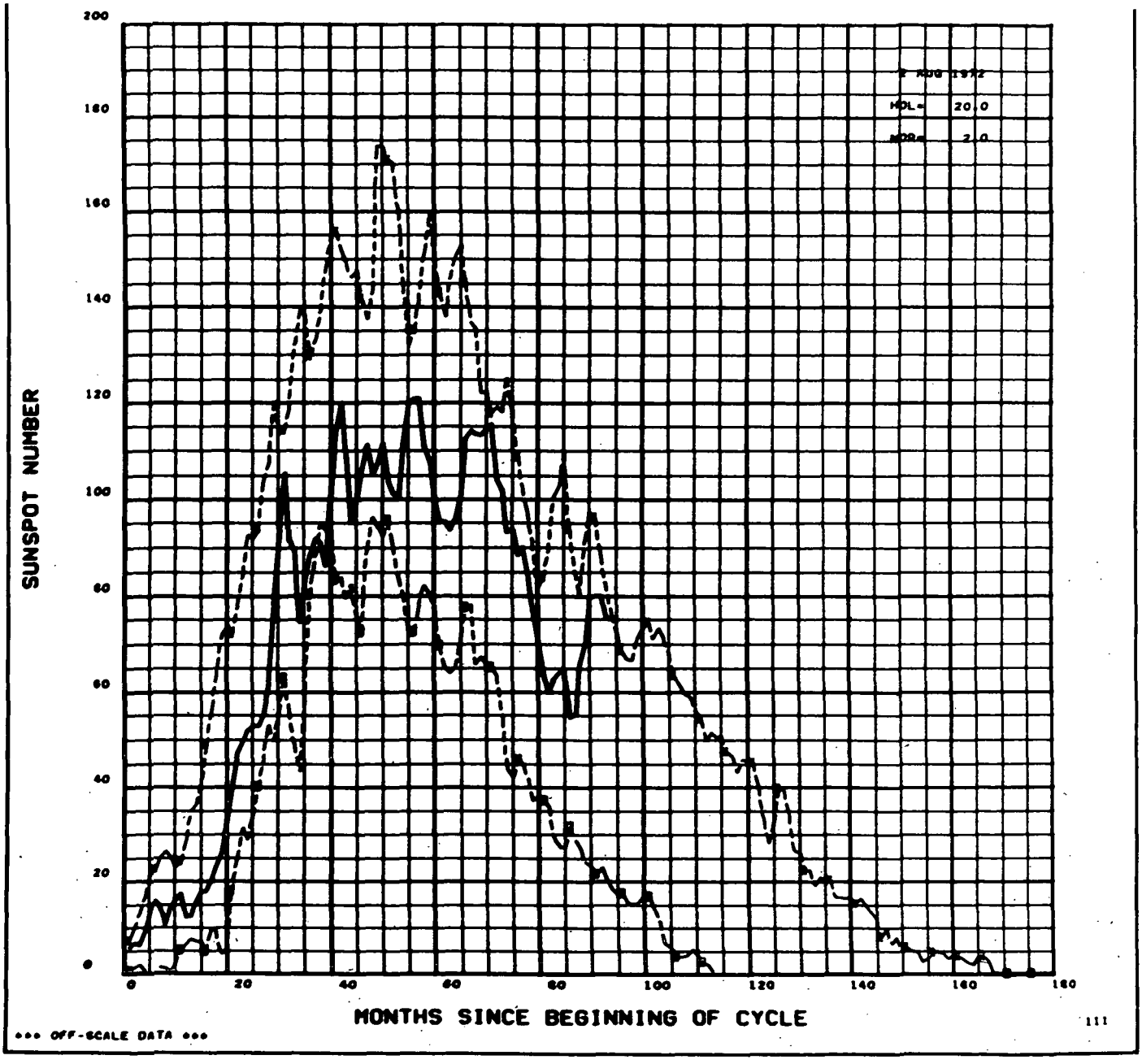


FIGURE 4.37 EFFECT OF DIMENSIONALITY ON THE RMS ERROR OF THE SUNSPOTS EXTRAPOLATION

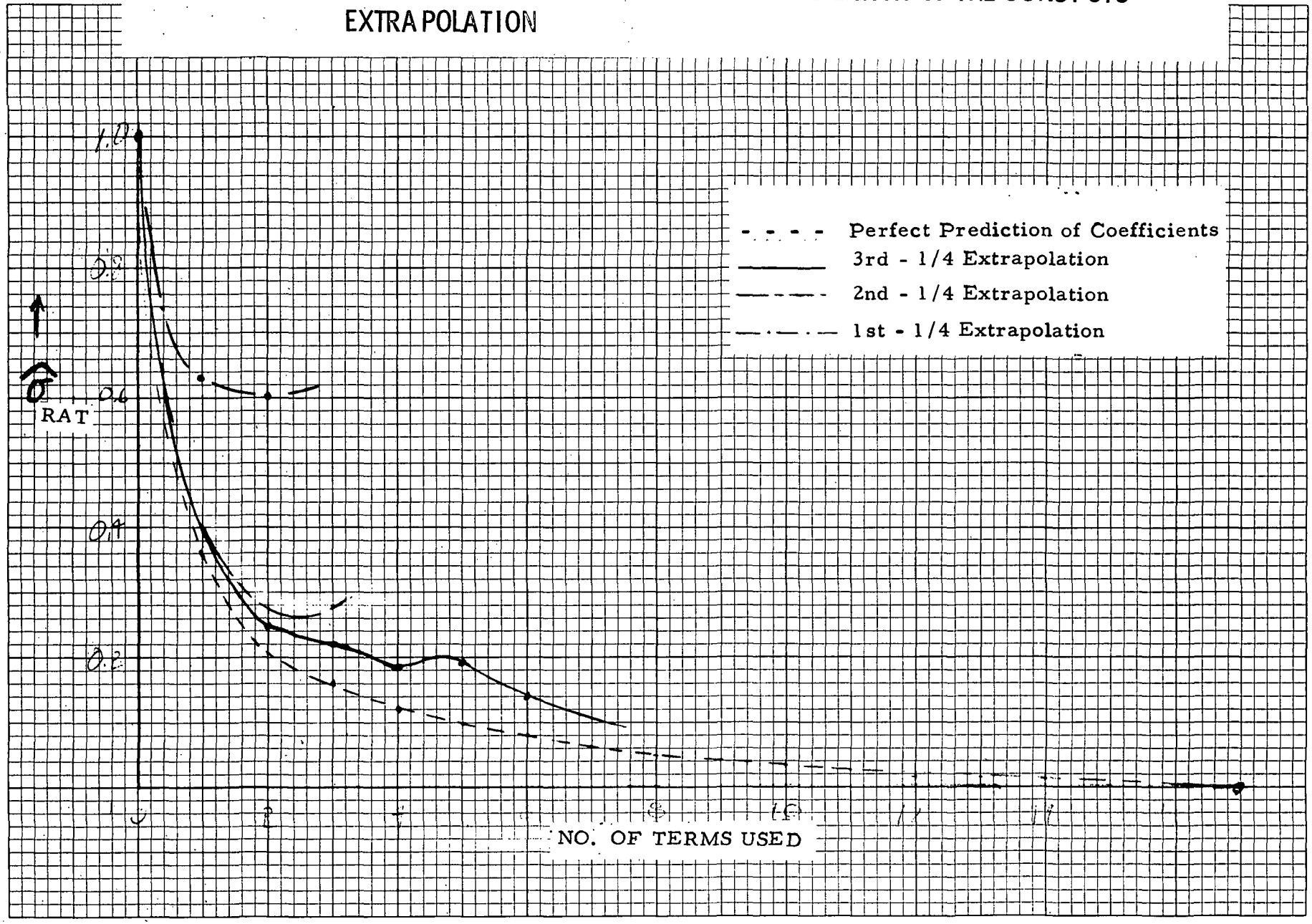


FIGURE 4.38 SCATTER PLOT COMPARISON OF THE PREDICTED AND ACTUAL FIRST COEFFICIENT

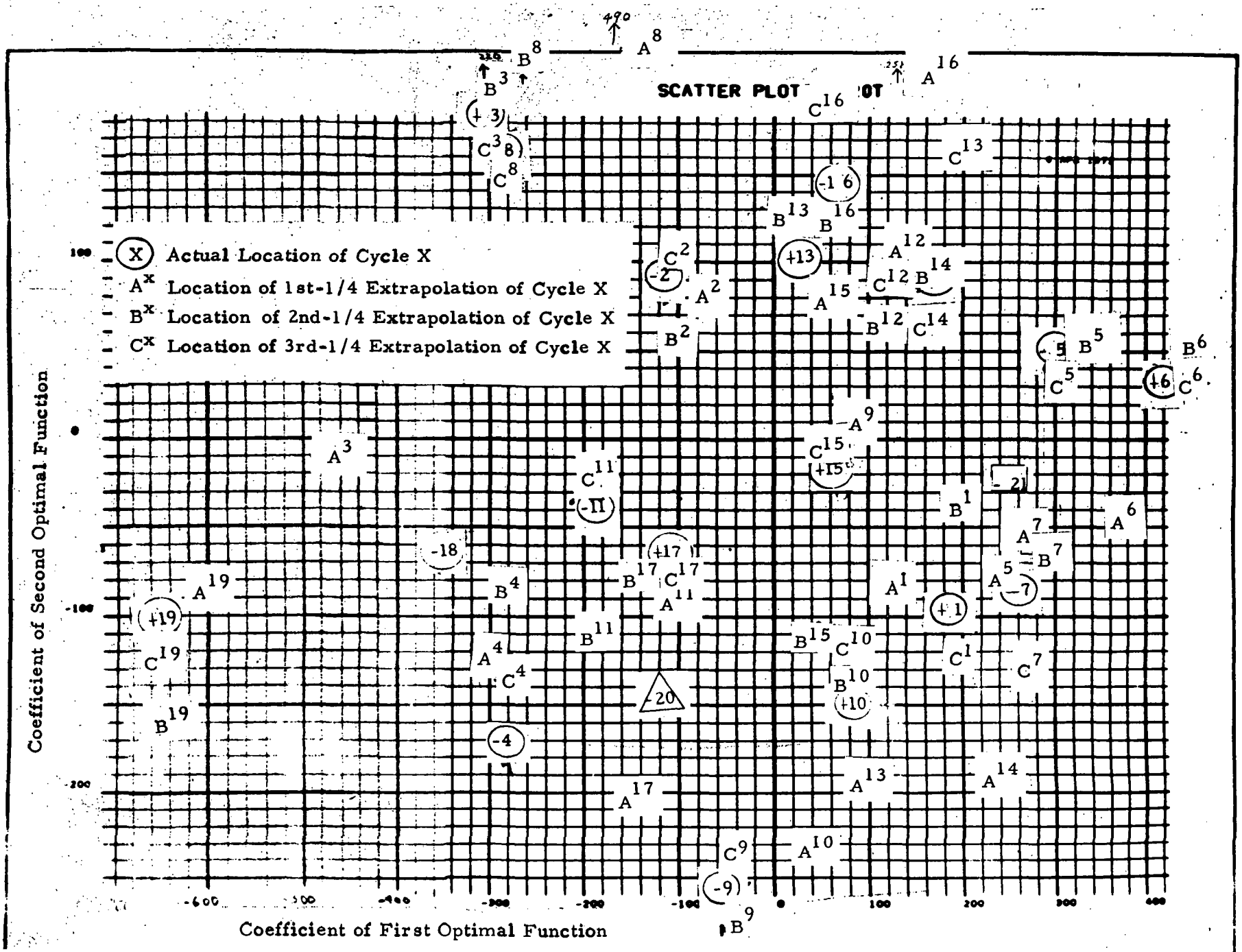


FIGURE 4.39 1ST QUARTER EXTRAPOLATION OF SUNSPOT NUMBER AND 2-SIGMA ERROR BOUNDS FOR CYCLE 19

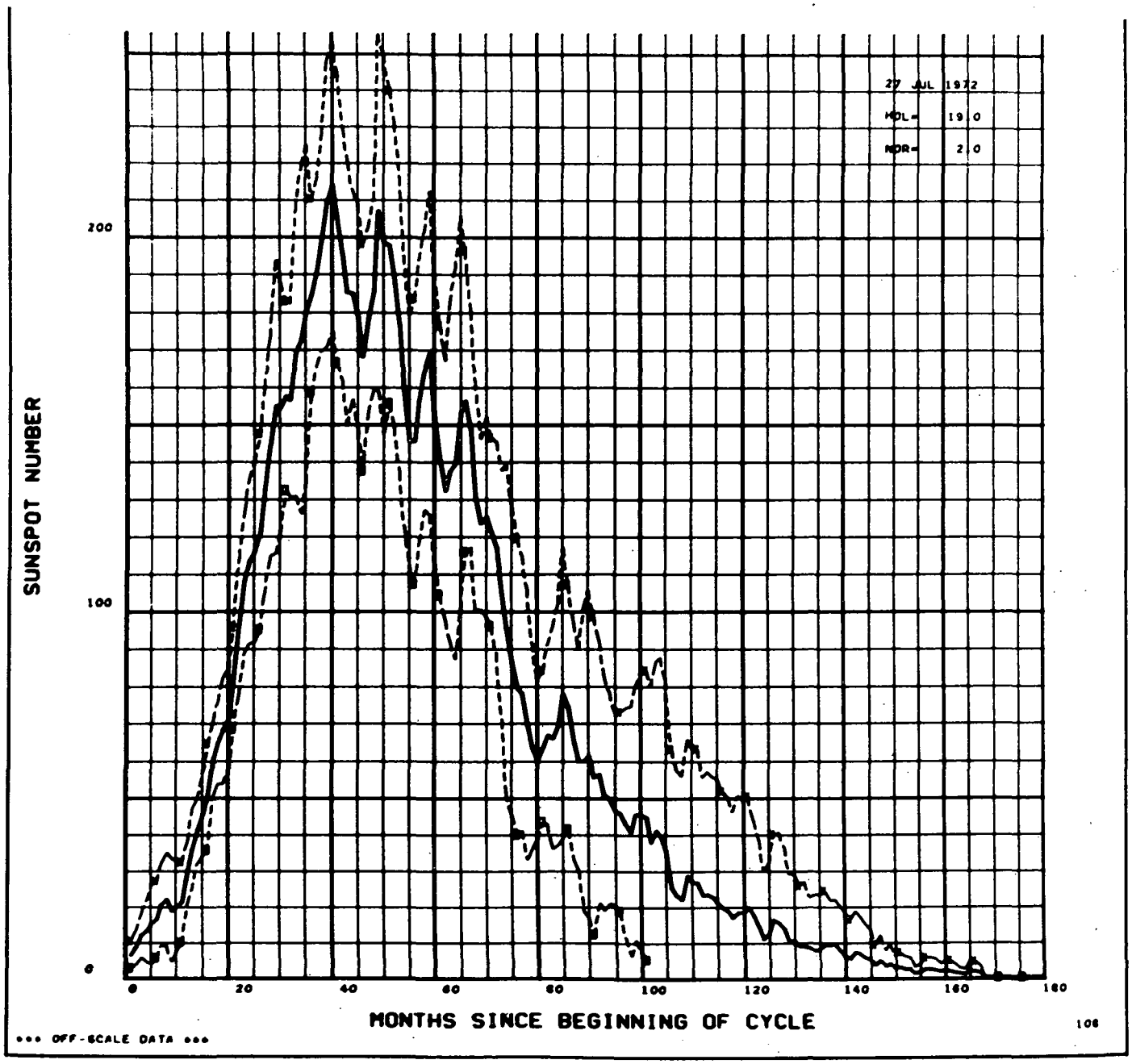


FIGURE 4.40 COMPARISON OF ACTUAL AND 1ST QUARTER EXTRAPOLATION OF SUNSPOT CYCLE 19

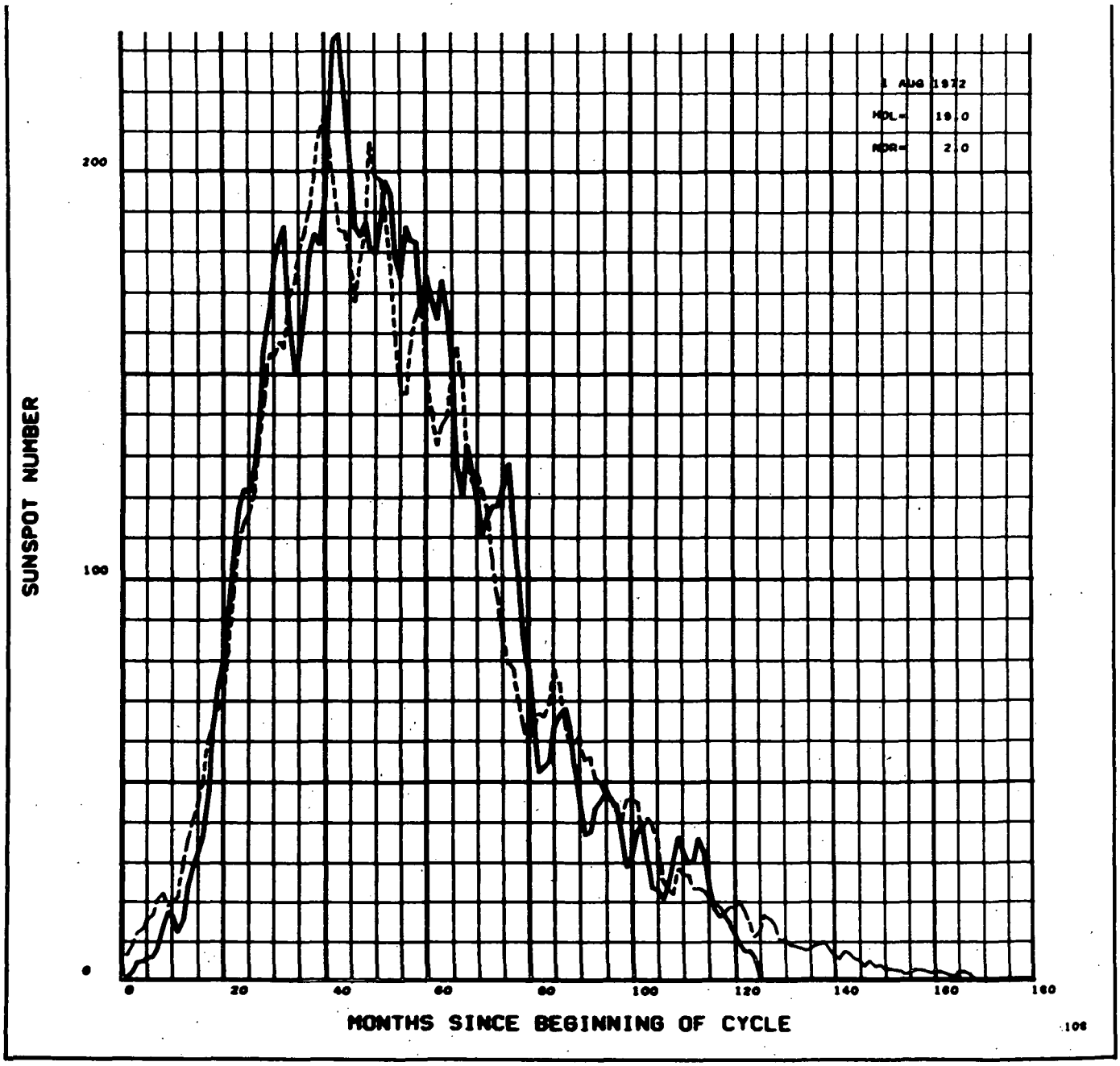


FIGURE 4.41 COMPARISON OF ACTUAL AND 2-SIGMA ERROR BOUNDS FOR 1ST QUARTER EXTRAPOLATION OF SUNSPOT CYCLE 19

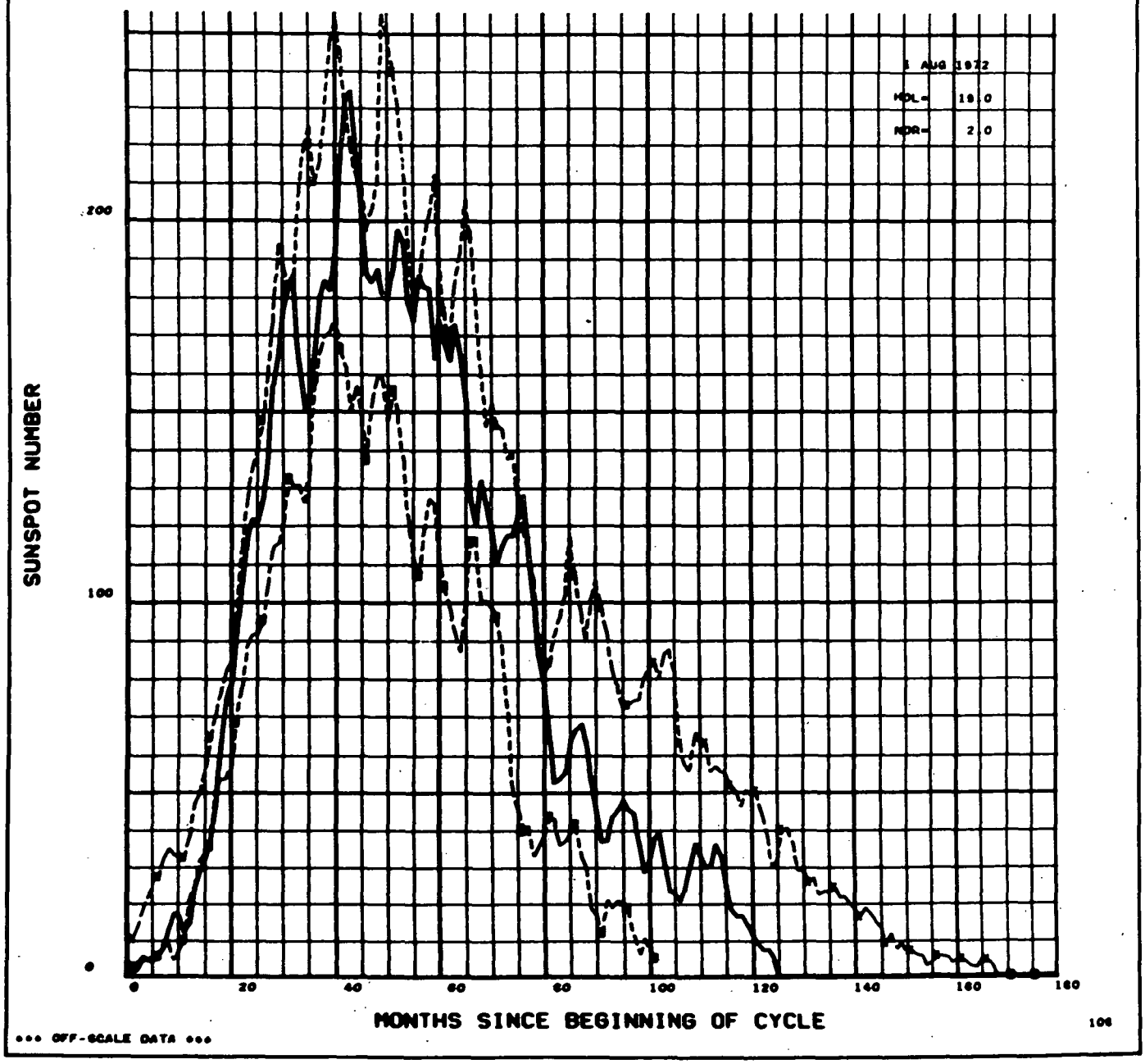


FIGURE 4.42 1ST QUARTER EXTRAPOLATION OF SUNSPOT NUMBER AND 2-SIGMA ERROR BOUNDS FOR CYCLE 20

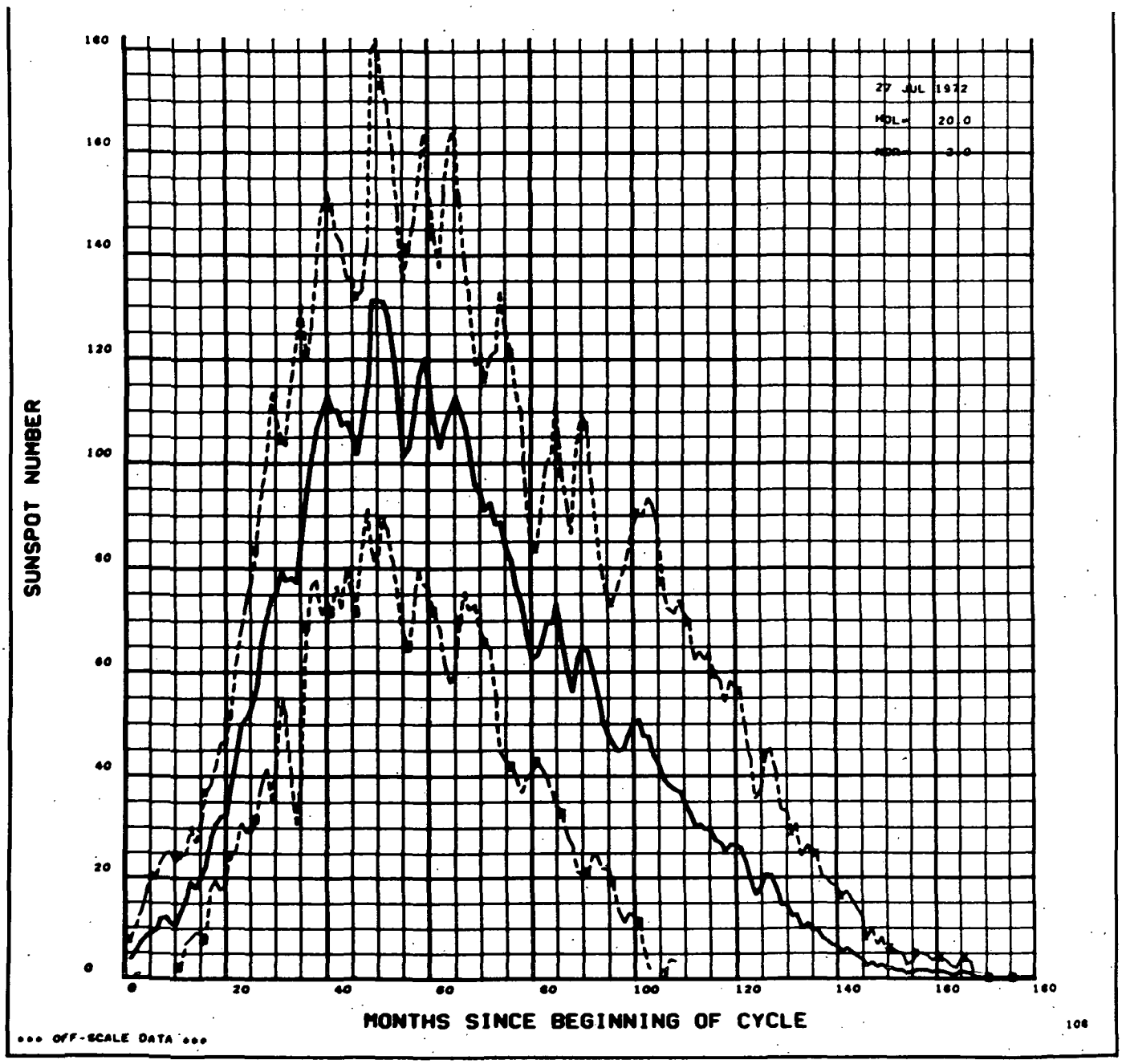


FIGURE 4.43 COMPARISON OF ACTUAL AND 1ST QUARTER EXTRAPOLATION OF SUNSPOT CYCLE 20

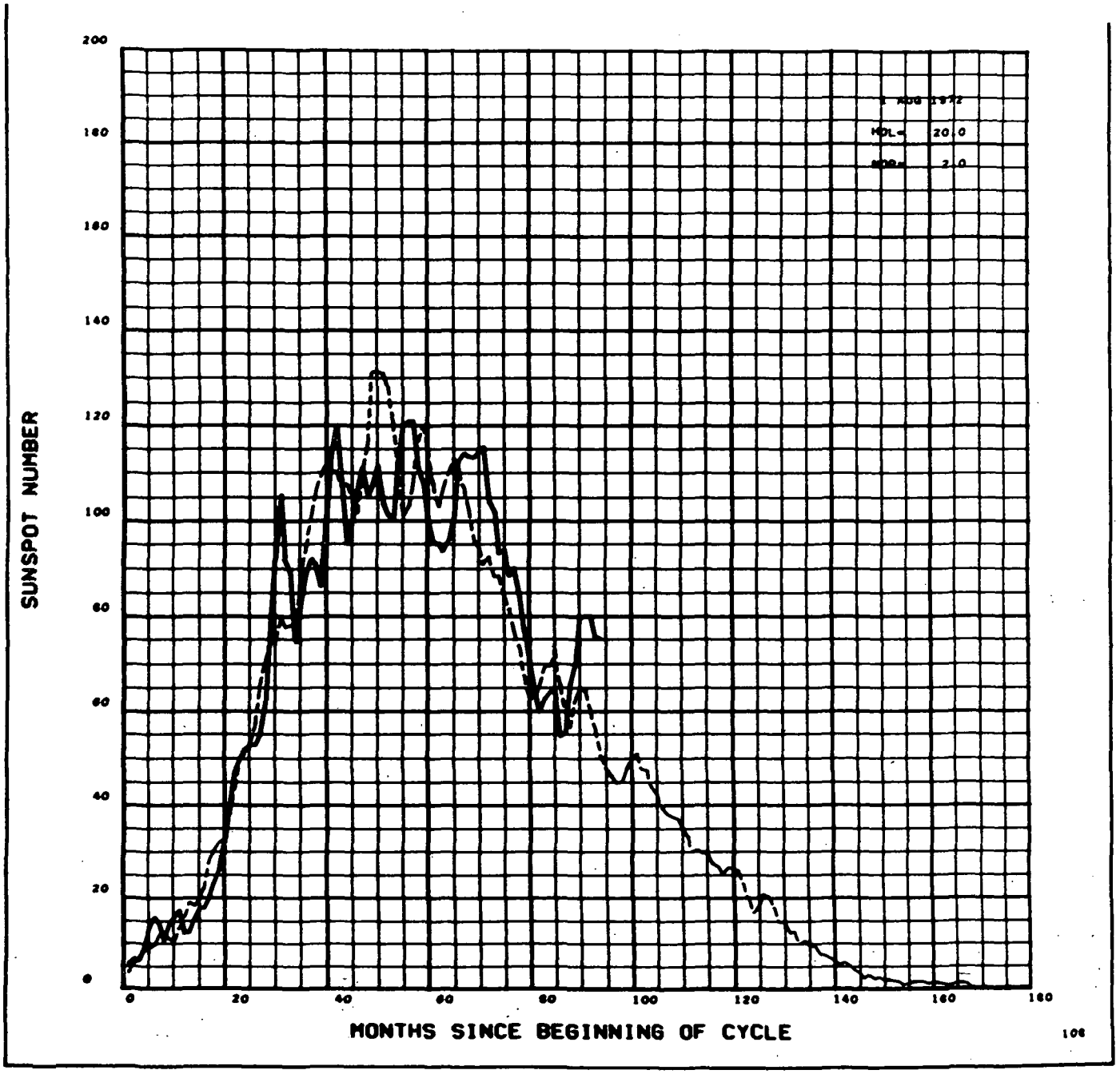


FIGURE 4.44 COMPARISON OF ACTUAL AND 2-SIGMA ERROR BOUNDS FOR 1ST QUARTER EXTRAPOLATION OF SUNSPOT CYCLE 20

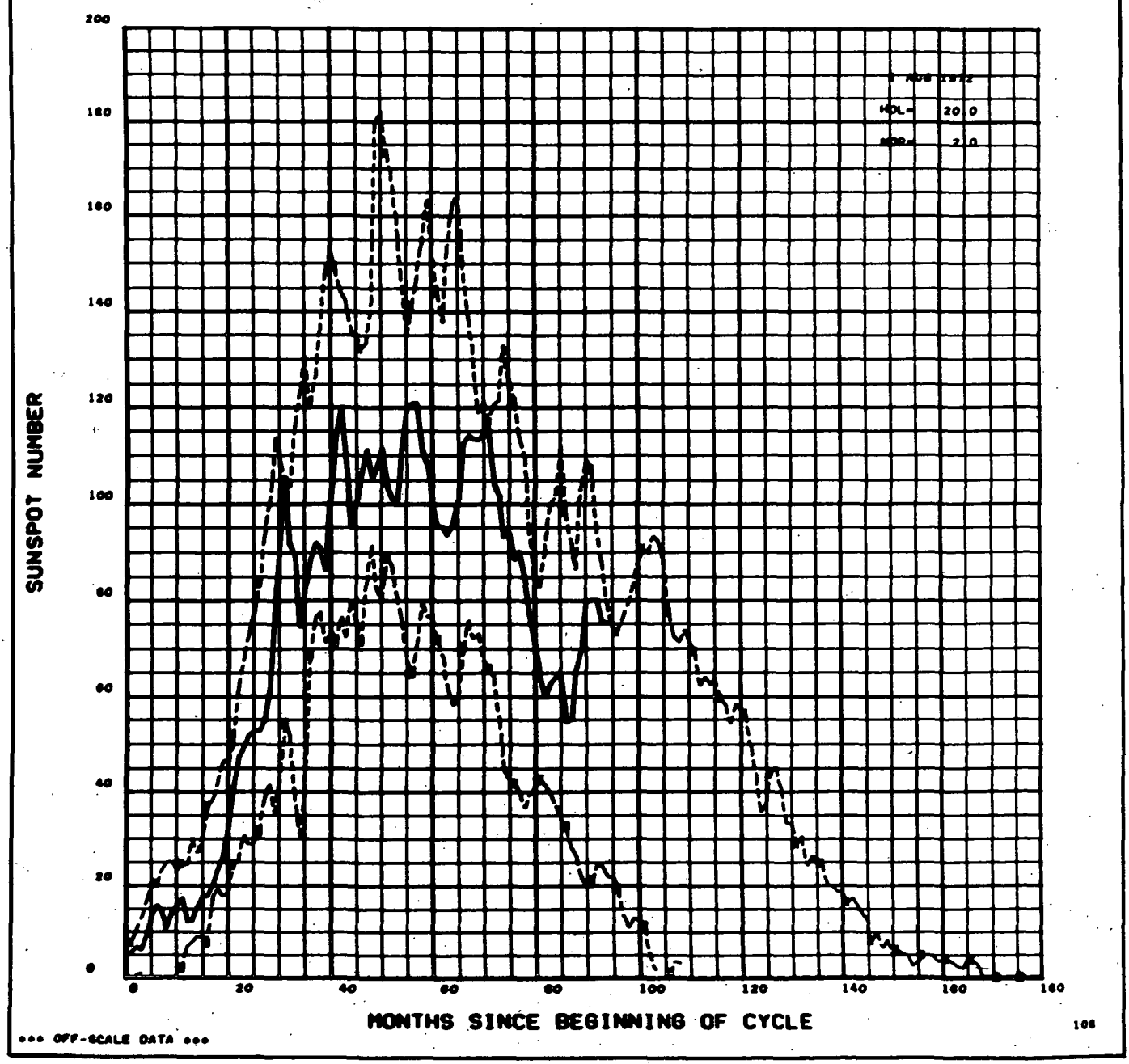


FIGURE 4.45 2ND QUARTER EXTRAPOLATION OF SUNSPOT NUMBER AND 2-SIGMA ERROR BOUNDS FOR CYCLE 19

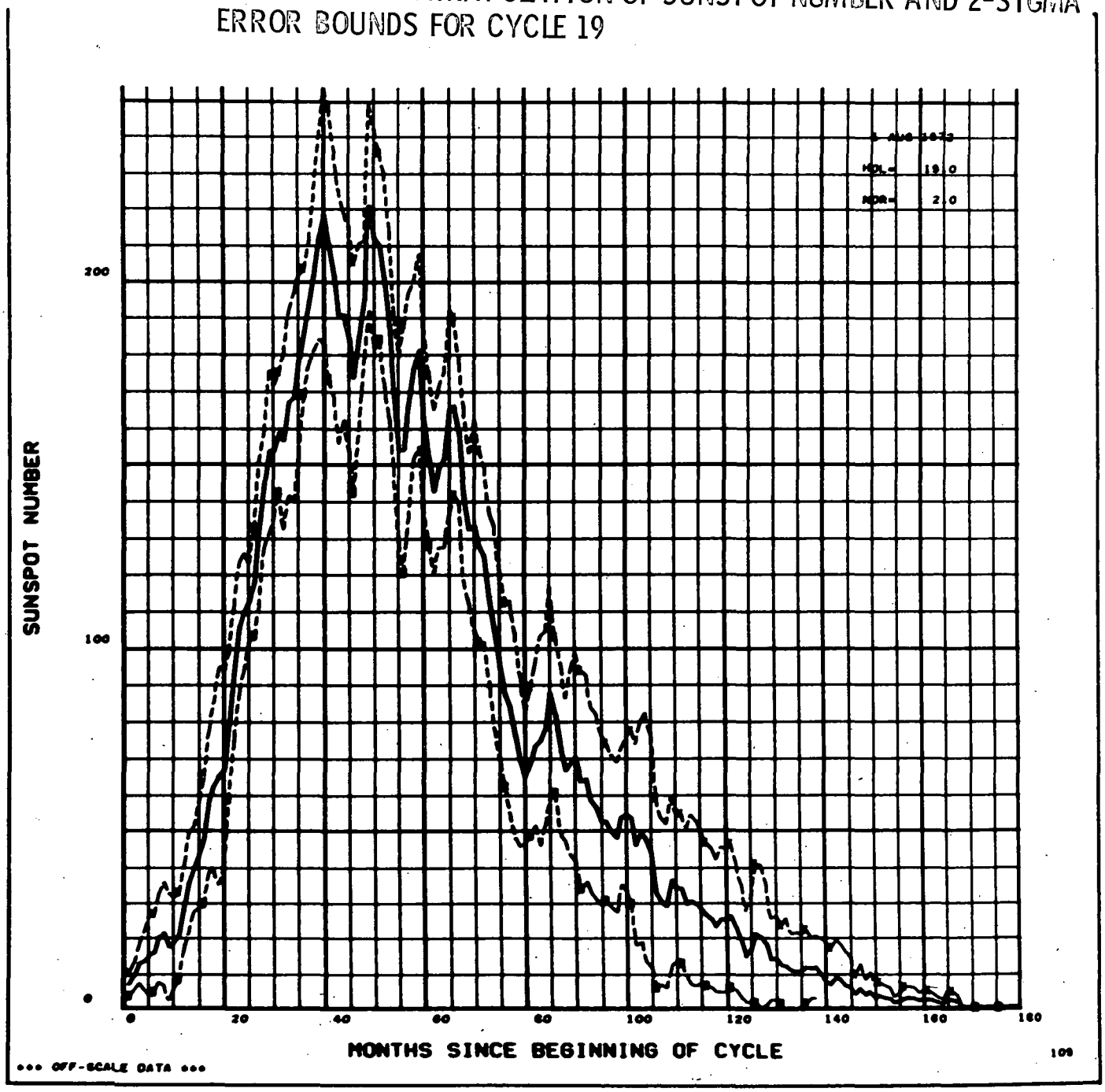


FIGURE 4.46 COMPARISON OF ACTUAL AND 2ND QUARTER EXTRAPOLATION OF SUNSPOT CYCLE 19

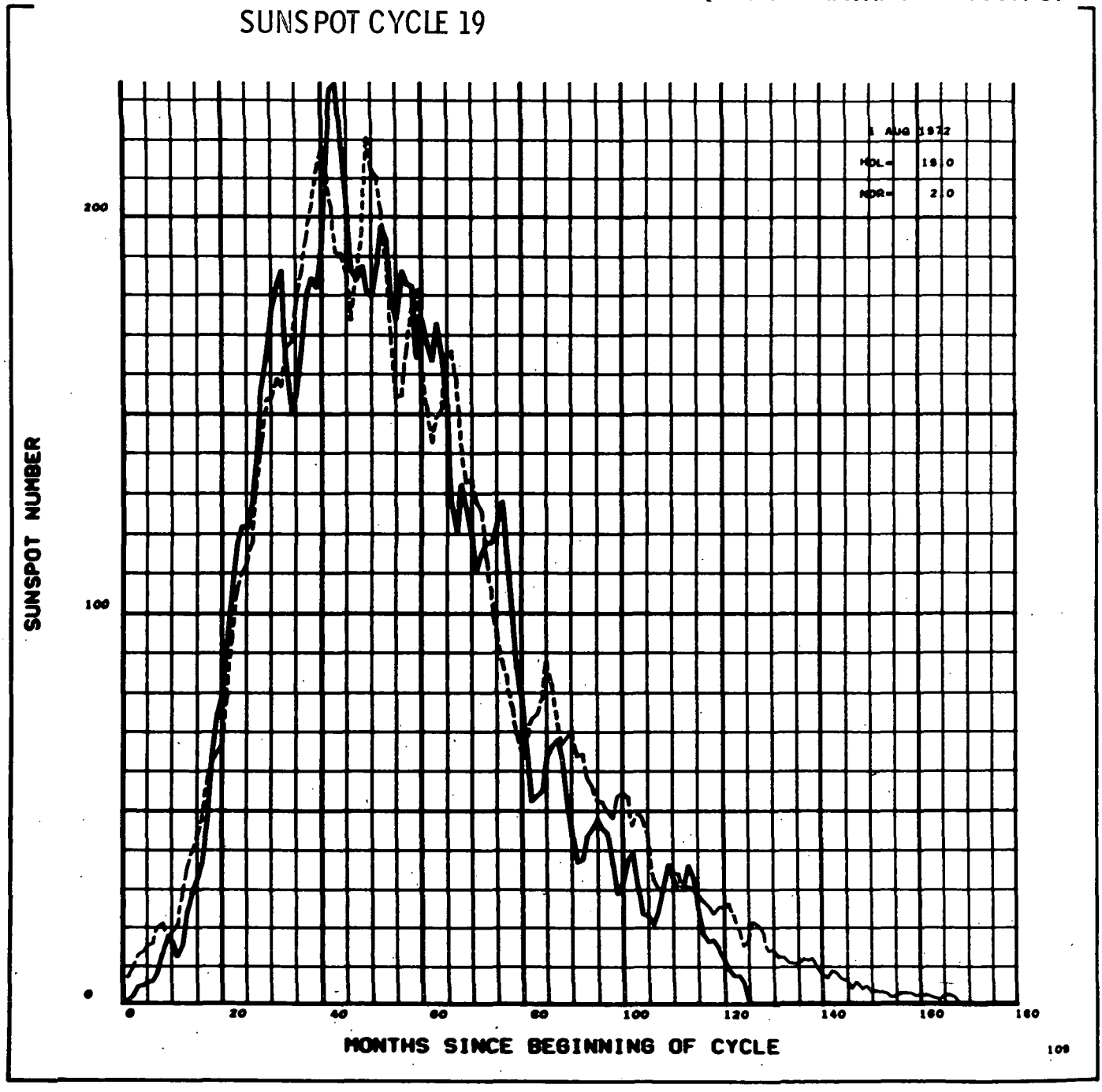


FIGURE 4.47 COMPARISON OF ACTUAL AND 2-SIGMA ERROR BOUNDS FOR 2ND QUARTER EXTRAPOLATION OF SUNSPOT CYCLE 19

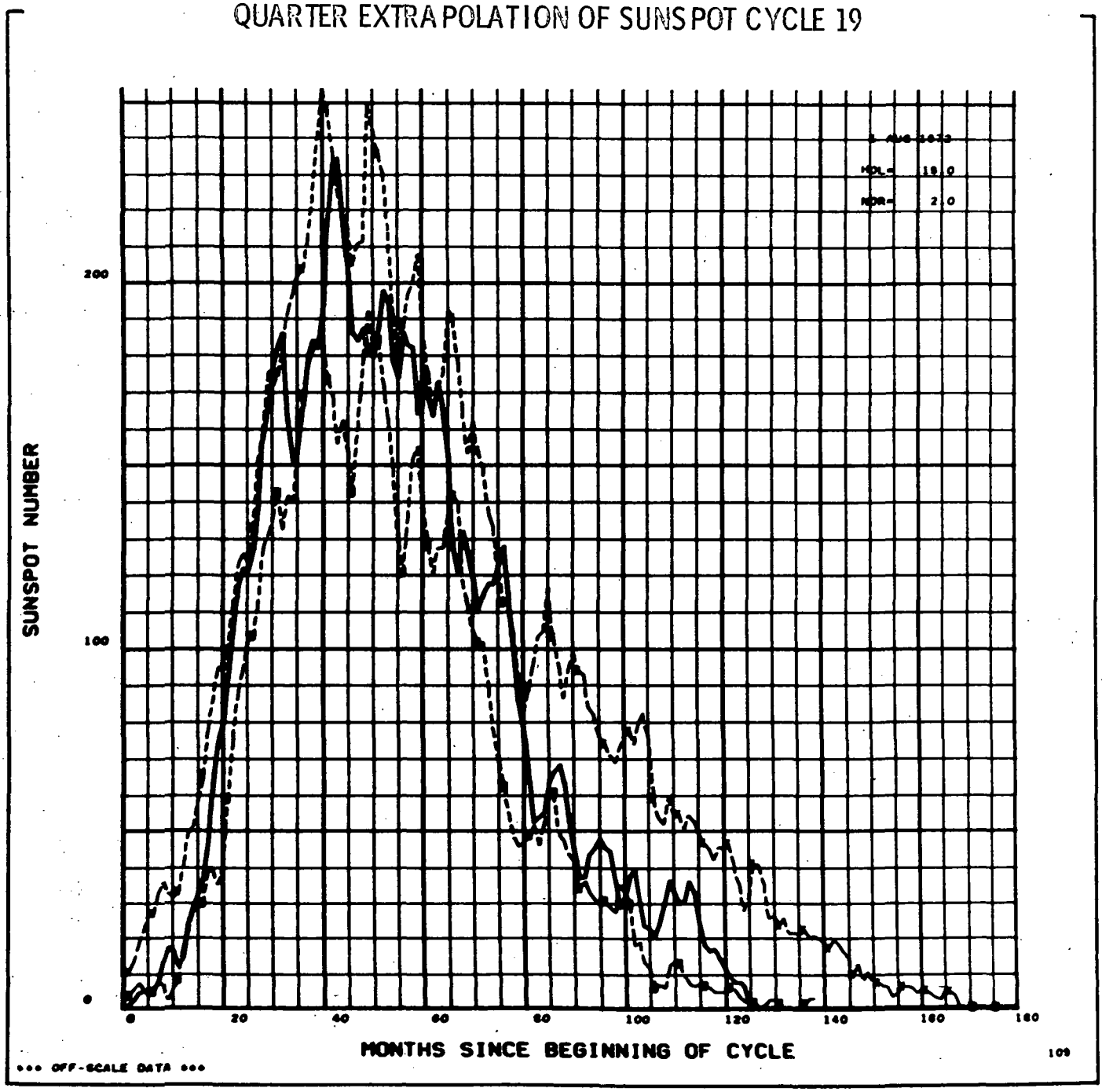


FIGURE 4.48 2ND QUARTER EXTRAPOLATION OF SUNSPOT NUMBER AND 2-SIGMA ERROR BOUNDS FOR CYCLE 20

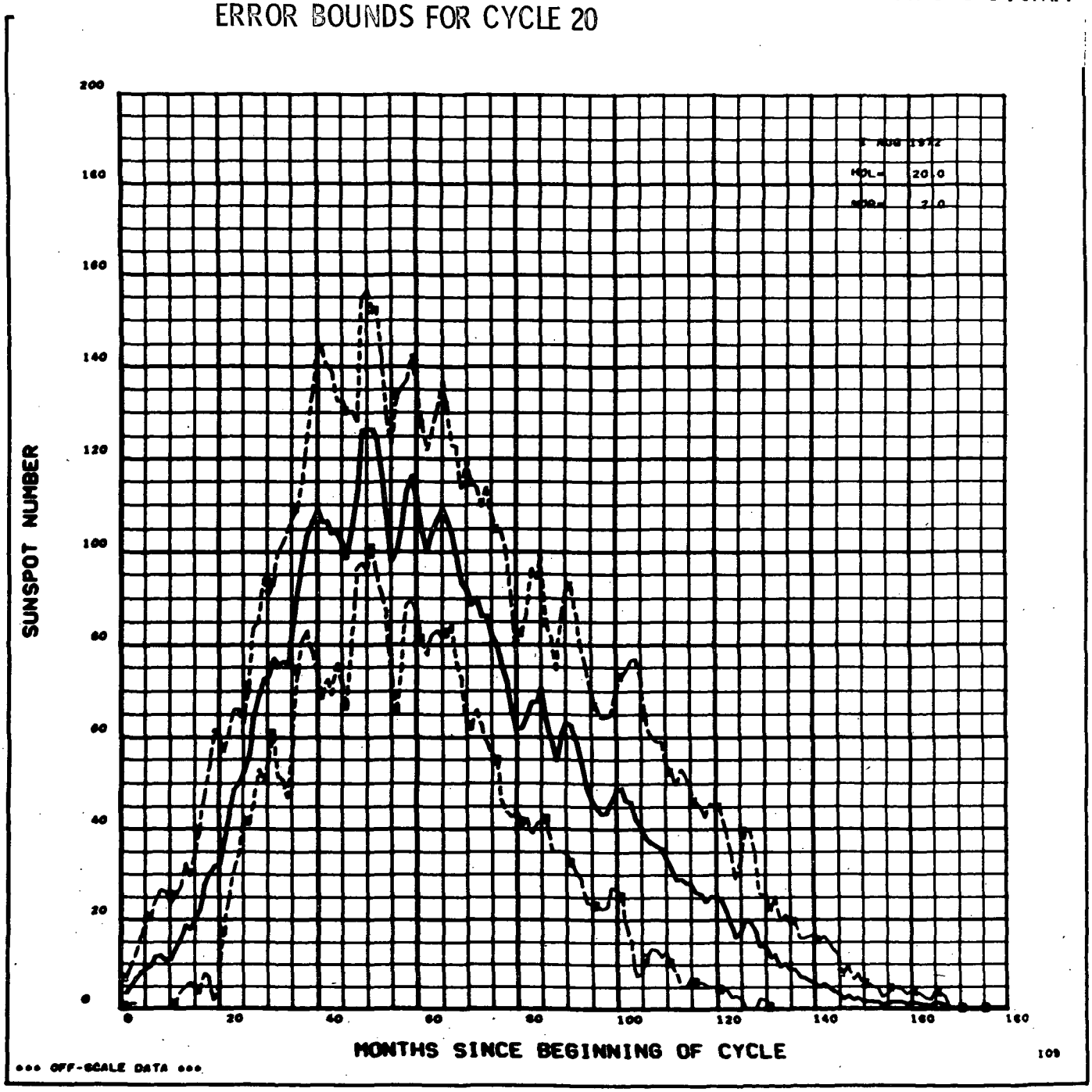


FIGURE 4.49 COMPARISON OF ACTUAL AND 2ND QUARTER EXTRAPOLATION OF SUNSPOT CYCLE 20

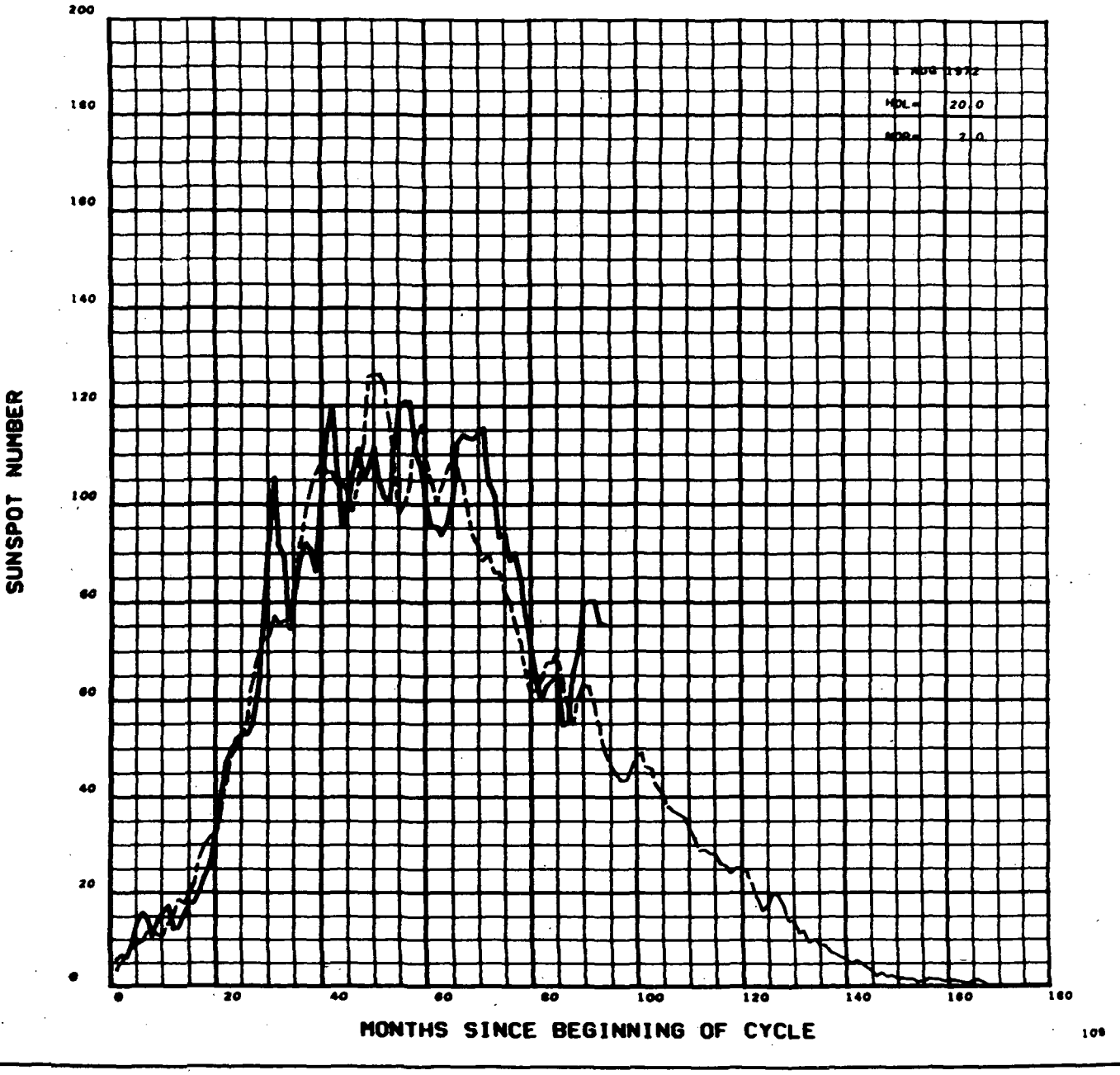


FIGURE 4.50 COMPARISON OF ACTUAL AND 2-SIGMA ERROR BOUNDS FOR 2ND QUARTER EXTRAPOLATION OF SUNSPOT CYCLE 20

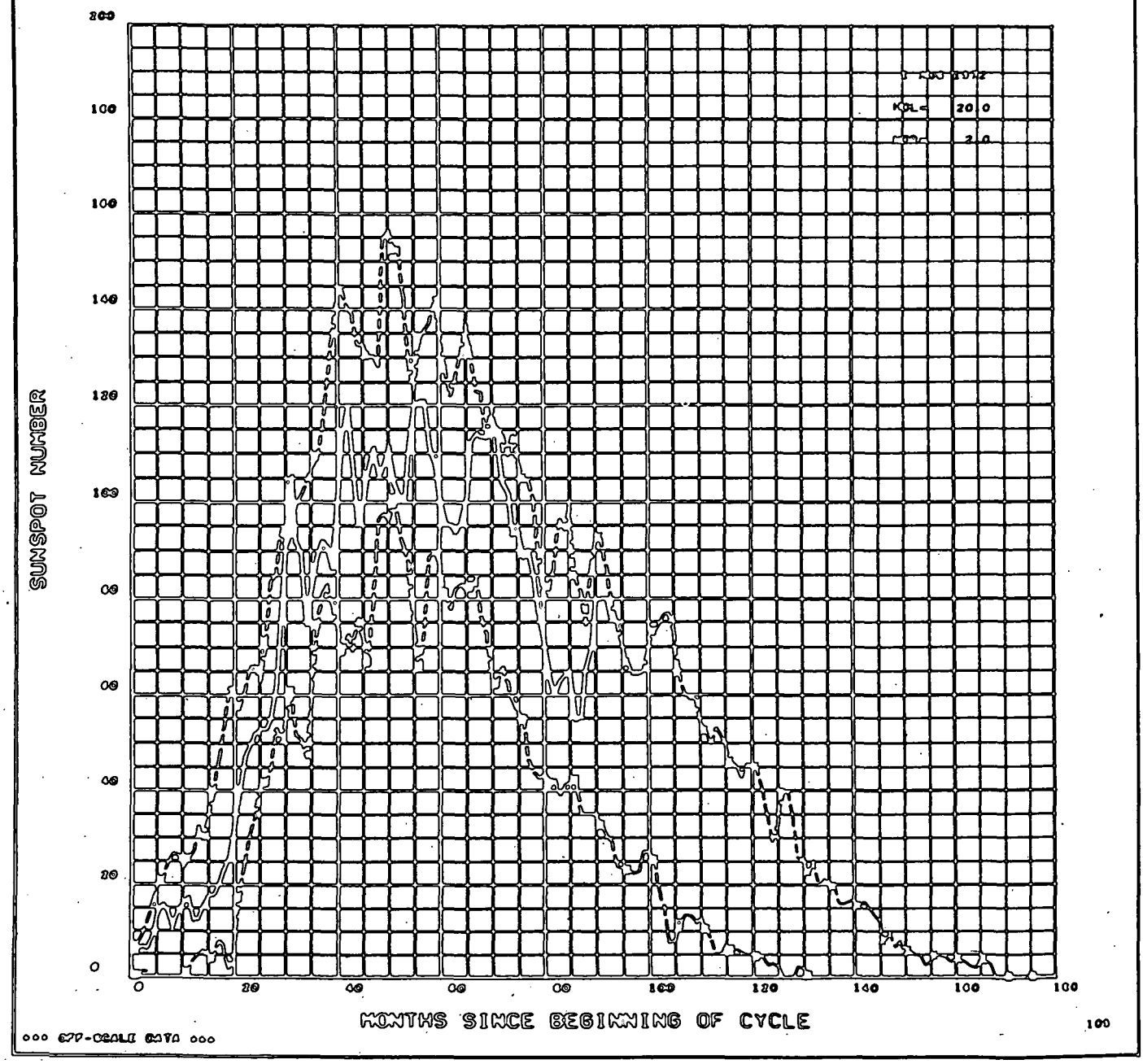


FIGURE 4.51 3RD QUARTER EXTRAPOLATION OF SUNSPOT NUMBER AND 2-SIGMA ERROR BOUNDS FOR CYCLE 19

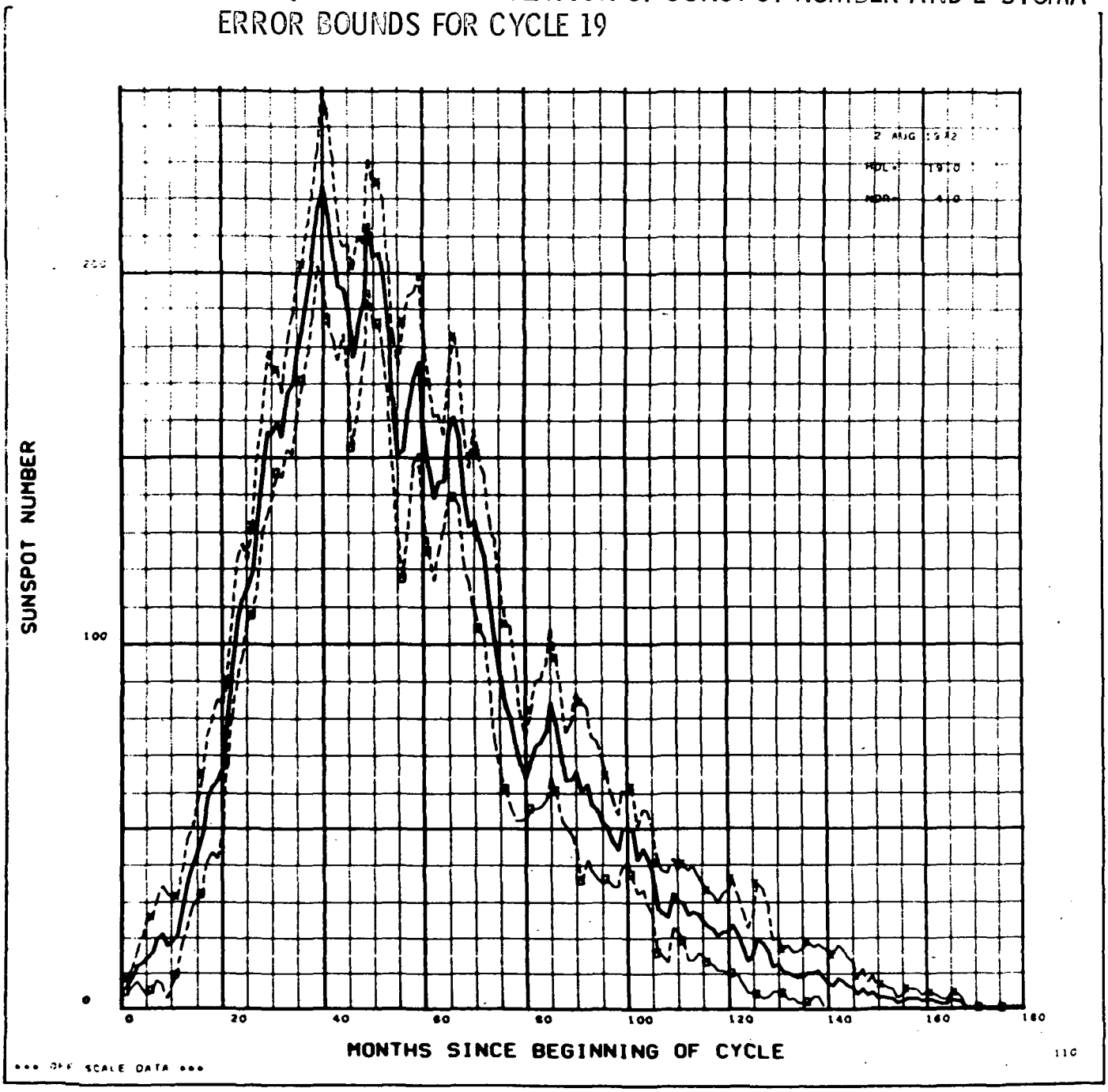


FIGURE 4.52 COMPARISON OF ACTUAL AND 3RD QUARTER EXTRAPOLATION OF SUNSPOT CYCLE 19

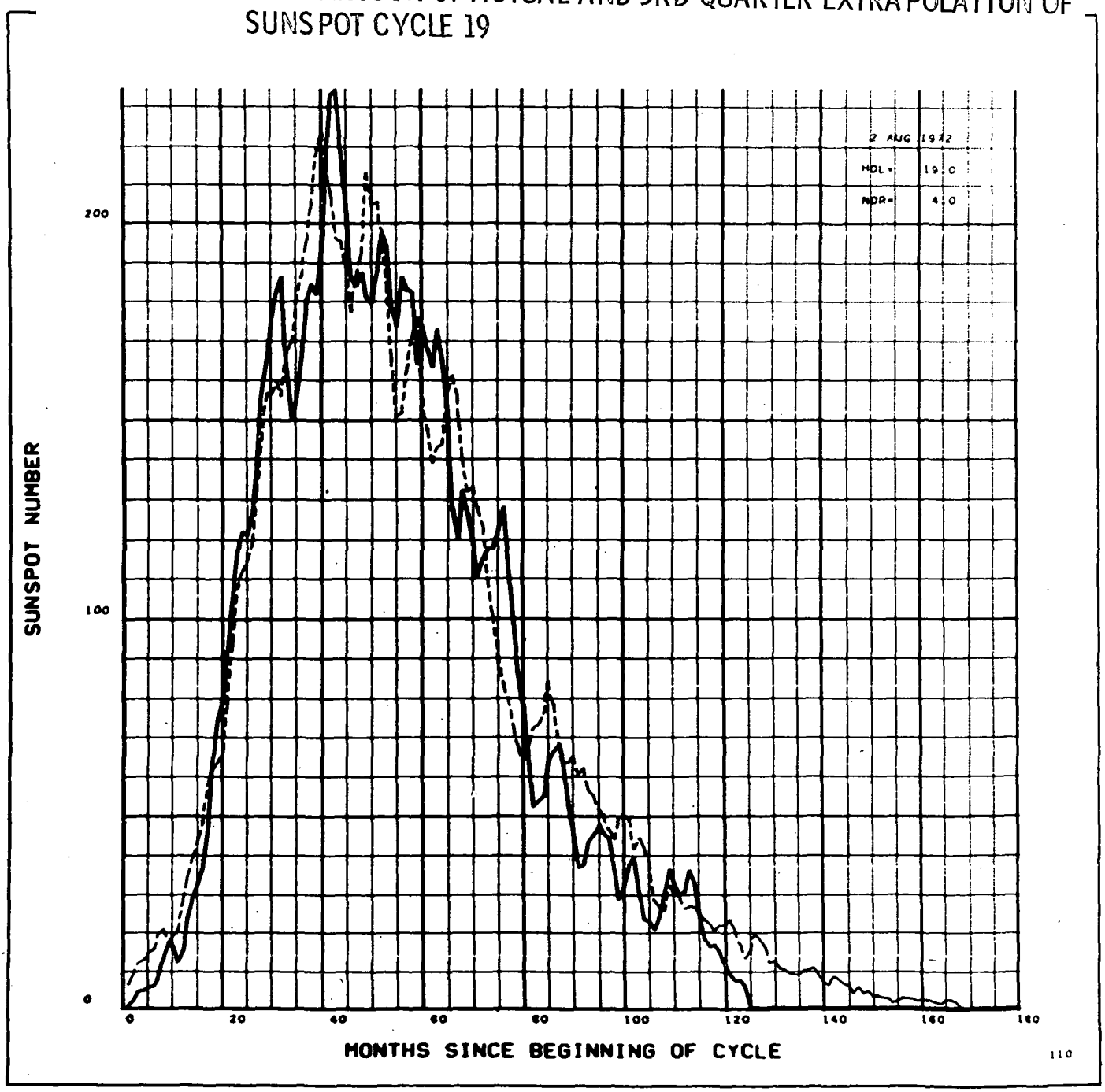


FIGURE 4.53 COMPARISON OF ACTUAL AND 2-SIGMA ERROR BOUNDS FOR 3RD QUARTER EXTRAPOLATION OF SUNSPOT CYCLE 19

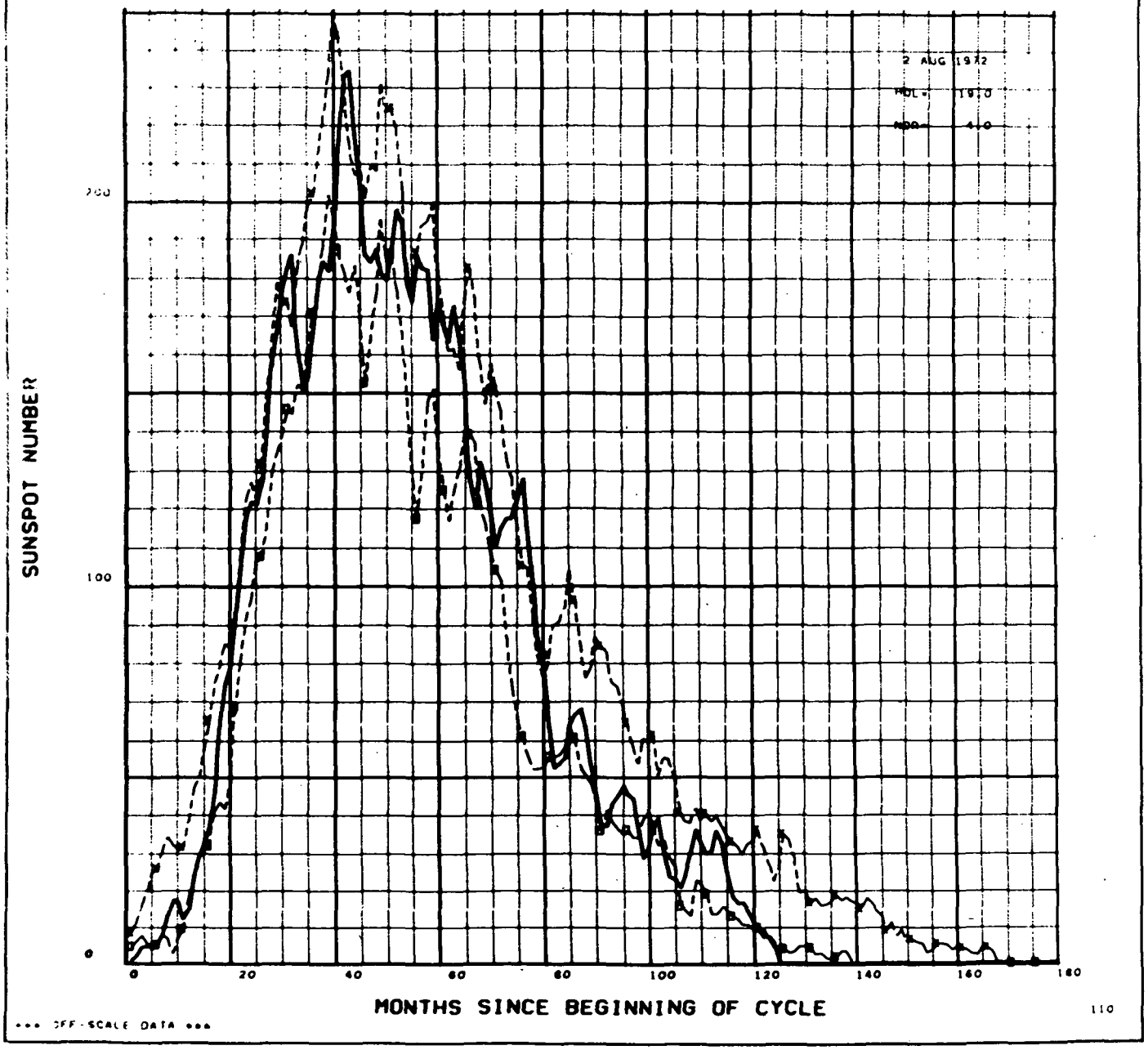


FIGURE 4.54 3RD QUARTER EXTRAPOLATION OF SUNSPOT NUMBER AND 2-SIGMA ERROR BOUNDS FOR CYCLE 20

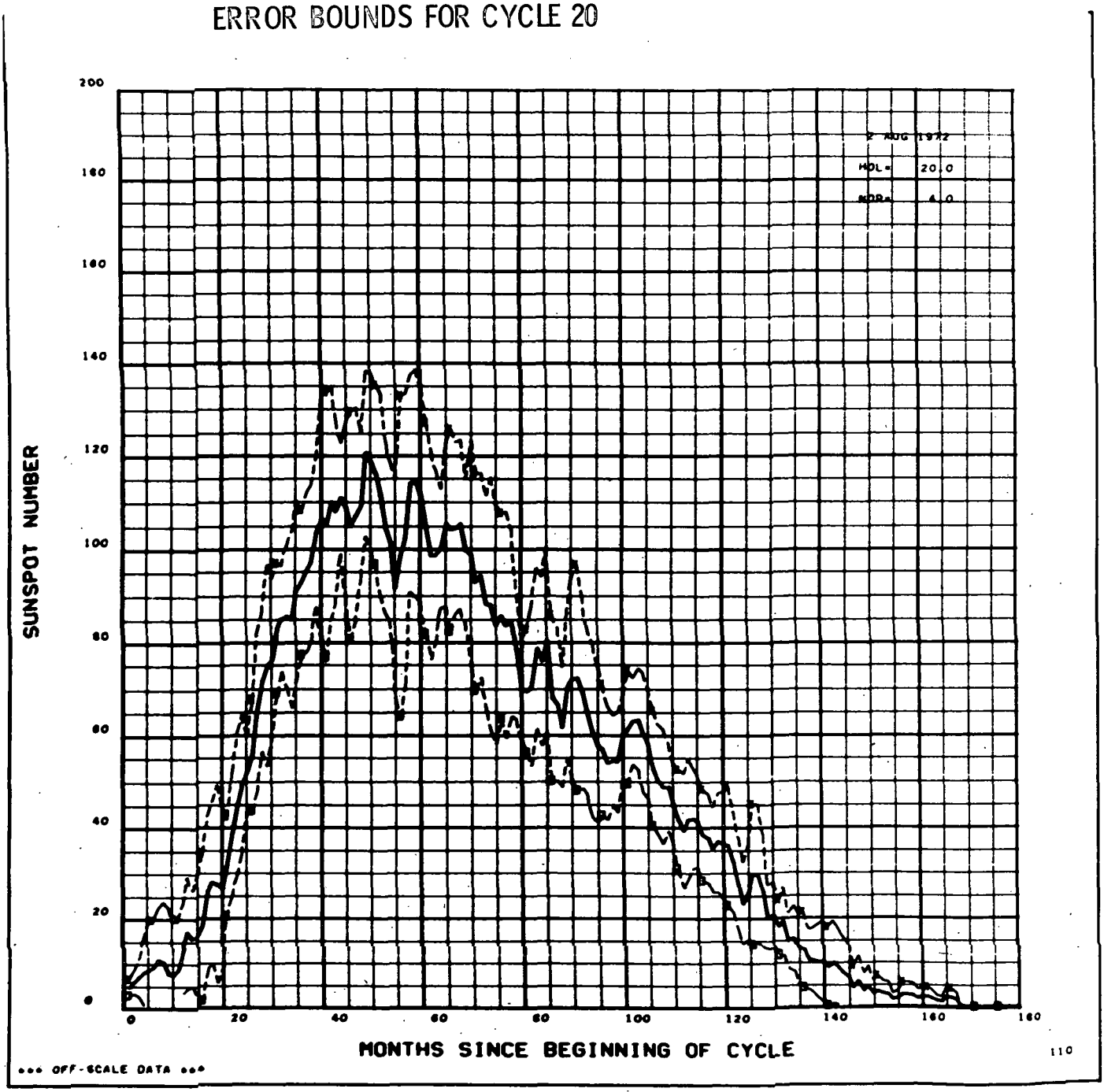


FIGURE 4.55 COMPARISON OF ACTUAL AND 3RD QUARTER EXTRAPOLATION OF SUNSPOT CYCLE 20

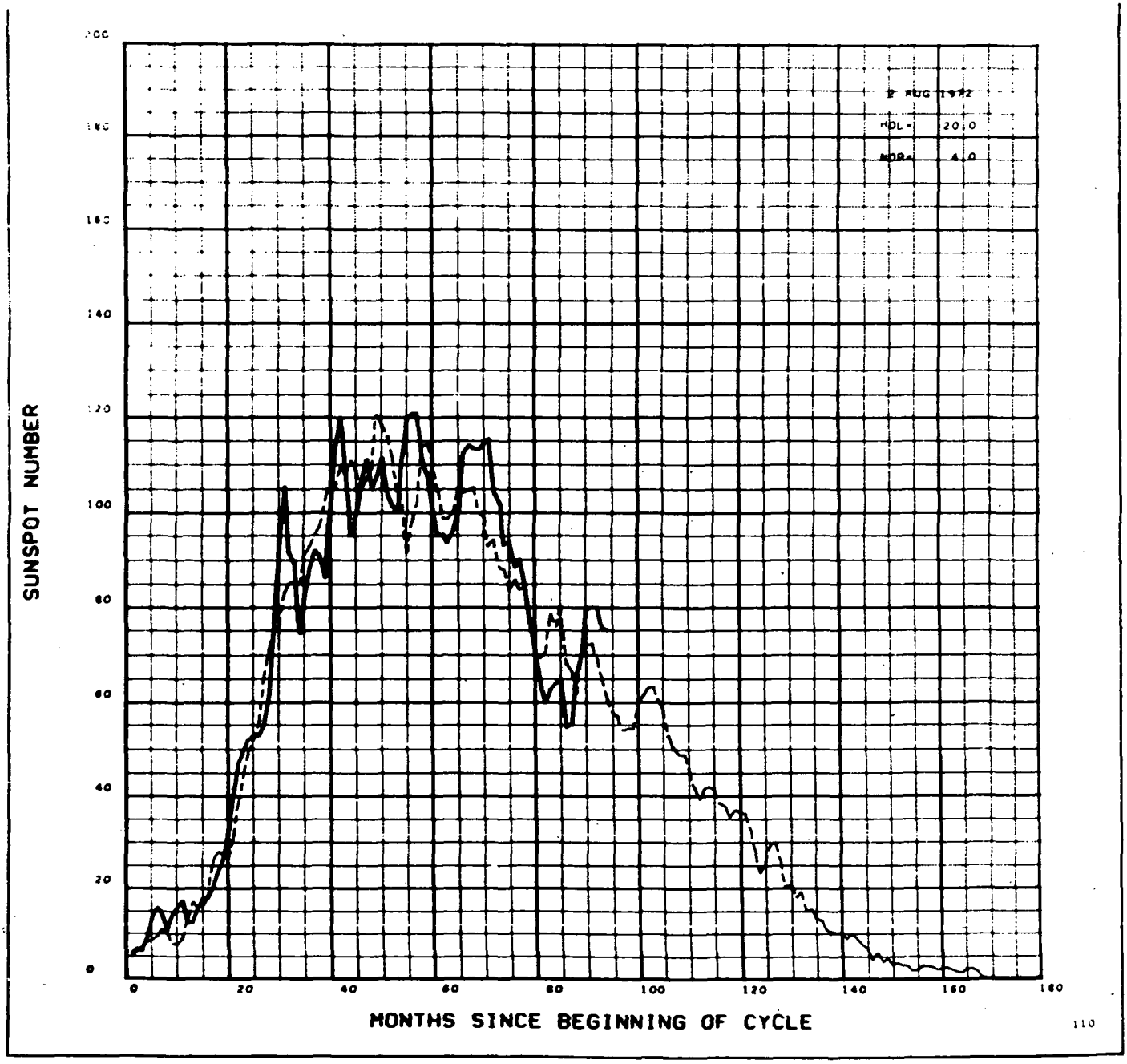


FIGURE 4.56 COMPARISON OF ACTUAL AND 2-SIGMA ERROR BOUNDS FOR 3RD QUARTER EXTRAPOLATION OF SUNSPOT CYCLE 20

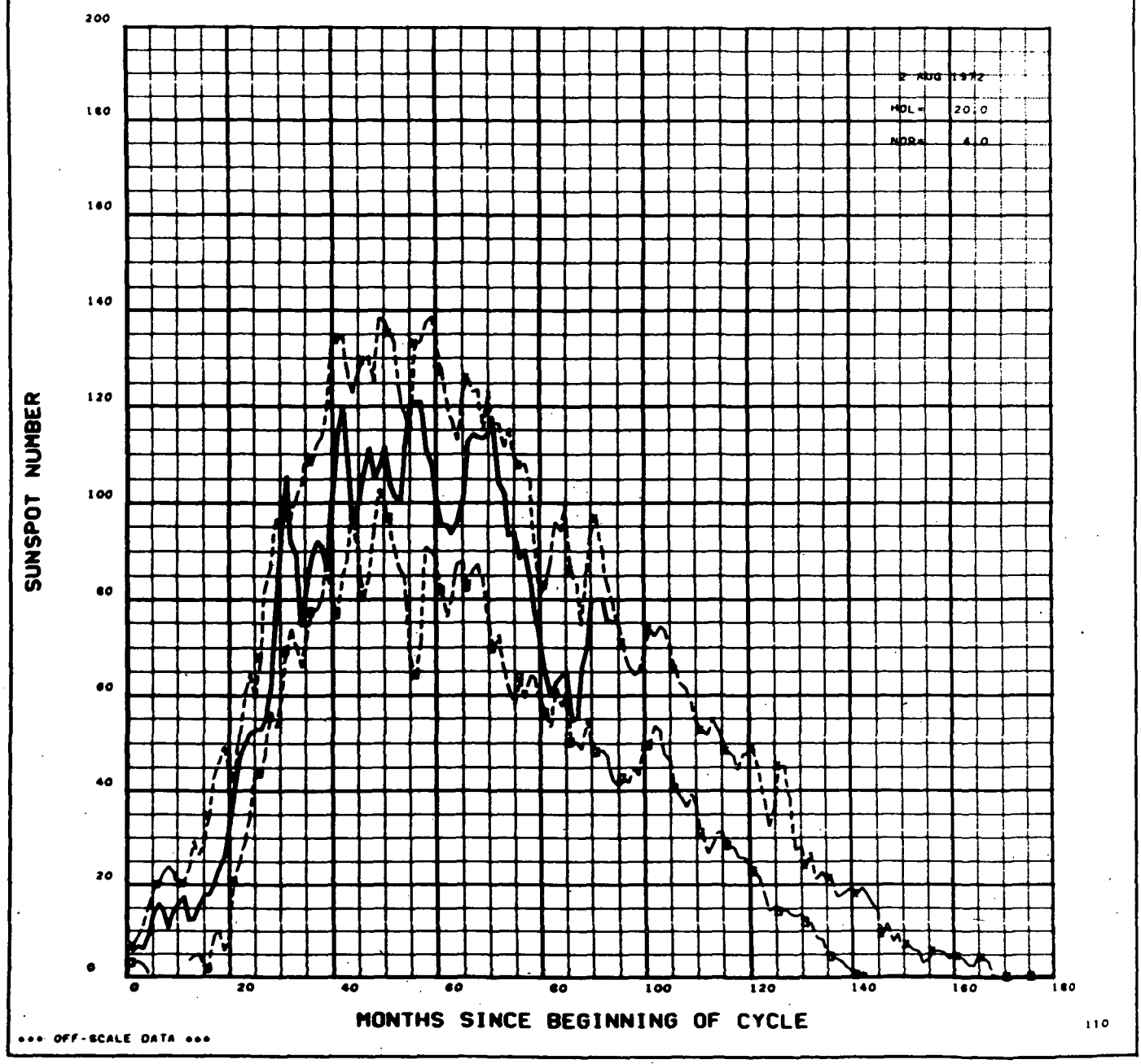
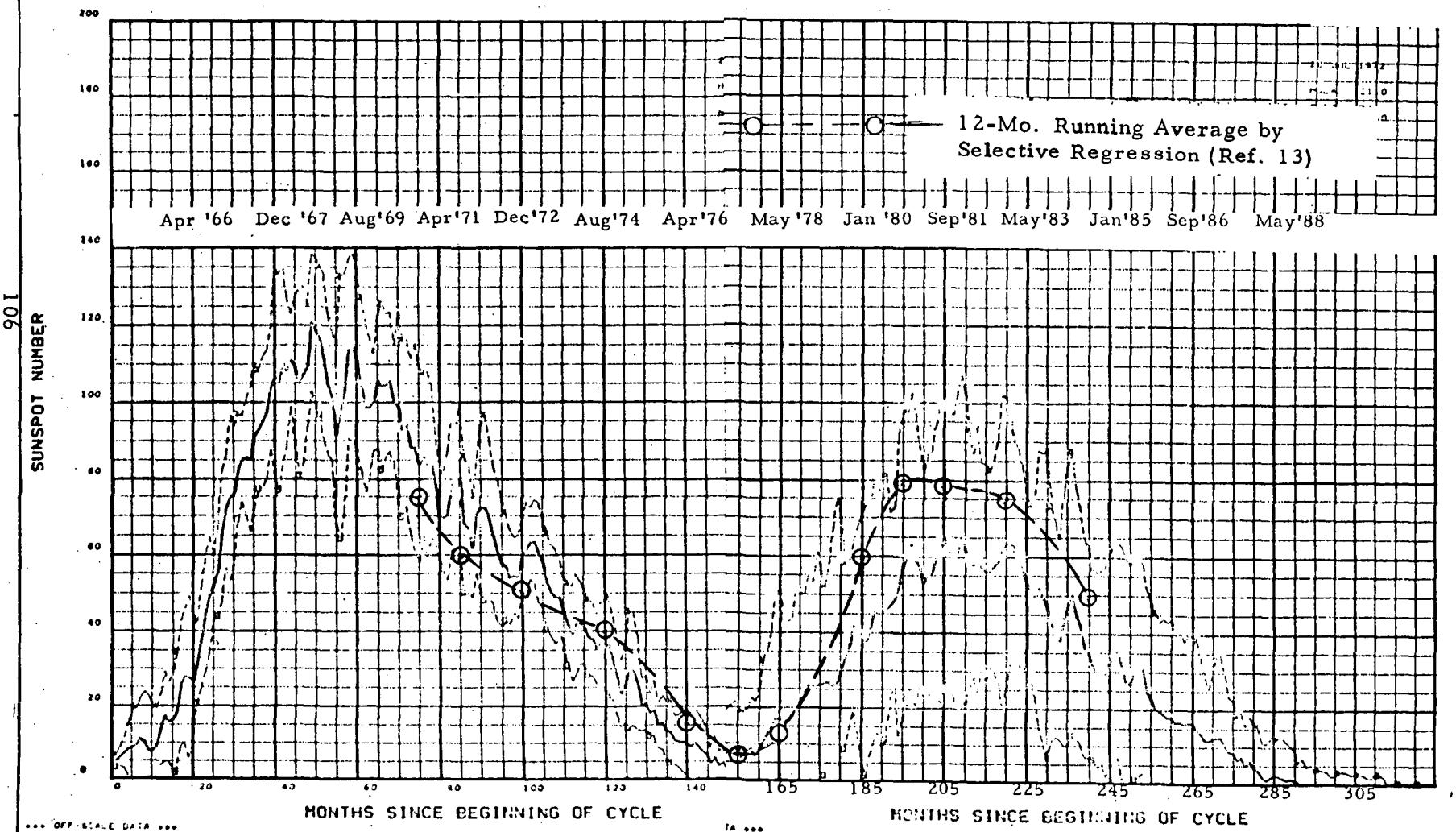


FIGURE 4.57 COMPARISON OF ADAPT AND SELECTIVE REGRESSION ESTIMATES FOR CYCLES 20 AND 21



5.0 ANALYSIS OF SUNSPOT DATA

5.1 Estimate of Sunspot Cycle Properties

The ADAPT regression techniques provide a capability to estimate characteristics of future, or past, sunspot cycles, based on the present or previous sunspot cycles and auxiliary information. Three characteristics which appear particularly useful to predict are the period, the maximum sunspot number, and the time of the maximum for the preceding or succeeding sunspot cycle. This information, although it may be extracted from the prediction of the sunspot numbers for the cycle, is useful as an independent prediction for two reasons. The first is that the selection of these parameters from the predicted sunspot number is often somewhat ambiguous. For example, consider the task of estimating the period of the cycle. The estimated cycle approaches zero and in some cases drops below zero sunspot number. The exact intercept which one should take as the end of the cycle is not absolutely clear. Similar problems occur when one adds the error bands to estimating the exact time of maximum or value of maximum. Thus, the direct prediction of these quantities could overcome some of the ambiguity in estimating them for cycles which have been predicted.

The second and perhaps a more important reason for this prediction capability is to predict characteristics of future or past cycles for which the sunspot numbers have not yet been predicted. Reference 13 has shown that certain characteristics of sunspot cycles behave in an orderly way. Figures 5.1 and 5.2 which have been taken from Reference 13 illustrate such behavior. Figure 2.1 shows the peak sunspot number as a function of date for the negative cycles. Examination of this figure shows a very orderly process over a 180 year cycle. A similar figure for the positive cycles is presented in Reference 13. However, the positive cycles have a much less orderly behavior. The solid squares on Figures 5.1 and 5.2 represent the updated positions for cycles 20 and 21 based on the most recent ADAPT predictions presented in the preceding sections. It should be recalled however, that the ADAPT predictions are for 81-day running averages rather than for 12-month running averages. The ADAPT points shown in Figures 5.1 and 5.2 have been corrected to 12-month averages. The polarity and mode for cycle 20 remains the same as estimated in Reference 13 and, in fact, the new values show better agreement with these correlations than the predictions of Reference 13.

The ability to predict the maximum sunspot number, the period, and location of the maximum allows one to place the next sunspot cycle on one or both of these types of figures. From this one can obtain an estimate of how typical the next cycle will be and thus have an additional validity criterion. In addition, the relationships between these quantities indicated by these figures allows one to correct

these predictions by moving the point to a region on the plots consistent with the behavior of the previous sunspot cycles. This improved estimate of the period, the maximum sunspot number and the time of maximum sunspot number could be used as input data to the prediction of the sunspot cycle. It would probably add information which is not used in the present ADAPT predictions since it involves a very nonlinear procedure.

The present analysis was concerned primarily with development of techniques for predicting future cycles, and thus very little effort was spent on predicting properties of the cycle in general. However, a good estimate of the period of a cycle is required to estimate the next succeeding cycle. The simplest method for predicting the period of a sunspot cycle for which one has estimated the sunspot numbers as a function of time is to examine this estimate and determine when it reaches zero. The difficulty with this approach is in the definition of when the predicted sunspot cycle actually reaches zero. Because there is always a finite error to be expected it is extremely unlikely that the sunspot cycle will reach a true zero at the end of the cycle; in fact the minimum for the 81-day running average over the first 18 cycles is a sunspot number of 5.3. Thus, it appears more reasonable to use the crossing of 5.3 as the nominal estimate of the period of the sunspot cycle. Table 5.1 presents a comparison of using the crossing of sunspot numbers 3, 5, and 10 as an estimate of the sunspot period over the learning set of cycles 1 thru 18. We see that the use of sunspot number of 3 tends to underestimate the threshold and use of a threshold of a sunspot number of 10 tends to overestimate the threshold. The standard deviation of the error in the estimate of the period actually tends to be a minimum between sunspot numbers of 5 and 10. Thus, we shall use the general ground rule for this procedure that the intercept of the prediction with the sunspot number of 5 constitutes the end of the sunspot cycle and defines the period of the sunspot cycle. A similar analysis using the third quarter extrapolated sunspot histories indicated that the best estimate for the extrapolated histories is obtained by using as the intercept 5 above the minimum sunspot number. This modification was required since some of the extrapolated histories produced small negative sunspot numbers.

Algorithms were also developed to predict the period of another cycle both from the preceding two cycles and from the coefficients of the cycle itself. The latter prediction algorithm was made for use with the extrapolation predictions where one has estimates of the coefficients of the cycle for which one would like to know the period. The prediction of the period of the next sunspot cycle is based on a two cycle base and allows one to predict the period of the future sunspot cycle for which the coefficients have not yet been estimated.

Figures 5.3 through 5.7 present the performance and relative importance information for the algorithms developed to predict the period of a sunspot cycle using the ADAPT estimate of that sunspot cycle. Note that one could either use the extrapolated estimate or the predicted estimate with this algorithm to find the period

of the sunspot cycle. Figure 5.3 presents a plot of the estimated and the actual values of the period for sunspot cycles 1 through 18 which were used as learning data for developing this algorithm in six dimensions. The solid line on Figure 5.3 is the line of perfect agreement. The dash lines indicate a one-year error in estimating the length of the sunspot cycles. Only three cycles have their period estimated in error by more than one year using this algorithm. The 2-sigma error for this algorithm is approximately 18 months. The relative importance vector for this algorithm is shown in Figure 5.4. As before this vector shown the importance of each region in the estimate of the sunspot cycle to determining the period of that sunspot cycle. The dot product of the relative importance vector with the data history yields a number which differs from the period estimate by a known constant. This algorithm is summarized in Table 5.2 which has been designed to allow one to implement the algorithm without further reference to this report. The relative importance spectrum associated with this algorithm is shown in Figure 5.5. Here we see that the dominant term in the optimal series for determining the period is the sixth term. The second, fourth, and fifth terms also make contributions to this estimate. Thus, we conclude that one can make a better estimate of the period using six dimensions than one would using only two dimensions. To verify this an algorithm was developed in two dimensions. The estimated versus actual predictions of the period is shown in Figure 5.6 for the two dimensional algorithm. Again the dash lines indicate an error of one year by more than a year. The 2-sigma variation of the estimate of the period using this algorithm is 24.4 months. Thus, we have verified the conclusion suggested by examination of the relative importance spectrum.

The algorithms derived on the single cycle base allows one to predict the period which should be associated with any estimated sunspot cycles. It is also desirable to be able to predict the period of the next sunspot cycle without making the prediction for the cycle itself. To do this it was decided that the best predictor would be the same predictor used to predict the next sunspot cycle, namely, the double cycle base. Thus, an algorithm was developed using the double cycle base to predict the period of the next cycle. This algorithm was developed in exactly the same way as the algorithm to predict the coefficients of the next sunspot cycle. The performance of this algorithm, when developed in six dimensions, is shown in the plot of the estimated versus actual periods in Figure 5.8. The dash line shows the one year error bands and we see that the predicted periods for four cycles have errors greater than one year. The 2-sigma error for this prediction is 20.4 months. Figure 5.9 shows the relative importance vector for predicting the period of the next sunspot cycle using the preceding two sunspot cycles. It is interesting to note that both of the preceding sunspot cycles make a significant contribution to the estimate of the period of the next cycle. The most dominant half cycle in the estimate is the first half of the second cycle preceding the cycle for which the estimate is being made. The least important half cycle is the second half of this same cycle. This relative importance vector

provides further evidence of the wisdom of selecting the two preceding cycles basis for predicting information regarding a sunspot cycle, rather than utilizing just the preceding cycle. Table 5.3 presents the detail instructions for applying this algorithm for predicting the period of future sunspot cycles. Figure 5.10 presents the relative importance spectrum associated with this algorithm for predicting the period of a sunspot cycle based on the preceding two sunspot cycles. This relative importance vector shows that the most important term in the double cycle base for predicting the period of the next cycle is the third term and that the first and fourth terms also make significant contributions to this prediction.

We have presented three general methods by which the period of a sunspot cycle may be estimated: 1) The intercept of the estimated cycle with the threshold sunspot number, 2) Utilizing the estimated cycle in the ADAPT single base period prediction algorithm presented in Table 5.2, and 3) The prediction of the period directly from the preceding two sunspot cycles using the algorithm presented in Table 5.3. Table 5.4 compares the performance of these three methods in terms of the standard deviation of the error based on the learning data and the performance in predicting the period of cycle 19. The predictions for the period of cycles 20 and 21 are also included. It is interesting to note that the period for both cycles 4 and 9 was underestimated by the prediction based on the preceding two cycles. Since both the analysis of Reference 13 and the scatter plots obtained in this study indicate cycles 4 and 9 are similar to cycle 20. This suggests that even the estimates of a significantly longer cycle 20 reported here, might actually be underestimates of the length of cycle 20.[1]

5.2 Rearward Predictions

Since the methods investigated in this study have shown better than a factor of two improvement in the ability to predict the sunspot numbers, and the addition of data such as the angular momentum of the solar system can be expected to significantly enhance this improvement, it is apparent that the recovery of earlier sunspot information can be significantly improved by the application of these techniques to estimating sunspot cycles in a rearward direction. This is extremely important since it will increase the amount of learning data available. There is confirming data for sunspot averages based on historical information such as the auroral displays. Others, for example see Reference 14, have shown that this information can be combined with even relatively crude estimation techniques for recovering estimates of the sunspot behavior as early as 600 B.C. Thus, the ability to accurately predict in a rearward direction would allow one to fill in the gaps between the available observations more accurately. It appears likely that the periods and perhaps maximum sunspot number would be easiest to recover. The next most likely quantity to be estimated in a rearward direction is

[1] See Appendix D for further discussion of this point.

the annual sunspot numbers. The ability to estimate the 81-day running averages to significantly early dates will depend greatly on the ability to develop algorithms which can make use of the annual information to estimate the monthly information. Since it is unlikely that there will be observations which can be useful in pinning down monthly values other than through the annual averages, it is unlikely that the learning data for the 81-day running averages can be extended much earlier than 1700. On the other hand, it is quite likely that at least 100 and maybe several thousand years of additional annual data can be obtained. If this is the case, and if it can be shown that the annual data is useful in predicting the monthly data, this annual data can then be used to predict forward beyond cycles 20 and 21 and then used as input to the monthly predictions for the forward-running information. Thus, the development of rearward prediction algorithms and the use of these algorithms to recover as much of the annual sunspot history prior to 1700 as possible should significantly improve the ability to make long range predictions of sunspot activity.

There are additional advantages to carrying through this rearward prediction over a significant length of time. For example, the availability of solar activity for a significant length of time (i.e., thousands of years) could provide sufficient information that this activity could be incorporated into stellar models. The verification of stellar model predictions of solar activity would be a major breakthrough in the understanding of stellar models and in the ability to project the effect of the sun on the solar system's environment for the distant past and the very distant future. Another potential benefit from studying the sunspot cycles over a significant length of time is the verification of the relationship between the angular momentum of the solar system and the sunspot cycles. If this proves to be valid, it offers an opportunity to use the stellar activity as a basis for inferring information about possible planetary systems beyond our solar system.

Although not required by the present study, it was possible to incorporate preliminary analysis of rearward predictions during the early exploratory studies as part of the development of exploratory representations. As a result of this a preliminary base using cycles 1 through 19 suitable for rearward predictions of sunspot cycles was developed and its major characteristics are presented in Figures 5.11 through 5.17. The major difference between this base and the single cycle base used for the forward predictions is that the start of the sunspot history is left open instead of the end, and the end point of the sunspot history were selected as month number 180. The sunspot cycle was then plotted with zeros in those month from zero until the first sunspot of the history. This had the effect of more highly organizing the variation of the back half of the sunspot cycle since they were all forced through point 180 and disorganizing the first half of the sunspot cycle. The average of the cycles 1 through 19 constructed in this manner is presented in Figure 5.11.

The ADAPT representation for the cycles 1 through 19 was then constructed by subtracting the average of cycles 1 through 19 presented in Figure 5.11 from each of the cycles and processing the resulting histories in the ADAPT programs to find the optimum representation. Figure 5.12 presents the information energy as a function of the number of terms retained in the optimal series representation. We see that for this base the first term contains approximately 50% of the information as compared to approximately 60% for the forward facing representation. The second term contains approximately 29% as compared to 18% for the forward facing sunspot cycles and the third term contains approximately 9%. Again there appears a second break at the fifth term in this history. By far the greatest amount of the information, namely 88%, is contained in the first three terms of this history. The first five optimal functions are presented in Figures 5.13 through 5.17. Comparison of Figure 5.13 with Figure 4.3 shows that the first term of the series now contains information over almost the entire cycle. The second term of the series shown in Figure 5.14 looks very much like the first term in the forward running sunspot cycles. This behavior is a direct result of the enhanced order of the second half of the sunspot cycle at the expense of the first half. The fact that the rearward representation has only 50% of the information in the first term is an indication that the most natural way to present the data is in the forward direction. The fourth and fifth optimal functions presented in Figures 5.16 and 5.17 show the characteristics of the higher numbered optimum functions for the forward running base, namely, they present the detailed structure information which is required to fill in the detailed oscillations occurring in the sunspot cycle.

Since the availability of an appropriate single cycle base is the only requirement for extrapolating a data history, it was possible, within the constraint of the present program, to apply the extrapolation program to this rearward facing base to complete cycle zero. Since the first optimum function now contains information over the entire cycle and the break point in the energy curve occurs at the fifth optimum function, it appears that any number of terms from 1 to 5 might be the best for extrapolating the sunspot histories in the rearward direction. Figure 5.18 presents the results of the rearward extrapolation for 1, 2 and 18 terms. The 18-term extrapolation is clearly overdetermined and we have a clear illustration of the effect of overdetermination here. Namely, the 18-term representation is a very poor estimate of the future although it does a reasonably good job of matching the input values. The 3, 4 and 5 term predictions lie between the 2 term and the 18 term prediction and thus have not been included in this figure.

Examination of Figure 5.18 indicates that either the 1 or 2 dimensional reconstructions probably represent the best results. The 2 dimensional reconstruction already has the difficulty that it has negative sunspot numbers although the maximum negative value is only minus 10. Also, if we use the sunspot number of 5 above zero intercept method of predicting the period, the 2 term reconstruction extrapolation predicts a period approximately a year less than that

would be indicated by Reference 14. The 1 term reconstruction has no negative values of sunspot numbers but tends to overpredict the Schove minimum by slightly more than a year. If we use the 5 sunspot number above the minimum as the intercept the 1 term representation still overpredicts the period by a year, but the 2 term now only underpredicts by about half a year. Thus, the two term estimate ending at about month 60 appears to be the best estimate of cycle 0. The advantage of adding the 2 preceding cycles to the forward predictions suggests the rearward predictions can be improved by adding the 2 following cycles. In fact, there may even be significantly greater gain in the rearward predictions because of the greater amount of information contained in the rearward terms 2 through 18 as compared to the forward terms 3 through 18. Thus, it is clear that the ADAPT techniques would significantly improve the recovery of information from the earlier cycles. In particular, it has already achieved a somewhat better estimate of the 81-day running averages from March of 1749 to early 1744.

5.3 Clustering Studies

The ADAPT programs provide as by-products to any analysis of data a series of outputs which are extremely useful for finding natural groups or clusters in the data. For example, a plot of the first coefficient versus the second coefficient of the optimal Fourier series representing each history is the best two dimensional representation of the data which can be made. This follows from the fact that the first coefficient explains the greatest amount of variation that one can explain in any single term representation and that the first two terms in the optimal series explain the greatest amount of variation in any possible two terms representation of this data. Since this latter amount of variation is displayed graphically in a two dimensional scatter plot when these two coefficients are plotted as a function of each other, one has the best two dimensional representation possible. Figure 5.19 presents such a plot for the single cycle forward facing base. Each of the cycles designated by the numbers enclosed in circles, triangles or squares are located according to the values of the first coefficient and second coefficient of the optimal Fourier series representation. For example, consider sunspot cycle 1 which is located at an NP 1 coordinate value of approximately 190 and an NP 2 coordinated value of approximately -95. This means that to reconstruct sunspot cycle 1, one takes the average of the single sunspot cycles presented in Figure 4.2 and adds to it at each point or for each month 190 times the value of the corresponding month in the first optimal function shown in Figure 4.4 and to that sum adds the product of -95 and the corresponding value of the second optimal function shown in Figure 4.5. The relationships between the sunspot cycles as displayed on this scatter plot, accounts for approximately 80% of the variation in the data.

The circled sunspot cycles on this figure represent the actual values obtained by projecting the observed sunspot numbers for that cycle on the single cycle base discussed in Section 4.1. The triangle around cycle 20 indicates that this position for cycle 20 is based on the third quarter extrapolation of cycle 20 as described in Section 4.3. Since the cycle has not yet been completed, this is the best available estimate of cycle 20's location. Similarly, the square around 21 indicates that this is the best estimate based on the predictions using the cycles 19 and 20 in the algorithms presented in Table 4.3.

The scatter plot can be used to obtain the same classification of sunspots according to mode that was reported in Reference 13 by plotting the maximum sunspot number versus period independently for the positive and negative sunspot cycles. If one considers the scatter plot to be divided into two regions by line c-d one notices that all of the sunspot cycles to the right line c-d are mode 1 and to the left of line c-d are mode 2 sunspot cycles. However, a more careful examination shows that one may also draw the line a-b and then consider the region A to the right of a-b, B between lines ab and cd, and C between lines cd and ef and D to the left of line ef.

If one now considers the negative and positive cycles independently, one sees that there is an even stronger separation between mode 1 and mode 2 for fixed polarity. That is, the negative mode 2 cycles all lie in region A and the negative mode 1 cycles all lie in region C. Regions A and C are separated by the entire expanse of region B. Furthermore, no positive mode 1 cycles lie in region C so the positive mode 1 cycles are separated from cycle 19 which has been tentatively identified as the only known positive mode 2 cycle, by the entire expanse of region C. Thus, we see that the scatter plot was capable of identifying the separation between mode 1 and mode 2 as a weak separation even when the polarity was ignored and that when the polarity was considered the separation became very strong. Although these classifications into mode 1 and mode 2 have been found independently by Sleeper using more conventional analysis it is hoped that this example will illustrate how the ADAPT scatter plot can be used to accomplish this analysis.

The 2-dimensional representation provided by the scatter plot is not the only useful form of clustering analysis. It is often desirable to perform clustering analysis in higher dimensional spaces. This is especially true when the first two dimensions do not explain the great majority of the variation, it is not expected that a higher dimensional cluster analysis will yield more significant results. However, to illustrate the technique, the single cycle data was processed through the ADAPT nearest neighbor program and the nearest neighbor of each of the sunspot cycles determined. This information is plotted in Figure 5.20. Figure 5.20 may be used to construct nearest neighbor trees as follows. Since Figure 5.20 plots the sunspot cycle as the abscissa and the nearest sunspot to each of the sunspot cycles as the ordinate, one may now read the curve

in the other direction and answer the question: For which cycles is cycle X the nearest neighbor? Each of these cycles for which a given cycle is the nearest neighbor are assumed to be a member of a grouping containing cycle X and provide the first branch in the tree presented in Figure 5.21. The process is then repeated for each element of the branch. There are one of three possible results: 1) One might find that a given cycle is not the nearest cycle to any other cycle in which case the procedure terminates for that path, 2) One might find that a given cycle is the nearest cycle to the cycle which produced that branch in which case the procedure terminates for that path, or 3) One may find that a given cycle introduces an entire new branch and the procedure may be continued. This procedure has been carried out and as shown in Figure 5.2 where three other groups of four or more sunspot cycles are defined. These four groups are enclosed in the dash lines shown in Figure 5.19. The groups are logical groups on this figure as could be expected from the fact that Figure 5.19 actually contains 80% of the variation. Other examples of this nearest neighbor analysis are given in References 1 and 4 and the reader is referred to these references for more details on this analysis.

As pointed out, these clustering outputs are by-products of the ADAPT analysis and have been included in the present report to illustrate some of the potential of the ADAPT programs for further analysis. There is no intention that this report be a complete clustering analysis as the major objective of this study was the development of advanced prediction techniques. However, to provide the reader with capability to perform clustering analysis which may be useful for other purposes the scatter plots for the other two bases which have been used namely the single cycle rearward prediction and the double cycle base are given in Figures 5.22 and 5.23 with the location of the sunspot cycles indicated on these figures. The numbers for the cycles on Figure 5.23 is the number associated with the first cycle in the pair for the double cycle. For example, No. 2 on Figure 5.23 is the location of double cycle 2-3.

5.4 Recommendations

This section briefly summarizes the recommended analysis suggested by the preceding three sections. The result most pertinent to the present study is that the best estimate of the periods of a sunspot cycle for which one has estimated the sunspot numbers is given by applying the algorithm presented in Table 5.2. This algorithm yields a period of approximately 150 month for cycle 20 which has been incorporated in the prediction shown in Figure 2.1.

The studies carried out here have shown that there is great potential for recovering additional sunspot data from historical records by applying the ADAPT techniques. To accomplish this it is recommended that the same developments outlined in Section 4.5 to improve the forward predicting algorithms be incorporated into developing algorithms for predicting rearward cycles. In addition, it

is recommended that these algorithms be developed to predict the individual properties of sunspot cycle such as period, maximum sunspot number and time of maximum in both the forward and rearward directions. It is also recommended that algorithms be developed in both the forward and rearward directions to predict the annual average sunspot numbers. The annual sunspot numbers for the cycle being predicted should also be included in the data vector for predicting the 81-day running average sunspot number, since it is likely that the result of the analysis suggested here will be that one can predict the annual sunspot cycles significantly better than the monthly cycles. Having developed the algorithms to predict the properties of the sunspot cycles and the annual values it is recommended that these algorithms be incorporated in an analysis similar to that carried out in Reference 14 to determine the best estimate of sunspot activity to the onset of available records or at least 600 BC. When this has been accomplished, it is recommended that the implications of both the sunspot activity and its relationship to the angular momentum of the solar system be applied to the construction of stellar models and to developing observational techniques for gaining information about planetary systems.

It is also recommended that the ADAPT clustering analysis be used for "a scientific fishing trip" to determine if there are any groupings of interest. This analysis should be carried out using both single and double bases, scatter plots and nearest neighbor analysis. Any groupings which are found such as those enclosed in dashed circles in Figure 5.19 should be studied individually. The ADAPT programs may be used for this both to construct average sunspot cycles for each of the groups and to construct relative importance vectors for separating each of the groups from one another and from all of the remaining data. These relative importance vectors will show exactly what portions of the cycles make each of the groups stand out as a group and can be used as a basis for trying to understand the reason for the groupings observed.

TABLE 5.1

EFFECT OF THRESHOLD ON ESTIMATE ON SUNSPOT CYCLE PERIODS USING INTERCEPT METHOD USING 2 TERM RECONSTRUCTIONS WITH ADAPT PREDICTED COEFFICIENTS FROM PRECEDING TWO CYCLES

		TH = 3		TH = 5		TH = 10	
Cycle	Actual Period (Mo.)	Estimated Period (Mo.)	Error (Mo.)	Estimated Period (Mo.)	Error (Mo.	Estimated Period (Mo.)	Error (Mo.)
3	106	138	-32	132	-26	123	-17
4	168	145	+23	140	28	131	37
5	146	145	1	141	5	135	11
6	153	135	18	133	20	130	23
7	131	133	-2	130	1	127	4
8	113	114	-1	113	0	108	5
9	143	145	-2	140	3	130	13
10	137	142	-5	137	0	130	7
11	145	137	8	130	15	120	25
12	133	136	-3	130	3	115	18
13	146	137	9	135	11	128	18
14	132	143	-11	135	-3	128	4
15	129	135	-6	130	-1	122	7
16	117	144	-27	137	-20	128	-11
17	126	143	-17	138	-12	130	-4
18	120	139	-19	129	-9	121	-1
	AVG		-4.1		+.94		+8.7
	STD Deviation		14.43		13.40		13.3

TABLE 5.2

TO PREDICT THE PERIOD τ OF A CYCLE FROM THE SAME CYCLE, CONSTRUCT A DATA VECTOR CONTAINING 180 ELEMENTS MADE UP OF THE SUNSPOT NUMBER FOR THE CYCLE IN QUESTION AND ZERO VALUES IF THE CYCLE IS LESS THAN 180 MONTHS, AND APPLY THE FOLLOWING ALGORITHM:

$$\tau = A^3 \cdot SS + 137.668$$

Mo. After Start of Cycle	A^3	Mo. After Start of Cycle	A^3	Mo. After Start of Cycle	A^3	Mo. After Start of Cycle	A^3
1	2.7267276D-03	46	2.8467645D-02	91	-5.2399065D-03	136	1.7788164D-02
2	3.7101400D-03	47	1.0860039D-02	92	-1.1919516D-03	137	1.4112384D-02
3	4.6651230D-04	48	-8.7769947D-03	93	8.2385869D-03	138	1.3402902D-02
4	-5.0282742D-04	49	-1.5861767D-02	94	2.2240754D-02	139	1.4256003D-02
5	-4.0825603D-03	50	2.4613980D-03	95	2.8112721D-02	140	1.4838706D-02
6	-9.3426416D-04	51	1.1581253D-02	96	2.6536577D-02	141	1.4132022D-02
7	-6.3783423D-03	52	2.0456161D-02	97	1.4380143D-02	142	1.5802947D-02
8	-9.0649674D-03	53	9.0097775D-03	98	1.1473450D-02	143	1.4102266D-02
9	-1.2971002D-02	54	1.3559628D-02	99	1.2317789D-03	144	1.4053314D-02
10	-5.7840591D-03	55	-3.4574846D-03	100	2.9735875D-03	145	1.1891876D-02
11	-3.0866427D-03	56	-4.4719073D-03	101	6.4383457D-03	146	8.4444659D-03
12	-5.4187325D-03	57	-2.9479644D-02	102	1.3096284D-02	147	8.8822510D-03
13	-4.9222215D-03	58	-2.3711776D-02	103	1.7422564D-02	148	5.5063529D-03
14	-3.7362196D-03	59	-1.9023479D-02	104	1.6611193D-02	149	6.7201346D-03
15	5.2813827D-03	60	-2.5251726D-03	105	1.6651975D-02	150	4.3734310D-03
16	9.9429011D-03	61	-4.9988299D-03	106	1.7815239D-02	151	4.9859186D-03
17	2.5714417D-03	62	-5.6802112D-03	107	1.7805031D-02	152	4.2302372D-03
18	-4.9087731D-03			108	2.0341904D-02	153	4.0258998D-03
19	-7.0356539D-03	63	-4.7692082D-03	109	1.9040361D-02	154	2.1867103D-03
20	3.3659290D-04	64	-7.2351729D-03	110	2.0711865D-02	155	2.7201543D-03
21	-6.3966166D-03	65	-4.3478554D-03	111	1.7559485D-02	156	4.0000803D-03
22	-6.9795563D-03	66	-5.0110349D-03	112	1.9205988D-02	157	4.0534144D-03
23	-8.0042086D-05	67	7.7949880D-03	113	1.9290344D-02	158	3.7867424D-03
24	-1.2945439D-03	68	1.5308514D-02	114	2.5602953D-02	159	2.8267235D-03
25	-1.1672299D-03	69	1.7096522D-02	115	2.7519225D-02	160	3.3067329D-03
26	-1.4720977D-02	70	3.8010458D-03	116	3.0443767D-02	161	3.2533983D-03
27	-8.8997834D-03	71	-1.3652338D-02	117	2.9198951D-02	162	2.7733889D-03
28	-5.5146954D-03	72	-2.1333752D-02	118	2.5975660D-02	163	1.9201382D-03
29	9.1732474D-03	73	-1.8430133D-02	119	2.9609779D-02	164	1.6000321D-03
30	9.3993924D-03	74	-1.5059857D-02	120	3.1023064D-02	165	3.2533983D-03
31	5.1645657D-03	75	-1.7017332D-02	121	3.6137840D-02	166	3.0933950D-03
32	-9.4617224D-03	76	-1.0612891D-02	122	3.2043383D-02	167	2.4000482D-03
33	-2.2034164D-02	77	-8.4499448D-03	123	2.5443588D-02	168	2.1333760D-04
34	-3.2096603D-02	78	4.2073174D-03	124	2.0955153D-03	169	0.0
35	-2.8922139D-02	79	5.1684851D-03	125	2.6101094D-03	170	0.0
36	-3.3953529D-02	80	5.3173796D-03			171	0.0
37	-3.0499380D-02	81	-1.6535064D-03	126	3.2738724D-03	172	0.0
38	-2.7638200D-02	82	-2.8243692D-03	127	3.2761380D-02	173	0.0
39	-7.0326213D-03	83	-9.8273200D-03	128	2.7866040D-03	174	0.0
40	5.4174912D-03	84	-1.7257564D-02	129	1.7557679D-02	175	0.0
41	1.3276343D-02	85	-2.8982179D-02	130	1.4767484D-02	176	0.0
42	2.5391051D-02	86	-2.6218668D-02	131	1.1674964D-02	177	0.0
43	2.5582497D-02	87	-2.0568176D-02	132	1.3264397D-02	178	0.0
44	2.5482519D-02	88	-1.8829446D-02	133	1.4227088D-02	179	0.0
45	2.9743977D-02	89	-1.6559251D-02	134	1.6337540D-02	180	0.0
		90	-2.1540316D-02	135	1.6654873D-02		

TABLE 5.3

TO PREDICT THE PERIOD OF A CYCLE FROM THE PRECEEDING TWO CYCLES, CONSTRUCT A DATA VECTOR AS OUTLINED IN TABLE 4.3 AND APPLY THE FOLLOWING ALGORITHM:

$$\uparrow = A^4 \bullet SS + 108.80786$$

Mo. After Start of Cycle	A^4	Mo. After Start of Cycle	A^4	Start of Cycle	A^4	Mo. After Start of Cycle	A^4	Mo. After Start of Cycle	A^4	Mo. After Start of Cycle	A^4	Mo. After Start of Cycle	A^4	Start of Cycle	A^4
1	2.45458750-04	46	9.2308202D-03	91	1.1435032D-03	136	1.0127956D-03	181	-2.8685780D-04	226	3.7523783D-03	271	-4.2676635D-03	316	-1.3415776D-03
2	4.4126955D-04	47	8.2122749D-03	92	2.0783261D-03	137	1.1666351D-03	182	-2.7006292D-04	227	1.6828691D-03	272	-3.2185984D-03	317	-7.6137583D-04
3	4.1489547D-04	48	7.7962336D-03	93	1.9267525D-03			183	-2.6477900D-04	228	-2.4611651D-03	273	-3.7712323D-03	318	-5.6836304D-04
4	6.0591504D-04	49	7.3725030D-03	94	2.5541197D-03	138	1.2479837D-03	184	-1.2890177D-04	229	-6.4416060D-03	274	-3.1441504D-03	319	-4.7554541D-04
5	6.0470196D-04	50	8.8157867D-03	95	3.0771564D-03	139	1.0658262D-03	185	5.3688596D-04	230	-6.4967272D-03	275	-3.0061388D-03	320	-5.0536228D-04
6	5.6900815D-04	51	1.0533574D-02	96	2.5169116D-03	140	8.8468567D-04	186	6.1274545D-04	231	-5.3759451D-03	276	-2.9279383D-03	321	-2.2866244D-04
7	2.2783039D-04	52	1.2404570D-02	97	7.3635868D-04	141	9.1450990D-04	187	-2.5809908D-04	232	-4.1870671D-03	277	-3.9133876D-03	322	3.6888961D-06
8	1.2769332D-04	53	1.2031420D-02	98	1.3574888D-04	142	1.3501692D-03	188	-7.4136422D-04	233	-4.9496352D-03	278	-4.2106077D-03	323	1.3672626D-04
9	4.0954042D-04	54	9.8087167D-03	99	7.6072895D-04	143	1.3035213D-03	189	-2.3794947D-04	234	-4.0505763D-03	279	-4.1454739D-03	324	1.8356504D-04
10	8.7158843D-04	55	6.6563546D-03	100	6.8357802D-04	144	1.2182212D-03	190	5.6148759D-04	235	-2.5129046D-03	280	-4.1051656D-03	325	2.5753175D-04
11	1.0141347D-03	56	4.6987197D-03	101	2.8585351D-04	145	1.1206285D-03	191	1.6827200D-04	236	-7.8300972D-04	281	-3.7555017D-03	326	1.7470614D-04
12	1.3043240D-04	57	3.3380860D-03	102	-8.0292203D-04	146	8.1750620D-04	192	1.3780829D-04	237	-6.3623528D-04	282	-3.9174319D-03	327	2.6116621D-04
13	3.2867085D-04	58	4.5169085D-03	103	-4.3820803D-04	147	9.9562164D-04	193	2.7777707D-05	238	-2.5221595D-03	283	-4.5959152D-03	328	1.7637905D-04
14	8.2939008D-04	59	6.4596421D-03	104	-6.1706283D-05	148	6.3916559D-04	194	9.5087571D-04	239	-4.2984167D-03	284	-4.9316467D-03	329	2.1762524D-04
15	1.8720429D-03	60	7.8186041D-03	105	5.2834721D-04	149	7.8648329D-04	195	6.1512383D-04	240	-4.5902138D-03	285	-5.2219016D-03	330	1.4162913D-04
16	2.5664353D-03	61	7.2677619D-03	106	9.0390884D-04	150	5.1183834D-04	196	8.3305383D-04	241	-2.3889510D-03	286	-4.9383179D-03	331	1.5758626D-04
17	1.0734235D-03	62	7.0178625D-03	107	1.0687245D-03	151	5.7299300D-04	197	-3.3004350D-04	242	-2.7273927D-03	287	-4.4986394D-03	332	1.3340568D-04
18	-3.3531234D-04	63	6.5542706D-03	108	1.1279753D-03	152	4.8560598D-04	198	-6.0776602D-04	243	-5.2019420D-03	288	-4.1550840D-03	333	1.2649695D-04
19	-1.8671264D-03	64	3.9651046D-03	109	9.6228506D-04	153	4.6063828D-04	199	-7.2898822D-04	244	-8.0789610D-03	289	-3.9583555D-03	334	7.0814555D-05
20	-2.7558932D-04	65	2.9044004D-03	110	6.8705246D-04	154	2.5591914D-04	200	9.9107236D-04	245	-7.2710414D-03	290	-4.1394545D-03	335	8.8086400D-05
21	1.2457599D-03	66	3.5083637D-03	111	1.4565377D-05	155	3.1833845D-04			246	-4.8151116D-03	291	-4.2695652D-03	336	1.2953884D-04
22	3.4008415D-03	67	5.7664387D-03	112	9.4121778D-05	156	4.6314484D-04	201	1.4911032D-03	247	-3.0123600D-03	292	-3.5592067D-03	337	1.3126601D-04
23	4.8720125D-03	68	6.2528410D-03	113	3.4561862D-04	157	4.7438673D-04	202	1.7627344D-03	248	-1.8190111D-03	293	-3.5907697D-03	338	1.2263009D-04
24	3.6638737D-03	69	6.4409833D-03	114	9.4918283D-04	158	4.4317708D-04	203	1.7843479D-03	249	-1.1817504D-03	294	-3.3734784D-03	339	9.1540766D-05
25	3.3294487D-03	70	5.9559194D-03	115	1.7670862D-03	159	3.8082230D-04	204	2.9521919D-03	250	-5.1815913D-05	295	-3.9806414D-03	340	1.0708544D-04
26	1.9625108D-03	71	4.8980329D-03	116	1.9778891D-03	160	3.8699972D-04	205	3.6807332D-03	251	-8.9127642D-04	296	-3.5109119D-03	341	1.0535824D-04
27	3.7894305D-03	72	4.2069655D-03	117	1.7871288D-03	161	3.8075776D-04	206	3.1219099D-03	252	-2.0606898D-03	297	-3.3981751D-03	342	8.9813591D-05
28	5.9633153D-03	73	3.5374760D-03	118	1.3239080D-03	162	3.2458041D-04	207	2.1930063D-03	253	-4.1304688D-03	298	-2.7634600D-03	343	6.2178632D-05
29	9.3462217D-03	74	3.0667823D-03	119	1.3850089D-03	163	2.2470949D-04	208	3.5839659D-03	254	-5.4924876D-03	299	-2.8208120D-03	344	5.1815535D-05
30	8.9356558D-03	75	1.4464430D-03	120	1.6139330D-03	164	1.8725793D-04	209	5.8473814D-03	255	-4.7192418D-03	300	-2.3596083D-03	345	1.0535824D-04
31	7.5773591D-03	76	5.7345594D-04	121	2.0531329D-03	165	3.8075776D-04	210	6.3063268D-03	256	-2.5360613D-03	301	-2.1560674D-03	346	1.0017669D-04
32	2.7166725D-03	77	-1.1886856D-04	122	2.0052944D-03	166	3.6203196D-04	211	5.5634295D-03	257	-1.1098802D-03	302	-1.6883252D-03	347	7.7723303D-05
33	1.6357886D-03	78	-6.4496208D-05	123	1.5479898D-03	167	2.8083690D-04	212	4.1979363D-03	258	-1.3682827D-03	303	-1.5659252D-03	348	6.9087376D-06
34	3.5765556D-04	79	-3.3674596D-04	124	1.3122446D-03	168	2.4967723D-05	213	5.7349961D-03	259	-1.1122152D-03	304	-1.1069988D-03	349	0.0
35	3.8835423D-03	80	-1.7604795D-04	125	1.9274395D-03	169	0.0	214	5.8429429D-03	260	-1.0282772D-03	305	-8.6466605D-04	350	0.0
36	5.7559742D-03	81	5.3394920D-04	126	2.7574019D-03	170	0.0	215	5.9245334D-03	261	-5.3766716D-04	306	-5.3330903D-04	351	0.0
37	8.9781639D-03	82	1.7968380D-03	127	2.5085136D-03	171	0.0	216	3.1391634D-03	262	-1.9352911D-03	307	-3.0373998D-04	352	0.0
38	9.8204837D-03	83	1.8984026D-03	128	2.1090375D-03	172	0.0	217	3.6276527D-03	263	-2.4458856D-03	308	-8.6720263D-04	353	0.0
39	1.2577254D-02	84	1.2683142D-03	129	1.1207637D-03	173	0.0	218	4.4803381D-03	264	-3.3397713D-03	309	-1.5325569D-03	354	0.0
40	1.3562758D-02	85	1.6043317D-04	130	1.2156169D-03	174	0.0	219	6.7186594D-03	265	-3.0799524D-03	310	-1.2705270D-03	355	0.0
41	1.4111366D-02	86	-6.0591481D-04	131	9.7353372D-04	175	0.0	220	5.7914426D-03	266	-2.3293529D-03	311	-1.0665261D-03	356	0.0
42	1.3235772D-02	87	-5.6568879D-04	132	9.8723413D-04	176	0.0	221	5.1071583D-03	267	-1.4002654D-03	312	-8.6899386D-04	357	0.0
43	1.3120353D-02	88	-1.8725753D-04	133	9.6405883D-04	177	0.0	222	4.8555381D-03	268	-2.0579780D-03	313	-6.6307583D-04	358	0.0
44	1.1691591D-02	89	-2.6044477D-04	134	7.8713797D-04	178	0.0	223	5.3501761D-03	269	-4.1143646D-03	314	-1.1320385D-03	359	0.0
45	1.1028722D-02	90	3.1681336D-04	135	8.0618041D-04	179	0.0	224	5.2213987D-03	270	-4.7157599D-03	315	-1.4416842D-03	360	0.0
						180	0.0	225	4.7279787D-03						

TABLE 5.4

COMPARISON OF THREE METHODS FOR ESTIMATING THE PERIOD OF A SUNSPOT CYCLE

Perform Meas. \ Method	Actual	Period Determined By Month Where Predicted Sunspot Number = 5		Period Estimated From Cycle Using ADAPT Algorithm	Period Predicted from Preceding Two Cycles
		Prediction	3rd Quarter Extrapolation		
Std. Deviation Learn Data (Mo.)	-	13.4	12.7	8.7	10.2
Cycle 19	125	129	132	145.	143
Cycle 20	-	143	145	151	157
Cycle 21	-	134	-	-	165

FIGURE 5.1 EFFECT OF ADAPT PREDICTIONS ON TREND OF PEAK MAGNITUDE FOR NEGATIVE CYCLES

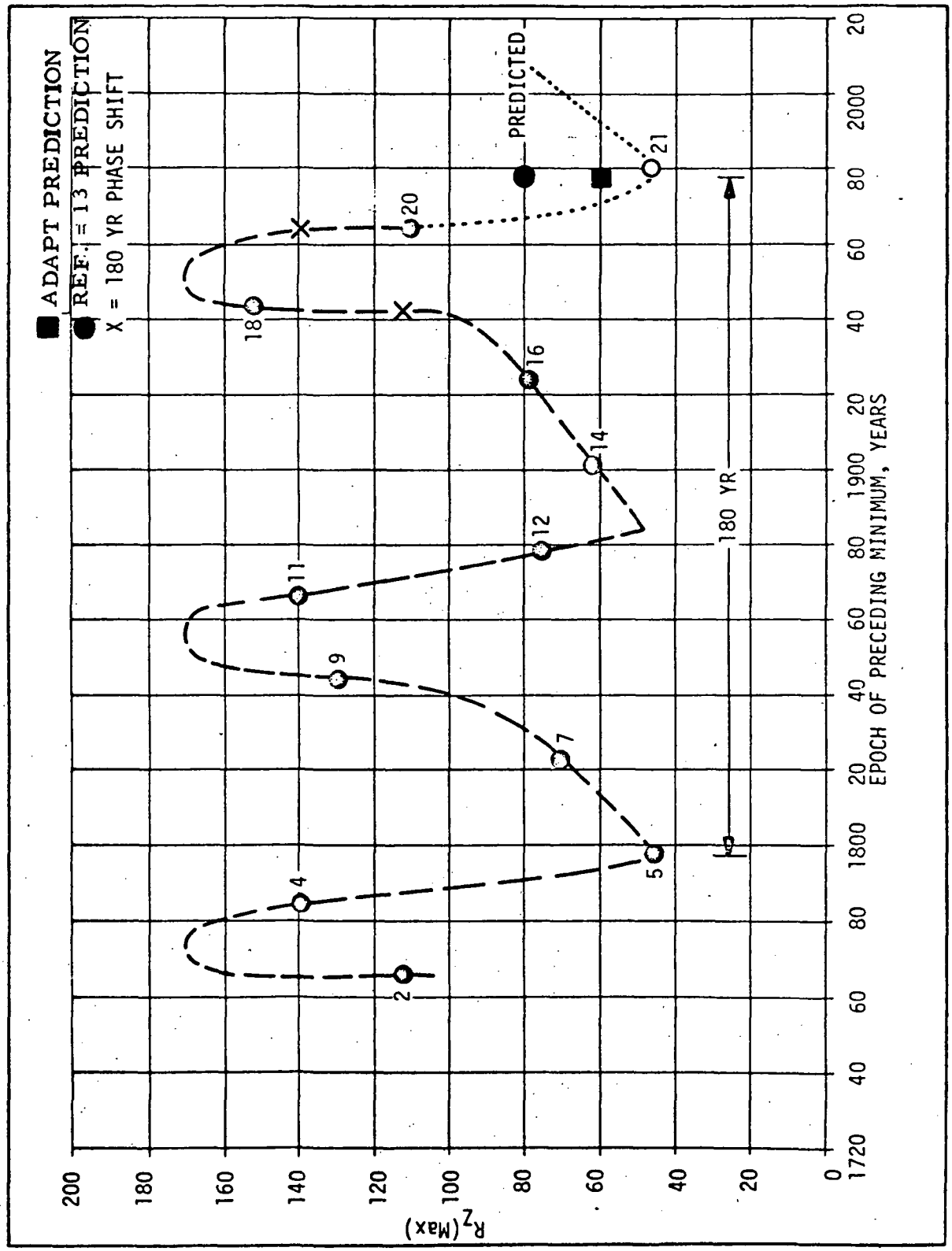


FIGURE 5.2 LOCATION OF ADAPT PREDICTIONS ON MAX SUNSPOT NUMBER VERSUS PERIOD PLOT

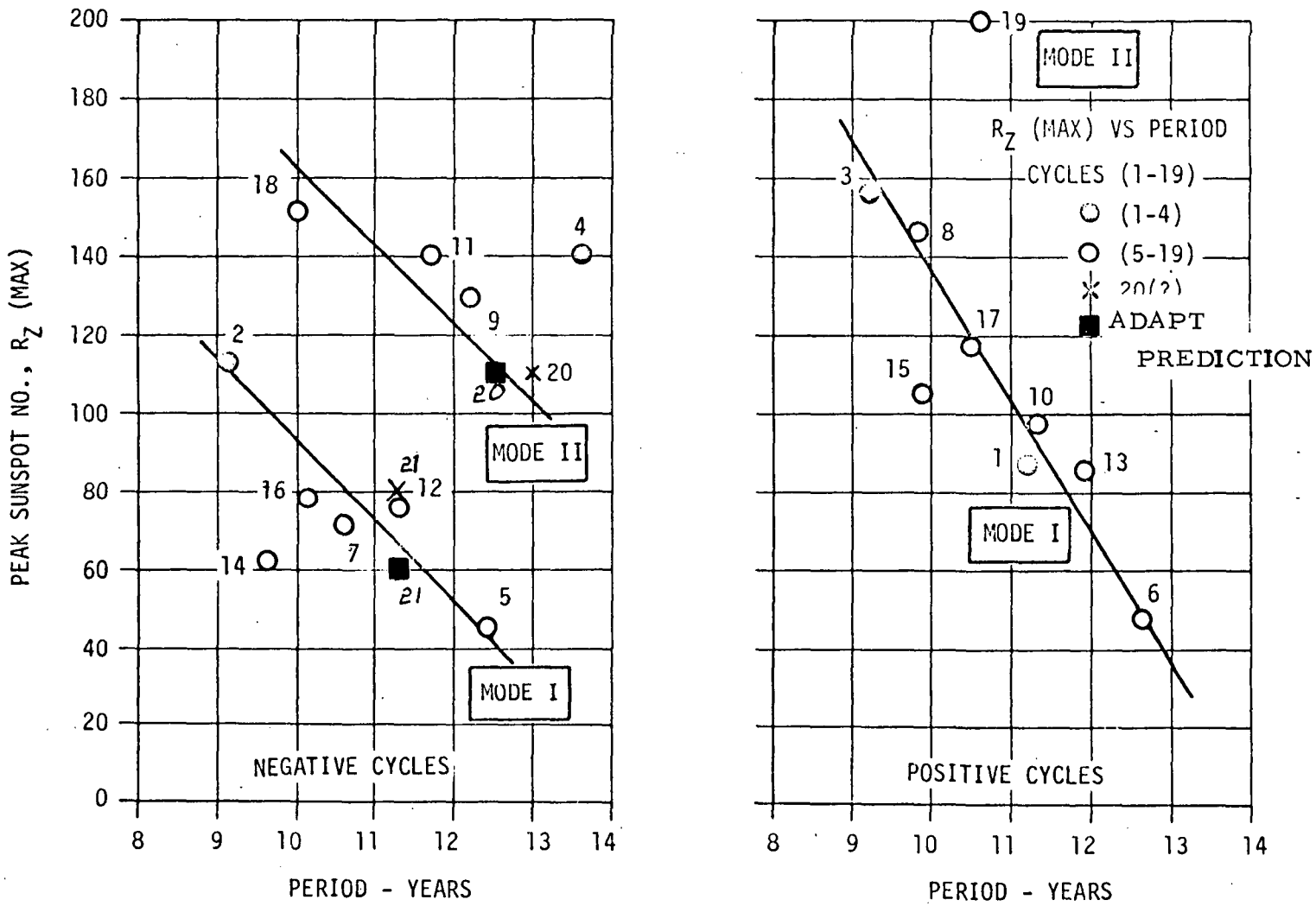


FIGURE 5.3 ESTIMATED VERSUS ACTUAL PERIODS USING SIX DIMENSIONS OF THE SINGLE CYCLE BASE

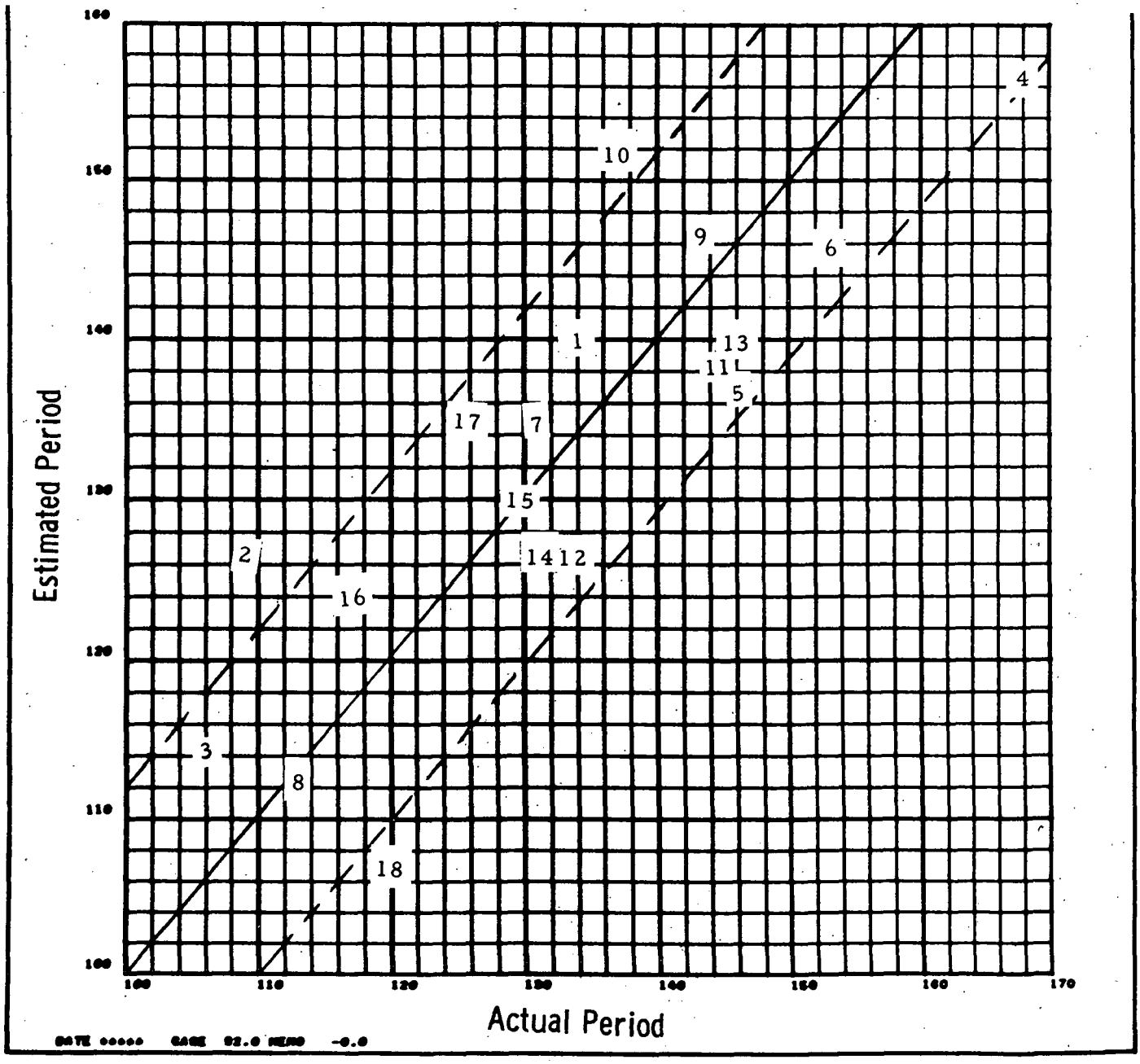


FIGURE 5.4 RELATIVE IMPORTANCE VECTOR FOR PREDICTING THE LENGTH OF A CYCLE FROM THE ESTIMATED CYCLE USING SIX DIMENSIONS

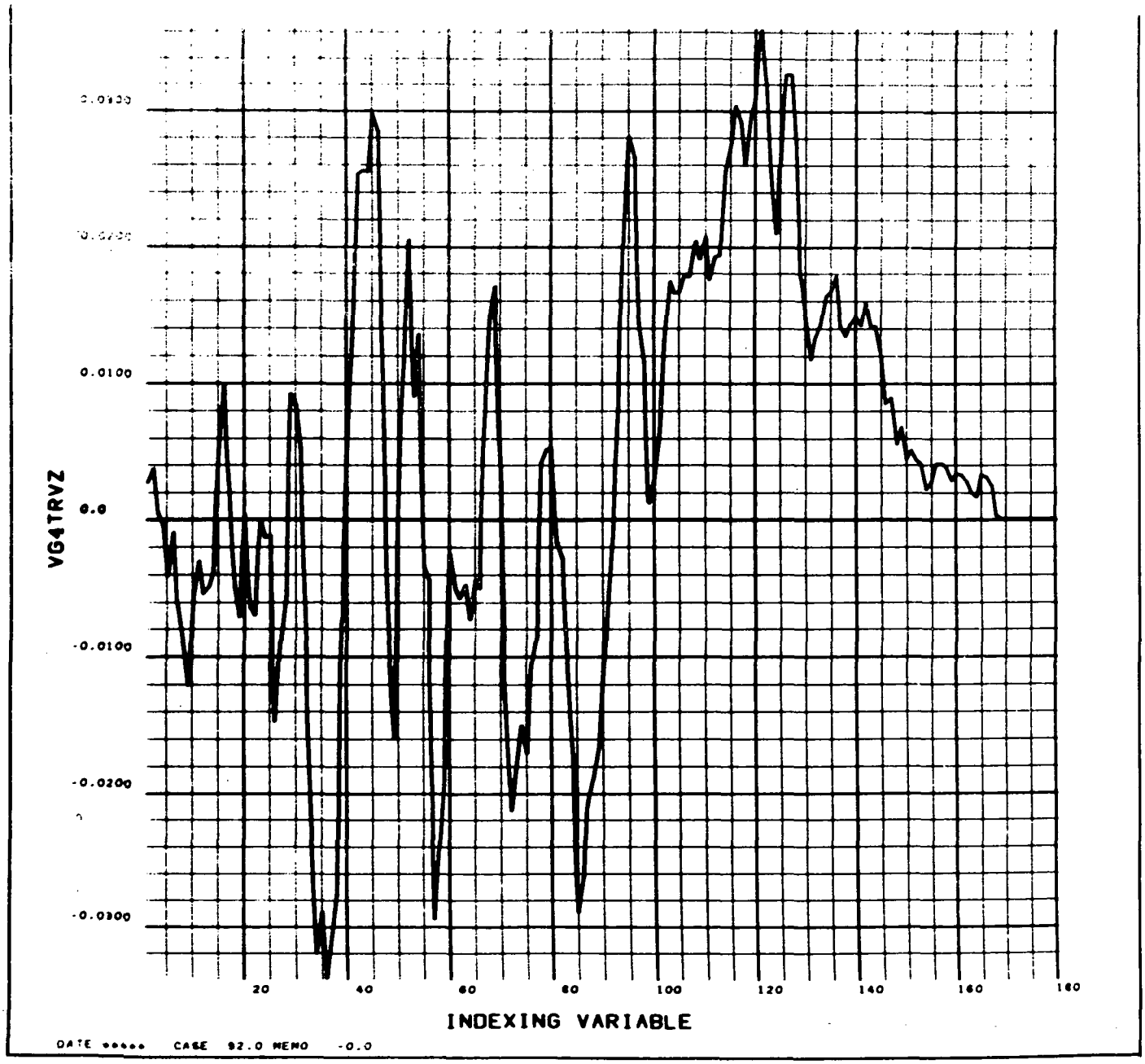


FIGURE 5.5 RELATIVE IMPORTANCE SPECTRUM FOR PREDICTING THE PERIOD OF A CYCLE FROM ITS COEFFICIENTS USING SIX DIMENSIONS

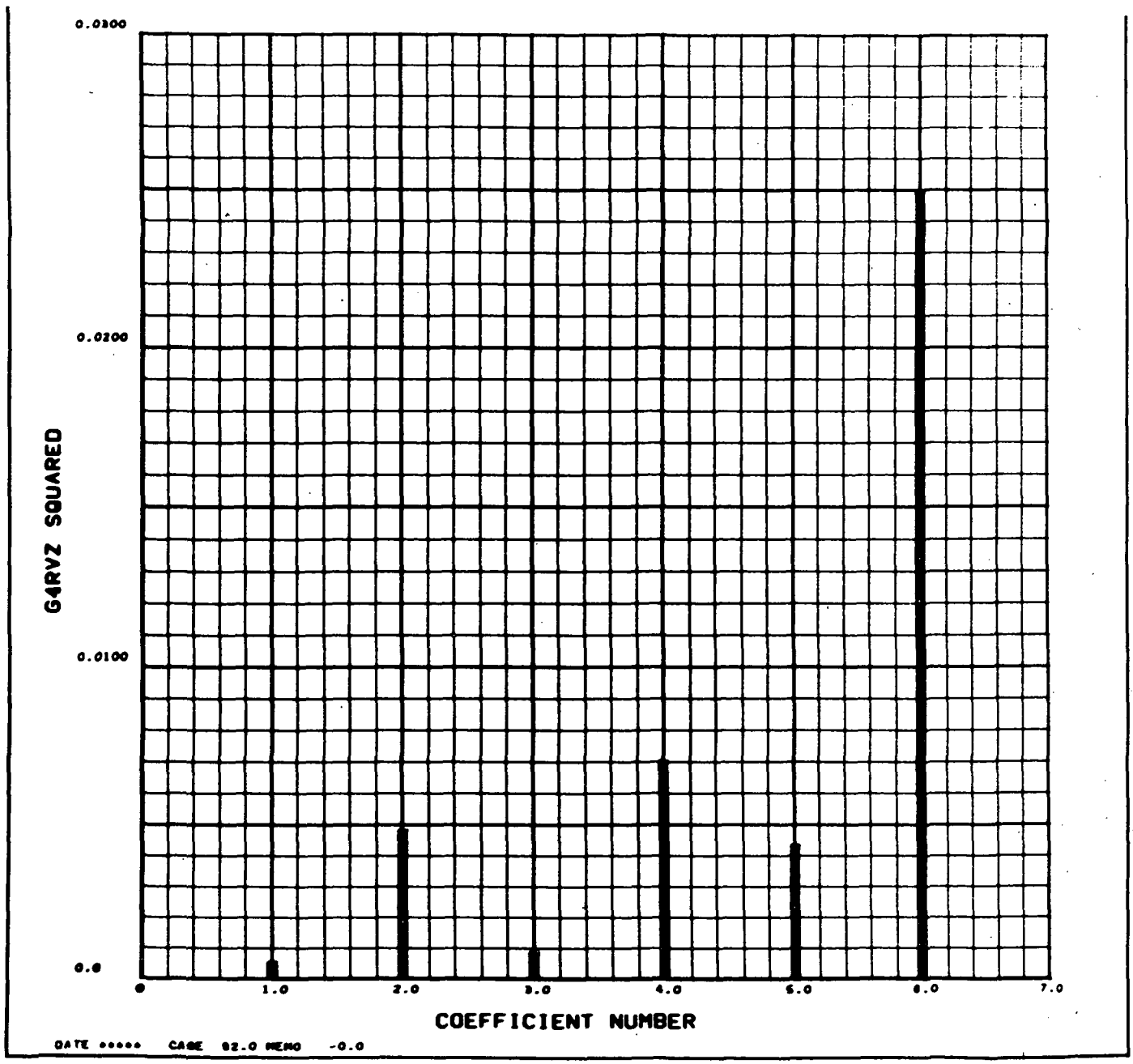


FIGURE 5.6 ESTIMATED VERSUS ACTUAL PERIOD USING TWO DIMENSIONS OF THE SINGLE CYCLE BASE

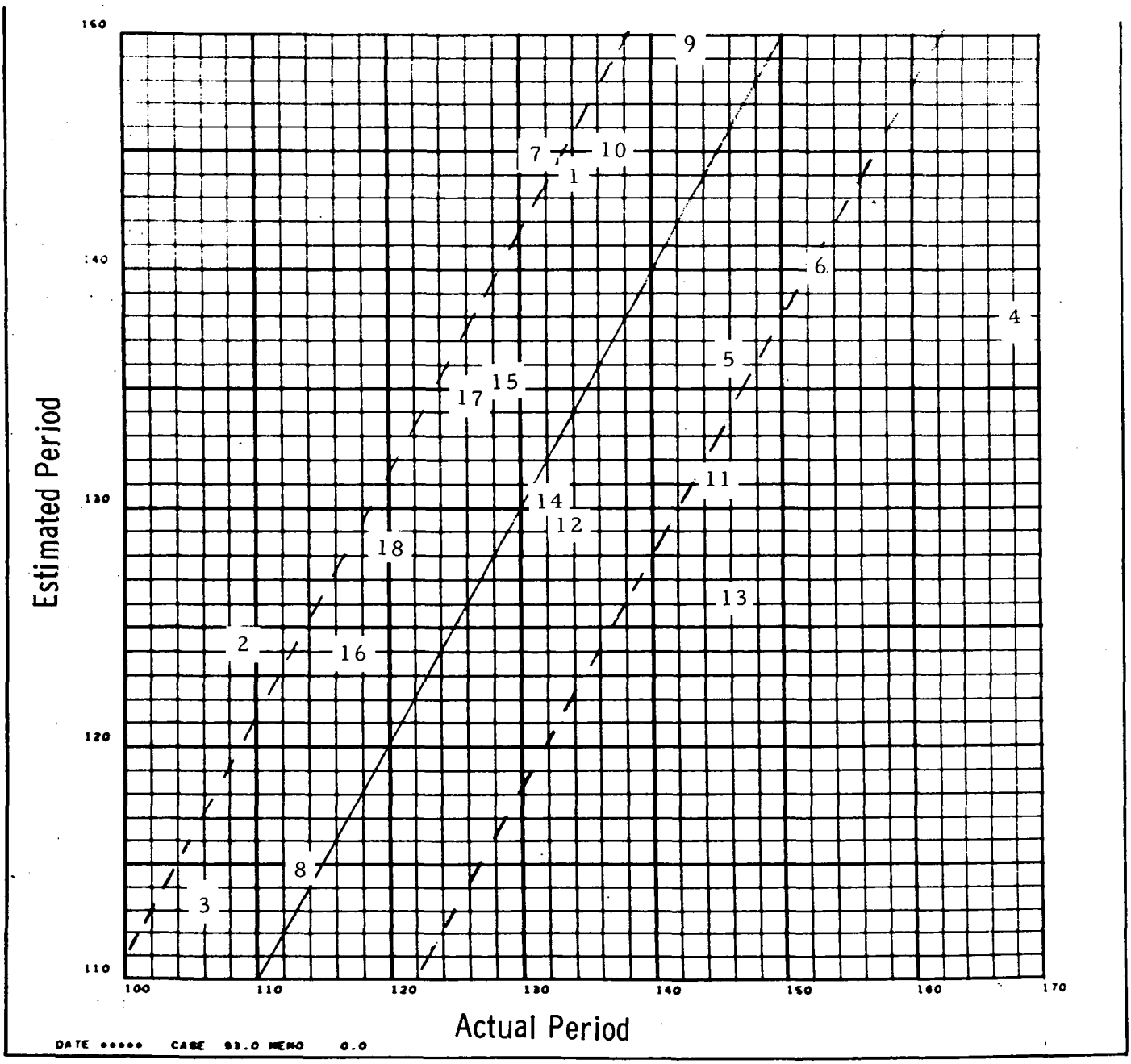


FIGURE 5.7 RELATIVE IMPORTANCE FOR PREDICTION OF PERIOD FROM ESTIMATE OF CYCLE USING TWO DIMENSIONS

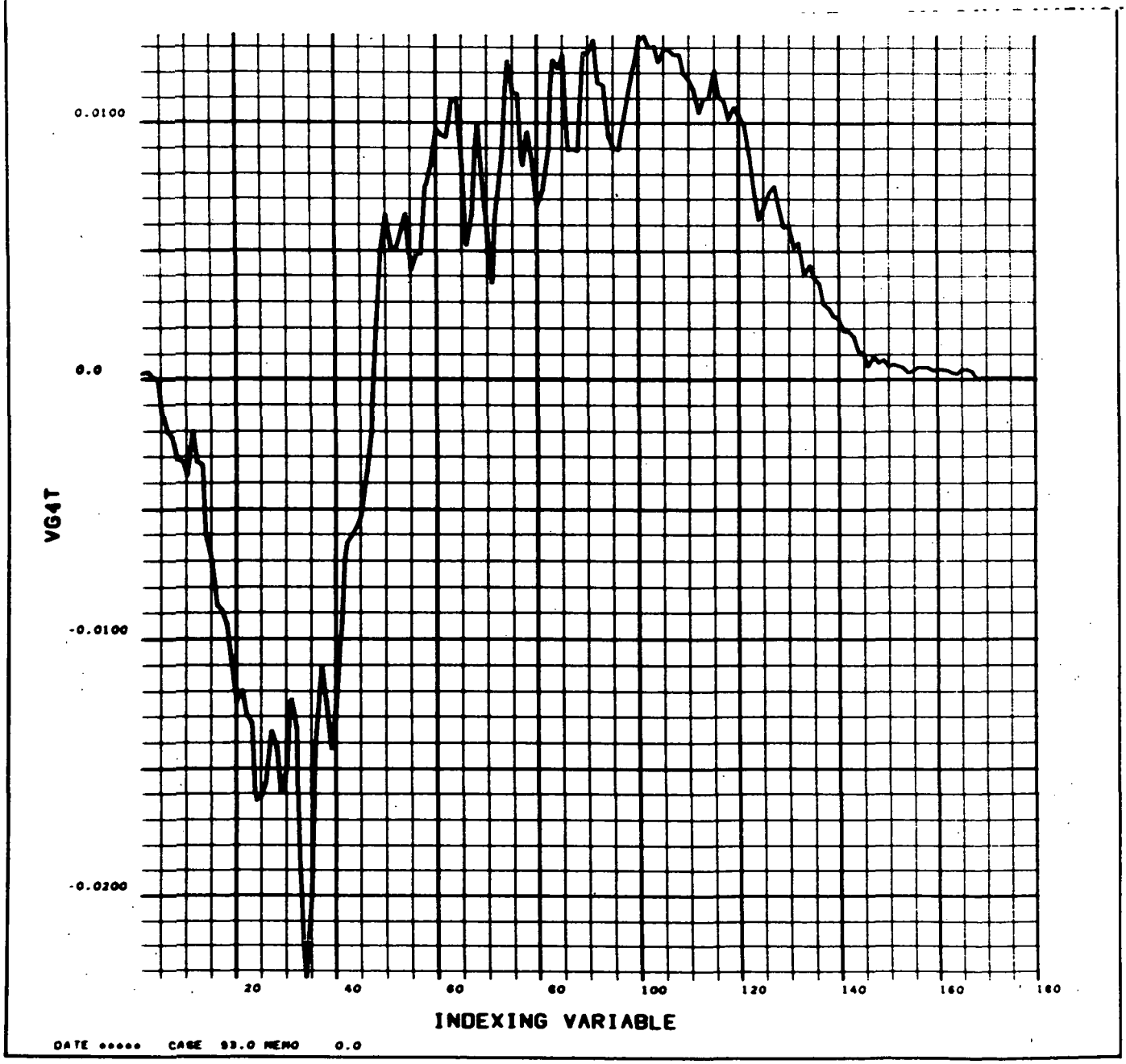


FIGURE 5.8 ESTIMATED VERSUS ACTUAL PERIOD PREDICTED FROM SIX DIMENSIONS OF THE PRECEDING TWO CYCLES

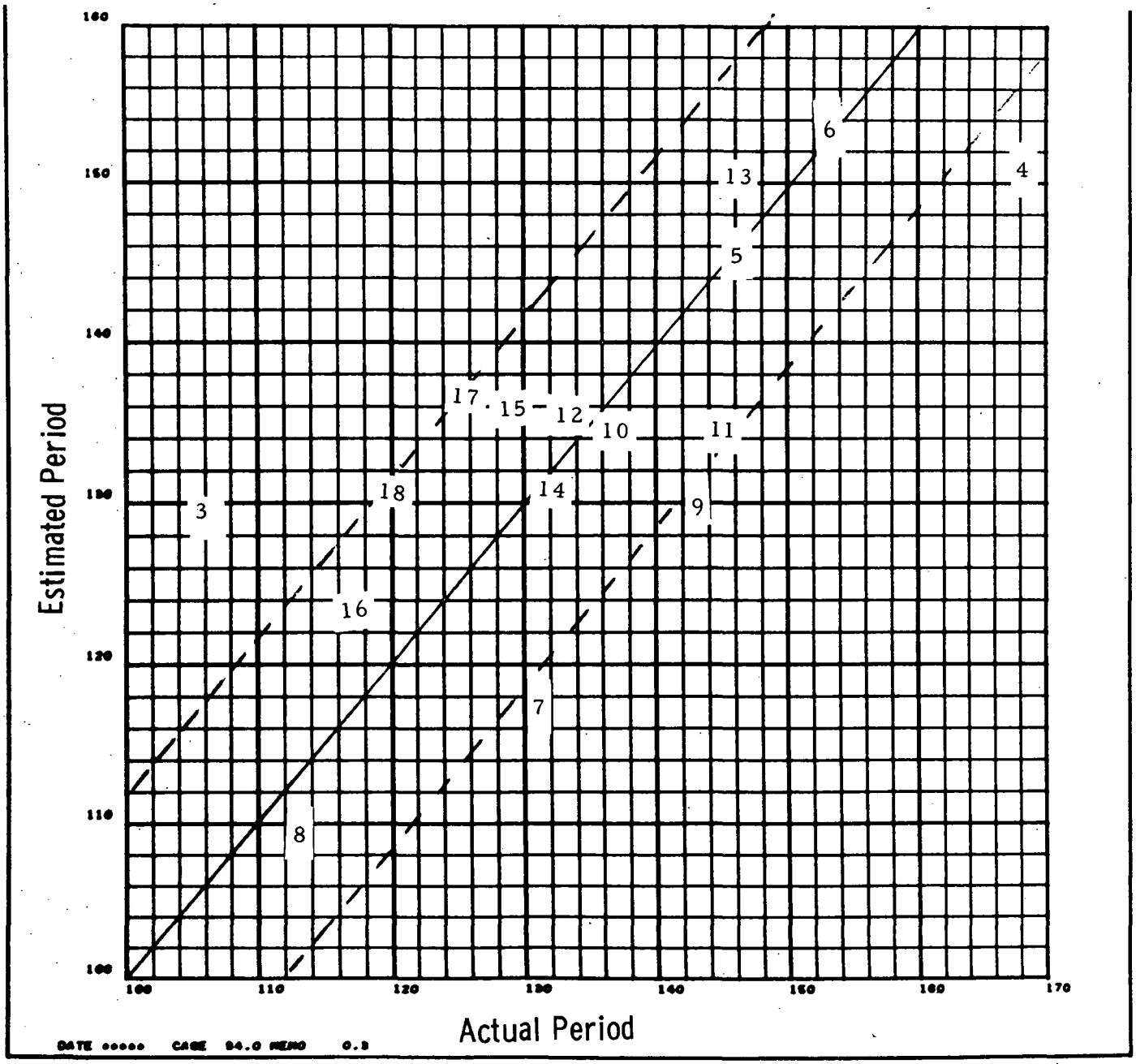


FIGURE 5.9 RELATIVE IMPORTANCE VECTOR FOR PREDICTING PERIOD FROM THE PRECEDING TWO CYCLES

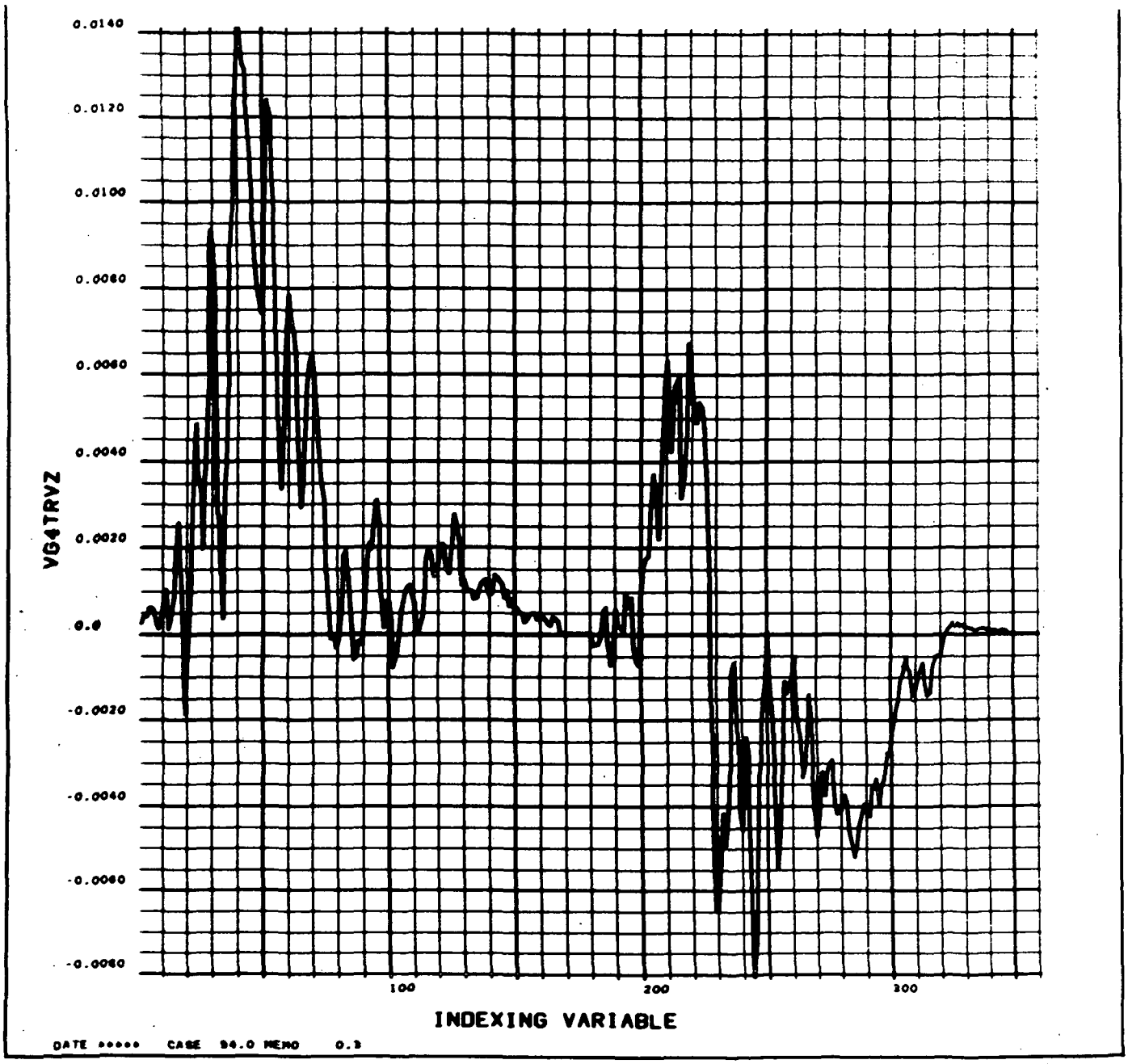


FIGURE 5.10 RELATIVE IMPORTANCE SPECTRUM FOR PREDICTING THE PERIOD FROM THE COEFFICIENTS OF THE PRECEDING TWO CYCLES

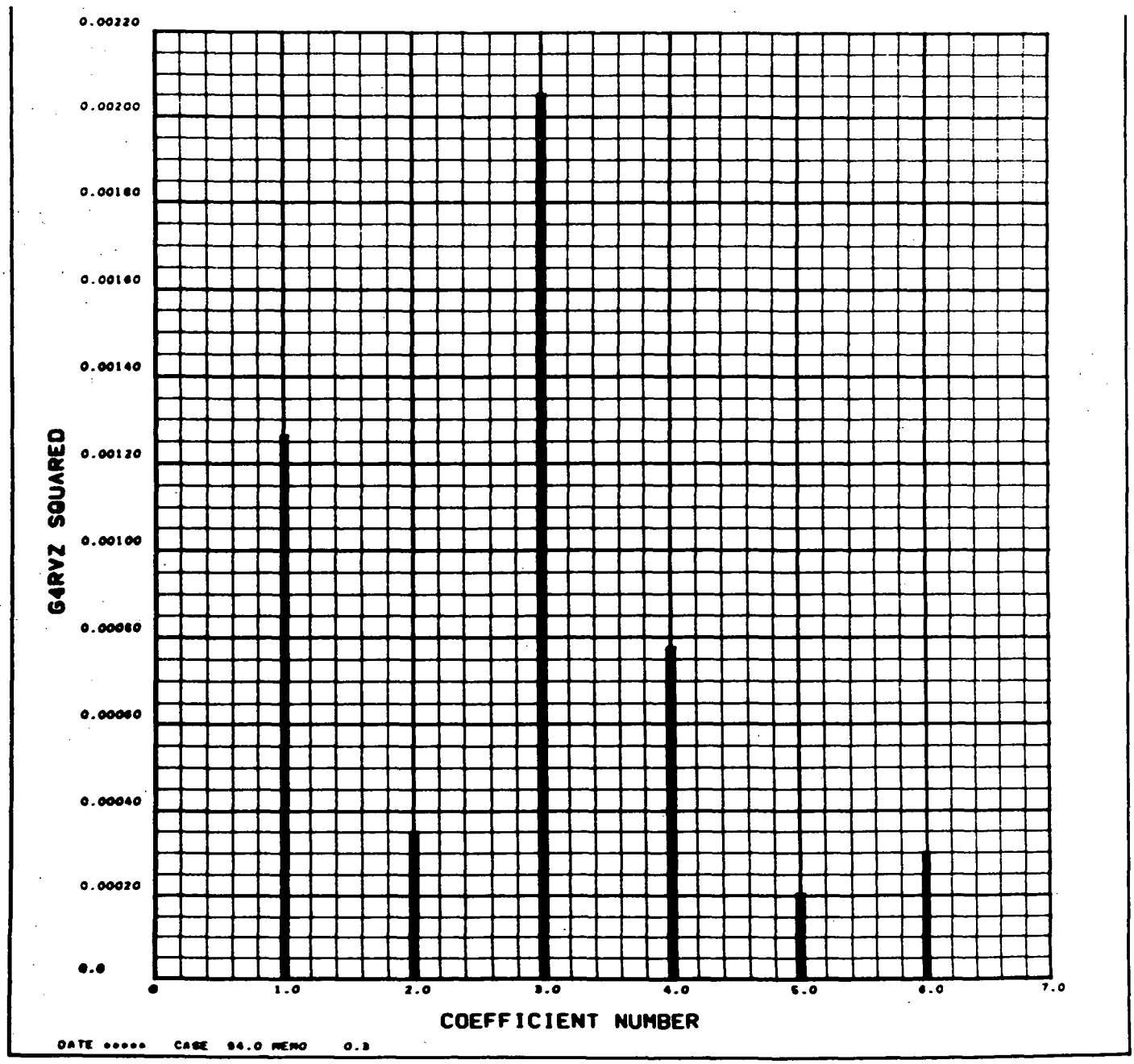


FIGURE 5.11 REARWARD PREDICTION-SUNSPOT AVERAGE INPUT VECTOR THREE-MONTH RUNNING AVERAGE-CYCLES 1 THRU 19

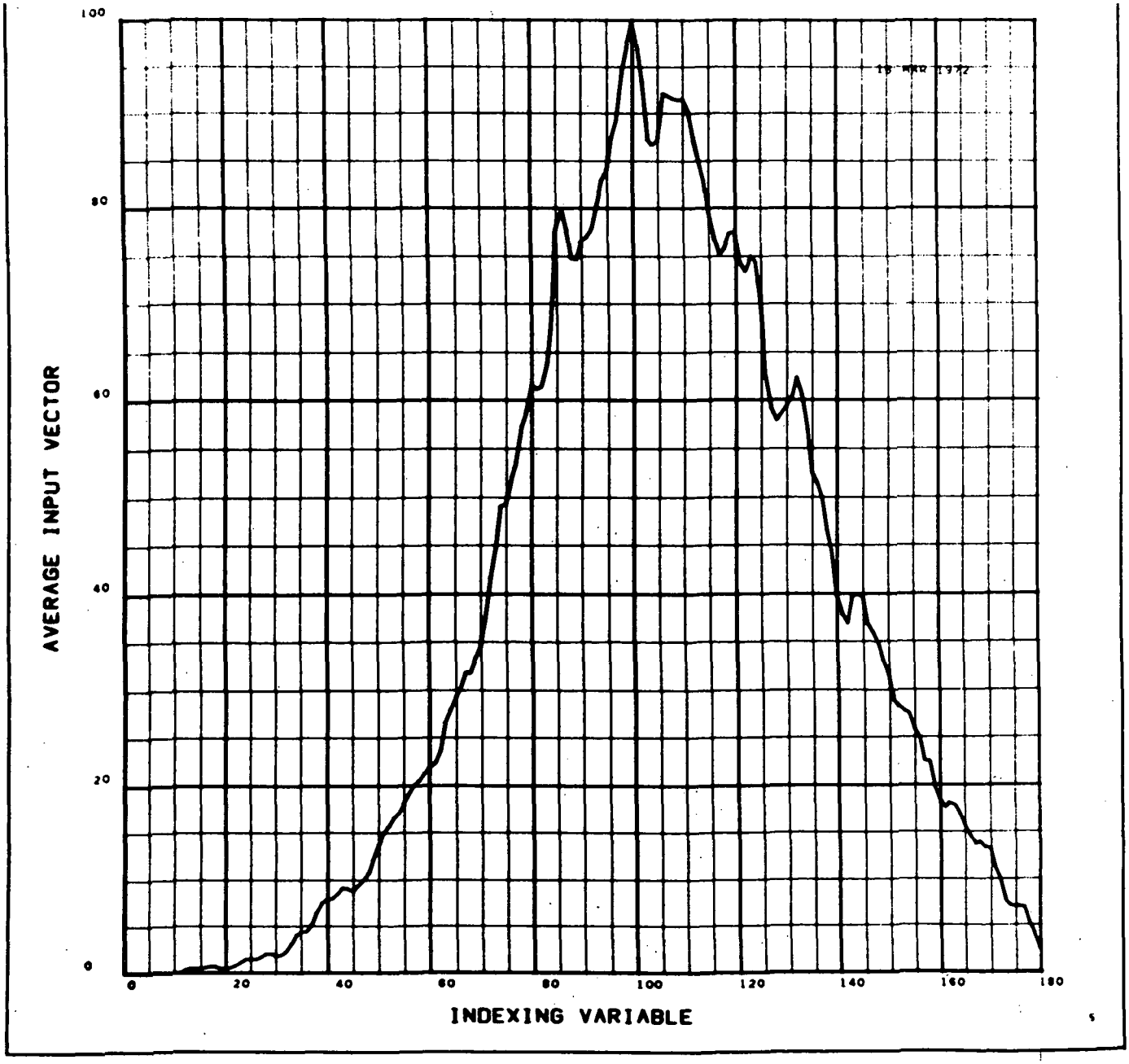


FIGURE 5.12 REARWARD PREDICTION VECTOR-SUNSPOT INFORMATION ENERGY-THREE-MONTH RUNNING AVERAGE-CYCLES 1 THRU 19

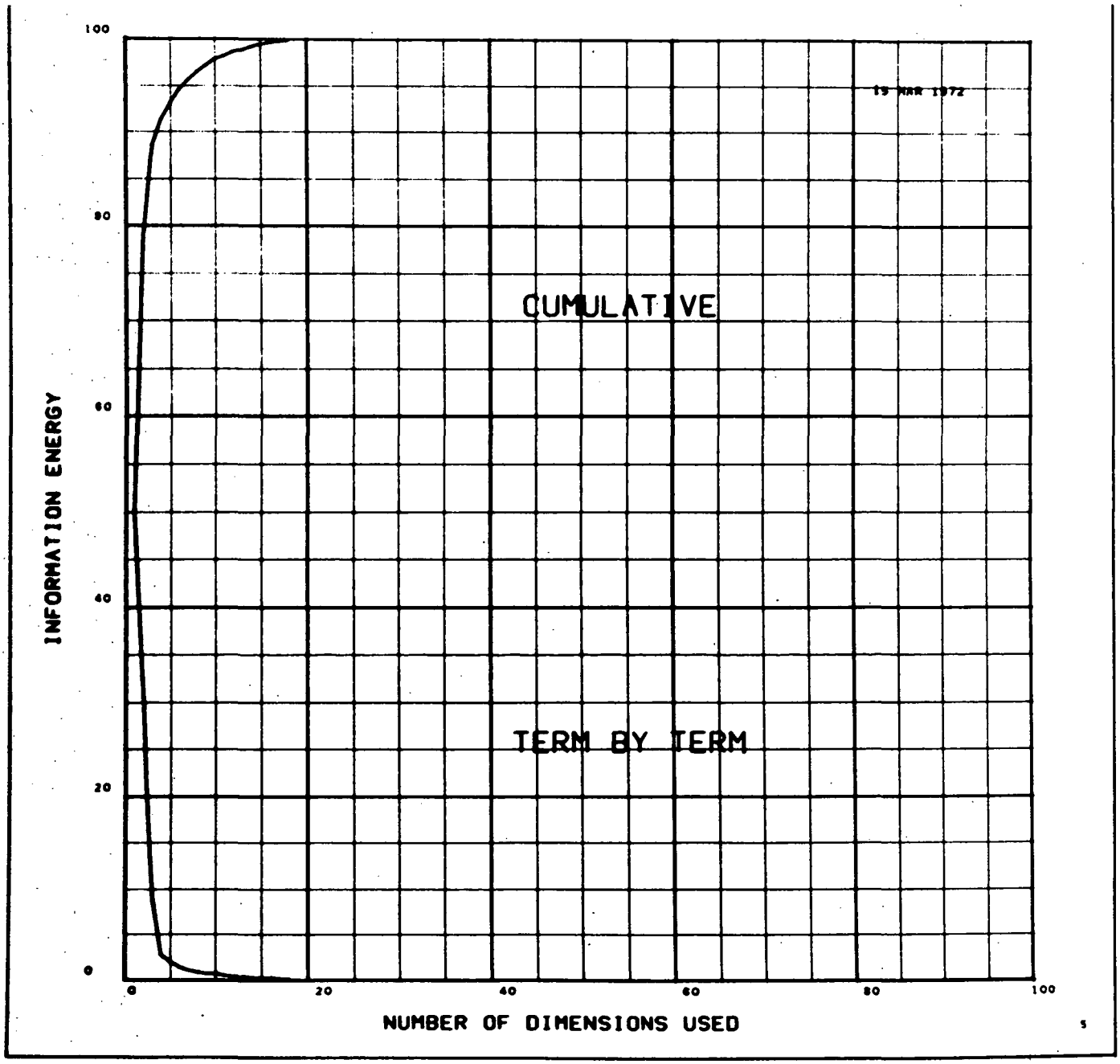


FIGURE 5.13 REARWARD PREDICTION-FIRST OPTIMAL FUNCTION-THREE MONTH RUNNING AVERAGE-CYCLES 1 THRU 19

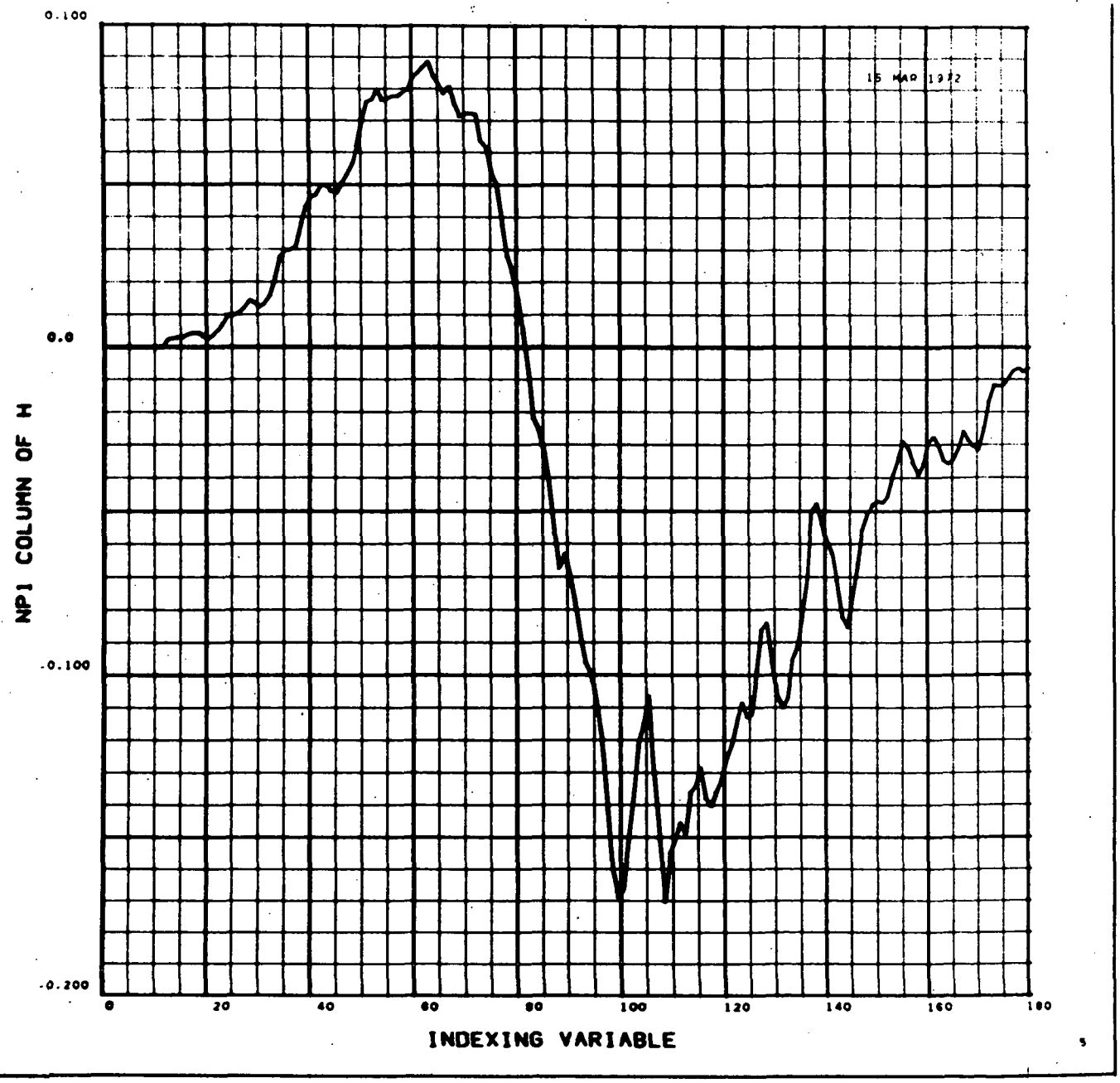


FIGURE 5.14 REARWARD PREDICTION-SECOND OPTIMAL FUNCTION-THREE-MONTH RUNNING AVERAGE-CYCLES 1 THRU 19

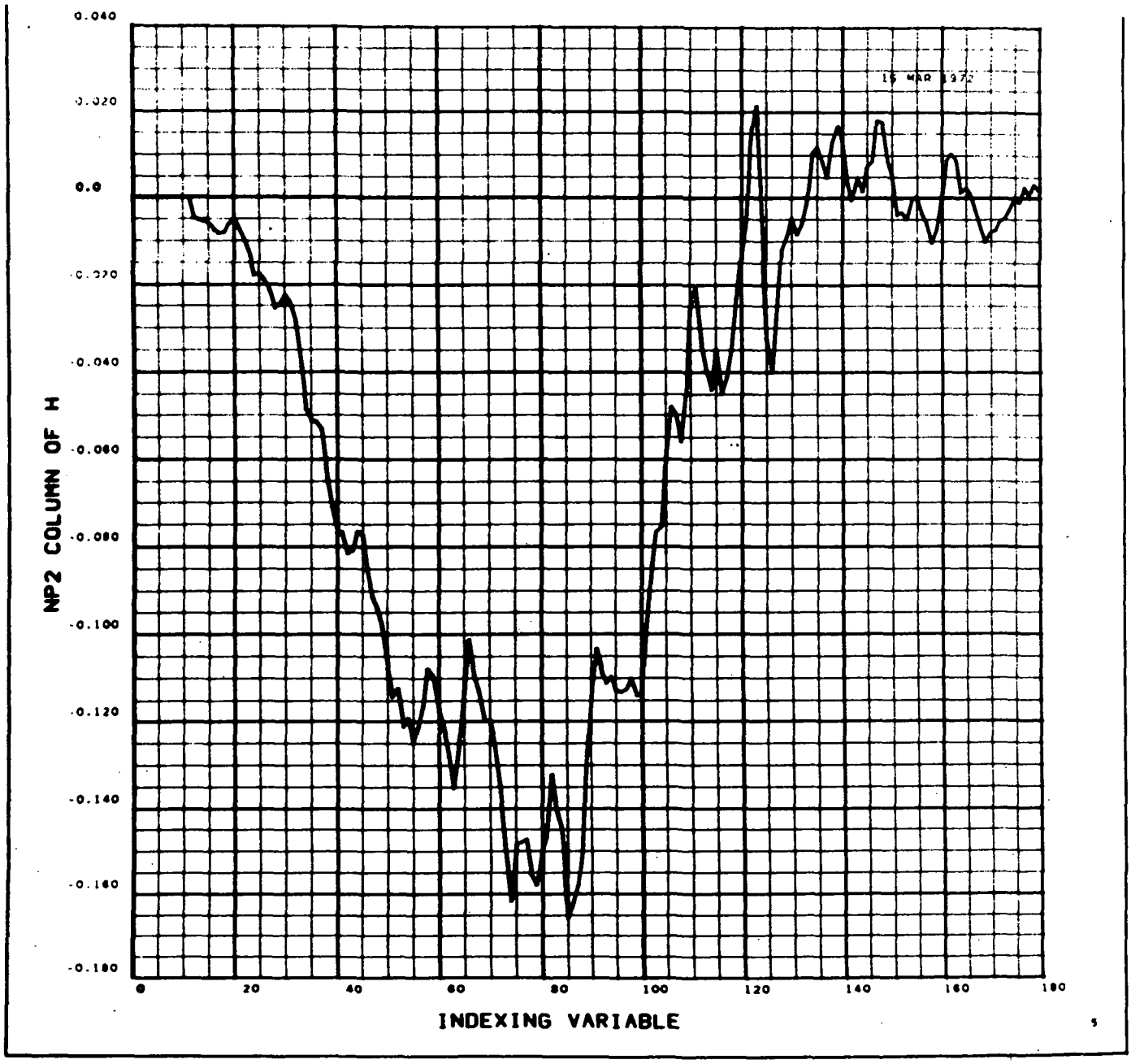


FIGURE 5.15 REARWARD PREDICTION-THIRD OPTIMAL FUNCTION-THREE MONTH RUNNING AVERAGE-CYCLES 1 THRU 19

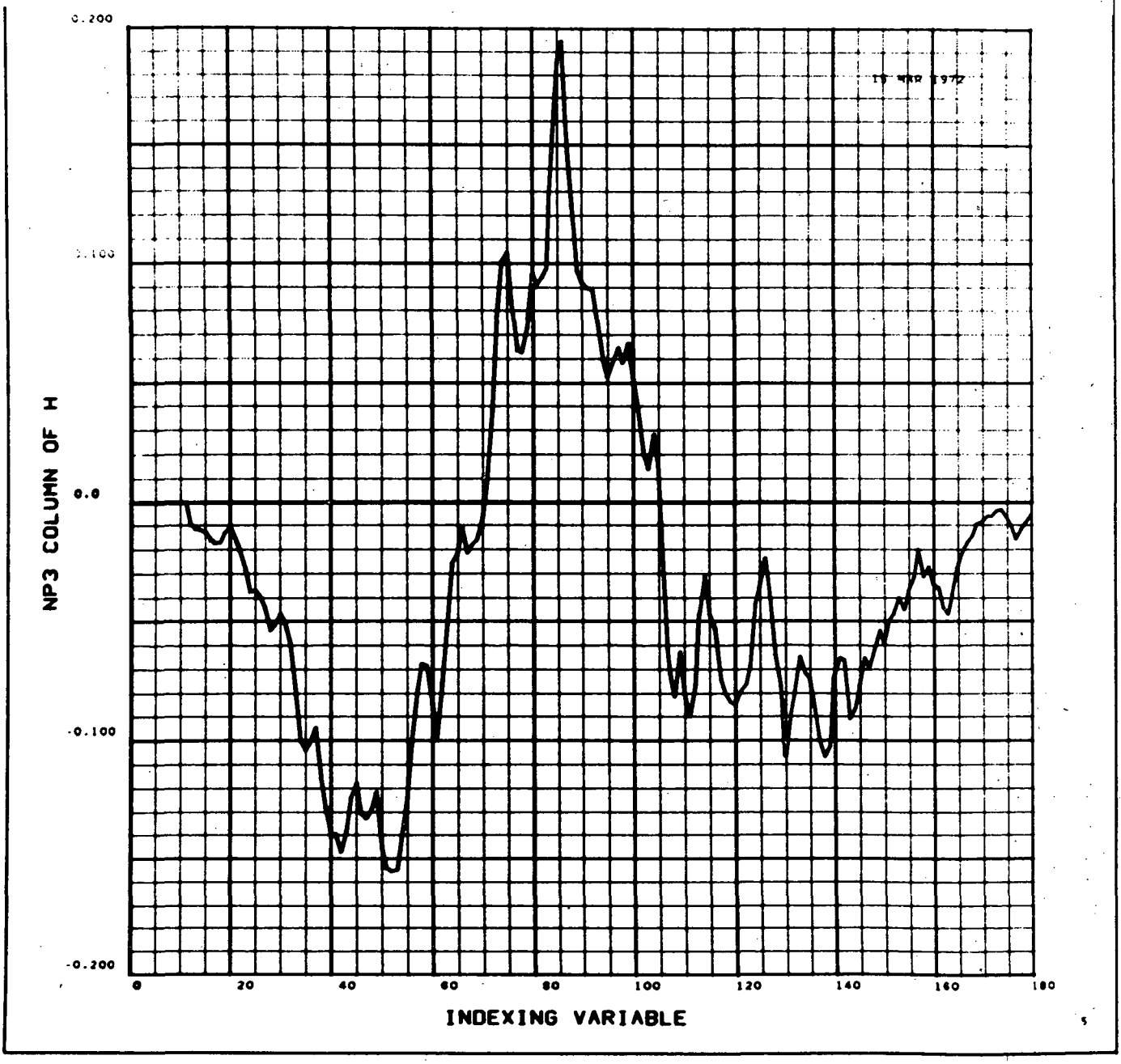


FIGURE 5.16 REARWARD PREDICTION-FOURTH OPTIMAL FUNCTION-THREE MONTH RUNNING AVERAGE-CYCLES 1THRU 19

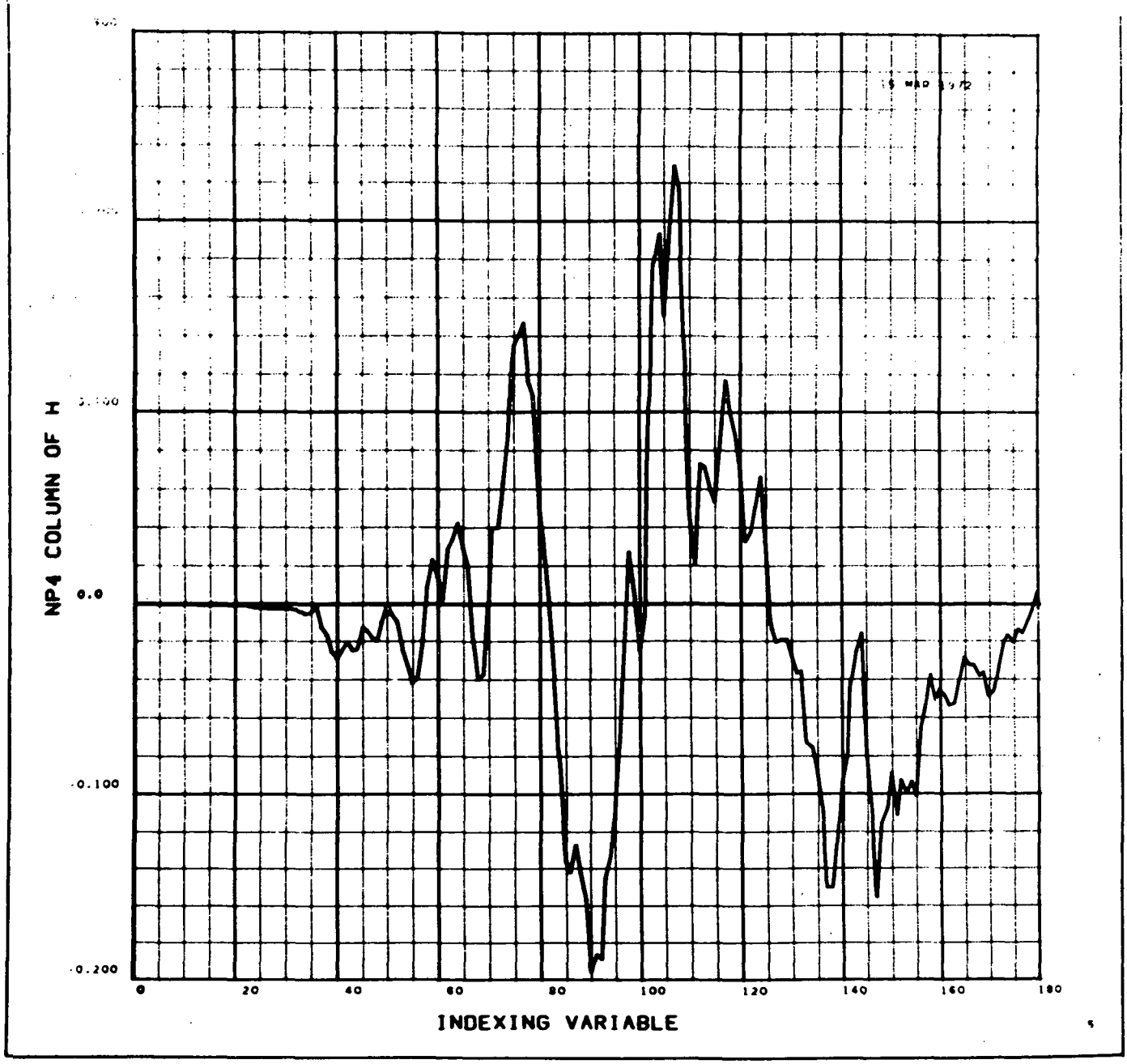


FIGURE 5.17 REARWARD PREDICTION-FIFTH OPTIMAL FUNCTION-THREE MONTH RUNNING AVERAGE-CYCLES 1 THRU 19

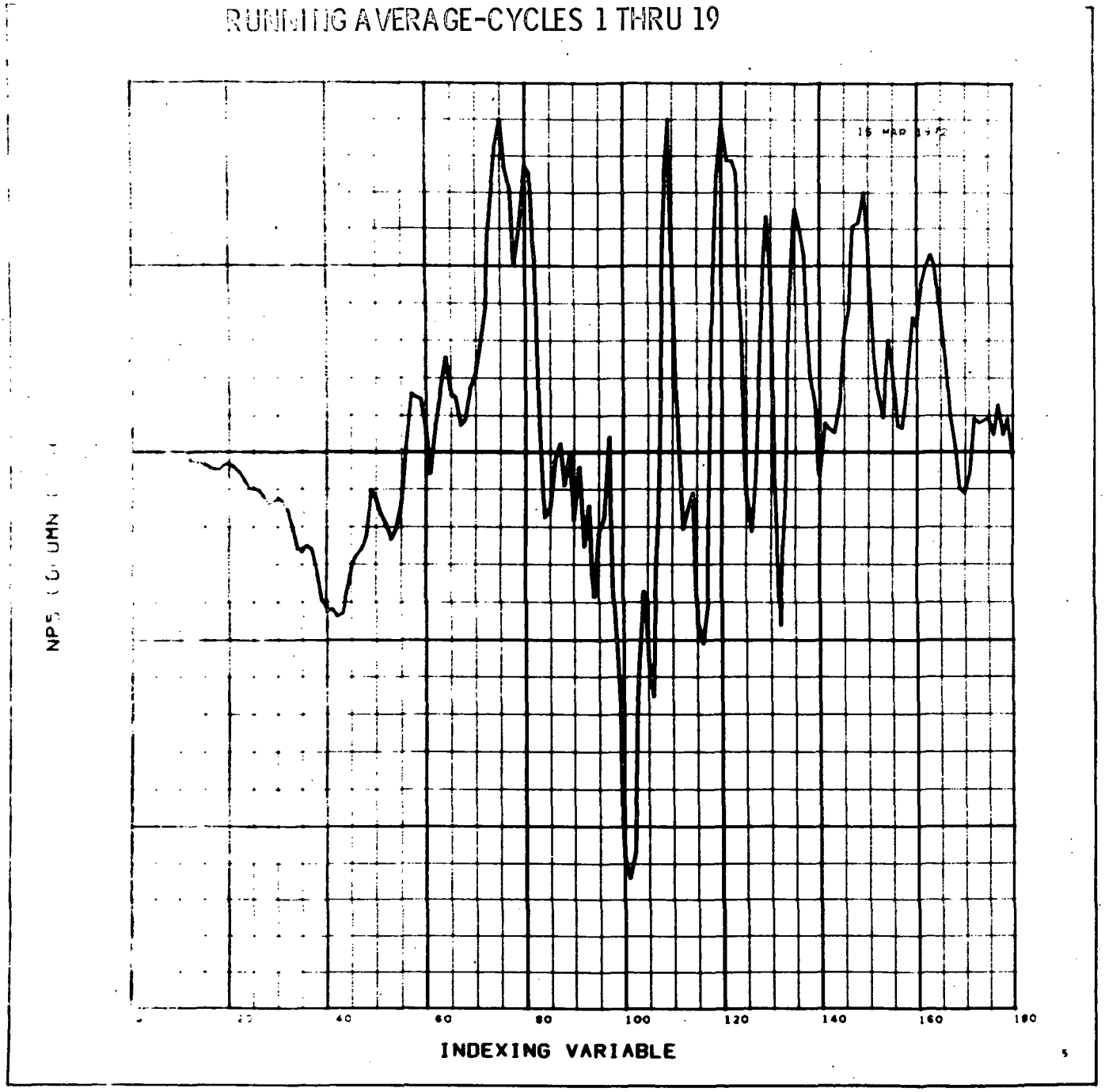


FIGURE 5.13 ESTIMATE OF CYCLE ZERO USING DATA FROM MARCH 1749 TO JUNE 1755 AS PREDICTORS

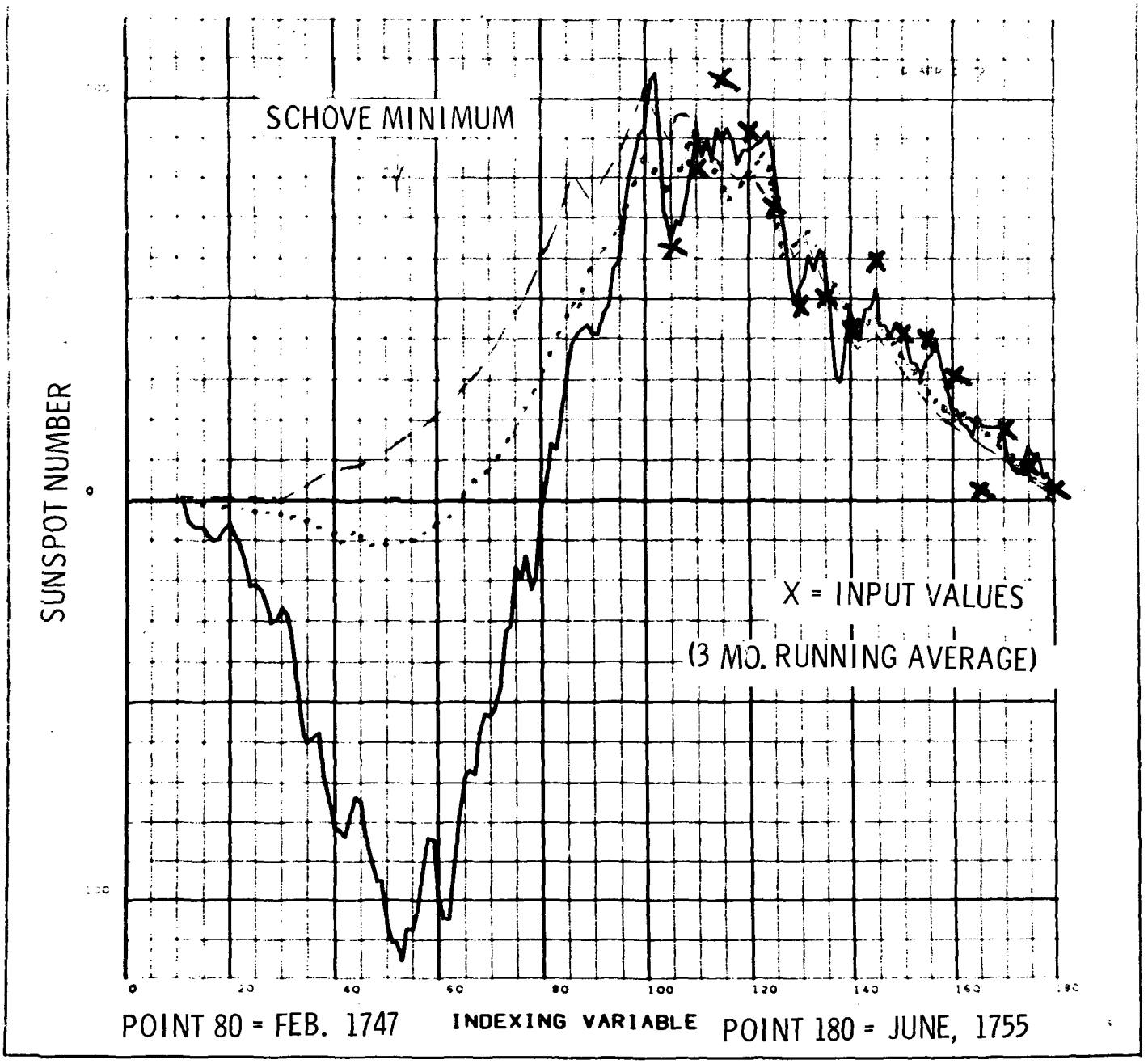


FIGURE 5.19 SCATTER PLOT FORWARD SINGLE CYCLE BASE

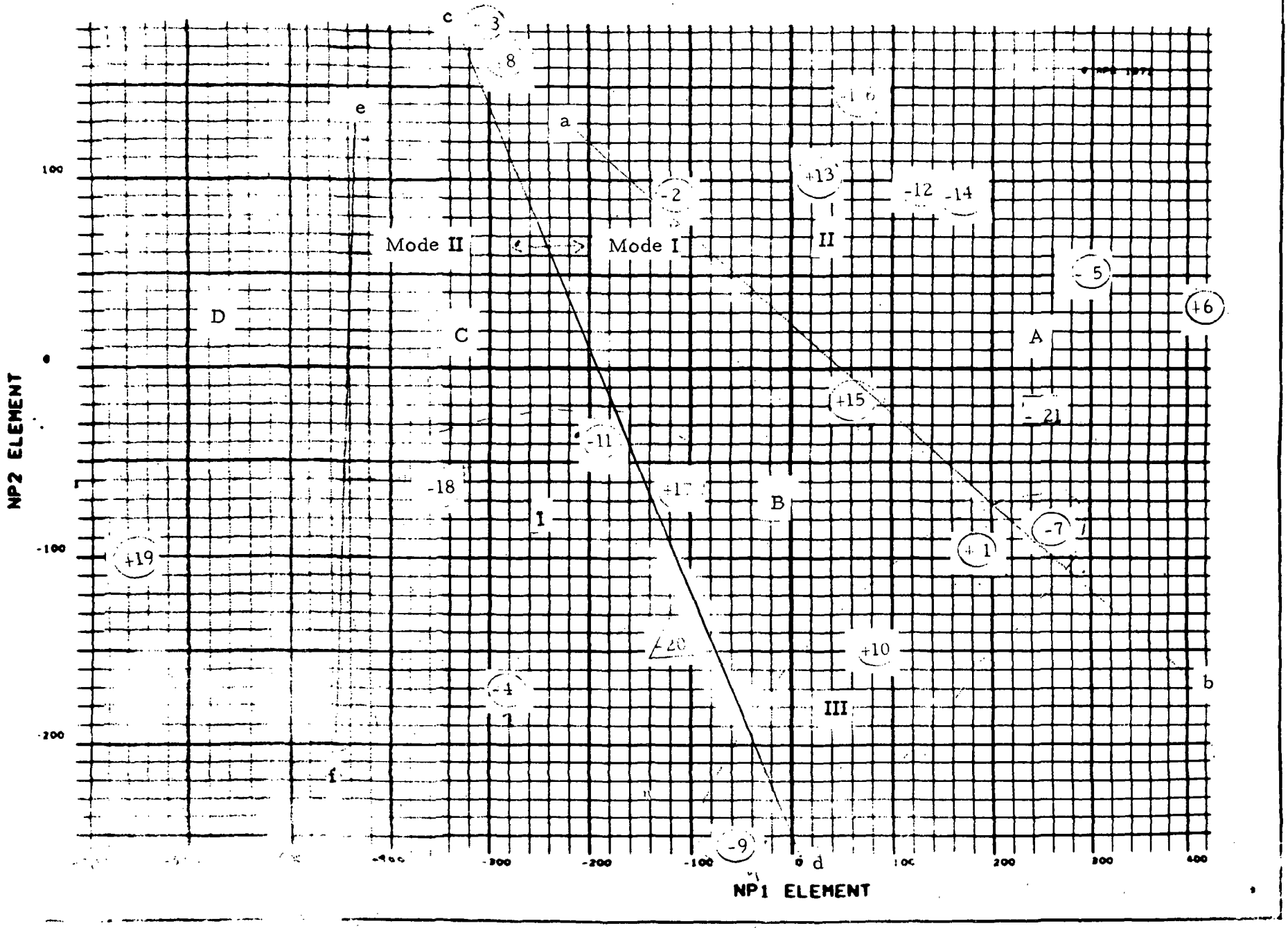


FIGURE 5.20 NEAREST NEIGHBOR PLOT SINGLE CYCLE BASE

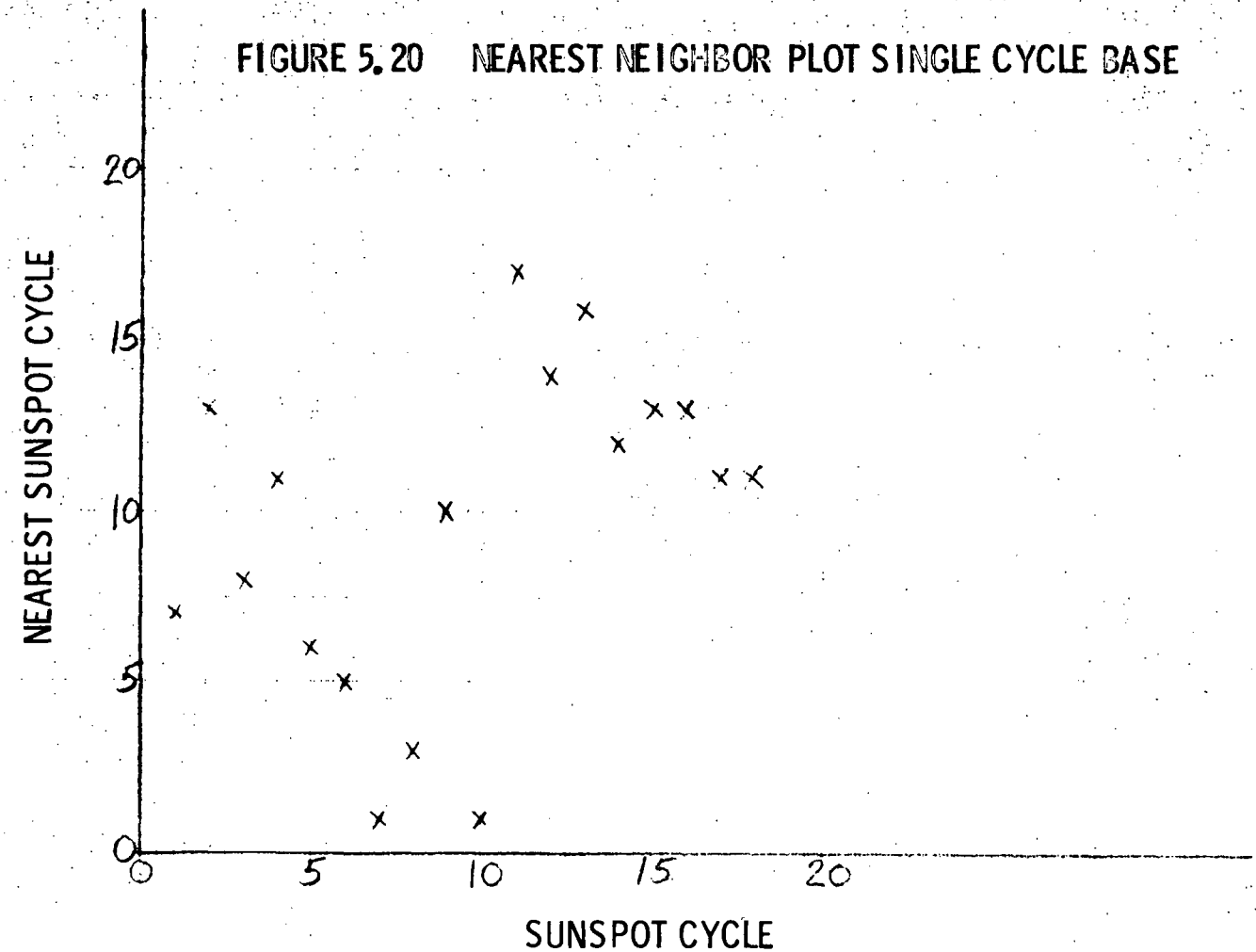


FIGURE 5.21 NEAREST NEIGHBOR TREE SINGLE CYCLE BASE

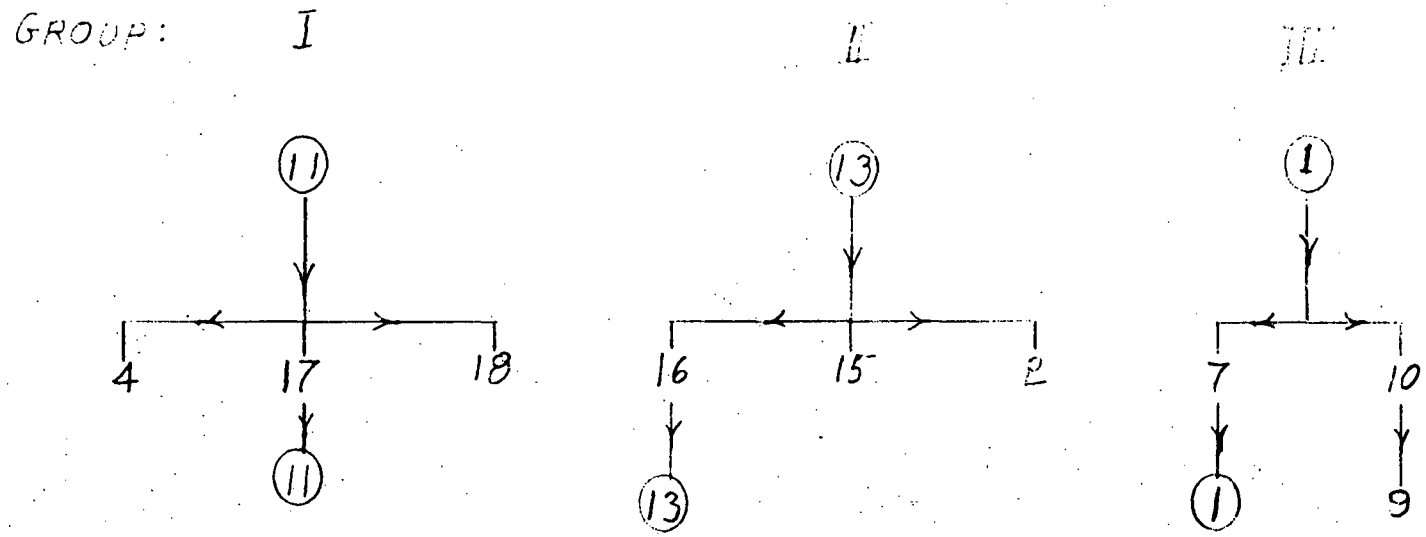


FIGURE 5.22 SCATTER PLOT REARWARD PROJECTION OF CYCLES 1 THRU 19

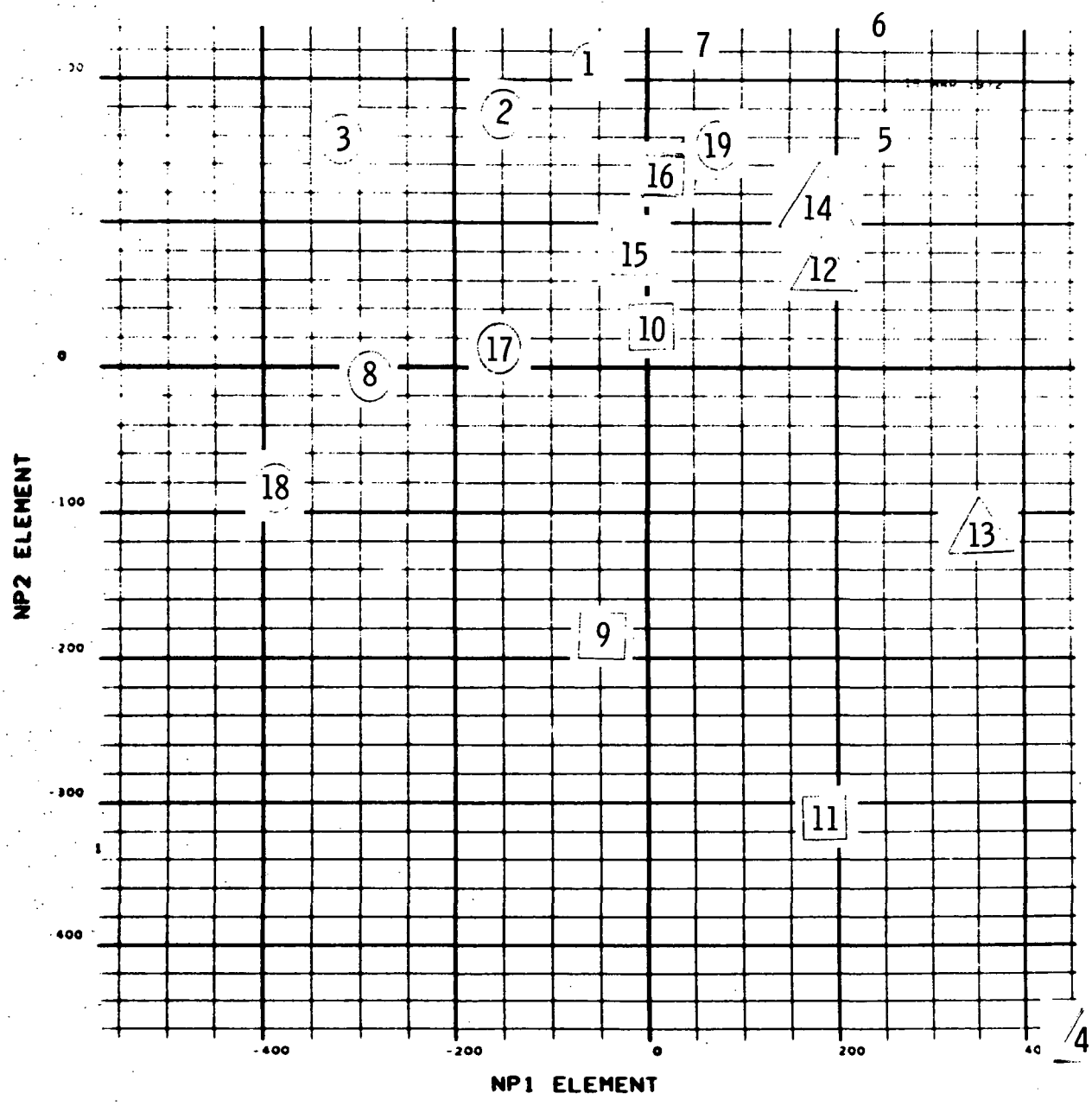
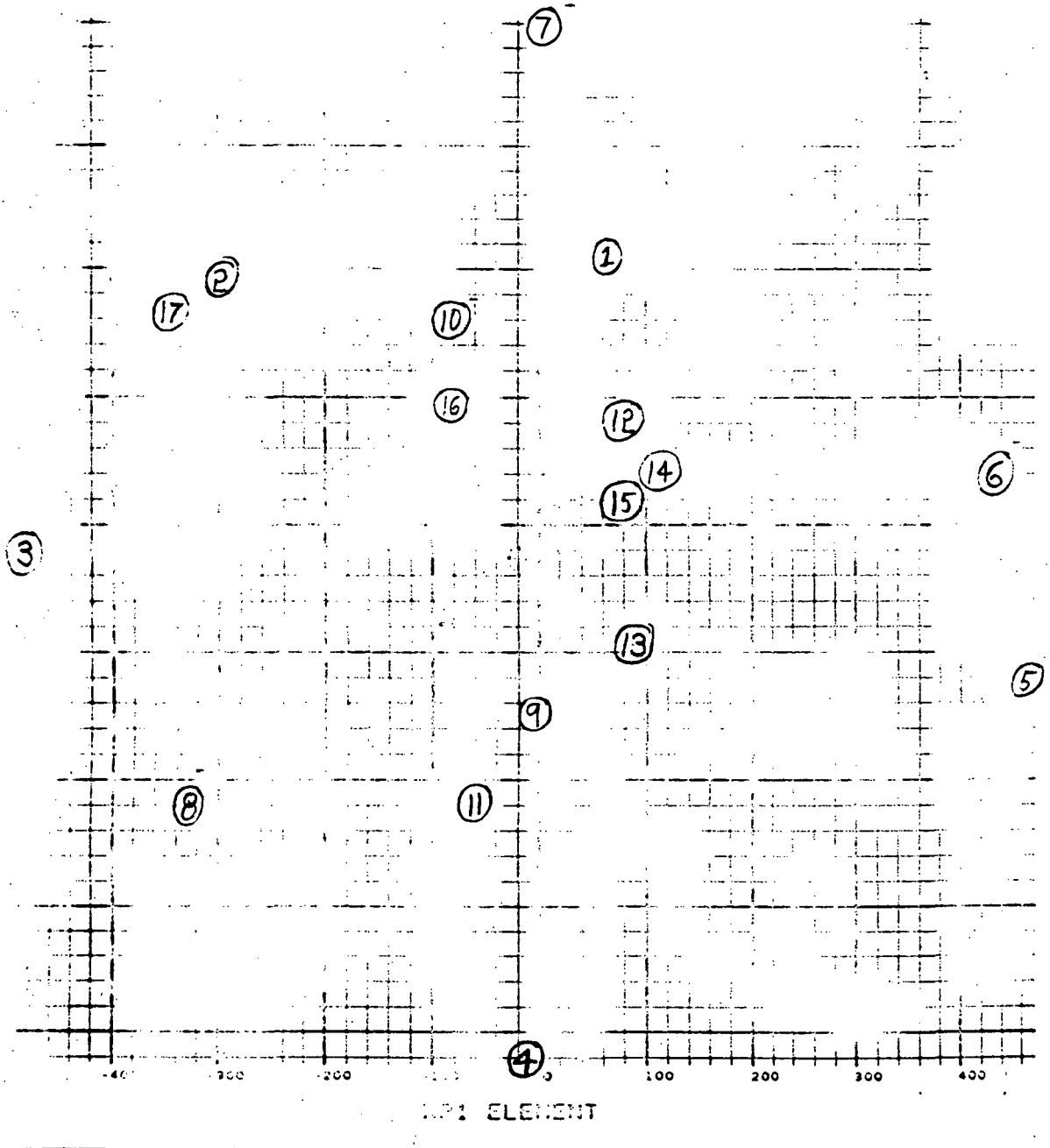


FIGURE 5.23 SCATTER PLOT DOUBLE CYCLE BASE



19

-772

REFERENCES

1. H. E. Hunter & N. H. Kemp, "Final Report for Application of Avco Data Analysis and Prediction Techniques (ADAPT) to Prediction of Cyclone Central Pressure and Its Derivatives Using NIMBUS HRIR Data", AVSD-0362-70-RR, Avco Systems Division, August 1970.

2. H. E. Hunter & N. H. Kemp, "Final Report for Application of Avco Data Analysis and Prediction Techniques (ADAPT) to Prediction of Cyclone Central Pressure and Its Derivatives Using NIMBUS HRIR Data Part II", AVSD-0142-71-RR, Avco Systems Division, March 1971.

3. H. E. Hunter & N. H. Kemp, "Final Report for Application of Avco Data Analysis and Prediction Techniques (ADAPT) to Prediction of Cyclone Present Motion and 12 Hour Motion, and Re-Centering Effects Using NIMBUS HRIR Data", AVSD-0334-71-RR, Avco Systems Division, July 1971.

4. N. H. Kemp, H. E. Hunter, H. E. Hunter, & R. A. Amato," Final Report for Application of Avco Data Analysis and Prediction Techniques (ADAPT) to a Multi-Spectral Extra Tropical Cyclone Accuracy Investigation Using NIMBUS HRIR Data", AVSD-0128-72-CR, Avco Systems Division, March 1972.

5. H. E. Hunter & J. Conway, "Final Report Demonstration of Feasibility of Using the Avco Data Analysis and Prediftion Techniques (ADAPT) to Develop Algorithms for Automating the Identification of Solar Bursts," AVSD-0255-71-RR, Avco Systems Division, May 1971.

6. H. E. Hunter, L. M. Meixsell, R. A. Amato, "Final Letter Report, ADAPT Solar Burst Compacting Study", AVSD-0209-72-CR, Avco Systems Division, May 1972.

7. N. H. Kemp, H. E. Hunter & R. A. Amato, "Final Report for Application of Avco Data Analysis and Prediction Techniques (ADAPT) to a Gauss-in-Gauss Detection Study," AVSD-0260-72-RR, Avco Systems Division, August 1972.

8. A. G. McNish & J. V. Lincoln, "Prediction of Sunspot Numbers", Trans. Am. Geophys. Union. 30, No. 5, p. 673-685, (1949).

9. D. K. Weidner, "Space Environment Criteria Guidelines for Use in Space Vehicle Development (1969 Revision)", 2nd Edition, NASA Technical Memorandum, TMX-53457, G. C. Marshall Space Flight Center, Aug. 1970.

10. R. E. Smith, "Space and Planetary Environment Criteria Guidelines for Use in Space Vehicle Development, 1971 Revision", NASA Tech. Memo., TMX-64627, G. C. Marshall Space Flight Center, Nov. 1971.

11. R. E. Smith, "Solar Activity Indices Memo," S & E-Aero-YS-14-72, G. C. Marshall Space Flight Center, March 15, 1972.

12. R. E. Smith, "Solar Activity Indices Memo", S & E-Aero-YS-24-72, G. C. Marshall Space Flight Center, June 28, 1972.

13. H. P. Sleeper, Jr., "Planetary Resonances, Bi-Stable Oscillation Modes and Solar Activity Cycles", TR-241-1053, Northrop Servicos, Inc. Feb. 1972. (NASA CR 2035)

14. D. J. Schove, "The Sunspot Cycle, 649 B. C. to A. D. 2000," J. GR, Vol. 60, No. 2, p 127-146, June 1955.

APPENDIX A

FEATURES OF ADAPT ANALYSIS

The unique aspect of the ADAPT approach to empirical data analysis is preceding the analysis with the derivation of the optimal representation for the particular data set. The ADAPT programs provide a unique capability for determining this optimum representation for large data sets. However, regardless of the size of the data set, the availability of this optimum representation provides many significant benefits to any further empirical analysis. These benefits include: 1) definition of which variables dominate the variation, 2) ordering of the data by its general usefulness for extracting information, 3) reduction in the computation required to perform further analysis, 4) reduction in the amount of learning data required to perform any given analysis, 5) an improved ability to establish performance and validity criteria, and 6) the ability to perform special functions such as clutter subtraction and extrapolation.

The availability of the optimum functions for representing any given data set is analogous to having the governing differential equations for a classical physics problem. These optimum functions provide information regarding the nature of the physics which govern the phenomena associated with this data. In particular, these functions will define exactly where the greatest and most highly correlated variation from case to case occurs. This information can be extremely useful in selecting data to be used for the analysis and in understanding the mechanism governing the phenomena which produced this data.

In addition to simply having the optimum functions for representing the data, these functions are ordered such that each function explains successively less variation in the data. This provides the user with a capability to reject variables in an intelligent rather than a random manner, if the resources or available learning data require the use of fewer dimensions than would naturally be used to describe the data. This ordering allows one to throw away those variables which explain the smallest amount of variation and therefore in general should be least useful to any analysis. Although it might be more desirable to be selective based on the particular analysis to be performed, this is not usually possible until after the analysis has been performed, when it is obviously no longer useful. Thus, it is almost axiomatic that the apirori rejection of data for a particular analysis cannot be based on that particular analysis, so the rejection based on explained variation is an attractive approach to eliminating data when realities of the resources or available learning cases makes such an elimination necessary.

Regardless of any prior decision to reduce the dimensionality, the ADAPT approach to any real problem will automatically lead to a significant reduction

in dimensionality. When the information energy curves which are produced by the ADAPT programs are examined, it is almost always possible for the analyst to select some dimensionality after which it is inconceivable that further useful information is incorporated in the data. This criteria alone usually results in a reduction of dimensionality of more than an order of magnitude.

A reduced dimensionality obviously allows one to perform computations with smaller computer capabilities. Furthermore, the orthogonality of the optimum representation also provides simplifications in the computation. For example, in the optimal ADAPT space one can in some cases derive the Fisher discriminant without inverting the covariance matrix. This combination of reduction in quantity of computation required and simplification due to orthogonality also makes it feasible to update classification and regression algorithms in real time for cases where this might otherwise be impossible.

A more significant aspect of the lower dimensionality of the learning space follows from the requirement that the amount of learning data be large compared to the dimensionalty of the learning data. This requirement arises from the situation analogous to fitting a third order polynomial through a series of points. If the third order polynomial is to be fitted to three points, it will always fit perfectly and no physical relationship need be involved. However, if the third order polynomial is to fit a hundred points well then one knows that this third order polynomial must be related to the data in some physical manner. The same is true for empirical analysis in general. If the number of dimensions of the learning space is equal to the number of learning cases one can expect most empirical algorithms to provide a perfect fit to the learning data. However, this fit is normally based on differences between the population and the sample statistics and is not based on the physics of the problem. Experience has shown that the number of learning cases required to derive an empirical algorithm varies from 2 to 6 times the number of dimensions of the learning space. Thus, the usual ADAPT reduction of an order of magnitude or more in dimensionality of the learning space translates immediately into an equivalent reduction in the requirement for learning data. Since obtaining learning data is one of the most expensive aspects of empirical data analysis, this attribute of the ADAPT approach is often sufficient by itself to make the difference between feasibility and infeasibility of solving a given problem.

The ADAPT representation also provides an opportunity for establishing a necessary, although not sufficient, validity criteria. Validity criteria provide a method of determining whether a particular test case is from the same population as the learning data, and therefore determine the validity of applying the algorithms derives on the learning data to that particular test case. The ADAPT validity criteria consists of comparing the length of the test data vector in the original data space and in the ADAPT optimum space. If this transformation from the original data space to the optimum space results in a shortening

of the test data vector by a factor considerably greater than the shortening which the learning data vectors suffered, one has an indication that the test data and learning data are from different populations. In addition to providing this validity criteria, the ADAPT programs have been designed to calculate performance criteria as part of the learning process. These performance criteria provide the analyst with a basis for immediately evaluating how well he can expect a given algorithm to perform on test data. The ADAPT programs provide the analyst with both the performance criteria and the experience factor required to determine whether the algorithm derived is overdetermined. If the algorithm is overdetermined, the analyst must adjust the dimensionality of the problem or increase the quantity of learning data to derive a physically meaningful algorithm.

The ADAPT approach of obtaining the optimum representation of the data prior to performing the analysis introduces the capability to perform clutter subtraction on the data prior to performing the analysis. The clutter subtraction can be used to eliminate any characterizable aspect of the signature from the data histories. This is accomplished by subtracting the coordinate directions corresponding to those characteristics to be eliminated from the space prior to the optimization procedure. Another unique capability resulting from the optimum representation step is the ability to do an extrapolation making use of both historical data from previous data histories and the available portion of any given data history. Conceptually this is equivalent to utilizing historical information to guide the interpolation over missing data points.

In addition to these advantages which accrue from the optimal representation, the ADAPT programs have been operational since approximately 1965. They have been applied to a great many different problems, and during this period part of the practical pit falls associated with empirical analysis have been encountered, overcome and the programs improved to take advantage of this experience. This experience has also provided Avco with the understanding of what diagnostic outputs are required to enhance the ability of the analyst to develop the required algorithms, and to provide the data necessary to reintroduce the physics to the problem at as many points as possible. The key areas where the physics may be reconsidered as part of the analysis are: 1) at the time of data selection and preprocessing decisions; 2) after the development of the optimum representation,it may be examined to insure that the variation is consistent with the expected variation based on the physics of the problem; 3) after the development of the algorithm, the relative importance vector may be examined to determine if the variables which appear important to the decision are consistent with the analyst's understanding of the physics and the relative importance spectrum may be examined to determined if the difficulty in obtaining the algorithm is consistent with the difficulty which would be expected based on the physics of the problem.

In summary, the capability to find the optimum representation for large data vectors has been combined with many years of experience in using this representation as a preliminary step preceding empirical data analysis. This unique

combination has been used to prepare a set of computer programs for performing empirical data analysis. These programs provide the user with a fast and economical way to generate simple empirical algorithms for classification, regression, clustering and extrapolation and/or analysis of any given set of learning data.

APPENDIX B

OPTIMAL ORTHOGONAL EXPANSION FOR TWO FUNCTIONS

We wish to carry though the ADAPT expansion of each of two given functions in the series of the optimal orthogonal functions defined by these two functions, as described in the Introduction.

Suppose we are given the functions $u_1(t)$ and $u_2(t)$ of the independent variable t, over some domain $t_1 \leq t \leq t_2$. Let the functions be normalized, so that

$$\int u_1^2 dt = \int u_2^2 dt = 1$$

Then the only parameter is the product integral

$$r \equiv \int u_1 u_2 dt, \quad |r| \leq 1$$

the last inequality being Schwarz' inequality for normalized functions.

First we construct an orthonormal set of 2 functions v_1, v_2 from the given ones by the Gram-Schmidt procedure. These functions are easily seen to be

$$v_1 = u_1, \quad v_2 = (u_2 - r u_1)/\sqrt{1-r^2}$$

We now find the expansion coefficients of u_1, u_2 in a series of v_1, v_2:

$$u_i = x_{i1} v_1 + x_{i2} v_2, \quad x_{ij} = \int u_i v_j dt$$

$$x_{11} = 1, \quad x_{12} = 0, \quad x_{21} = r, \quad x_{22} = \sqrt{1-r^2}$$

The optimal orthogonal functions are now obtained by finding the eigenvalues and eigenvectors $\underline{d}$ of the two-by-two matrix

$$S = \frac{1}{2}\left[x_{1i}x_{1j} + x_{2i}x_{2j}\right]$$

(the factor in front corresponds to weighing by dividing by the number of functions, in our case 2.) They are easily found to be

$$\lambda_1 = \frac{1}{2}(1+|c|) \quad , \quad \lambda_2 = \frac{1}{2}(1-|c|)$$

$$\underline{d}_1 = (\sqrt{\lambda_1}, \sqrt{\lambda_2}) \quad , \quad \underline{d}_2 = (\sqrt{\lambda_2}, -\sqrt{\lambda_1})$$

The eigenvectors are the expansion coefficients of the optimal orthogonal functions h_1, h_2 in a series in v_1, v_2, i.e.,

$$h_i = d_{i1}v_1 + d_{i2}v_2 \quad , \quad \underline{d}_i = (d_{i1}, d_{i2})$$

Returning to the original u functions we find the associated optimal functions to be

$$h_1 = \frac{1/2}{\sqrt{\lambda_1}}\left(u_1 + \frac{c}{|c|}u_2\right) \quad , \quad h_2 = \frac{1/2}{\sqrt{\lambda_2}}\left(u_1 - \frac{c}{|c|}u_2\right)$$

and the expansions of the u functions in them are

$$u_1 = \sqrt{\lambda_1}\,h_1 + \sqrt{\lambda_2}\,h_2 \quad , \quad u_2 = \frac{c}{|c|}\left(\sqrt{\lambda_1}\,h_1 - \sqrt{\lambda_2}\,h_2\right)$$

It is sufficient to discuss the case of $\rho \geq 0$ because if $\rho < 0$, a change in the sign of u_2 returns to the first case. We note that the optimal function h_1 is proportional to the average of the input functions. The average is intuitively the best single function to represent two functions, so we see the best single function is associated with the larger eigenvalue λ_1. The optimal function associated with λ_2 is proportional to the difference of the given functions.

We also note that

$$\lambda_1 + \lambda_2 = 1, \quad \lambda_1 - \lambda_2 = \rho = \int u_1 u_2 \, dt \leq 1$$

The decrease in the eigenvalue from the first to the second is the product integral of the two functions. If the functions are closely correlated one would expect ρ to be near unity, and λ_2 would be much less than λ_1. But if the functions are nearly uncorrelated one would expect ρ to be small, and there is only a slight decrease in the eigenvalue, going from the larger to the smaller. Thus the rate of decrease of eigenvalues can be associated with the degree of correlation of the input functions.

APPENDIX C

DESCRIPTION OF ADAPT HISTORY EXTRAPOLATION PROGRAM

This appendix contains a mathematical description of the method used in this project to extrapolate a complete sunspot cycle from a partial one and a set of optimal vectors derived by ADAPT analysis of past cycles.

Given T values of an input history vector which is normally N (N < T) values in length i.e.

$$U^S = U_i^S \qquad i = R, R+1, \ldots, R+T-1$$

and a set of optimal data vectors NR in number, with each vector containing N values.

$$H = h_{il} \qquad \begin{matrix} i = 1, 2, \ldots, R, \ldots R+T-1, \ldots, N \\ l = 1, 2, \ldots, NR \end{matrix}$$

and an assumption that the data vector u of which u^S is a segment is well represented by these optimal functions, this program will calculate the entire vector $(u_i \quad i = 1, 2, \ldots, N)$ by estimating the coefficients of the vector u from the given segment of the history and the corresponding region of the optimal vectors. Two cases can be distinguished depending upon the value of

$$T/NR$$

and the mathematics for each is described below.

Case 1: $T/NR < 1$

Let $H^S = (h_{il})$ $\quad i = R, R+1, \ldots, R+T-1$

$\quad l = 1, \ldots, T$

Setup $U^S = H^S y^S$

which is T equations in T unknowns (Y^S)

Now this can be solved exactly for Y^S

$$Y^S = (H^S)^{-1} U^S$$

and the history can be estimated by

$$U_t' = \sum_{i=1}^{T} h_{ti} Y_i^S \qquad t = 1, 2, \ldots, N$$

In this case the T points of the estimated history U_t' will equal that data points of the segment U^S.

Case 2: $T/NR > 1$

Let $H^P = (\quad)$ $\quad = R, R+1, \ldots, R+T-1$

$\quad = 1, \ldots, NR$

Setup $U^S = H^P Y^P$

which is T equations in NR unknowns (Y^R)

Now this cannot be solved exactly. One method of solution is by least squares

$$Y^P = (H^{PT} H^P)^{-1} H^{PT} U^S$$

where H^{PT} = transpose of H_P

Again we can reconstruct the history

$$U_t' = \sum_{i=1}^{NR} h_{ti} Y_i^P \qquad t = 1, 2, \ldots, N$$

In this case the T points of the estimated history U^E which correspond to the T points of the segment U^S are not exactly equal, but rather differ by an amount which has been minimized by the least square technique.

For the case $T/_{NR} = 1$, both methods are exactly identical and therefore, either can be used.

Also attached to this appendix is a copy of the FORTRAN listing of the program which performs the analysis just described.

```
      COMPILER OPTIONS - NAME=  MAIN,OPT=01,LINECNT=50,SOURCE,BCD,NOLIST,NODECK,LOAD,MAP,NOEDIT,ID,XREF
           C 5070A
           C MAIN
ISN 0002                IMPLICIT REAL * 8 (A-H , O-Z )
ISN 0003              COMMON /STORAGE/                    H(2000)
                     1 , DUM(2050), ALT(2050), ADJSV(2050), SV(2050), O(2050)
ISN 0004                N10000= 2000
ISN 0005                N2050= 2050
ISN 0006              CALL SUBMAN( N10000 , H
                     1 ,N2050, DUM, ALT, ADJSV, SV, O)
ISN 0007              CALL EXIT
ISN 0008              RETURN
ISN 0009              END
```

```
         COMPILER OPTIONS - NAME=  MAIN,OPT=01,LINECNT=50,SOURCE,BCD,NOLIST,NODECK,LOAD,MAP,NOEDIT,ID,XREF
ISN 0002              SUBROUTINE SUBMAN( N10000  , H
                     1 ,N2050, DUM, ALT, ADJSV, SV, U)
                    C  PROG. 5070
                    C HISTORY CONTINUATION PROGRAM JAY GOTTESFELD FOR KEMP, HUNTER
                    C H MATRIX TAPE ON N19,(15). INPUT TAPE.........
                    C   OUTPUT TAPE N9(9) HAS STANDARD VELOCITIES (1E20T INPUT TAPE)
                    C                              OUTPUT TAPE Y VECTORS ON N4(4).
ISN 0003              IMPLICIT REAL * 8(A-H, O-Z )
ISN 0004             DIMENSION   TABLE(300)  , NDCSTP(20)  , H(10000 )  , DUM(2050)
                    C MAKE  H =  42000  PRODUCTION.......
                         1 , ALT(2050), ADJSV(2050),  HS(70,70) , NTAP(20) , SV(2050),U(2050
                         2) , YS(70)
                         3 , XNAME(2)  , DATE12(2)
ISN 0005                 INDX(I,J)=  (I-1) * NN & J
ISN 0006                DATA  PIN, XNAME, PROGRAM /'14      ', '        ','        ',
                        1 'CONTINUE'/
ISN 0007                DATA  TENLOG / 2.3025850929940 6  /
ISN 0008                 CALL WHERE( TABLE )
ISN 0009                    CASE=0.D0
ISN 0010                   IL= 8888
ISN 0011                     CALL DATE( DATE12)
ISN 0012                  N19= 15
ISN 0013                  CASE=1.D0
ISN 0014                  HOL=1.D0
ISN 0015                 CALL SETUP(     'N19      '  , 4  ,  N19          )
ISN 0016                 CALL SETUP(     'MR       '  , 4  ,  MR           )
ISN 0017                 CALL SETUP(     'NT       '  , 4  ,  NT           )
ISN 0018                 CALL SETUP(     'U        '  , 8  ,  U ,2050      )
ISN 0019                 CALL SETUP(     'OPNRM2   '  , 8  ,  OPNRM2       )
ISN 0020                 CALL SETUP(     'SIGADJ   '  , 8  ,  SIGADJ       )
ISN 0021                 CALL SETUP(     'OPLOG    '  , 8  ,  OPLOG        )
ISN 0022                 CALL SETUP(     'HPRINT   '  , 8  ,  HPRINT       )
ISN 0023                 CALL SETUP(     'OPNORM   '  , 8  ,  OPNORM       )
ISN 0024                 CALL SETUP(     'AVGMLT   '  , 8  ,  AVGMLT       )
ISN 0025                  CALL SETUP('OPLOG2  ', 8, OPLOG2 )
ISN 0026                  CALL SETUP('AVGML2  ',8, AVGML2   )
ISN 0027                  CALL SETUP('TAPEOP  ',8, TAPEOP  )
ISN 0028                  CALL SETUP('N9      ',4,   N9)
ISN 0029                  CALL SETUP('N4      ',4,   N4)
ISN 0030                    CALL SETUP('NOR     ', 4, NOR)
ISN 0031                   CALL SETUP('CASE    ', 8, CASE)
ISN 0032                   CALL SETUP( 'HOL    ', 8, HOL)
ISN 0033                    ONE=1.D0
ISN 0034                    ZERO= 0.D0
ISN 0035                   OPNRM2 = 0.D0
ISN 0036                   SIGADJ = 0.D0
```

```
ISN 0037              OPLOG  = 0.D0
ISN 0038             HPRINT  = 0.D0
ISN 0039              OPNORM = 0.D0
ISN 0040               AVGMLT=1.D 0
ISN 0041                MDR=100
ISN 0042              NS=9
ISN 0043                  NT=4
ISN 0044               TAPEIT = 1.D0
ISN 0045               OPLOG2=  0.D0
ISN 0046              AVGML2= 0.D0
ISN 0047                  NCOUNT= 1
ISN 0048              CALL IFERMV( XNAME, BIN, PRGRAM)
ISN 0049              NO=0
ISN 0050               N1= 1
ISN 0051                N2=2
ISN 0052                N12=N2
ISN 0053       1     CONTINUE
ISN 0054              CALL FRAMIN(IL ,$  2  )
ISN 0055                 WRITE(6,1009)  NCOUNT, HOL
ISN 0056       1009       FORMAT('0  THIS IS U VECTOR NUMBER'  ,I4,' BEING PROCESSED.
                     1 HOL NO',1F12.4
                     2           ,' IS GOING OUT ON TAPE 9'         ////)
ISN 0057                 REWIND N19
ISN 0058                 READ( N19)  NN  , KR , M1 , M2
ISN 0059                    NR= MIN0(KR, NOP, 70)
ISN 0060                  MP1 =  NP +1
ISN 0061                 NNNR =  NN * NR
ISN 0062                 WRITE(6,1001) N19,KR,NN,NR ,NT , NNNR , N10000
ISN 0063       1001  FORMAT(    '     WE ARE READING H TAPE ON UNIT ',I2,  ' IT HAS ',
                     1 I3, ' COLUMN VECTORS EACH WITH ' , I4,  ' ELEMENTS, WE ARE USING
                     2', I3,' VECTORS.....' /10X,' NT='
                     2, I3,'  NOTE IF NT GREATER THAN NR,SOLUTION IS LEAST SQUARE....'/
                     3' H MATRIX =',I5,  ' WHICH IF GREATER THAN',I5,' WE STOP...'///)
ISN 0064                 IF( NN .LT. N2050) GO TO 47
ISN 0065                    WRITE(6, 1010) NN, N2050
ISN 0066       1010         FORMAT(///'   NUMBER OF ELEMENTS IN H MATRIX IS ', I7,' WHI
                     1CH IS GREATER THAN ', I4,' SO WE STOP.....'////20X,' USE LARGER C
                     2ORE.....' )
ISN 0067                 GO TO 1
ISN 0068       47         CONTINUE
ISN 0069                 IF( N10000 .LT. NNNR ) GO TO 1
ISN 0070                 MRT1 = MR  & NT -1
ISN 0072                DO  26  I = 1, KR
ISN 0073                 READ( N19) (COLUM(J) ,J = 1, NN )
ISN 0074              IF( MERMIT .EQ.  1.)WRITE(6,1002)I  , (COLUM(J),J=1,NN )
ISN 0075       1002  FORMAT(    '  H-COLUMN VECTOR ',I3 / (1PRE16.7) )
ISN 0077                    IF( I.GT. NR) GO TO 26
ISN 0078                DO  3   J = 1,  NN
ISN 0079
```

```
ISN 0081        3     H( INDX(I,J))=  DUM(J)
ISN 0082              IF( NT .GT.  NR )  GO TO 26
ISN 0084               IF( I.GT. NT) GO TO 26
ISN 0085              DO  7   J = MR , MRT1
ISN 0087        7     HS( J&1-MR  ,I) =  DUM(J)
ISN 0088       26      CONTINUE
ISN 0089              READ(N1S) (ALT(I) , I= 1,NN )
ISN 0090              READ(N1S) ( ADJSV(I) , I=1,NN )
ISN 0091              REWIND N1S
ISN 0092               IF( NT  .GT. NR )  GO TO 8
ISN 0094                 WRITE(6,1004)((HS(I,J),J=1,NT ), I=1,NT )
ISN 0095             CALL MATXIN( HS , NT , IERR )
ISN 0096                 WRITE(6,1004)((HS(I,J),J=1,NT ), I=1,NT )
ISN 0097               IF( IERR  .LT.  2  ) GO  TO  4
ISN 0099       11     WRITE(6 , 1000 )  IERR
ISN 0100       1000   FORMAT(  '     NO INVERSE FOR HS MATRIX. ERR NO', I2 / )
ISN 0101               GO TO 1
ISN 0102        8    CONTINUE
ISN 0103               DO  10  I = 1, NR
ISN 0104               DO  10  K = 1, NR
ISN 0105               HS(I,K) =  0.D0
ISN 0106              DO  10   J = 1,  NT
ISN 0107                 JJ = MR1 &J
ISN 0108       10     HS(I,K)  =  HS(I,K)  &  H(INDX(I,JJ))*  H( INDX( K,JJ) )
ISN 0109                 WRITE(6,1004)((HS(I,J),J=1,NR ), I=1,NR )
ISN 0110       1004     FORMAT(//'0  HS MATRIX'/ ( 1P10D13.4))
ISN 0111              CALL  MATXIN( HS ,  NR , IERR  )
ISN 0112                 WRITE(6,1004)(( HS(I,J) , J= 1,NR) ,I=1, NR )
ISN 0113               IF( IERR .GE. 2 ) GO TO  11
ISN 0115        4    CONTINUE
ISN 0116               DO  6  I =  1,  NT
ISN 0117        6       SV(I) =  U(I)
ISN 0118              ERR=0.D0
ISN 0119               IF( OPNRM2 .NE. 0. )
                  1    CALL NORM2( SV , NT , OPNRM2 , ERR )
ISN 0121               IF( ERR .NE.  0. )              GO TO1
ISN 0123              CALL LOGARI( OPLOG , SV , NT )
ISN 0124              CALL NORM2( SV , NT , OPNORM , SV )
ISN 0125               DO  12  I = 1, NT
ISN 0126       12      SV(I) = SV(I) - AVGMLT *  ADJSV(MR1 &I)
ISN 0127                IF(  NT .GT. NR ) GO TO 13
ISN 0129                DO  15  I = 1,  NR
ISP 0130                        YS(I)= 0.D0
ISN 0131                  IF(I .GT. NT ) GO TO 15
ISN 0133                DO  14  J = 1, NT
ISN 0134       14        YS(I)= YS(I) &  HS(I,J) * SV(J)
ISN 0135       15      CONTINUE
ISN 0136                GO  TO  16
```

```
ISN 0137        13   CONTINUE
ISN 0138                DO  17  I = 1, NR
ISN 0139                   YS(I)= 0.D0
ISN 0140                DO  17  J = 1, NT
ISN 0141               CON          =  0.D0
ISN 0142                DO  16     K = 1, NR
ISN 0143       16         CON    = CON    &  HS(I,K)  *   H( INDX( K, J&MP1 ) )
ISN 0144       17         YS(I) = YS(I)  &    CON   * SV(J)
ISN 0145       16    CONTINUE
ISN 0146                 DO  21  J = 1, NN
ISN 0147                  DUM(J) =  0.D0
ISN 0148                 DO  21  K =  1, NR
               C  DUM IS UI-EO
ISN 0149       21      DUM(J) =  DUM(J)   &  H(  INDX(K,J) )  *   YS(K)
ISN 0150               NR1= NR - 1
ISN 0151                DO 31 I = 1,NN
ISN 0152                  IF(AVG-L2 .EQ. 1.) GO TO 42
ISN 0154                TEMP = AVGMLT* ADJSV(I)
ISN 0155               DUM(I)= DUM(I) &TEMP
ISN 0156       42              IF( DPLOG2.EQ. 1.) GO TO 31
ISN 0158                IF(DUM(I) .EQ. 0.) GO TO 31
ISN 0160               DUM(I)= DEXP( TENLOG *DUM(I)          )
ISN 0161       31        CONTINUE
ISN 0162                WRITE(6,1005)
ISN 0163        1005   FORMAT( '1          NO          U-EO                U
                    1AVG. INPUT VECTOR'  /                    )
ISN 0164              IF( NR .GT. 1) GO TO 35
ISN 0166              WRITE(6,1007) (I, DUM(I), U(I), ADJSV(I), I=1,NT)
ISN 0167        1007    FORMAT( 1X, I9, 2X, 1P3D18.6) )
ISN 0168                IF( NT  .EQ. NN) GO TO 36
ISN 0170               NT1= NT+1
ISN 0171                  WRITE(6,1008) ( I, DUM(I), ADJSV(I), I=NT1, NN)
ISN 0172                GO TO 36
ISN 0173        1008    FORMAT(   1X,I9,2X, 1P1D18.6, 18X, 1P1D18.6 )
ISN 0174        35      CONTINUE
ISN 0175                WRITE(6,1008) ( I, DUM(I),  ADJSV(I), I=1, MR1)
ISN 0176                WRITE(6,1007)(I, DUM(I), U(I-MR1), ADJSV(I), I=MR, MRT1)
ISN 0177                IF( MRT1 .GE. NN) GO TO 36
ISN 0179                   MRTP= MR+NT
ISN 0180               WRITE(6,1008) (I, DUM(I), ADJSV(I), I=   MRTP, NN)
ISN 0181        36      CONTINUE
ISN 0182                WRITE(6,1006)( I, YS(I)          , I=1, NR )
ISN 0183        1006  FORMAT(    '1          NO          Y-EO   '/
                    1 (1X,I9,2X 1P1D18.6 ) )
ISN 0184                DO  30  I=1,NN
ISN 0185        30         ALT(I)= I
                   CALL EZPLOT(S,D0, N1, N1, ALT, DUM, NN, N1, N2, N42,
                    1 22,       '7EROTH ORDER ESTIMATE ' , 19,   ' INDEXING VARIABLE ',
```

```
                2 6,        ' U-EO ', N1, N1, XL, XUP, N1, YL, YUP, N1, NO, N1)
ISN 0187            CALL PRINTV( 12 , DATE12 , 800 , 900)
ISN 0188            CALL LABLV ( CASE , 970, 30, 4, 2, 4 )
ISN 0189              CALL PRINTV(4, 'HOL=', 800 , 930)
ISN 0190              CALL DECLV( HOL, 835, 930, 7, 2, 4)
ISN 0191            IF(TAPEOP .EQ. 0.) GO TO 43
ISN 0193              WRITE(N6) ZERO, ZERO, HOL , NN
ISN 0194            CALL WRITE4(N6, ALT, DUM, NN)
ISN 0195             HOL=HOL&1.E0
ISN 0196             IF( NCOUNT .GT. 2) GO TO 45
ISN 0198             WRITE(N4) (YS(I),I=1, NP),( ONE,(ZERO, J=1, NF),I=1,NR1),ONE
ISN 0199              GO TO 43
ISN 0200     ! 45     WRITE(N4) (YS(I),I=1, NF)
ISN 0201     ! 43      CONTINUE
ISN 0202                NCOUNT=1& NCOUNT
ISN 0203                 GO TO 1
ISN 0204        2      CONTINUE
ISN 0205            CALL RETNF
ISN 0206              IF( TAPEOP .EQ. 0.) GOTO 44
ISN 0208            ENDFILE N6
ISN 0209            ENDFILE N4
ISN 0210        44    CONTINUE
ISN 0211              RETURN
ISN 0212               END
```

C-9

LEVEL 17 (1 NOV 68) OS/360 FORTRAN H DATE 72.105/14.02.14

```
        COMPILER OPTIONS - NAME=  MAIN,OPT=01,LINECNT=50,SOURCE,EBCD,NOLIST,NODECK,LOAD,MAP,NOEDIT,ID,XREF
ISN 0002          SUBROUTINE LOGADJ ( OPLOG,  SV , NON )
ISN 0003          IMPLICIT REAL * 8 (A-H, O- Z)
ISN 0004          DIMENSION  SV(2000)
ISN 0005          DATA  TENLOG / 2.3025850929940&6  /
ISN 0006          IF( OPLOG ) 2, 3, 2
ISN 0007        2 DO  1 I=1, NON
ISN 0008        1 SV(I)= DLOG( SV(I) ) / TENLOG
ISN 0009        3 RETURN
ISN 0010          END
```

```
LEVEL 17 ( 1 NOV 68)                    OS/360  FORTRAN H                    DATE  72.105/14.02.25

        COMPILER OPTIONS - NAME=  MAIN,OPT=01,LINECNT=50,SOURCE,ECD,NOLIST,NODECK,LOAD,MAP,NOEDIT,ID,XREF
ISN 0002              SUBROUTINE NORM2( SV  , NN , OPNRM2 , ERR)
ISN 0003              IMPLICIT REAL * 8 (A-H , O -Z )
ISN 0004                 DIMENSION  SV(1000)
ISN 0005                 IF( OPNRM2 .EQ. 0.D0 ) RETURN
ISN 0007              IF( SV(1) .NE. 0.D0 )     GO TO 2
ISN 0009                 ERR=1.D0
ISN 0010                                    WRITE(6,1000)
ISN 0011      1000 FORMAT(  '1OPNRM2  = 1. AND FIRST ELEMENT OF THE CASE ON TAPE IS Z
                  1ERO. THUS DIVISION WILL BLOW UP.....  ' / )
ISN 0012              RETURN
ISN 0013      2       CONTINUE
ISN 0014                 DO  1 I = 2, NN
ISN 0015        1        SV(I) = SV(I) / SV(1)
ISN 0016                 SV(1) = 1.D0
ISN 0017                 RETURN
ISN 0018                 END
```

C110

```
LEVEL 17 ( 1 NOV 68)                    OS/360  FORTRAN H                    DATE  72.105/14.02.32

        COMPILER OPTIONS - NAME=  MAIN,OPT=01,LINECNT=50,SOURCE,ECD,NOLIST,NODECK,LOAD,MAP,NOEDIT,ID,XREF
ISN 0002              SUBROUTINE NORM(V,N,OPNORM,ANS)
ISN 0003                 IMPLICIT REAL*8 (A-H,O-Z)
ISN 0004              DIMENSION V(100),ANS(100)                          NOR40010
            CNORM                                                        NOR70010
            C  NORMALIZE VECTOR                                          NOR70020
ISN 0005              IF (OPNORM)3,4,3                                   NOR70030
ISN 0006           3  S=0.D0                                             NOR70040
ISN 0007              DO 1 I=1,N                                         NOR70050
ISN 0008           1  S=S+V(I)*V(I)                                      NOR70060
ISN 0009              S=DSQRT(S)
ISN 0010              DO 2 I=1,N                                         NOR70070
ISN 0011           2  ANS(I)=V(I)/S                                      NOR70080
ISN 0012           4  RETURN                                            NOR70100
ISN 0013              END                                                NOR70110
```

```
      COMPILER OPTIONS - NAME=  MAIN,OPT=01,LINECNT=50,SOURCE,EBCD,NOLIST,NODECK,LOAD,MAP,NOEDIT,ID,XREF
ISN 0002            SUBROUTINE MATXIN(A1,N7,INDEX)                                  AT20010
ISN 0003              IMPLICIT REAL*8 (A-H,O-Z)
ISN 0004              COMMON / JOINT / N1000 , NLUM, N12000 , N30
ISN 0005            DIMENSION EPS(100),KOL(100)                                     MAT40010
ISN 0006            DIMENSION A1( 70, 70)
                  CMATXIN  JAY GOTTESFELD                                           MAT70010
ISN 0007            N=N7                                                            MAT70020
ISN 0008            EPZ=1.000                                                       MAT70030
ISN 0009            DET=EPZ                                                         MAT70040
ISN 0010            KOL(1)=1                                                        MAT70050
ISN 0011          100 DO  1 I=2,N                                                   MAT70060
ISN 0012            1 KOL(I)=KOL(I-1)&1                                             MAT70070
ISN 0013          101 DO 12 K=1,N                                                   MAT70080
ISN 0014            L=N-K&1                                                         MAT70090
ISN 0015            M=KOL(1)                                                        MAT70100
ISN 0016            J=1                                                             MAT70110
ISN 0017          200 IF (N-K)17,6,2                                                MAT70120
ISN 0018            2 CONTINUE                                                      MAT70130
ISN 0019              IF (A1(I,1))300,301,301                                       MAT70140
ISN 0020          300 AMPY=-A1(I,1)                                                 MAT70150
ISN 0021              GO TO 302                                                     MAT70160
ISN 0022          301 AMPY=A1(I,1)                                                  MAT70170
ISN 0023          302 CONTINUE                                                      MAT70180
ISN 0024          102 DO 4 I=2,L                                                    MAT70190
ISN 0025              IF (A1(I,1))307,308,308                                       MAT70200
ISN 0026          307 ABS=-A1(I,1)                                                  MAT70210
ISN 0027              GO TO 201                                                     MAT70220
ISN 0028          308 ABS=A1(I,1)                                                   MAT70230
ISN 0029          201 IF (AMPY-ABS)3,310,310                                        MAT70240
ISN 0030            3 CONTINUE                                                      MAT70250
ISN 0031              J=1                                                           MAT70260
ISN 0032              IF (A1(I,1))304,305,305                                       MAT70270
ISN 0033          304 AMPY=-A1(I,1)                                                 MAT70280
ISN 0034              GO TO 306                                                     MAT70290
ISN 0035          305 AMPY=A1(I,1)                                                  MAT70300
ISN 0036          306 CONTINUE                                                      MAT70310
ISN 0037              M=KOL(I)                                                      MAT70320
ISN 0038          310 CONTINUE                                                      MAT70330
ISN 0039            4 CONTINUE                                                      MAT70340
ISN 0040          202 IF (KOL(1)-M)5,6,5                                            MAT70350
ISN 0041        5     CONTINUE                                                      MAT70360
ISN 0042              DET=-DET                                                      MAT70370
ISN 0043              KOL(J)=KOL(1)                                                 MAT70380
ISN 0044              KOL(1)=M                                                      MAT70390
ISN 0045        6     CONTINUE                                                      MAT70400
ISN 0046              IF (A1(J,1))7,19,7                                            MAT70410
```

```
ISN 0047          7 AMPY=A1(J,1)                                                MAT70420
ISN 0048            DET=DET*AMPY                                                MAT70430
ISN 0049        103 DO 8 I=2,N                                                  MAT70440
ISN 0050            ROW(I-1)=A1(J,I)/AMPY                                       MAT70450
ISN 0051            A1(J,I-1)=A1(1,I-1)                                         MAT70460
ISN 0052          8 CONTINUE                                                    MAT70470
ISN 0053            ROW(N)=DP1/AMPY                                             MAT70480
ISN 0054            A1(J,N)=A1(1,N)                                             MAT70490
ISN 0055        104 DO 10 I=2,N                                                 MAT70500
ISN 0056            AMPY=A1(I,1)                                                MAT70510
ISN 0057        105 DO 9 J=2,N                                                  MAT70520
ISN 0058            A1(I-1,J-1)=A1(I,J)-AMPY*ROW(J-1)                           MAT70530
ISN 0059          9 CONTINUE                                                    MAT70540
ISN 0060            A1(I-1,N)=-AMPY*ROW(N)                                      MAT70550
ISN 0061         10 CONTINUE                                                    MAT70560
ISN 0062        106 DO 11 J=1,N                                                 MAT70570
ISN 0063            KOL(J)=KOL(J&1)                                             MAT70580
ISN 0064            A1(N,J)=ROW(J)                                              MAT70590
ISN 0065         11 CONTINUE                                                    MAT70600
ISN 0066         12 KOL(1)=M                                                    MAT70610
ISN 0067        107 DO 16 K=1,N                                                 MAT70620
ISN 0068        203 IF (KOL(K)-K)17,16,13                                       MAT70630
ISN 0069         13 DO 18 I=K,N                                                 MAT70640
ISN 0070        204 IF (KOL(I)-K)17,14,18                                       MAT70650
ISN 0071         14 DO 15 J=1,N                                                 MAT70660
ISN 0072            ROW(J)=A1(J,I)                                              MAT70670
ISN 0073            A1(J,I)=A1(J,K)                                             MAT70680
ISN 0074            A1(J,K)=ROW(J)                                              MAT70690
ISN 0075         15 CONTINUE                                                    MAT70700
ISN 0076            M=KOL(K)                                                    MAT70710
ISN 0077            KOL(K)=KOL(I)                                               MAT70720
ISN 0078            KOL(I)=M                                                    MAT70730
ISN 0079            GO TO 16                                                    MAT70740
ISN 0080         18 CONTINUE                                                    MAT70750
ISN 0081            INDEX=3                                                     MAT70760
ISN 0082            GO TO 21                                                    MAT70770
ISN 0083         16 CONTINUE                                                    MAT70780
ISN 0084         20 INDEX=1                                                     MAT70790
ISN 0085        21  CONTINUE                                                    MAT70800
ISN 0086        9999 RETURN                                                     MAT70810
ISN 0087         17 INDEX=5                                                     MAT70820
ISN 0088            GO TO 21                                                    MAT70830
ISN 0089         19 INDEX=2                                                     MAT70840
ISN 0090            GO TO 21                                                    MAT70850
ISN 0091            END                                                         MAT70860
```

C-12

LEVEL 17 (1 NOV 68) OS/360 FORTRAN H DATE 72.105/14.02.48

```
        COMPILER OPTIONS - NAME= MAIN,OPT=01,LINECNT=50,SOURCE,BCD,NOLIST,NODECK,LOAD,MAP,NOEDIT,ID,XREF
ISN 0002              SUBROUTINEWRITE4( INTAP, ZZ4, SV4, NCN)
                C  NO  4 BYTE WORDS....
                C  NEEDS NO VARIABLE DIMENSIONS AS FIRST SUBSCRIP IS ALWAYS 2.....
ISN 0003                    DIMENSION  ZZ4(2,1000), SV4(2,1000)
ISN 0004                      WRITE (INTAP) ( ZZ4(1,I), SV4(1,I), I=1, NCN)
ISN 0005                    RETURN
ISN 0006                      END
```

```
F8E-LEVEL LINKAGE EDITOR OPTIONS SPECIFIED MAP,LET,LIST
          VARIABLE OPTIONS USED -  SIZE=(99304,8192)-          DEFAULT OPTION(S) USED
IEW0191
IEW0191
```

MODULE MAP

CONTROL SECTION

NAME		ORIGIN	LENGTH
MAIN		00	130
SUBMAN		130	BD4C
LOGACC		BEF0	1EA
NCRR2		BFE0	20C
NCRR		C200	1CC
MATRIX		C3D0	A7B
WRITEA		CE4E	1E2
DATEX	*	CFB0	1F8
IHCLEXP	*	D198	270
IHCFEXIT	*	D408	1C
EZFLCT	*	D428	2AEC
IHCFCOMH	*	FF18	F31
IHCCOMH2	*	10F50	5A5
PLTWX	*	11398	A94
CECK1	*	11F30	124
IDFRMV	*	11F58	121E
LABLV	*	13178	688
BIGV	*	13800	BE
IHCLLOG	*	138C0	230
IHCLSORT	*	13AF0	142
IHCERRM	*	13C38	5AC
NAV	*	141E6	2A6

ENTRY

NAME	LOCATION	NAME	LOCATION	NAME	LOCATION	NAME	LOCATION
DATL	CFB0						
DEXP	D198						
EXIT	D408						
IBCOM=	FF18	FDICCS=	FFD4	INTSWTCH	10E36		
SEGBASE	11080						
OPENER	11398	PLTW1	113C0	PLTW2	11400	PLTW3	1156C
PLTW3A	11614	PLTW4	11682	PLTW5	117A8	PLTND	118F8
SERSAV	11E30	SERREV	11F52	FLAGSV	11F6F	LIST	11FAC
ZDDPTA	11EB0	DXXSYY	11EB4	XXXXDD	11EB8	YYYYDD	11EBC
AXDD	11EC0	AYDD	11EC4	BXDD	11EC8	BYDD	11ECC
DHIGHD	11ED0	DLOWDD	11ED8	MLDD	11EE0	MRDD	11EE4
MBDD	11EE8	MIDD	11EEC	WIDELD	11EF0	HIGHED	11EF4
DELTYY	11EF8	DXXDYY	11EFC	DDDSLD	11F00	HOLDDD	11F04
CAMVDD	11F08	YIDPDD	11F0C	YPFGDD	11F10	FRMCNT	11F1C
SCEL	11F20	XERRDD	11F30	YERRDD	11F34	VCHXDD	11F38
VCHYDD	11F3C	ISPACE	11F40	IPDW	11F44	NTBL	11F48
BITE2X	11F4C	BITE2Y	11F50				
DLOG10	138C0	DLOG	138DC				
DSORT	13AF0						
ERRMON	13C38	IHCERRE	13C50				

NAME		ORIGIN	LENGTH
NYV	*	14490	2A6
PRINTV	*	14736	240
LA000000	*	14978	150D
SCARX	*	15FEE	108
SCRT	*	15F90	1FA
IHCSLCG	*	16180	16A
BNBCDV	*	16340	3FE
CAMRAV	*	16740	132
CNSIZV	*	16878	110
EXDYV	*	16988	70E
ENDJOB	*	17098	306
EZGETD	*	173A0	11EC
IHCFFXPI	*	18590	141
LINEV	*	186D8	4C4
SMXYV	*	18B70	110
FRAMEV	*	18C80	17E
PLOTVI	*	18E00	132
POLPLT	*	18F38	A62
RITSTV	*	19FF0	622
SETMIV	*	1A018	174
SETXCV	*	1A190	1A4
SMALLV	*	1A338	8E
TABLIV	*	1A3F8	9C0
XAXISV	*	1ADC8	1BA
IHCFCVTH	*	1AF88	1175
IHCFFNTH	*	1C100	512
IHCFFIOS	*	1C618	111C
IHCUCPT	*	1D738	3C8
YAXISV	*	1DBC0	1BA
BRITEV	*	1DCC0	8E
CORE=	*	1DD80	A8
IHCFLXPI	*	1DE28	14D
IHCLTRCH	*	1DF78	28E
JBNAME	*	1E208	D0
EMPTY	*	1E2D8	210
DCBS	*	1E4E8	1C0
DCBDUMP	*	1E628	F4

NAME	LOCATION	NAME	LOCATION	NAME	LOCATION	NAME	LOCATION
READIN	14978	HEDING	15E7E	WHERE	15BBA	SETUP	15960
ZERDIN	15A94						
SCAN	15E88						
ALOGIO	16180	ALOG	1619C				
FFXPI=	18590						
RITE2V	19F3C						
ADCON=	1AF88	FCVAOUTP	1B032	FCVLOUTP	1B0C2	FCVZOUTP	1B212
FCVIOUTP	1B59E	FCVEOUTP	1BAA0	FCVCOUTP	1BCBA	INTRSWCH	1BFA3
ARITH=	1C100	ADJSWTCH	1C46C				
FIOCS=	1C618	FIOCSHFP	1C61E				
CORE	1DD80						
FDXPI=	1DE28						
IHCTRCH	1DF78	ERRTRA	1DF80				
MYSTAE	1E31C	CLSPLT	1E35F	FIXUP	1E3A2		
ECBADD	1E4E8	MIBUNI	1E590				

NAME		ORIGIN	LENGTH
OUT2	*	1E720	81F
BINHEX	*	1EF40	D0
WTO	*	1F010	C0
MIDCOM	*	1F0E0	334
LOADPT	*	1F4CE	F8
PUTV	*	1F5C0	11E
TYPSET	*	1F720	98
XMGRV	*	1F8B8	11A
YMGRV	*	1F9D8	CE
ERRLNV	*	1FAA8	21C
ERRMLV	*	1FCC8	232
HOLLEV	*	1FF00	FE
LINEV	*	200FE	7FA
MASKX	*	205F6	A2
MSXYV	*	206A0	12C
SELCGUF	*	207D0	B4
SHFTIV	*	20F88	5C
IHCSSCN	*	20FE8	1ED
NONLNV	*	20AD6	60C
ANUMARG	*	210E8	28
POINTV	*	21110	2F8
POLARV	*	213C8	FC8
SETCIV	*	22390	12C
STOPTV	*	224C0	CE
XSCALV	*	22590	28C
YSCALV	*	22620	28C
ACCN	*	226B0	3C
ERRFRV	*	226F0	12E
IHCFCPT	*	22C18	304
HOLLV	*	22F20	1C0
IHCUATBL	*	230E0	E38
SCERRV	*	23718	12C
SNAP	*	23848	2D0
VCHARV	*	23B18	3E2
ANDVX	*	23F00	24
BCDNT	*	23F28	3E
SHFT2V	*	23F68	F0
STOPAG		23F80	66DD0
JCINT		8A0E0	10

NAME	LOCATION	NAME	LOCATION	NAME	LOCATION	NAME	LOCATION
DUMPCD	1E628						
OUTCD	1EEC4						
FULCMP	1F33F	IHCOMP	1F37C	IHCERR	1F3C0		
MASKS	205F8	MASKR	2061E	MASK	2063F		
CR	20700						
COS	208E8	SIN	20904				
NUMARG	210E8						
ACCON	22AB0						
ERRSET	22C18	ERRSAV	22F10	ERRSTR	22F36		
SNAPIT	23848	ABEND	239E4				
ANDV	23F00						

```
ENTRY ADDRESS       00
TOTAL LENGTH      8A090

****FHXXMAIN  DOES NOT EXIST BUT HAS BEEN ADDED TO DATA SET
```

DIAGNOSTIC MESSAGE DIRECTORY

IEW0151 WARNING - MAIN STORAGE REQUIREMENTS FOR OUTPUT LOAD MODULE HAVE EXCEEDED 512K BYTES.

APPENDIX D

COMPARISON OF ESTIMATED AND ACTUAL SUNSPOT CYCLES FROM THE LEARNING CASES

This appendix presents a comparison of the actual and predicted sunspot numbers for cycles 3 through 18 which were the learning data for developing the prediction algorithm discussed in this report. Figures D-1 through D-16 present the actual sunspot numbers as the dash lines. These figures present graphical interpretation of the RMS error of 17.4 of the learning data for ADAPT prediction algorithm.

Figures D-17 through D-35 present the information required to make a similar comparison for the ADAPT third quarter extrapolation. Figures D-17 and D-19 give the actual sunspot number for cycles 1 and 2 and Figures D-18 and D-20 through D-35 give the extrapolated sunspot numbers, based on the first 93 months of the cycle.

The predictions, for which Figures D-1 through D-16 indicate the performance are used for the recommended estimate of cycle 21. The third quarter extrapolations, for which Figures D-17 through D-35 indicate the performance are used for the recommended estimate of the remainder of cycle 20.

Comparison of the predicted and extrapolated cycles with the actual cycles show that the third quarter extrapolations are significantly better than the prediction. In both cases, it is interesting to note that the predictions of the location of the short term oscillation is quite good. The actual amplitude, and in some cases the phase of the short term oscillation is quite poor for the prediction algorithm. When the error in the underlying basic cycle is also considered, the extrapolated values for the short term oscillations are actually quite good. Thus, we conclude that one can use the estimate presented in Figure 2.1 of the report to infer the general characteristics of the short term (scale of several months) behavior for cycle 20, but only as an indication of when a spike might occur for the estimates of cycle 21.

Sleeper, in Reference 13, has pointed out that cycles 4 and 9 should be considered as models of cycle 20. Thus, a comparison of the extrapolation of these two cycles with their actual values may be used to anticipate the types of error which might be expected in the estimates of sunspot numbers presented in Figure 2.1. Note that the prediction technique compared in Figures D-2 and D-7 which significantly underpredicted both of these cycles was not used in the estimate of cycle 20. However, comparison of the extrapolation of cycle 4 shown in Figure D-22 with the actual cycle presented in Figure D-2, and assuming the cycle ends at a sunspot number of 5 shows that the period is still

underpredicted by approximately 1 1/2 to 2 years and in general the sunspot activity for the last quarter is underpredicted from 5 to 20 sunspot numbers. Comparison of the extrapolation of cycle 9 shown in Figure D-27 with the actual values of cycle 9 shown in Figure D-7 shows that both the period and sunspot numbers are estimated quite well for cycle 9. Thus, one concludes that the estimate of the current cycle presented in Figure 2.1 might range from quite good to slightly underestimating the sunspot activity and period of cycle 20.

FIGURE D-1

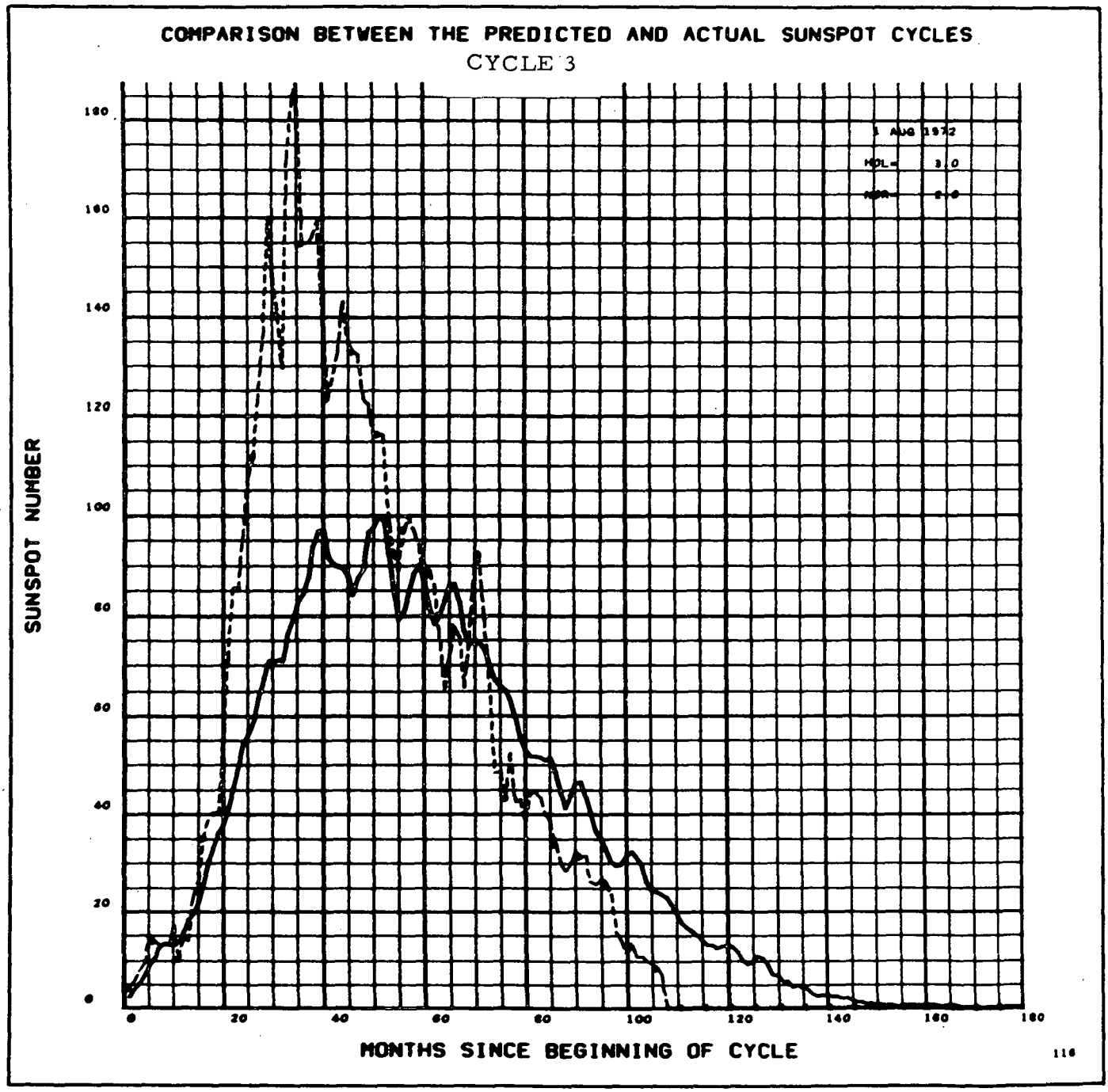

COMPARISON BETWEEN THE PREDICTED AND ACTUAL SUNSPOT CYCLES
CYCLE 3
1 AUG 1972
MDL= 3.0
SUNSPOT NUMBER
180
160
140
120
100
80
60
40
20
0
0
20
40
60
80
100
120
140
160
180
MONTHS SINCE BEGINNING OF CYCLE
116

FIGURE D-2

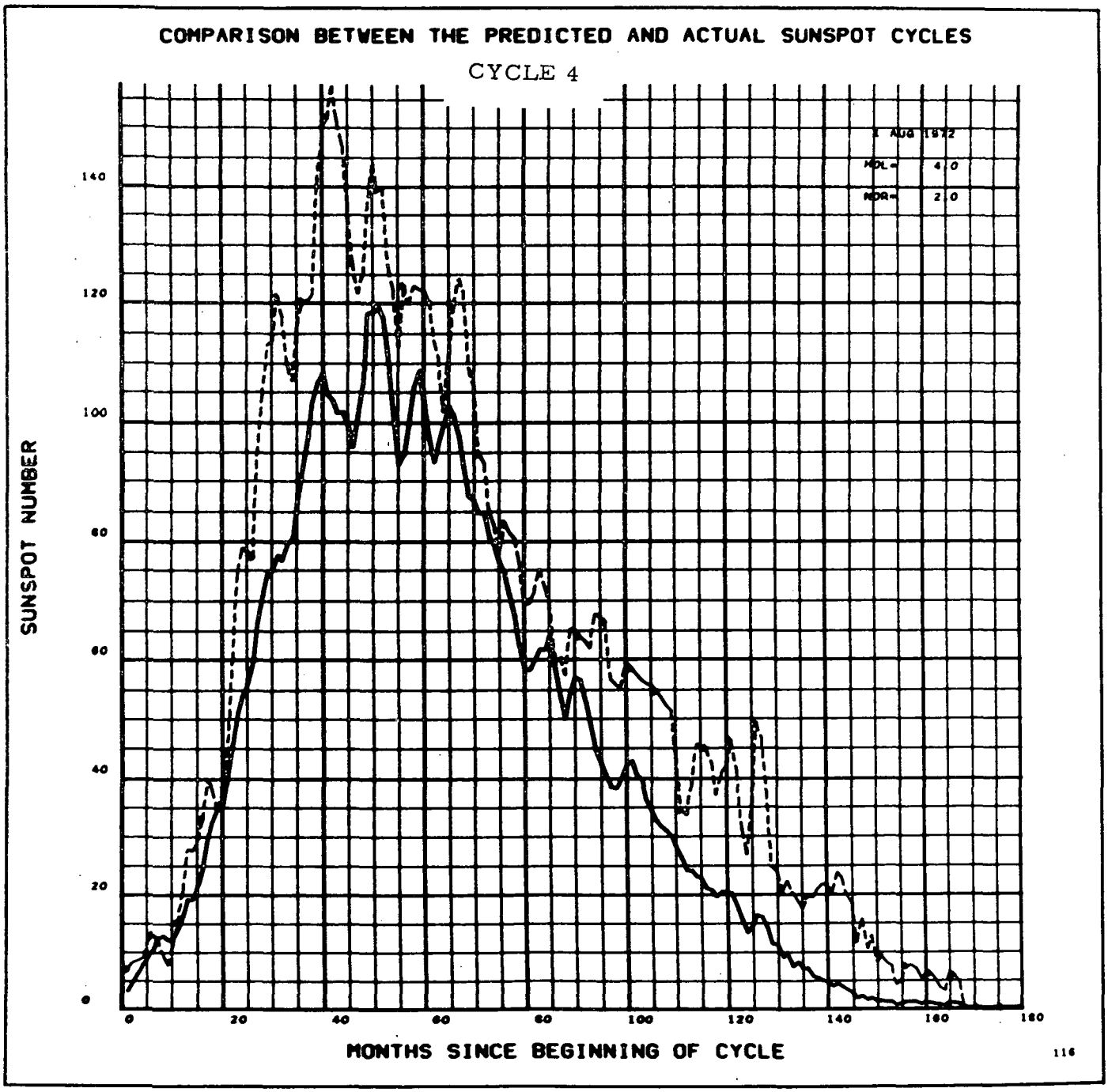

COMPARISON BETWEEN THE PREDICTED AND ACTUAL SUNSPOT CYCLES
CYCLE 4
1 AUG 1972
MOL= 40
NOR= 20
SUNSPOT NUMBER
0
20
40
60
80
100
120
140
160
180
MONTHS SINCE BEGINNING OF CYCLE
116

FIGURE D-3

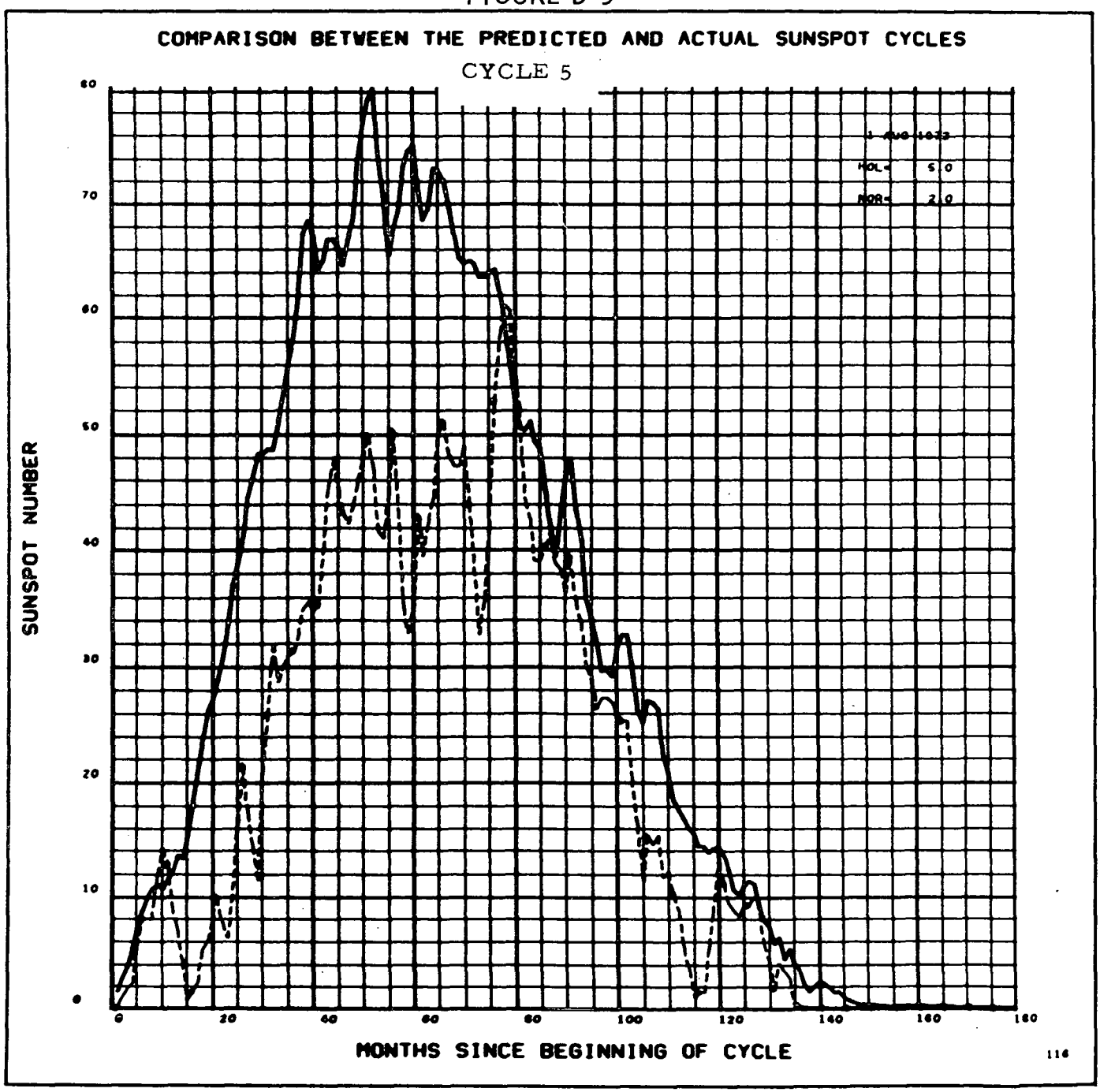

COMPARISON BETWEEN THE PREDICTED AND ACTUAL SUNSPOT CYCLES
CYCLE 5
SUNSPOT NUMBER
MONTHS SINCE BEGINNING OF CYCLE
0
10
20
30
40
50
60
70
80
0
20
40
60
80
100
120
140
160
180
116

FIGURE D-4

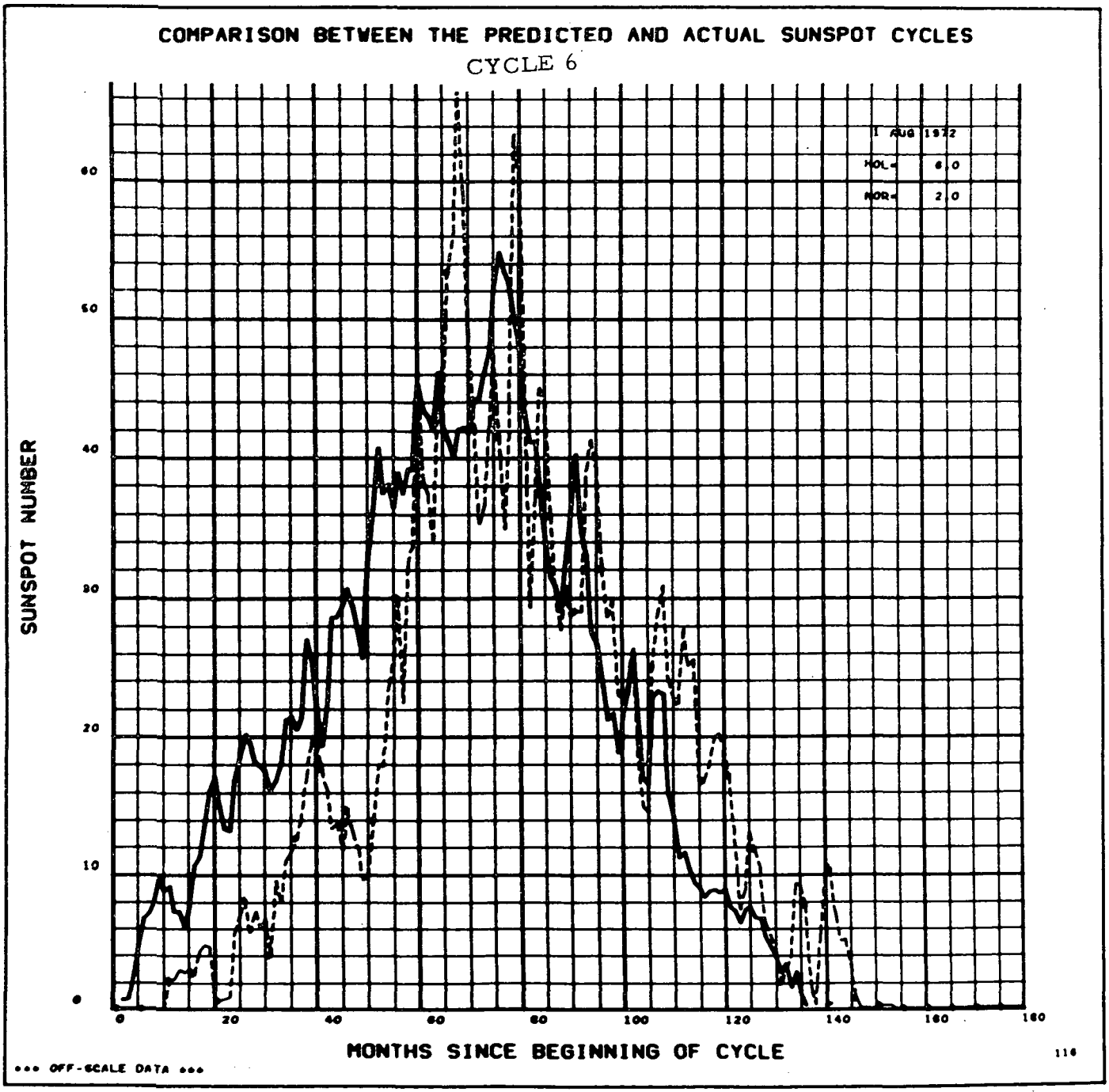

COMPARISON BETWEEN THE PREDICTED AND ACTUAL SUNSPOT CYCLES
CYCLE 6
1 AUG 1972
HOL= 8.0
NOR= 2.0
SUNSPOT NUMBER
0
10
20
30
40
50
60
0
20
40
60
80
100
120
140
160
MONTHS SINCE BEGINNING OF CYCLE
116
••• OFF-SCALE DATA •••

FIGURE D-5

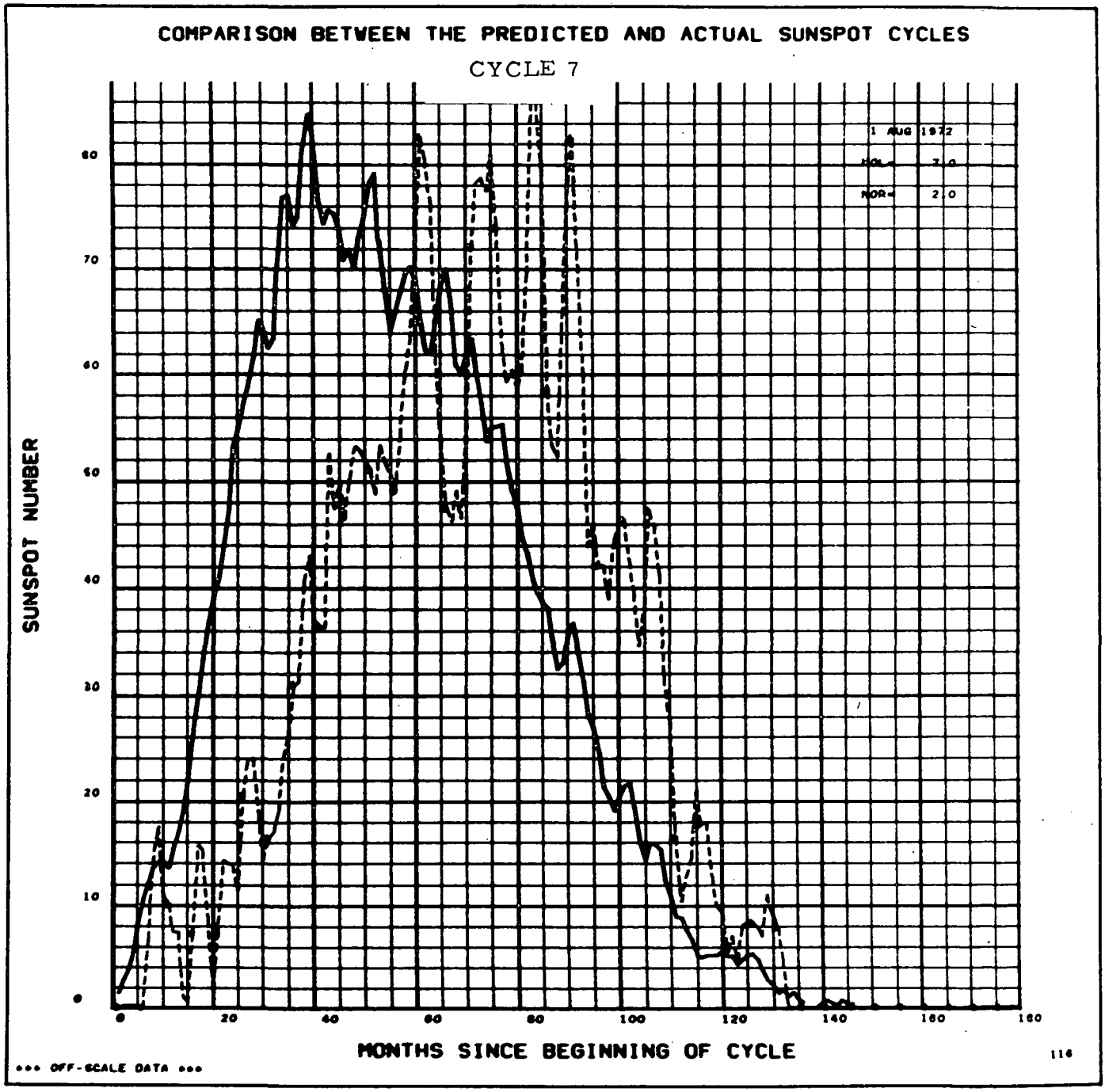

COMPARISON BETWEEN THE PREDICTED AND ACTUAL SUNSPOT CYCLES
CYCLE 7
1 AUG 1972
HOL- 7.0
NOR- 2.0
SUNSPOT NUMBER
0
10
20
30
40
50
60
70
80
0
20
40
60
80
100
120
140
160
MONTHS SINCE BEGINNING OF CYCLE
••• OFF-SCALE DATA •••
116

FIGURE D-6

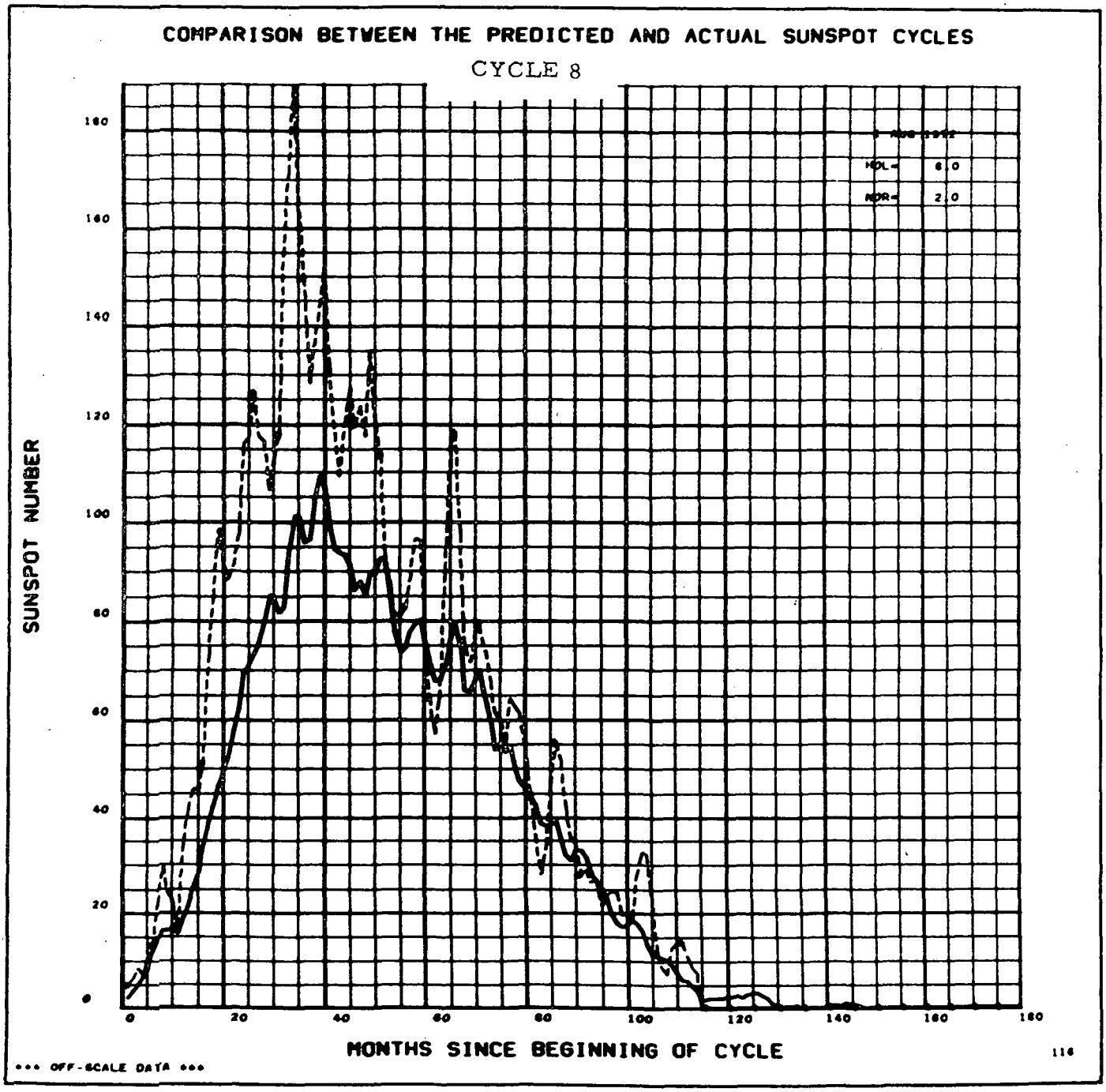

COMPARISON BETWEEN THE PREDICTED AND ACTUAL SUNSPOT CYCLES
CYCLE 8
SUNSPOT NUMBER
MONTHS SINCE BEGINNING OF CYCLE
0
20
40
60
80
100
120
140
160
180
MDL= 8.0
MDR= 2.0
*** OFF-SCALE DATA ***
116

FIGURE D-7

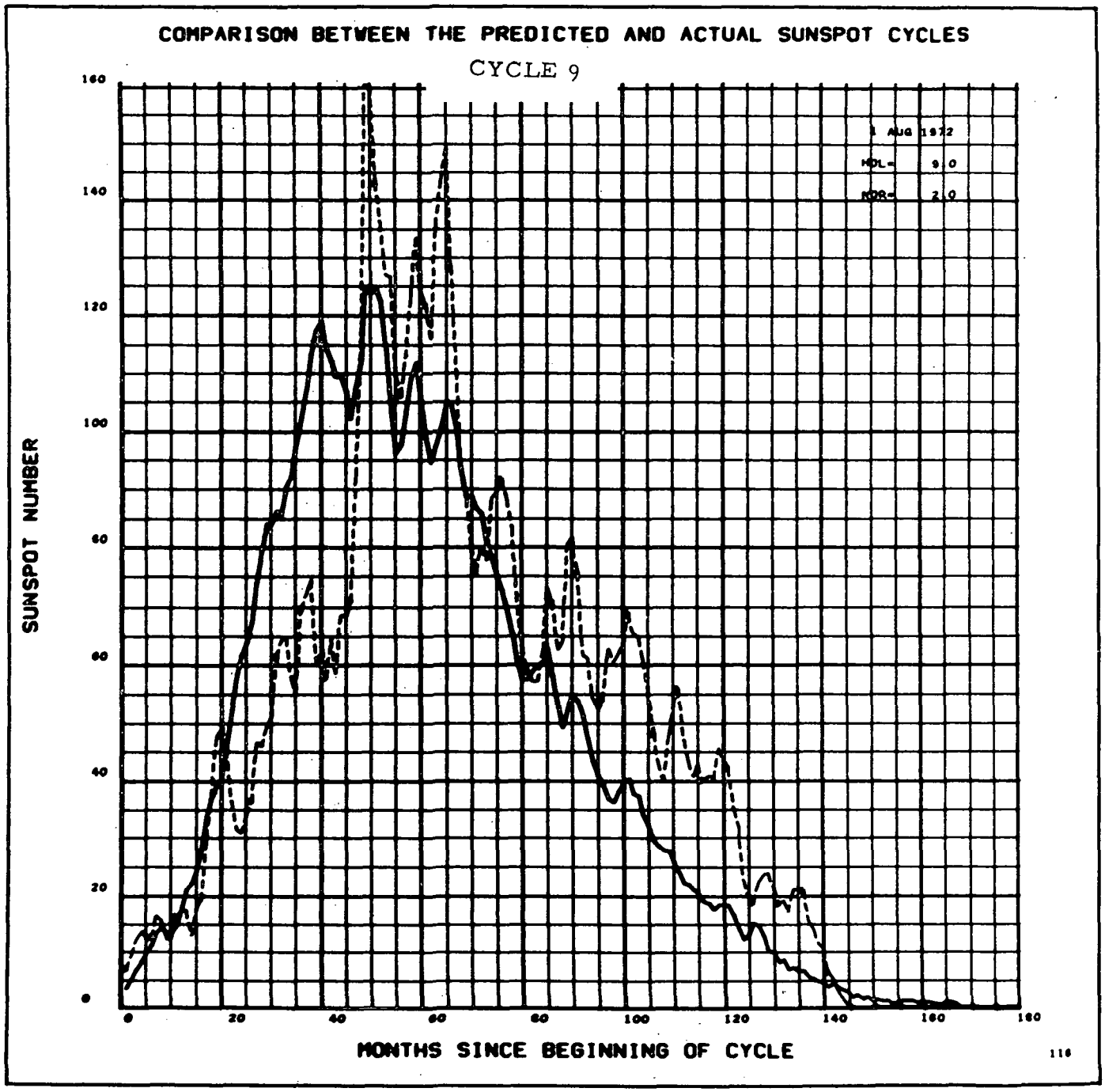

COMPARISON BETWEEN THE PREDICTED AND ACTUAL SUNSPOT CYCLES
CYCLE 9
1 AUG 1972
MOL= 9.0
NOR= 2.0
SUNSPOT NUMBER
MONTHS SINCE BEGINNING OF CYCLE
0
20
40
60
80
100
120
140
160
180

FIGURE D-8

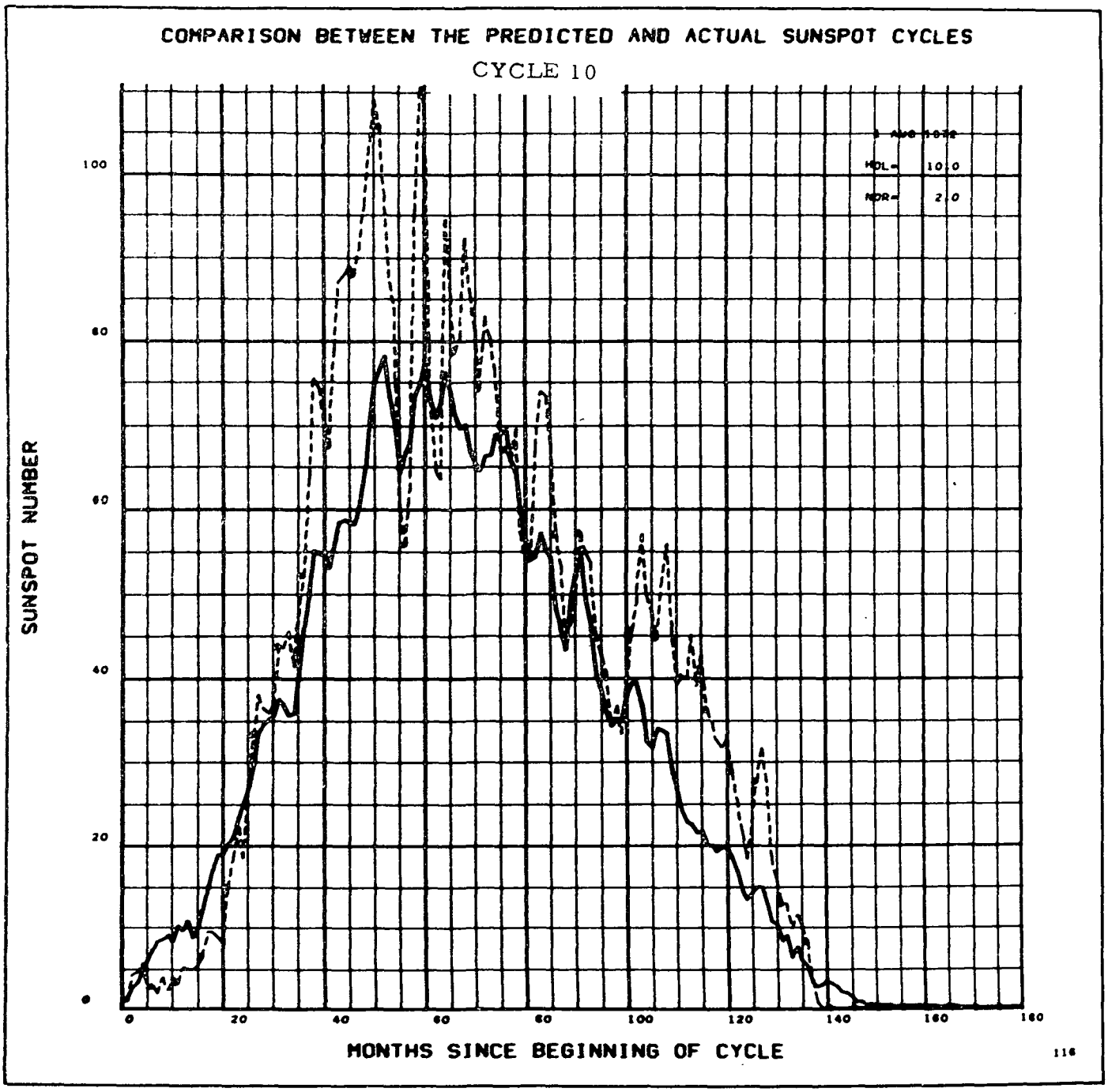

COMPARISON BETWEEN THE PREDICTED AND ACTUAL SUNSPOT CYCLES
CYCLE 10
MDL= 10.0
MDR= 2.0
SUNSPOT NUMBER
100
80
60
40
20
0
0
20
40
60
80
100
120
140
160
180
MONTHS SINCE BEGINNING OF CYCLE
116

FIGURE D-9

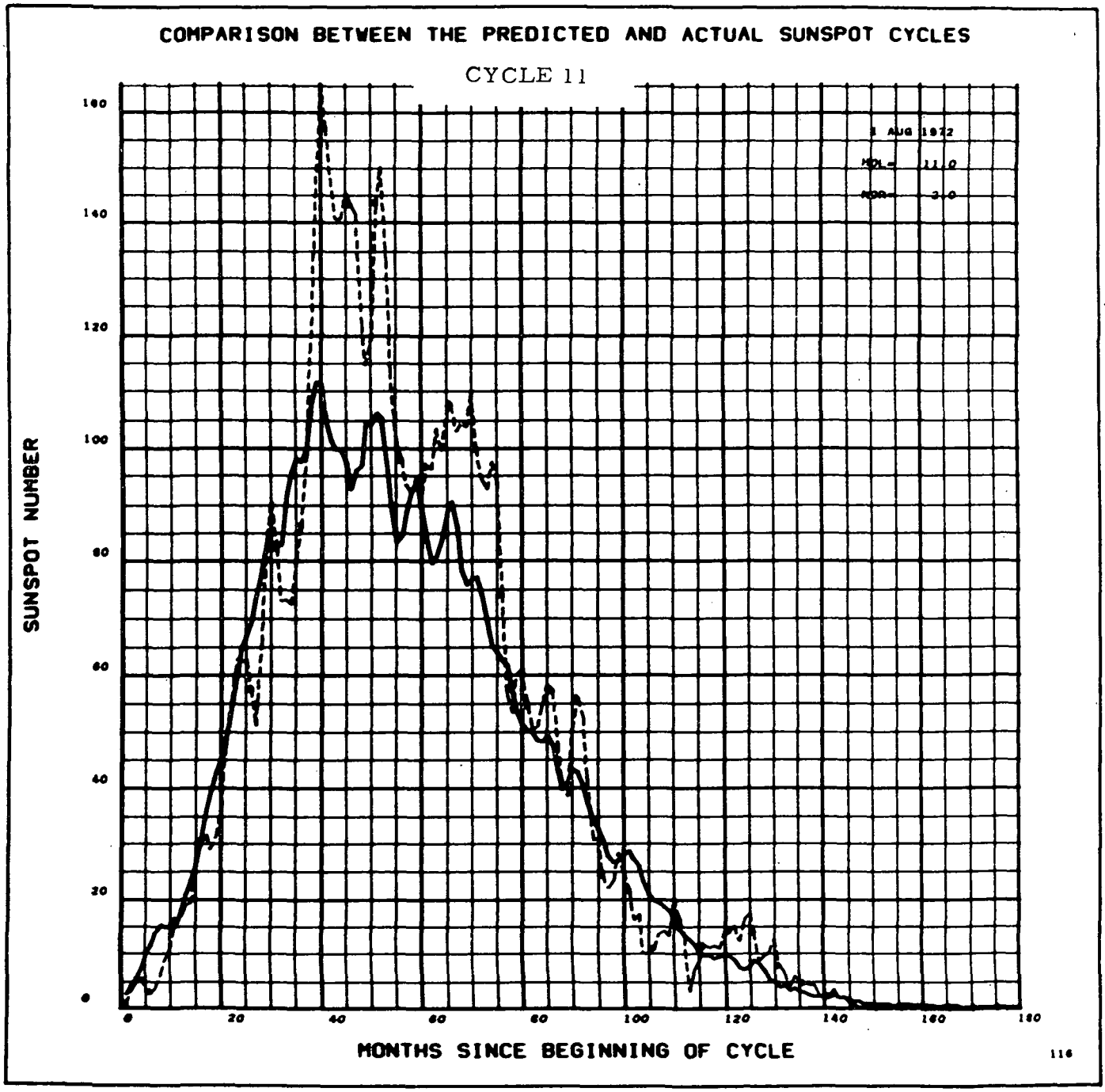

COMPARISON BETWEEN THE PREDICTED AND ACTUAL SUNSPOT CYCLES
CYCLE 11
1 AUG 1972
MDL= 11.0
MOD= 3.0
SUNSPOT NUMBER
0
20
40
60
80
100
120
140
160
0
20
40
60
80
100
120
140
160
180
MONTHS SINCE BEGINNING OF CYCLE
116

FIGURE D-10

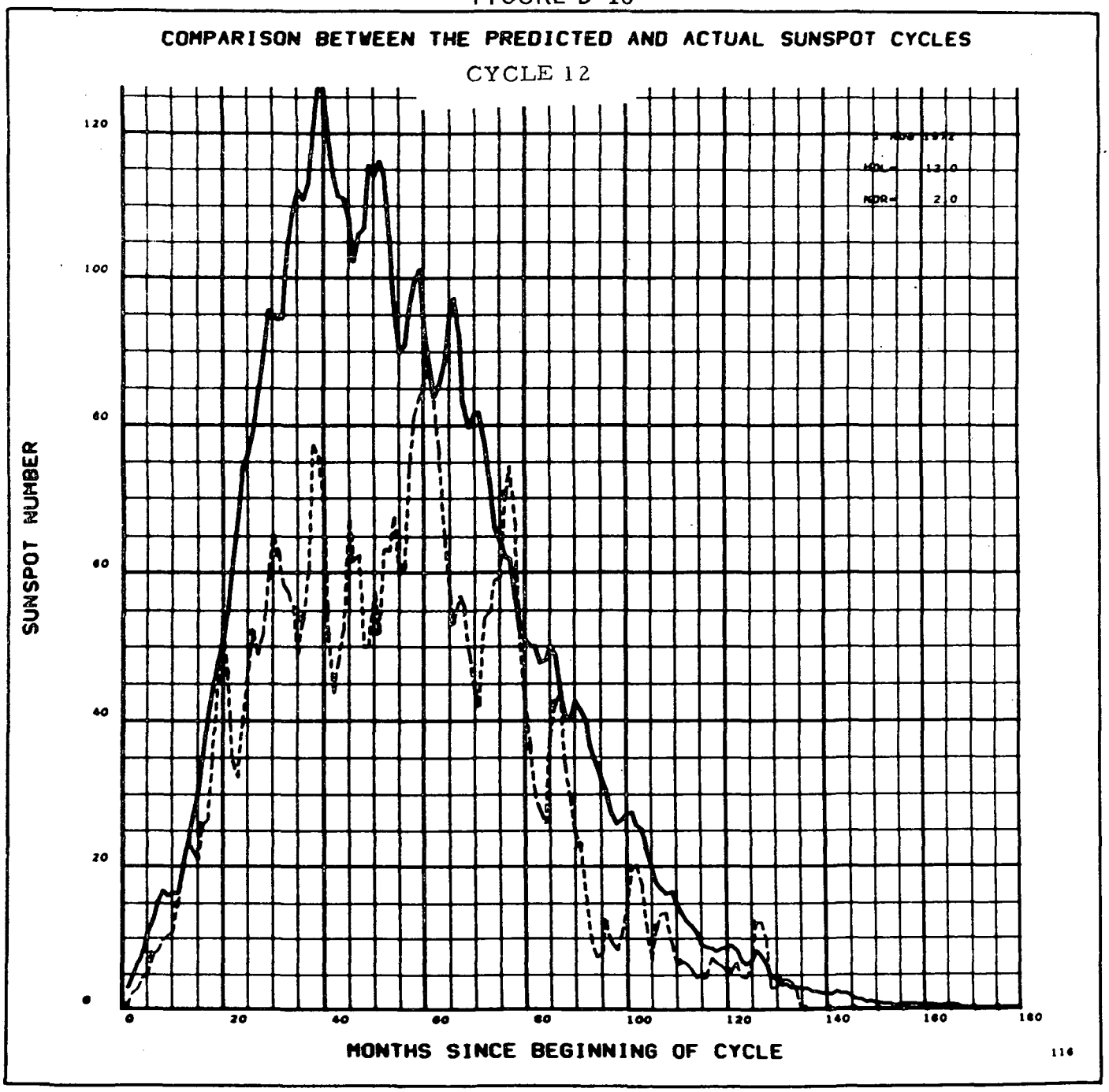

COMPARISON BETWEEN THE PREDICTED AND ACTUAL SUNSPOT CYCLES
CYCLE 12
SUNSPOT NUMBER
MONTHS SINCE BEGINNING OF CYCLE
0
20
40
60
80
100
120
140
160
180
116

FIGURE D-11

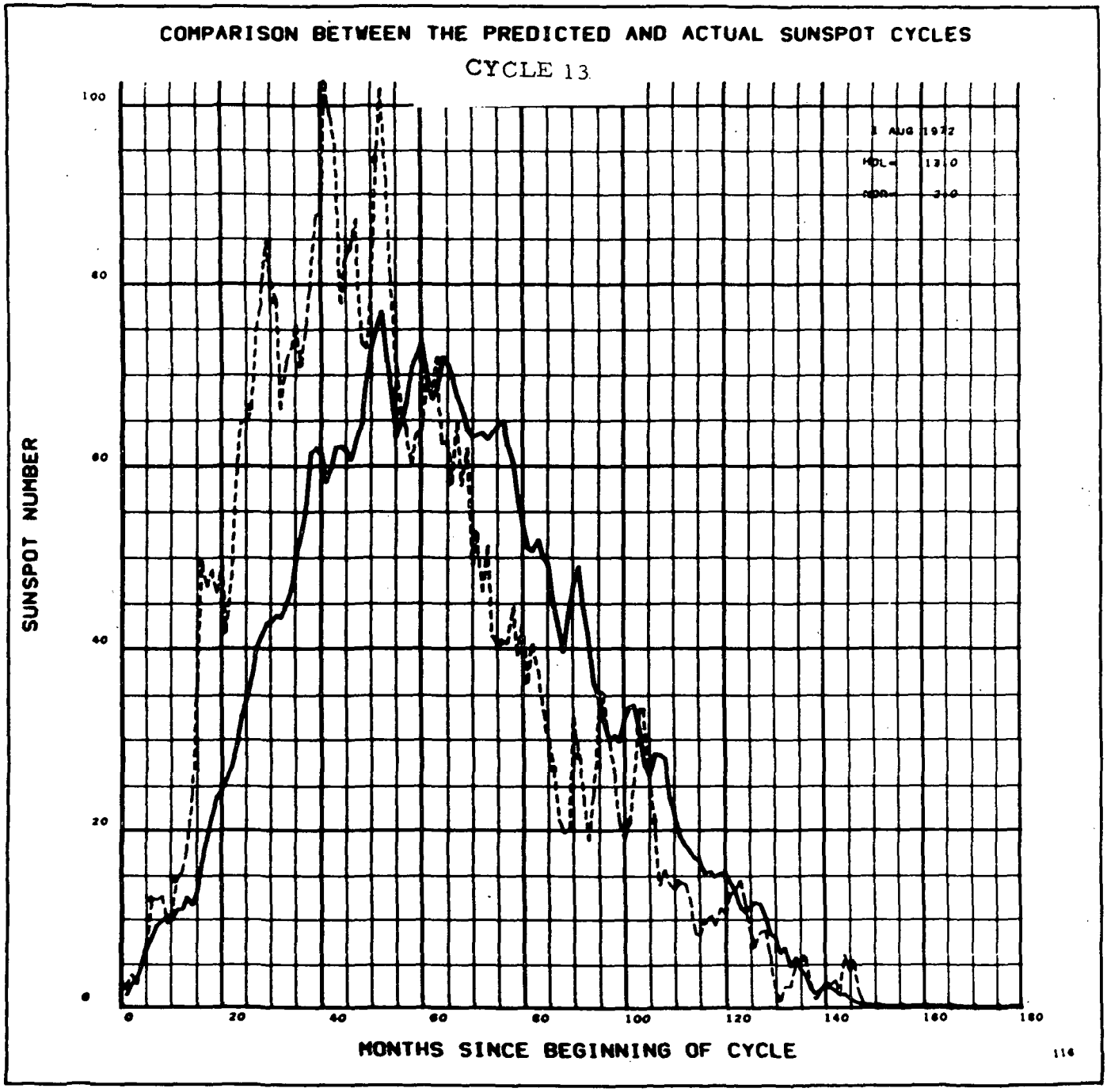

COMPARISON BETWEEN THE PREDICTED AND ACTUAL SUNSPOT CYCLES
CYCLE 13
1 AUG 1972
MDL= 13.0
RPR= 3.0
SUNSPOT NUMBER
100
80
60
40
20
0
0
20
40
60
80
100
120
140
160
180
MONTHS SINCE BEGINNING OF CYCLE
114

FIGURE D-12

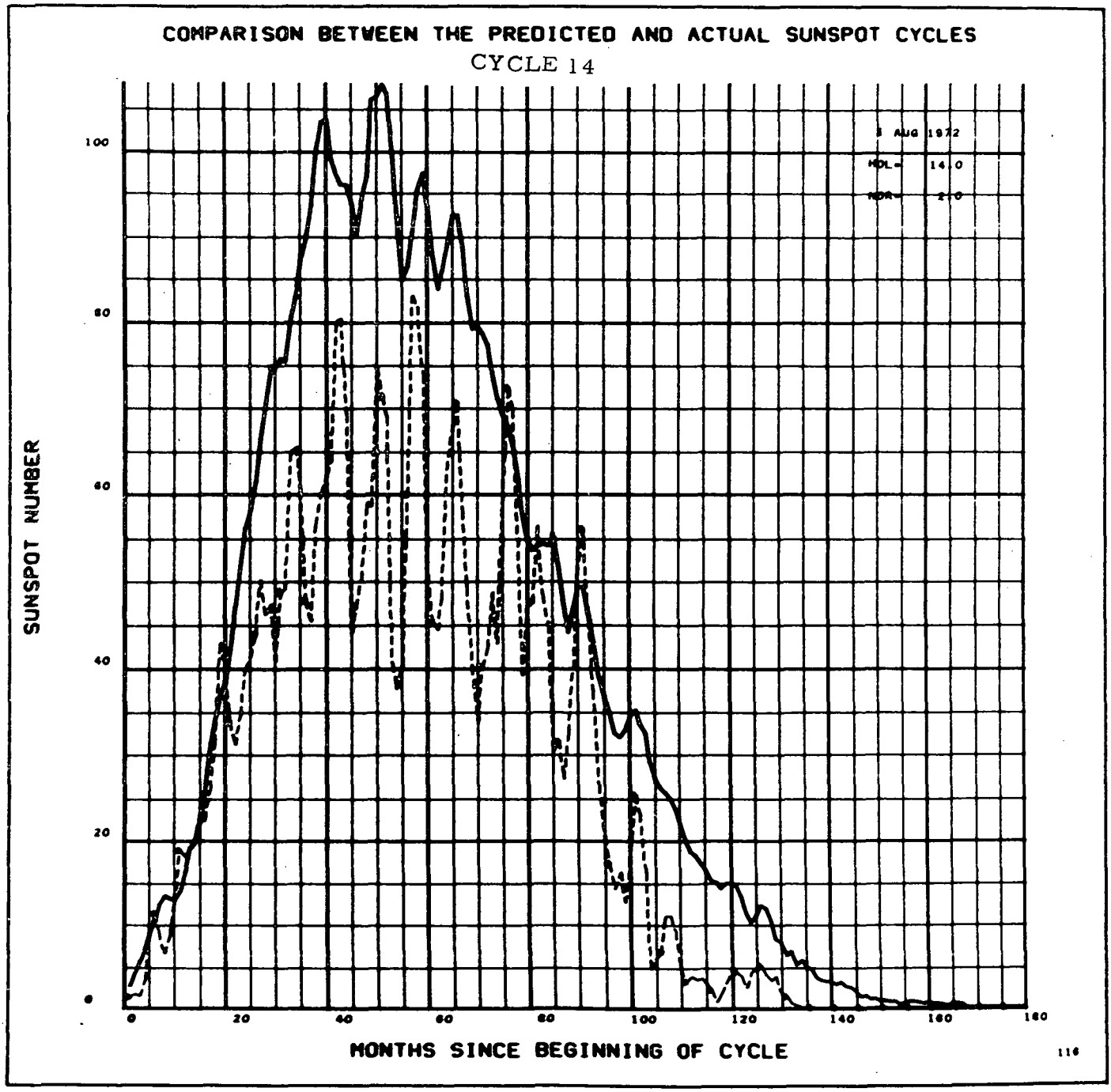

COMPARISON BETWEEN THE PREDICTED AND ACTUAL SUNSPOT CYCLES
CYCLE 14
SUNSPOT NUMBER
MONTHS SINCE BEGINNING OF CYCLE
1 AUG 1972
NOL= 14.0
NOR= 2.0
0
20
40
60
80
100
120
140
160
180
116

FIGURE D-13

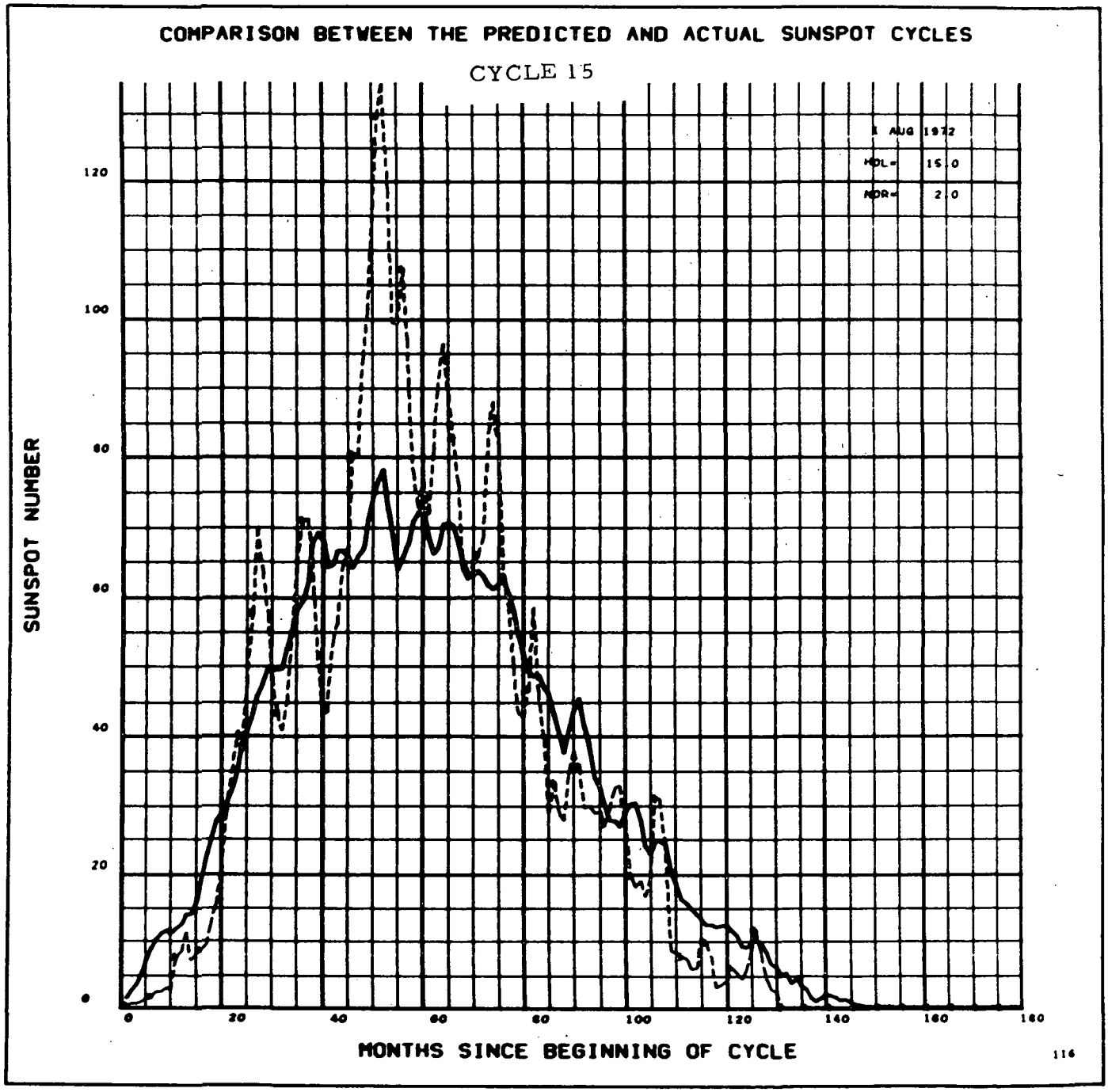

COMPARISON BETWEEN THE PREDICTED AND ACTUAL SUNSPOT CYCLES
CYCLE 15
1 AUG 1972
MDL= 15.0
MDR= 2.0
SUNSPOT NUMBER
0
20
40
60
80
100
120
0
20
40
60
80
100
120
140
160
MONTHS SINCE BEGINNING OF CYCLE
116

FIGURE D-14

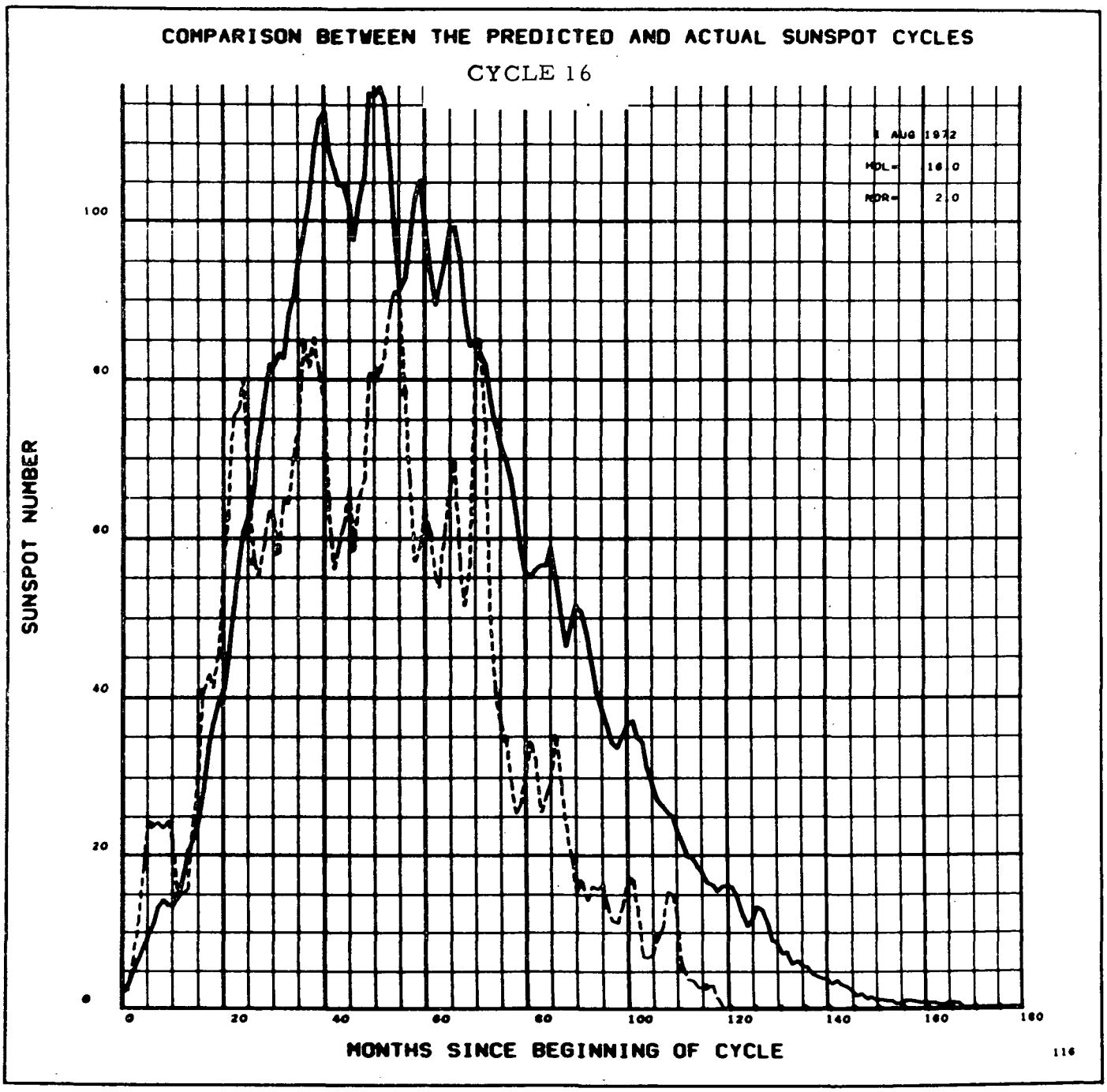

COMPARISON BETWEEN THE PREDICTED AND ACTUAL SUNSPOT CYCLES
CYCLE 16
1 AUG 1972
MDL= 16.0
MDR= 2.0
SUNSPOT NUMBER
100
80
60
40
20
0
0
20
40
60
80
100
120
140
160
180
MONTHS SINCE BEGINNING OF CYCLE
116

FIGURE D-15

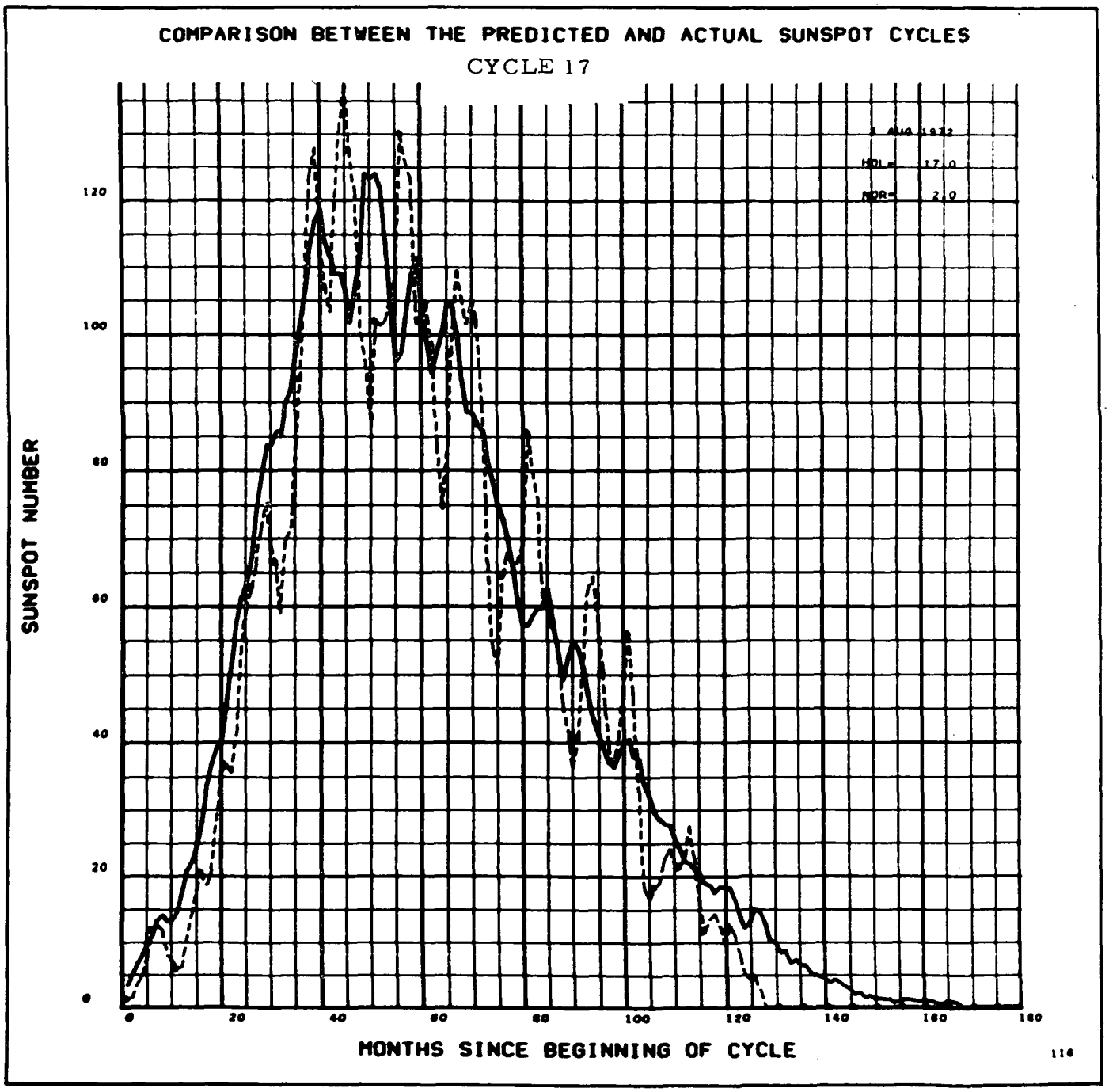

COMPARISON BETWEEN THE PREDICTED AND ACTUAL SUNSPOT CYCLES
CYCLE 17
SUNSPOT NUMBER
MONTHS SINCE BEGINNING OF CYCLE
0
20
40
60
80
100
120
140
160
180
116

FIGURE D-16

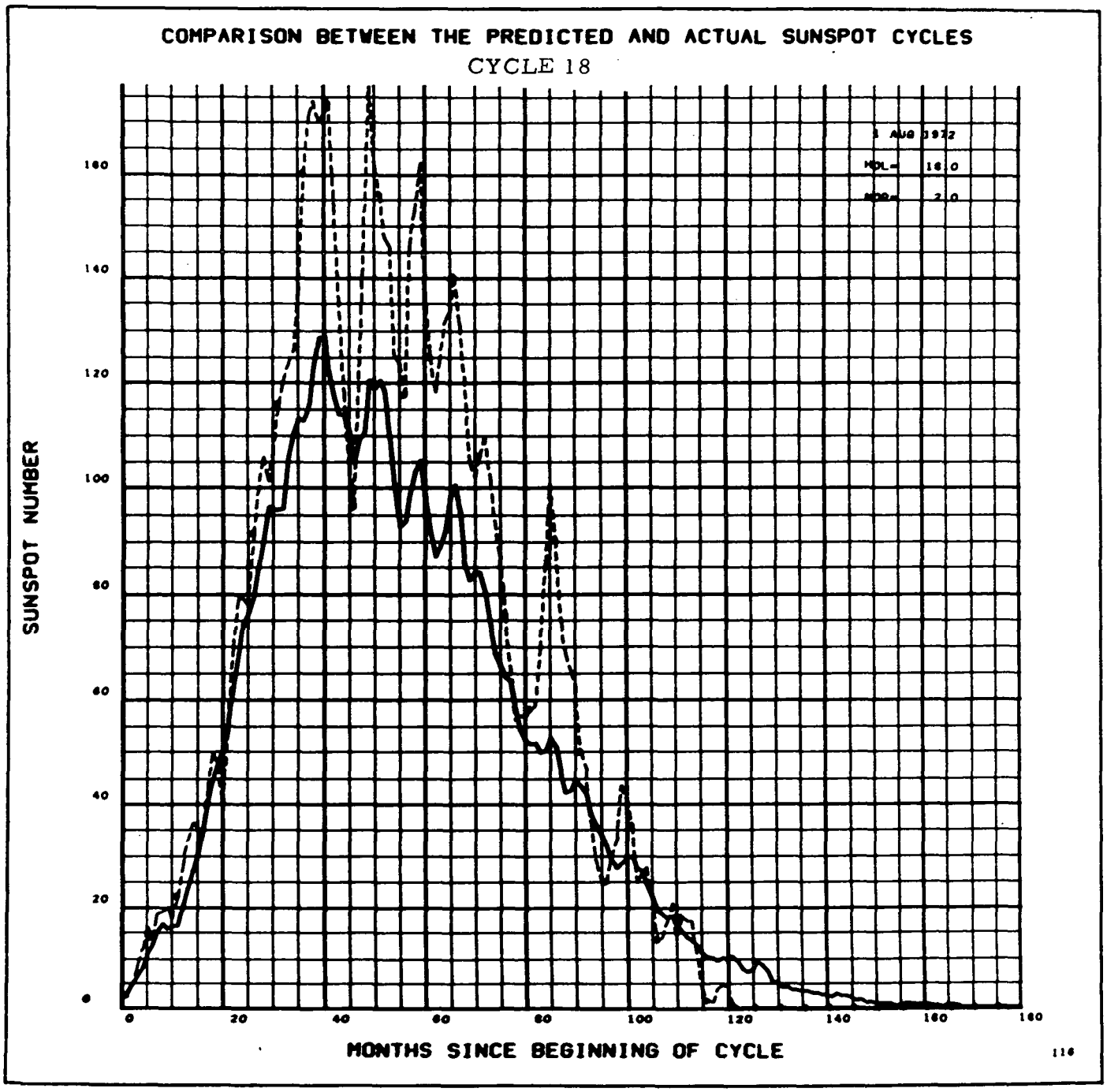

COMPARISON BETWEEN THE PREDICTED AND ACTUAL SUNSPOT CYCLES
CYCLE 18
1 AUG 1972
MDL= 18.0
RDR= 2.0
SUNSPOT NUMBER
160
140
120
100
80
60
40
20
0
0
20
40
60
80
100
120
140
160
180
MONTHS SINCE BEGINNING OF CYCLE
116

FIGURE D-17

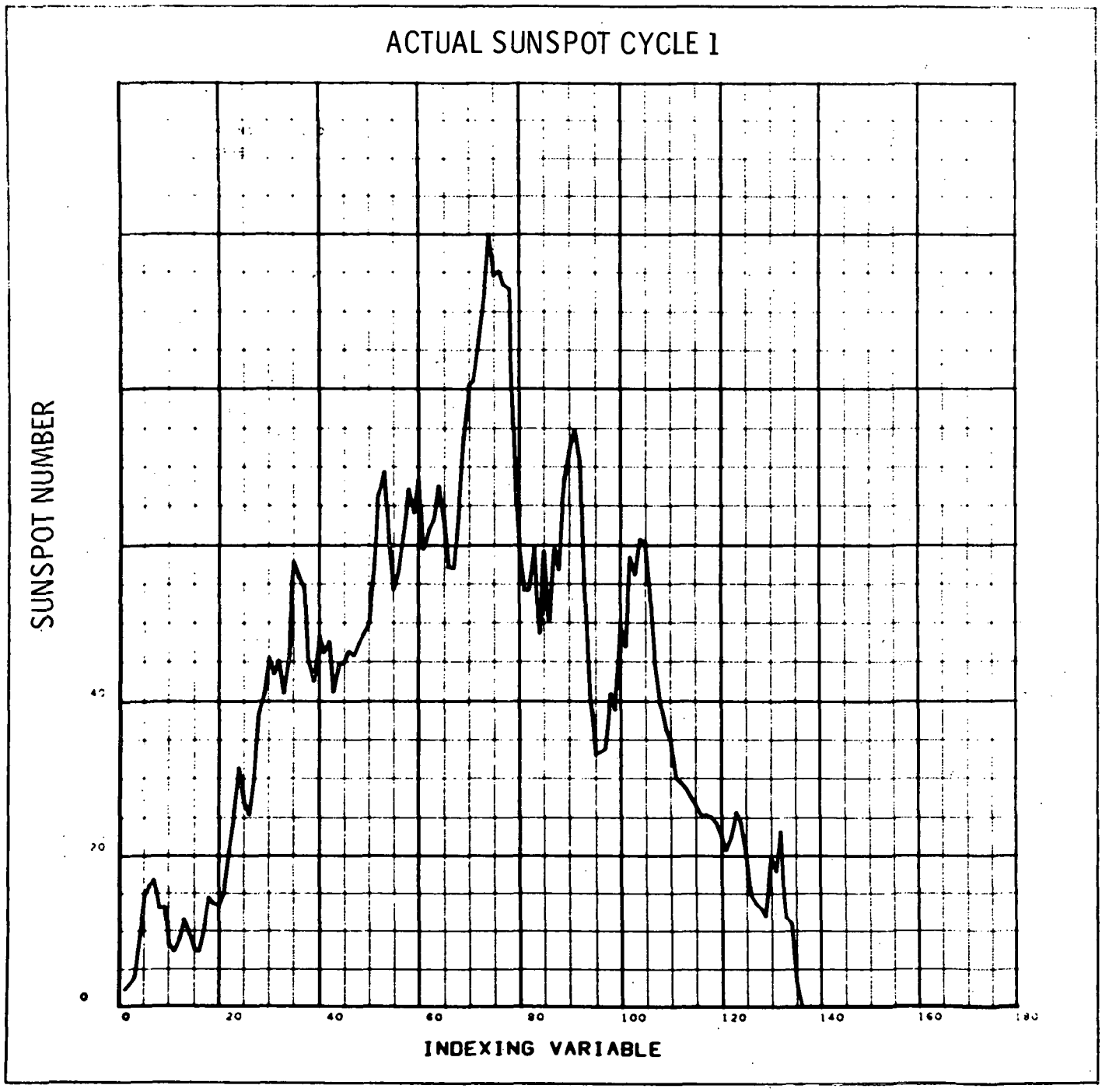

ACTUAL SUNSPOT CYCLE 1
SUNSPOT NUMBER
0
20
40
0
20
40
60
80
100
120
140
160
INDEXING VARIABLE

FIGURE D-18

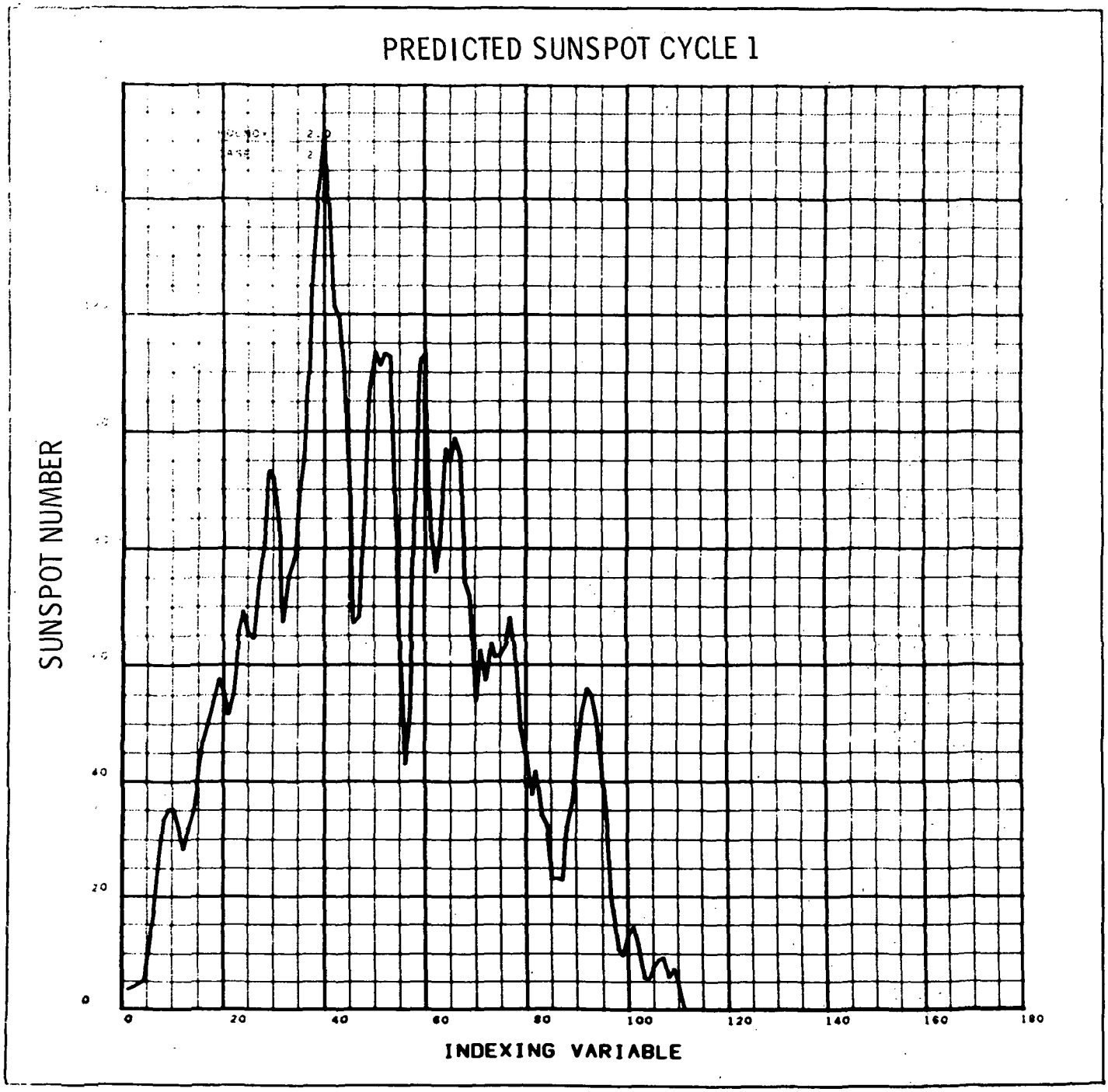

PREDICTED SUNSPOT CYCLE 1
SUNSPOT NUMBER
0
20
40
60
80
100
120
140
160
180
INDEXING VARIABLE

FIGURE D-19

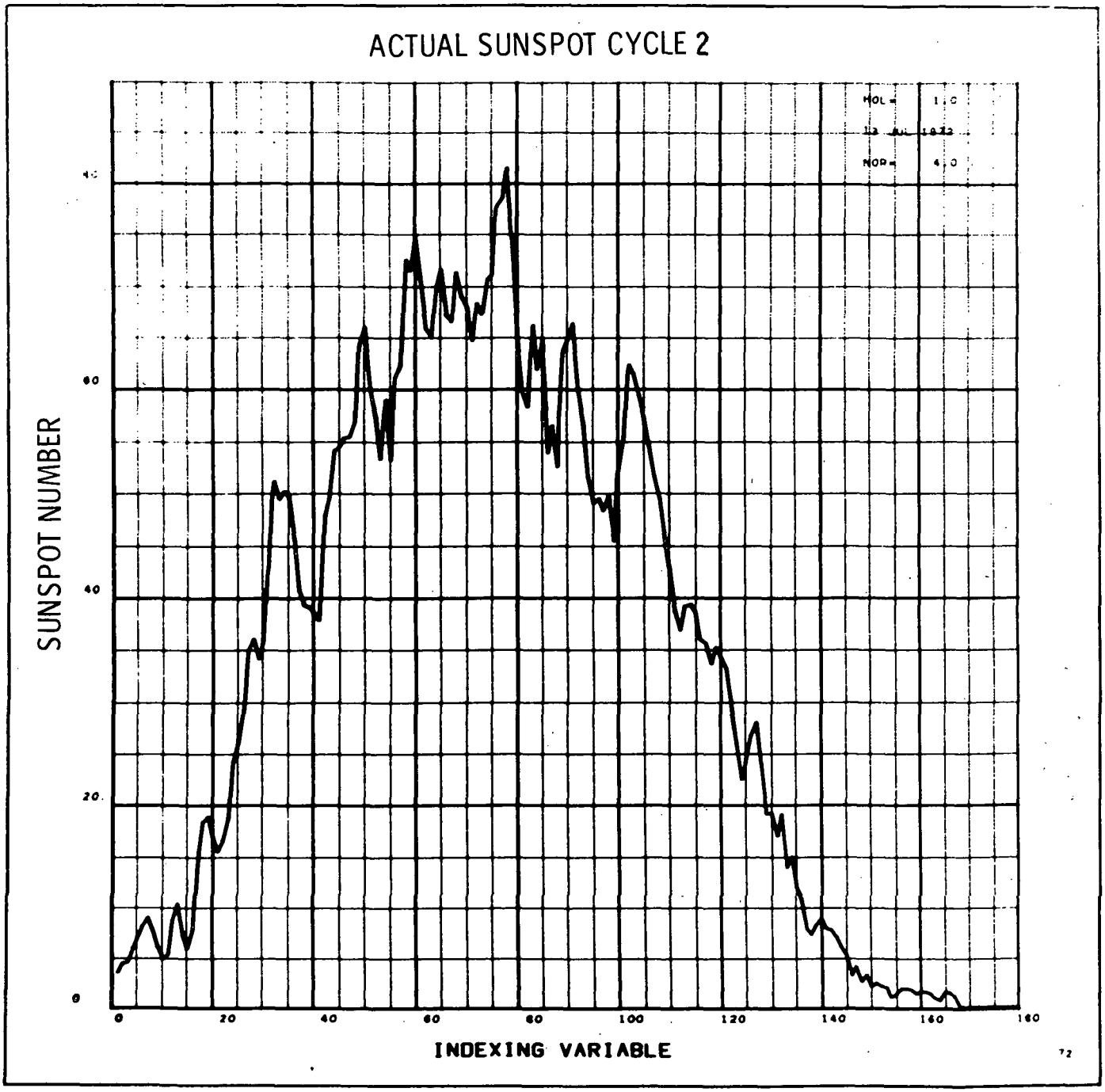

ACTUAL SUNSPOT CYCLE 2
SUNSPOT NUMBER
INDEXING VARIABLE
0
20
40
60
80
100
120
140
160
MOL= 1.0
NOR= 4.0

FIGURE D-20

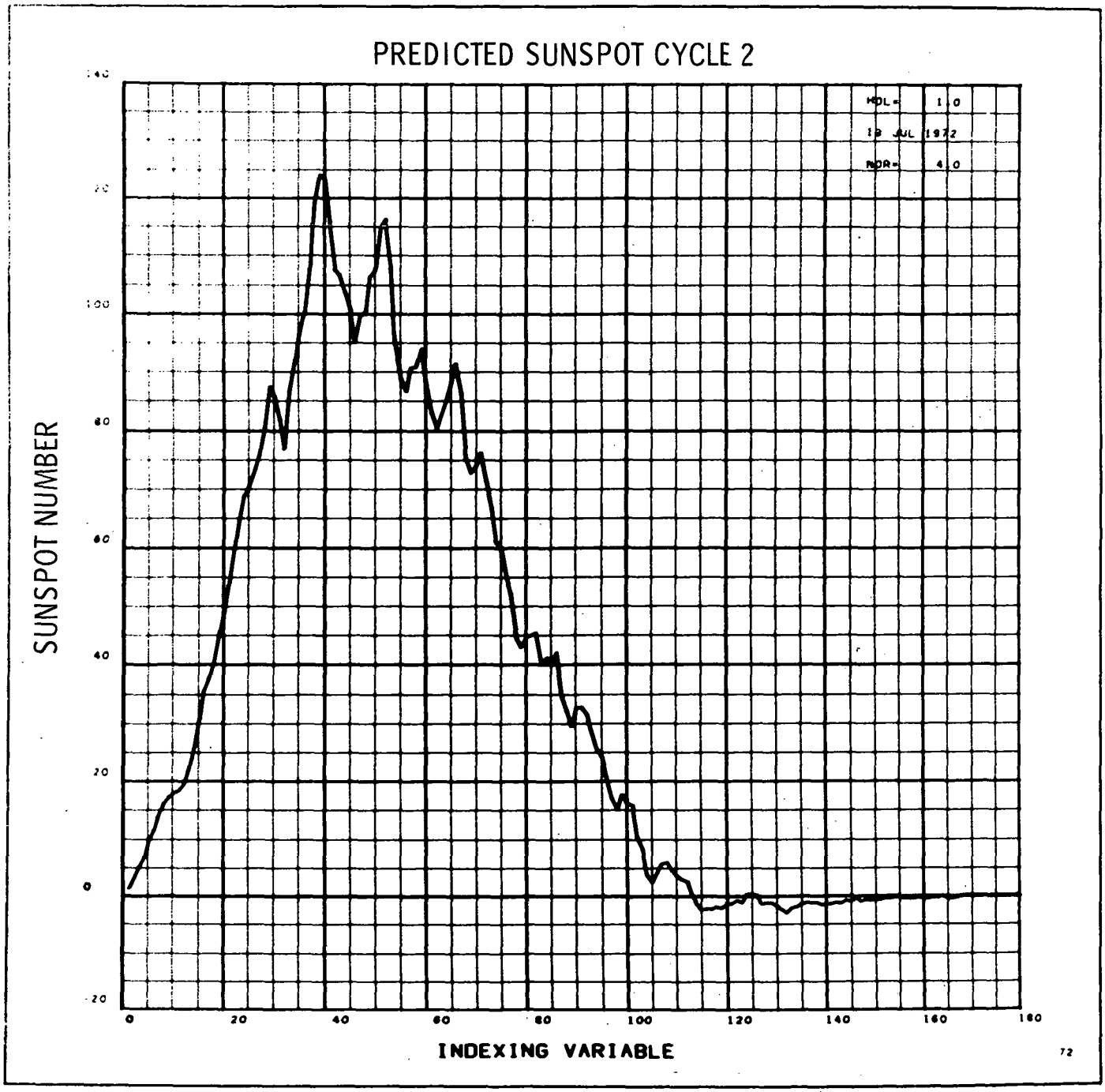

PREDICTED SUNSPOT CYCLE 2
MDL= 1.0
12 JUL 1972
MDR= 4.0
SUNSPOT NUMBER
INDEXING VARIABLE
0
20
40
60
80
100
120
140
160
180
72

FIGURE D-21

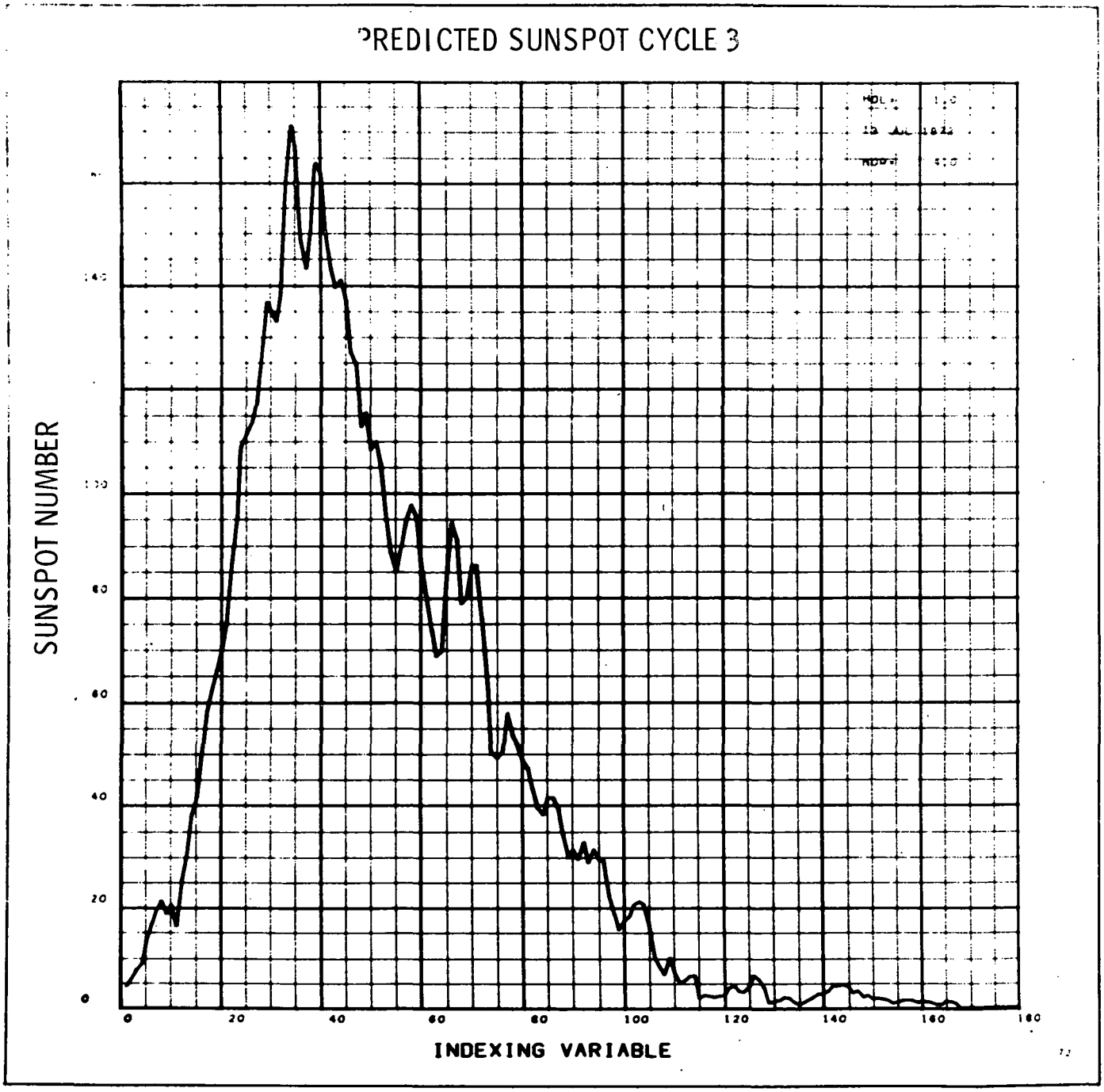

PREDICTED SUNSPOT CYCLE 3
SUNSPOT NUMBER
INDEXING VARIABLE
0
20
40
60
80
100
120
140
160

FIGURE D-22

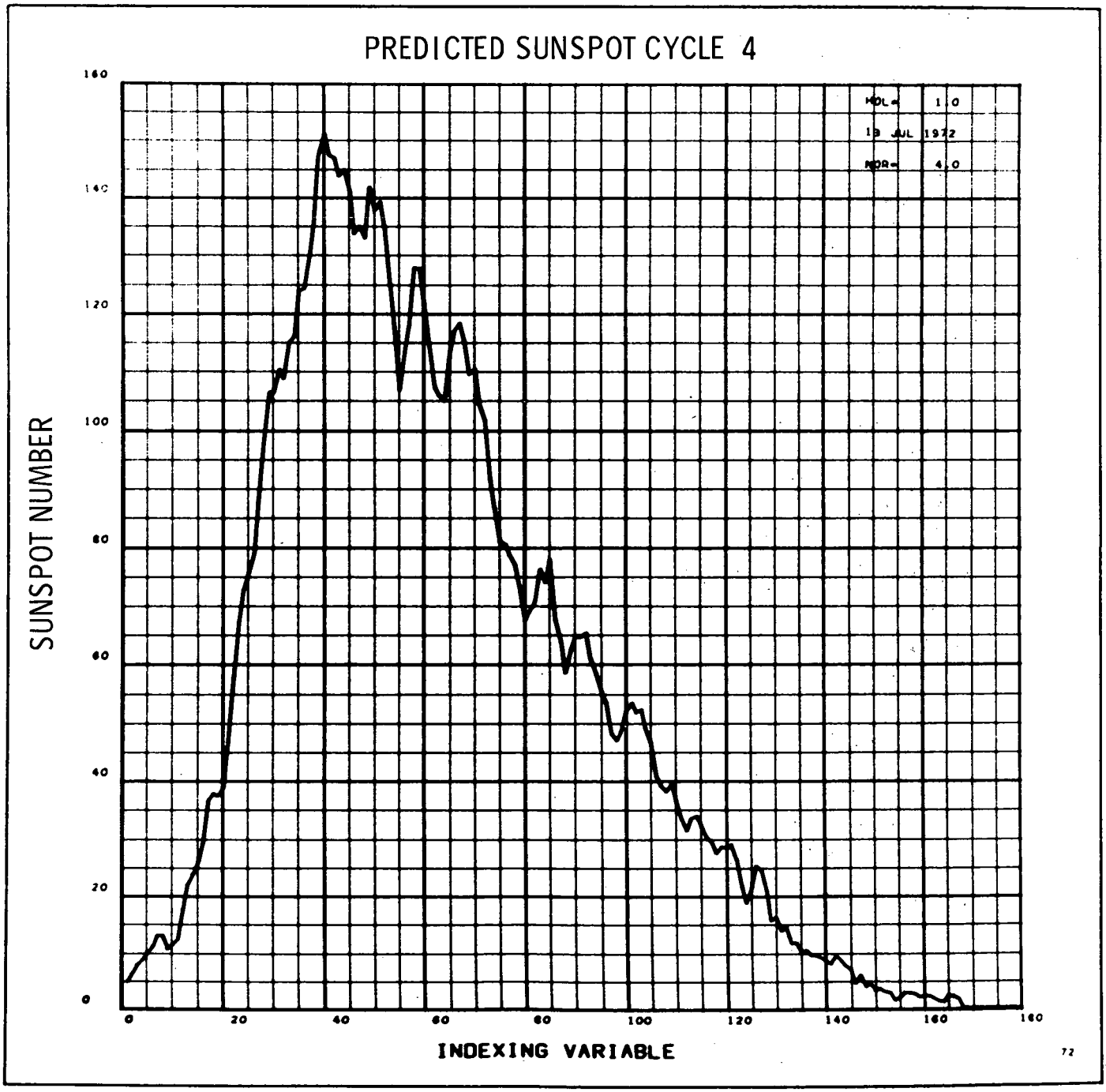

PREDICTED SUNSPOT CYCLE 4
SUNSPOT NUMBER
INDEXING VARIABLE
0
20
40
60
80
100
120
140
160
MDL= 1.0
19 JUL 1972
MDR= 4.0
72

FIGURE D-23

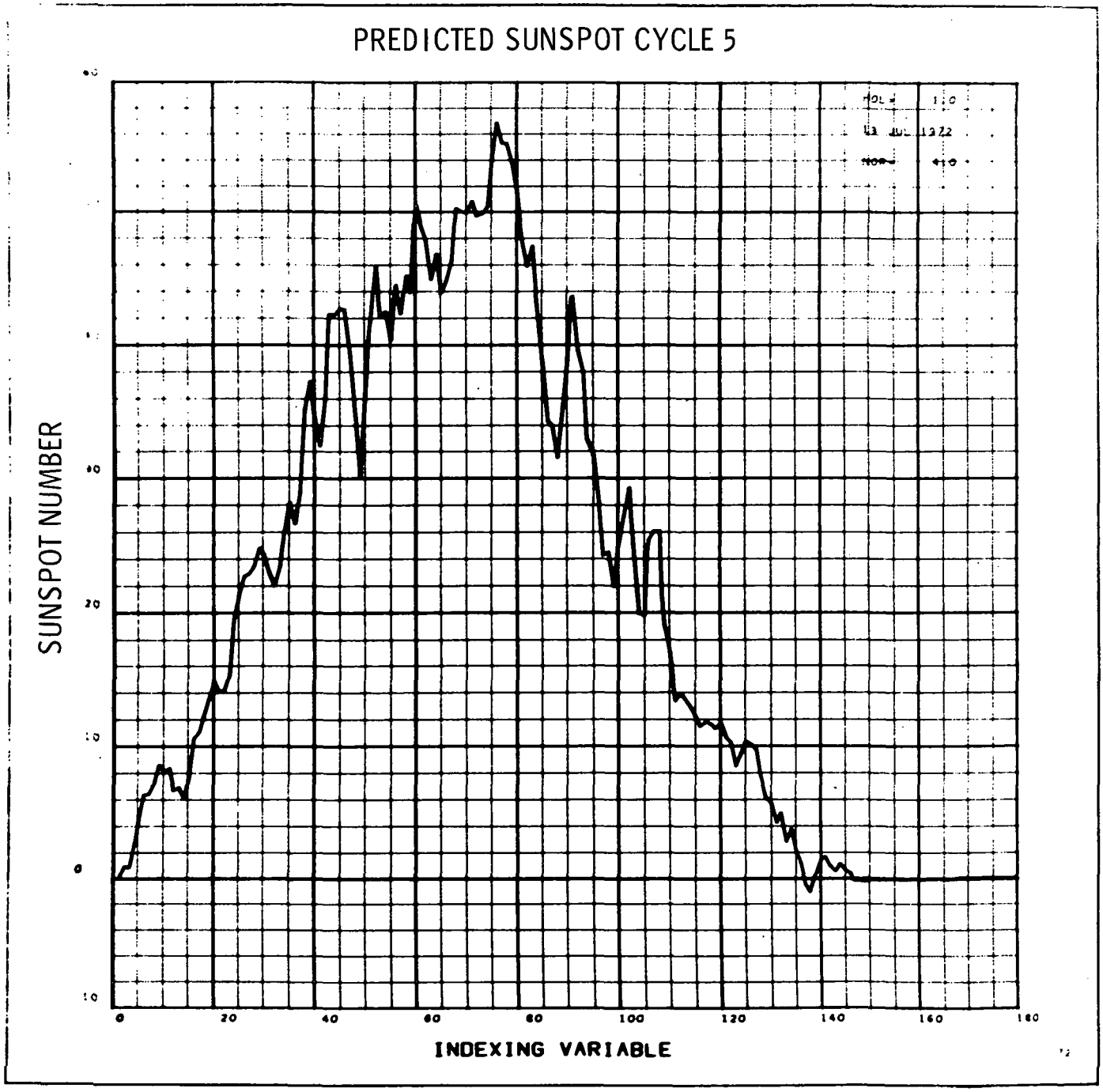

PREDICTED SUNSPOT CYCLE 5
SUNSPOT NUMBER
0
20
40
60
80
100
120
140
160
180
INDEXING VARIABLE

FIGURE D-24

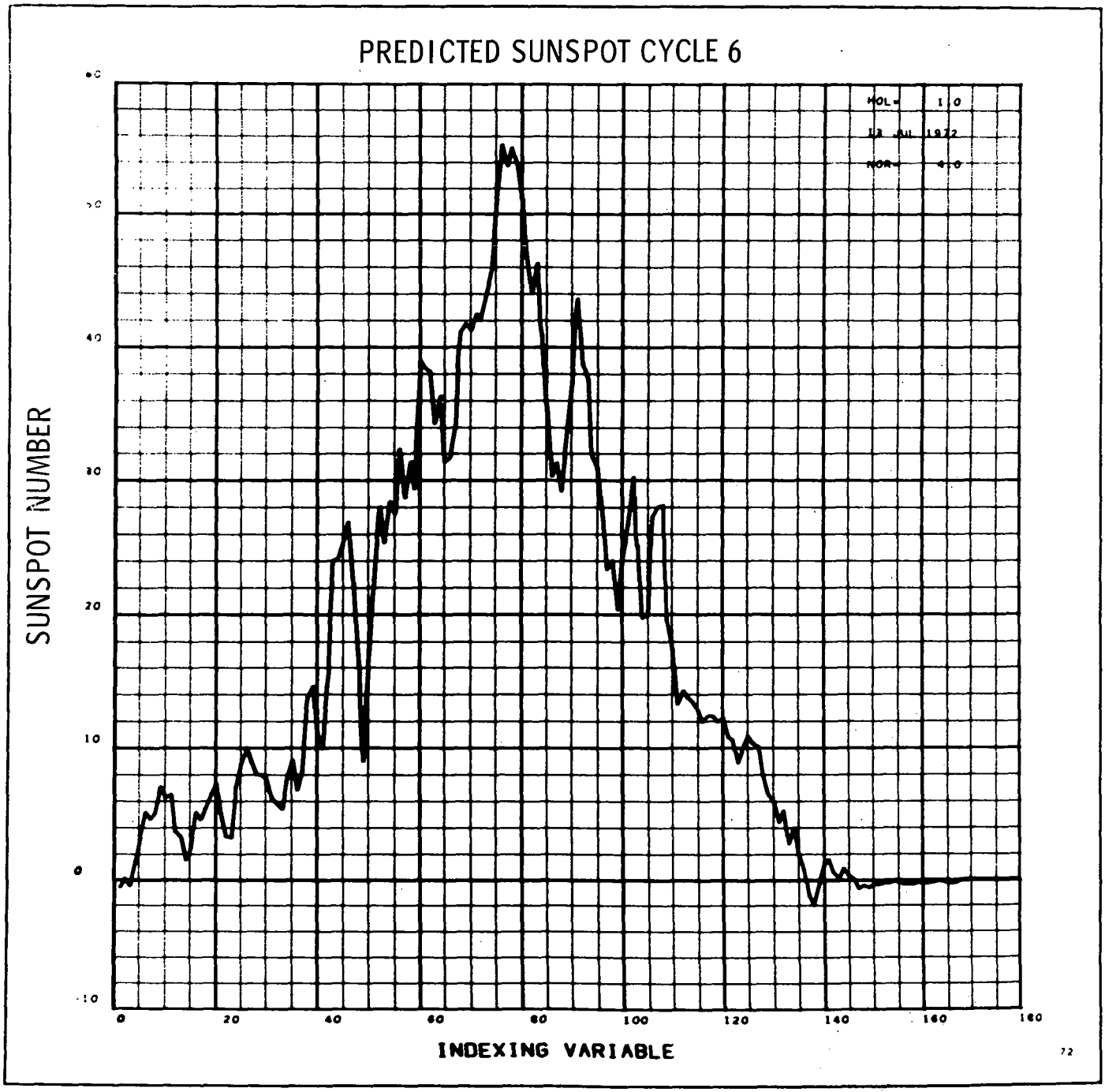

PREDICTED SUNSPOT CYCLE 6
SUNSPOT NUMBER
INDEXING VARIABLE
0
20
40
60
80
100
120
140
160
-10
10
20
30
40
72

FIGURE D-25

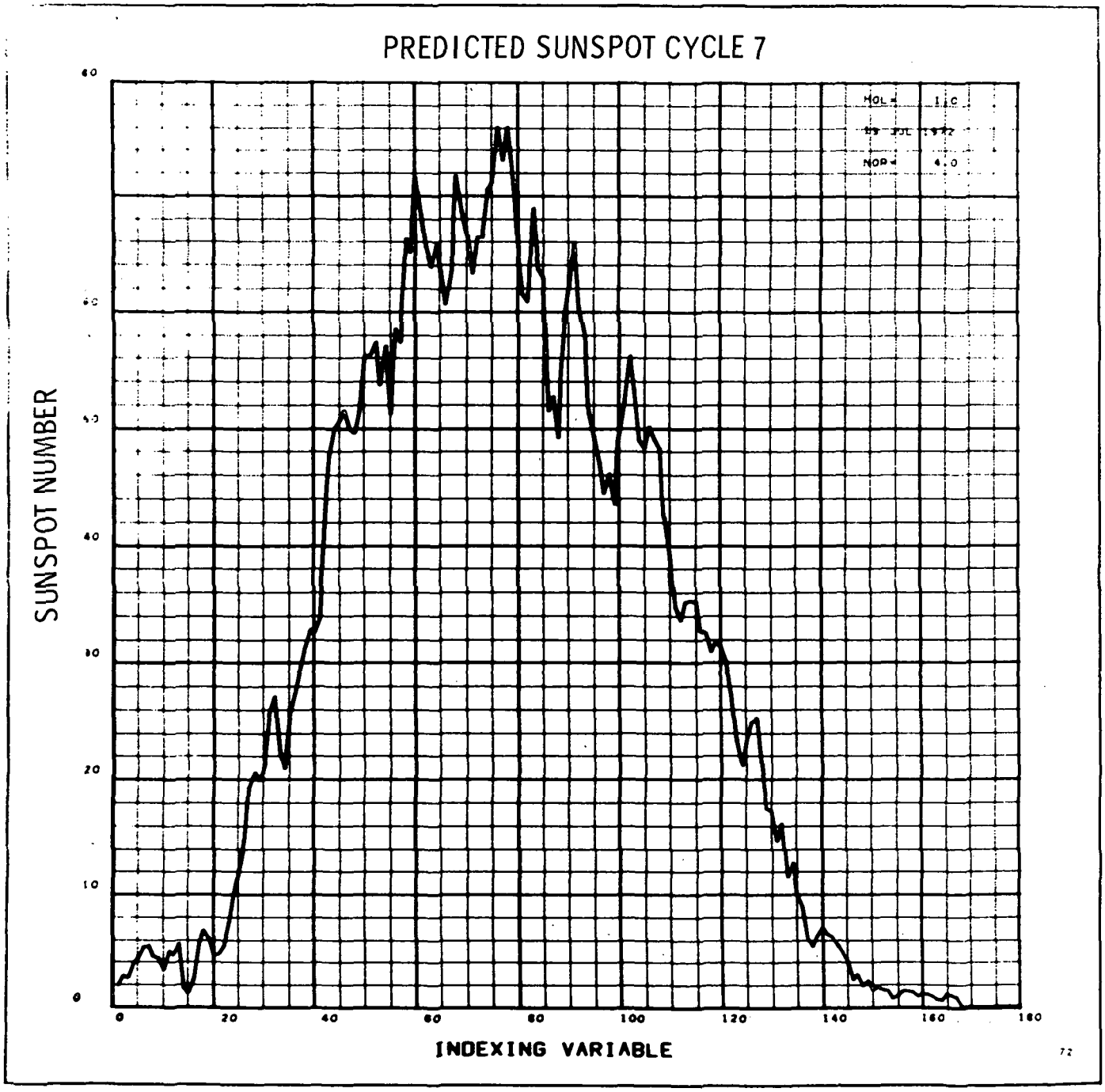

PREDICTED SUNSPOT CYCLE 7
SUNSPOT NUMBER
INDEXING VARIABLE
0
20
40
60
80
100
120
140
160
180
10
20
30
40
50
60
NOR = 4.0

FIGURE D-26

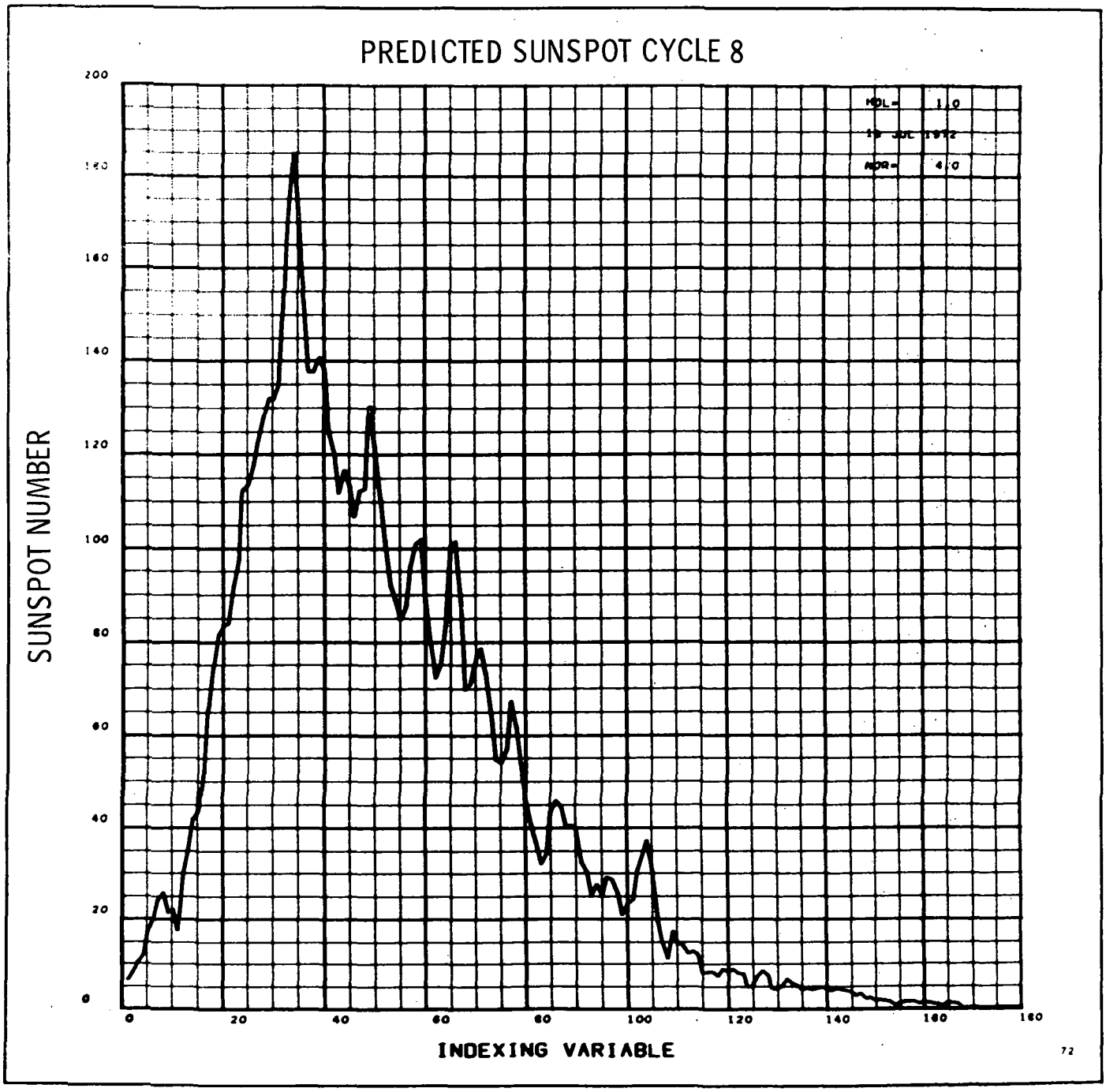

PREDICTED SUNSPOT CYCLE 8
SUNSPOT NUMBER
INDEXING VARIABLE
0
20
40
60
80
100
120
140
160
180
200
72

FIGURE D-27

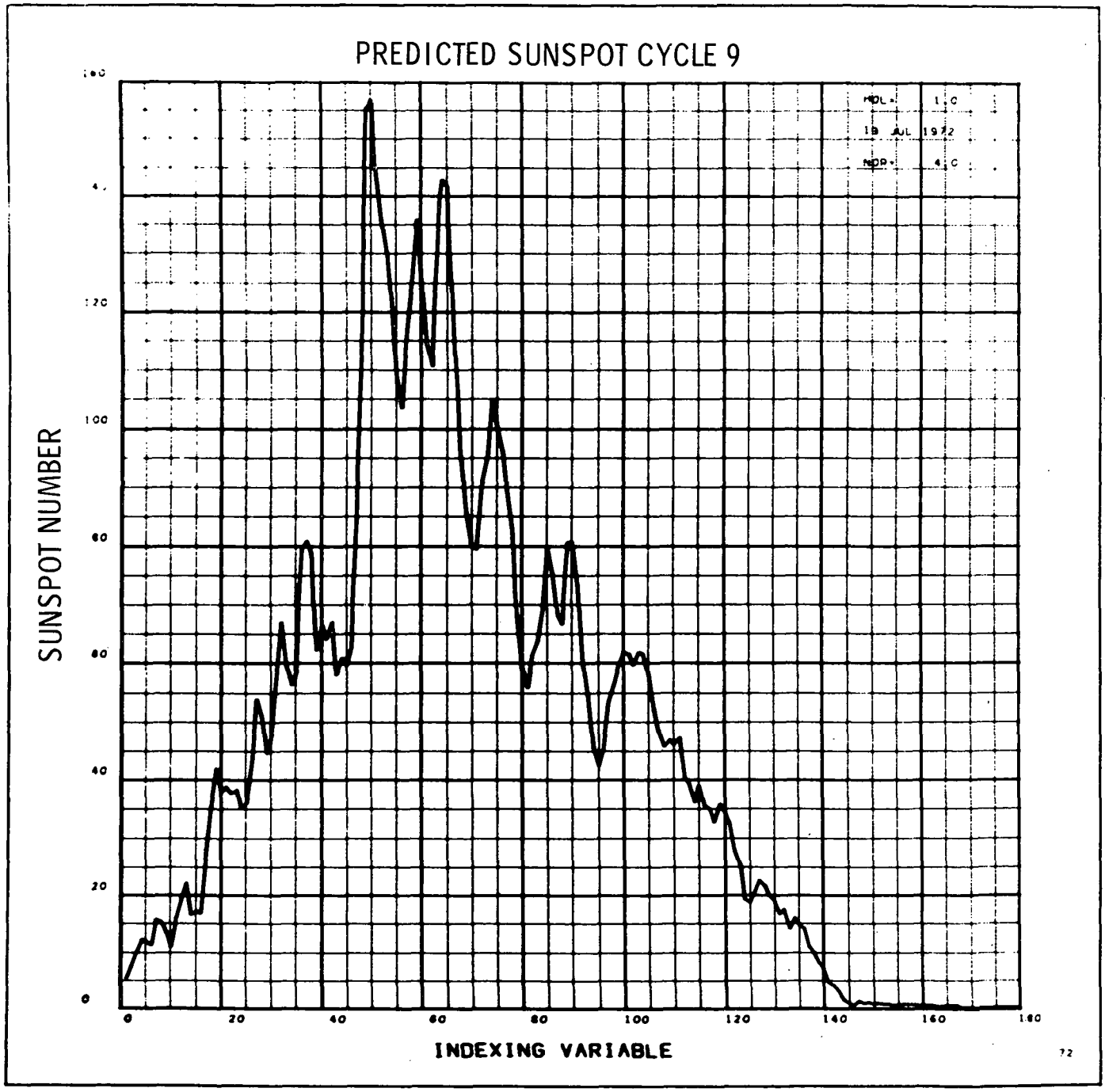

PREDICTED SUNSPOT CYCLE 9
SUNSPOT NUMBER
INDEXING VARIABLE
0
20
40
60
80
100
120
140
160
180

FIGURE D-28

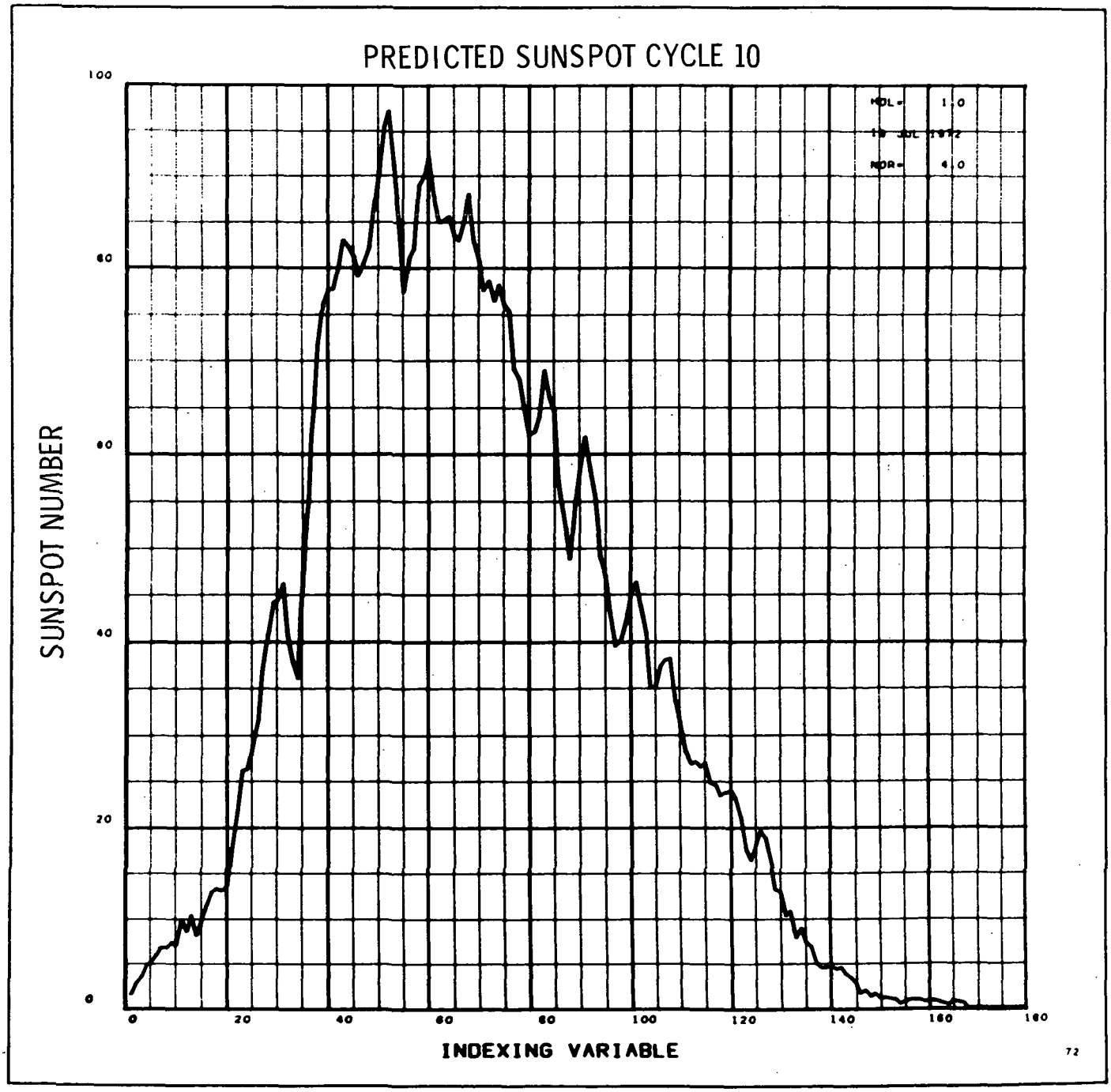

PREDICTED SUNSPOT CYCLE 10
SUNSPOT NUMBER
INDEXING VARIABLE
0
20
40
60
80
100
120
140
160
180
MDL= 1.0
MDR= 4.0
72

FIGURE D-29

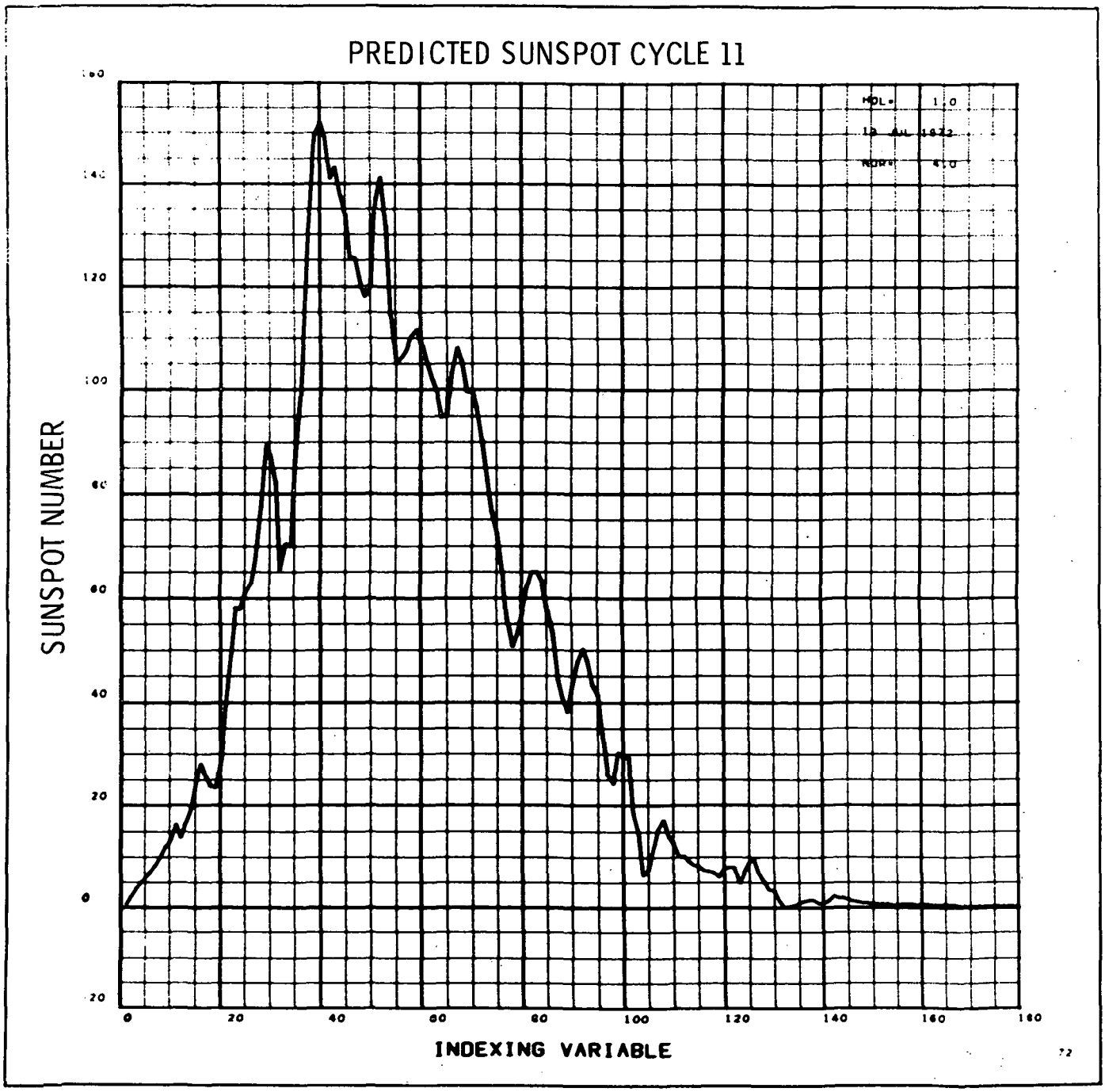

PREDICTED SUNSPOT CYCLE 11
SUNSPOT NUMBER
INDEXING VARIABLE
0
20
40
60
80
100
120
140
160
180
-20

FIGURE D-30

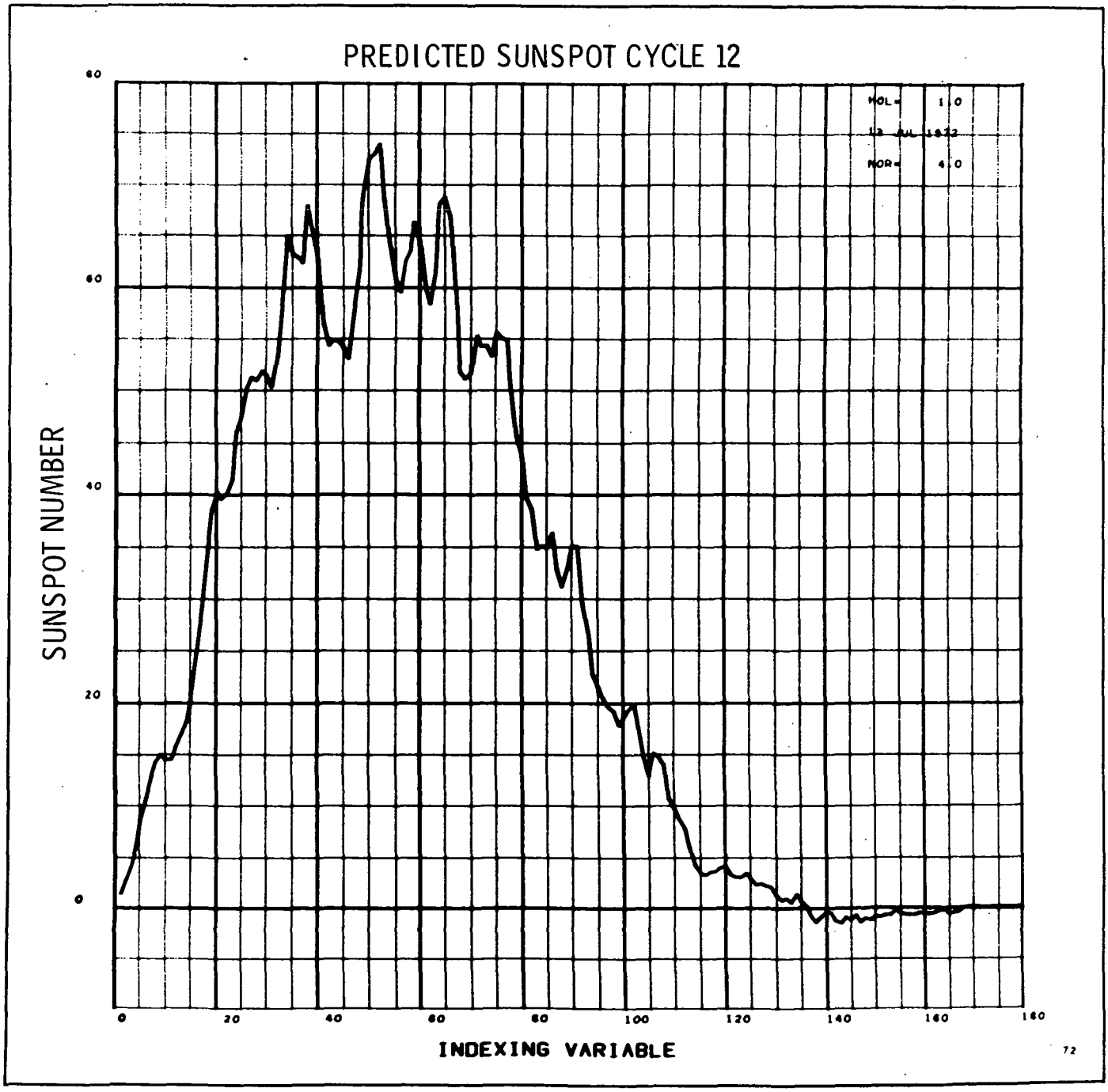

PREDICTED SUNSPOT CYCLE 12
SUNSPOT NUMBER
INDEXING VARIABLE
0
20
40
60
80
100
120
140
160
180
MOL= 1.0
NOR= 4.0
72

FIGURE D-31

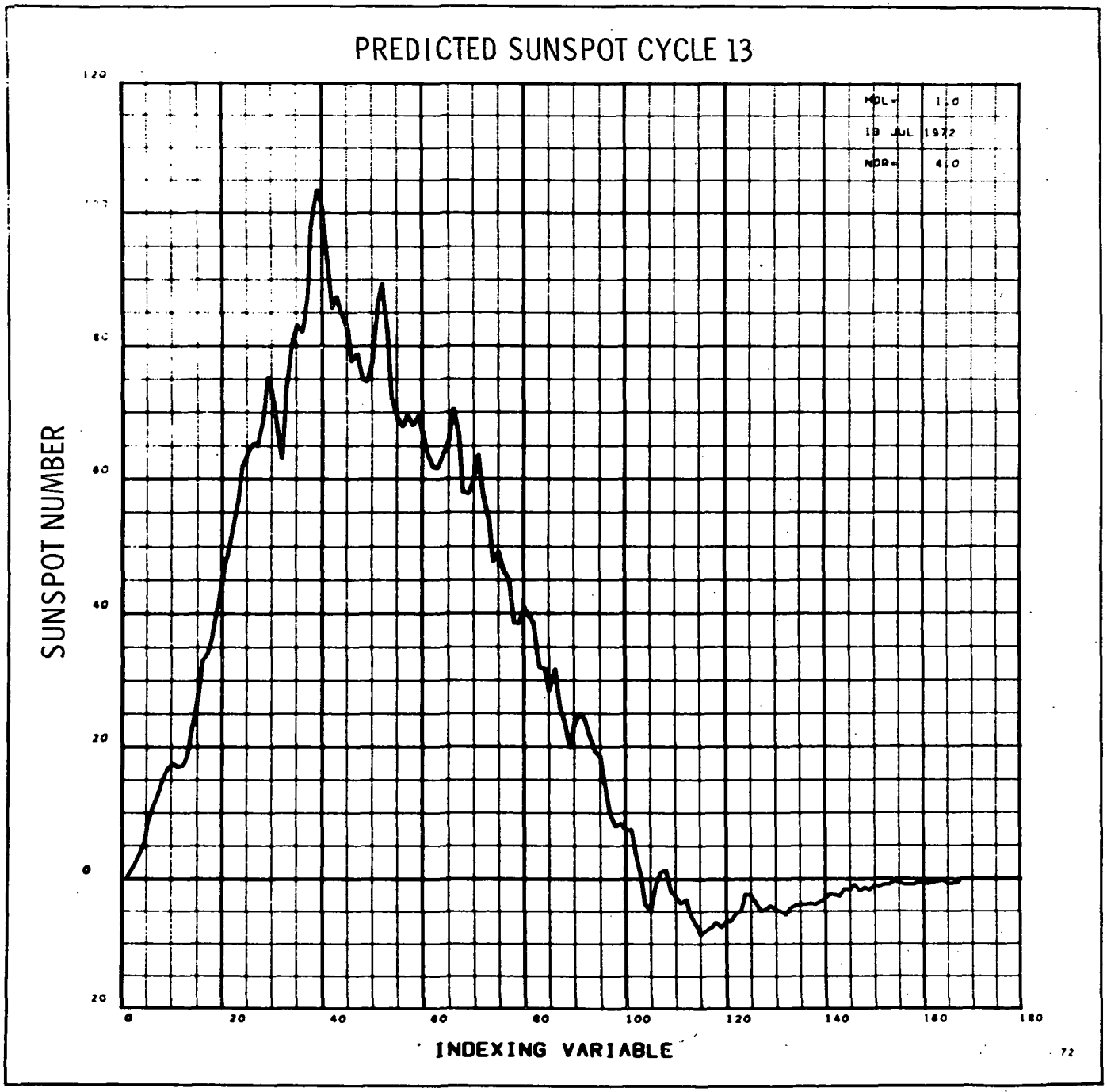

PREDICTED SUNSPOT CYCLE 13
MDL= 1.0
19 JUL 1972
NOR= 4.0
SUNSPOT NUMBER
120
80
60
40
20
0
20
0
20
40
60
80
100
120
140
160
180
INDEXING VARIABLE
72

FIGURE D-32

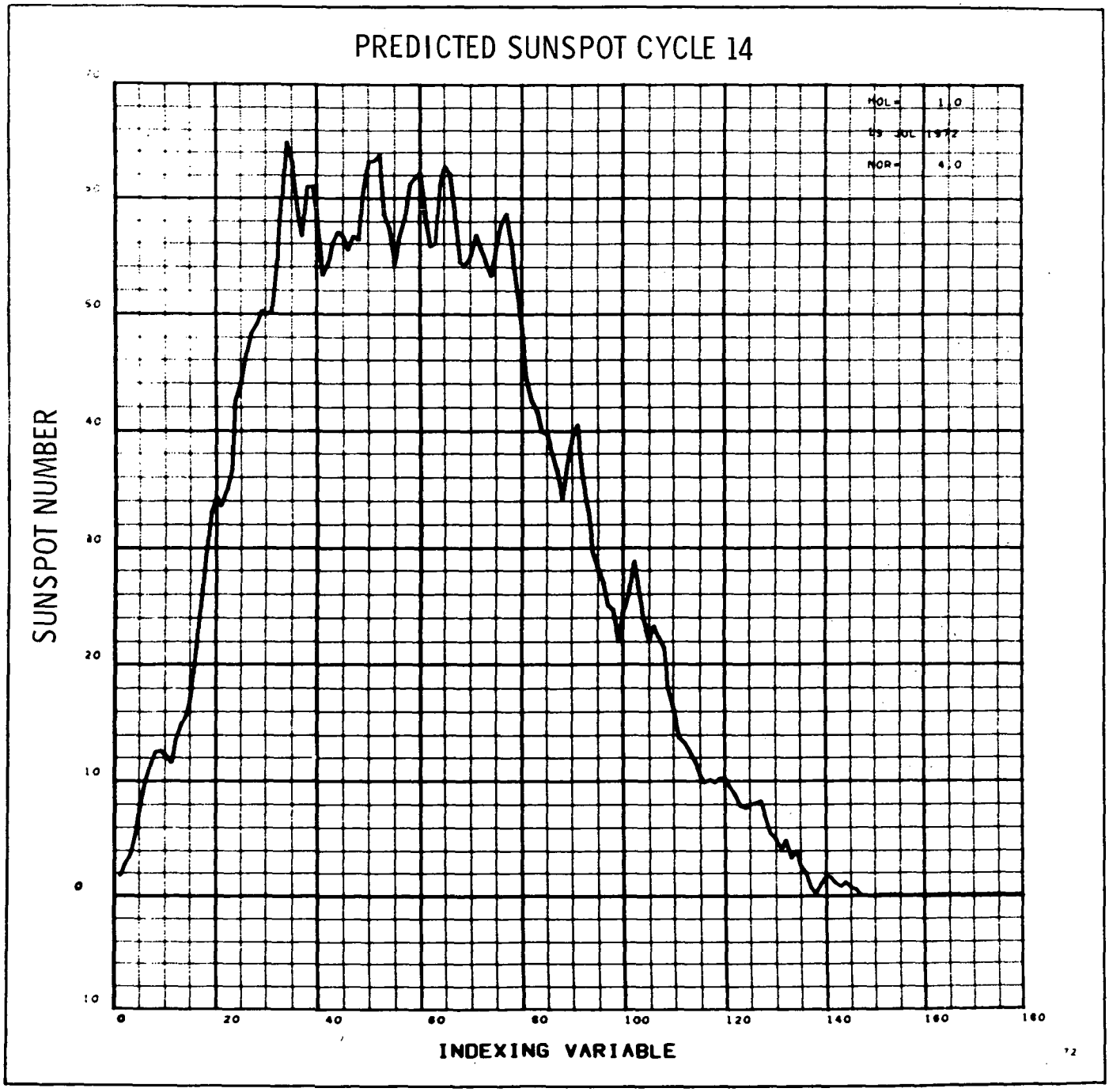

PREDICTED SUNSPOT CYCLE 14
SUNSPOT NUMBER
INDEXING VARIABLE
0
20
40
60
80
100
120
140
160
180
50
40
30
20
10
0
10

FIGURE D-33

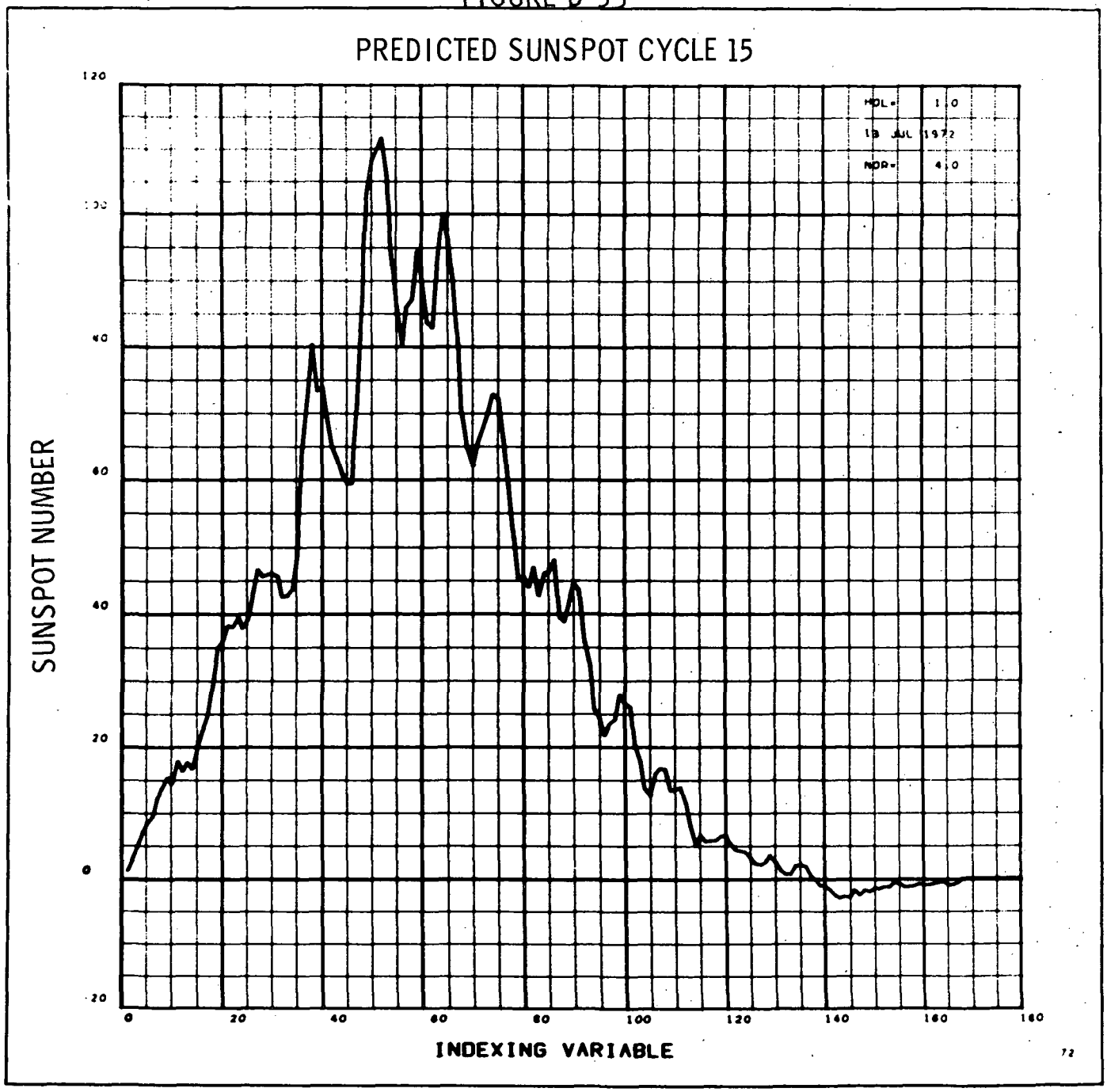

PREDICTED SUNSPOT CYCLE 15
SUNSPOT NUMBER
INDEXING VARIABLE
120
100
80
60
40
20
0
-20
0
20
40
60
80
100
120
140
160
MOL= 1.0
13 JUL 1972
NOR= 4.0
72

FIGURE D-34

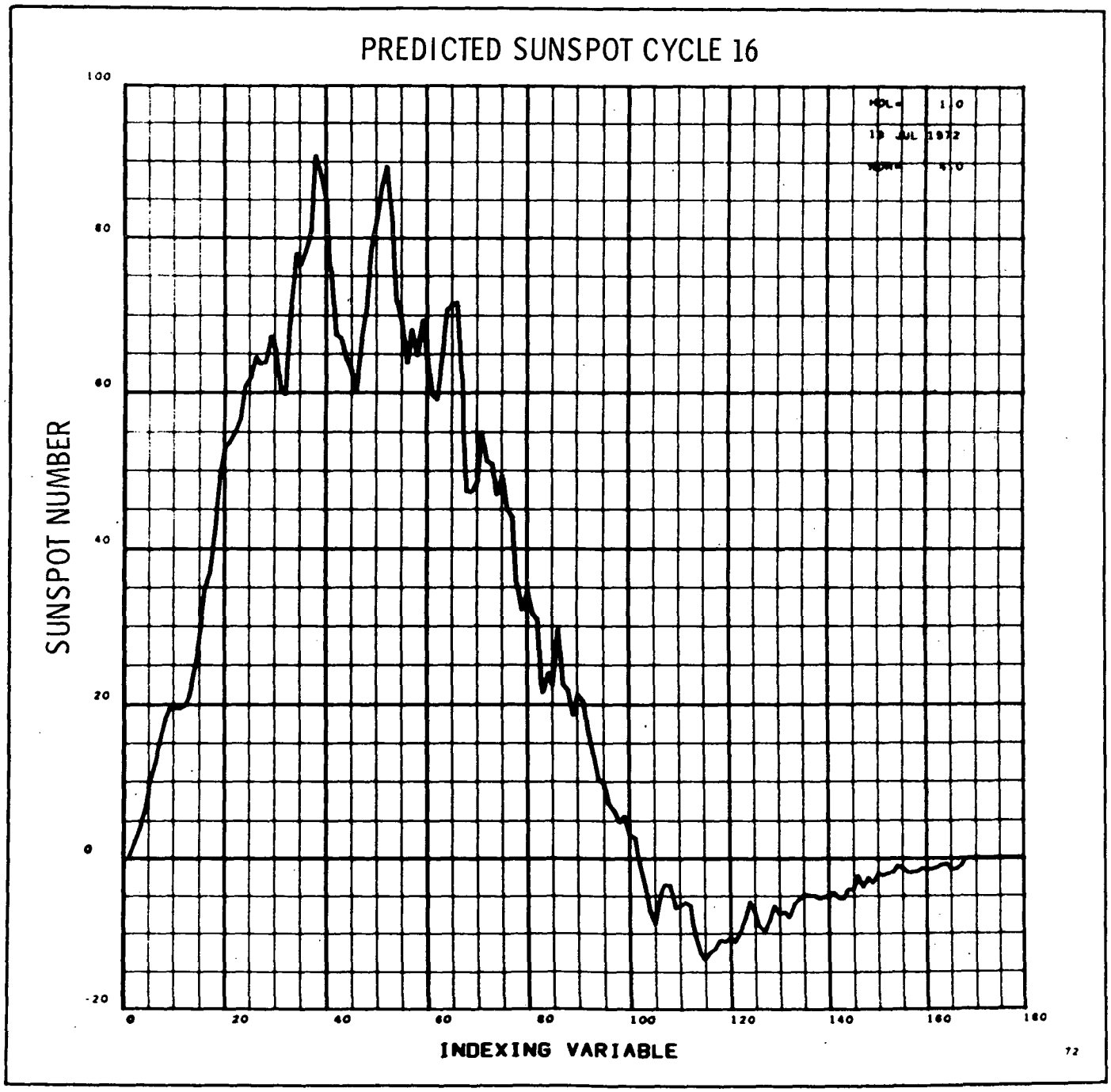
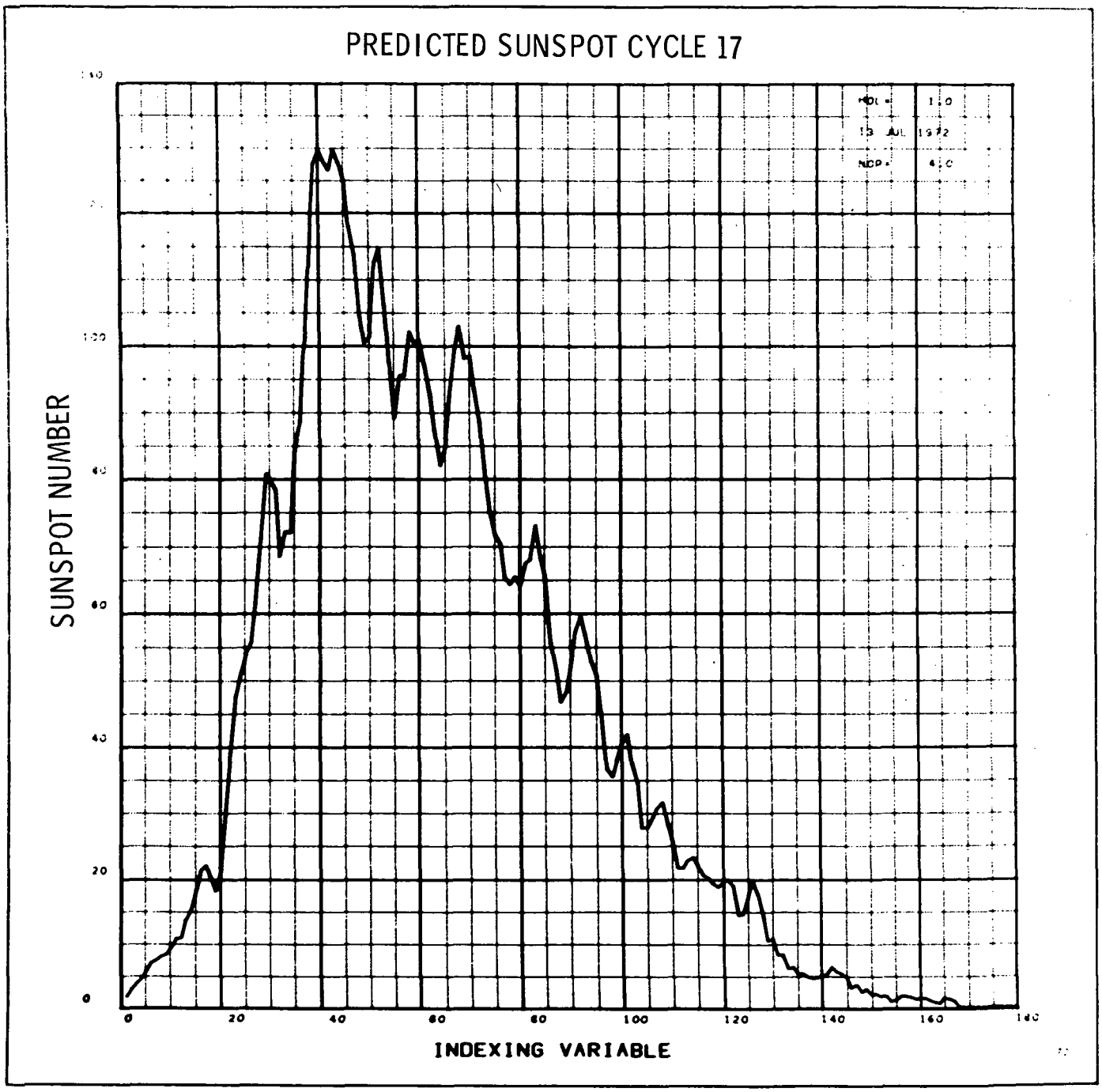

PREDICTED SUNSPOT CYCLE 16
SUNSPOT NUMBER
100
80
60
40
20
0
-20
0
20
40
60
80
100
120
140
160
180
INDEXING VARIABLE
72

FIGURE D-35


PREDICTED SUNSPOT CYCLE 17
SUNSPOT NUMBER
INDEXING VARIABLE
0
20
40
60
80
100
120
140
160
180