NRL Report 7435

Recent Computational Experience with the Method of Phase and Amplitude Diagrams

HOWARD L. WIENER

Operations Research Branch
Report 72-3
Mathematics and Information Sciences Division

August 15, 1972

NAVAL RESEARCH LABORATORY
Washington, D.C.

UNCLASSIFIED

CONTENTS

RECENT COMPUTATIONAL EXPERIENCE WITH THE METHOD OF PHASE AND AMPLITUDE DIAGRAMS

Howard L. Wiener

Operations Research Branch
Mathematics and Information Sciences Division

Abstract: Recent experience with observed time-series data have suggested the possibility of renewed interest in the analysis of periodic behavior. This report summarizes recent computational results using the method of phase and amplitude diagrams, which is an analytic and graphic technique for indicating the presence of periodic components and for determining period length, phase, and amplitude. The method has promise in areas of application involving relatively small numbers of data points or possibly small numbers of cycles (if a period is actually present). As the name implies, two sets of clues are used: phase diagrams and amplitude diagrams. The results reported here discuss the use of phase diagrams only. Findings and conclusions include the influence of an additive error component on the effectiveness of phase diagrams and recommended procedures for performing repeated confirmations of an indicated period length.

INTRODUCTION

This report discusses recent computational experience in the analysis of discrete time-series data using the method of phase and amplitude diagrams. This method is an analytic and graphic technique for investigating series in which periodic components are suspected or assumed to be present. Initial evidence for the existence of periods could have been obtained by such means as preliminary analyses or theoretical considerations. The method was described by Stumpff [1] and discussed by Blume [2], who has applied the method to data from the life sciences.

The method uses two sets of clues for both initial estimation of period length and subsequent confirmations. The analytical procedure can be modified by changing certain of the operating parameters. A given set of observed data thus may be subjected to a number of separate analyses to provide additional confirming evidence for the existence of a periodic component. The method seems useful in areas of application involving relatively small numbers of observed data or in cases where periodic components, if present, could have passed through only a small number of cycles.

Recent experiences in the analysis of the Beveridge Wheat Price Series [3] have suggested the possibility of a renewed interest in graphical periodogram-type methods. There have been few investigations dealing with the results of applying phase and amplitude diagrams to data with known statistical structures [4]. An investigation has been performed dealing with the development of a test of significance for amplitude diagrams when applied in a confirmation role [5]. This report is primarily concerned with phase diagrams; it discusses initial results of investigations into various ways of evaluating their effectiveness.

NRL Problem B01-10; Project RR 003-02-41-6152. This is a final report on one phase of the problem; work is continuing on other phases. Manuscript submitted April 24, 1972.

OUTLINE OF THE METHOD

This section describes briefly one version of the analytical method of phase and amplitude diagrams. It is assumed that there are $r + 1$ observed data points $u_0, u_1, \ldots, u_r$ which have been obtained at equally spaced time points $t = 0, 1, 2, \ldots, r$ and that a preliminary analysis has indicated the possible presence of a period of length T. The method is then applied to obtain confirming evidence on the existence of the period T. The following are the main steps of the procedure.

1. Select a positive integer n_s such that $n_s \geqslant 1$ and $n_s T < (r + 1)$.

2. Define the *analysis interval* $p = n_s T$.

3. Construct the *sample function* $f(t)$ by connecting consecutive observed data points with straight lines.

4. Select a positive integer m such that $1 \leqslant n_s \leqslant m$.

5. For each $n = 1, 2, \ldots, m$ and each $q = 0, 1, 2, \ldots, r - p$, compute

$$a(n; q; p) = \frac{2}{p} \int_q^{q+p} f(t) \cos \left[\frac{2\pi n}{p} (t - q)\right] dt$$

and

$$b(n; q; p) = \frac{2}{p} \int_q^{q+p} f(t) \sin \left[\frac{2\pi n}{p} (t - q)\right] dt.$$

6. Plot the values of

$$\varphi(n; q; p) = \tan^{-1} \left(\frac{a(n; q; p)}{b(n; q; p)}\right)$$

as an increasing function of both n and q, with the abscissa comprising the q values. The plotted values of φ vs q constitute the *phase diagram.* (In a general computational procedure, the following process is used to make φ an increasing function of q. Whenever the computation yields the relation $\varphi(n; q; p) \geqslant \varphi(n; q + 1; p)$, the smallest positive integer k such that $\varphi(n; q; p) < \varphi(n; q + 1; p) + k\pi$ is obtained, and $\varphi(n; q + 1; p)$ is redefined to have the value $\varphi(n; q + 1; p) + k\pi$. For the results discussed here, we have followed Blume [2] and computed φ in degrees; thus $k\pi$ is calculated in terms of multiples of $180°$.)

7. Using the φ values in the phase diagram, obtain related straight lines $\hat{\varphi}(n; q; p)$. From the slope of $\hat{\varphi}(n_s; q; p)$, obtain an *estimated period* $\hat{T}$ by $\hat{T} = \text{slope}/2\pi$ or $\hat{T} = \text{slope}/360°$, whichever is appropriate. For the computations discussed here, $\hat{\varphi}$ has been obtained by least-squares linear regression.

8. Plot the values of

$$c(n; q; p) = (a^2(n; q; p) + b^2(n; q; p))^{1/2}$$

for all q and n, with the q values along the abscissa. This plot is called the *amplitude diagram.*

9. The following clues are used as confirming evidence for the existence of the period T:

   9.1 the estimated period $\hat{T}$ obtained from $\hat{\varphi}(n_s; q; p)$ is reasonably near T;

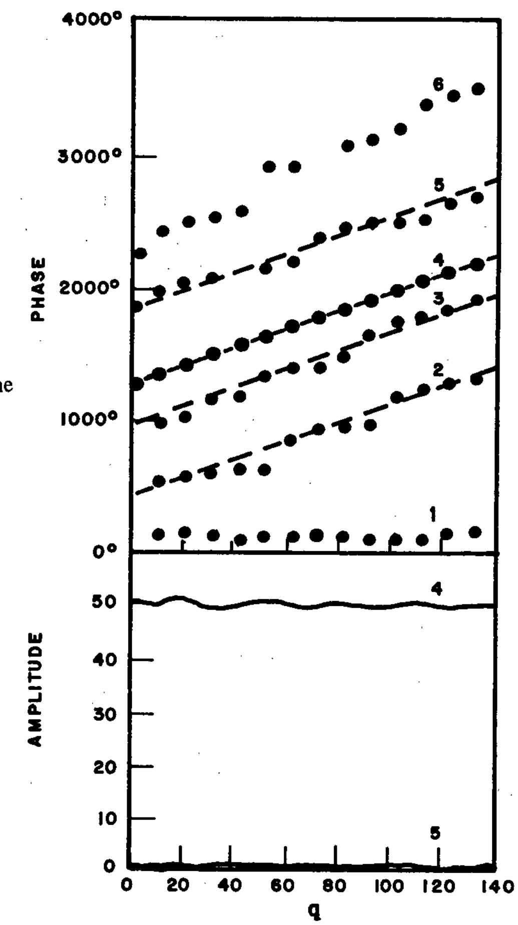

Fig. 1 – Phase and amplitude diagrams (adapted from Blume [2]); $g(t) = 50 \sin (2\pi t/56)$; $p = 224$

9.2 there is a reasonable fit between the computed phase values $\varphi(n_s; q; p)$ and the estimated line $\hat{\varphi}(n_s; q; p)$;

9.3 the values of $c(n_s; q; p)$ are relatively greater than the other c values.

When these criteria are assumed to be met, the presence of period T is accepted, the amplitude of the underlying periodic function is obtained from the $c(n_s; q; p)$ values, and the phase is estimated from $\varphi(n_s; 0; p)$. As mentioned previously, in the present case we mainly considered the phase diagram, so clues 9.1 and 9.2 were of most importance.

Figure 1, reproduced from [2], shows an analysis of data from the curve

$$g(t) = 50 \sin\left(\frac{2\pi}{56} t\right)$$

using an analysis interval of $p = 224$. The upper diagram is the phase diagram; the lower one is the amplitude diagram. Note the straight-line fit in the phase diagram for the line labeled 4 and the poor fits for the other lines.

UNDERLYING MODEL AND EVALUATION PARAMETERS

Data sets of size $r + 1 = 65$ were used for the investigation. The basic model used for generating the data was

$$u_t = \sin\left(\frac{\pi}{4} t\right) + \epsilon_t,\ t = 0, 1, \ldots, 64, \tag{1}$$

where ϵ_t was the random error, assumed to comprise independent and identically distributed random variables, each with a mean equal to zero and variance σ^2. The adequacy of the phase diagrams could thus be measured by the extent to which a period of $T = 8$ was indicated.

For each given set of conditions, the analysis was repeated for a number of computer-generated sets of data. Because many expected applications of the method will involve relatively few data sets, only relatively small samples were used. For the most part, samples of 15 sets were examined, although in a number of test cases, only 10 sets were generated. The use of a number of samples with the same underlying structure is relevant for analyzing cases where observational data of limited extent can be obtained at different times but under identical circumstances. In many instances – medical data for example – such repetition may not be possible, and a single limited data set will be all that is available.

Values assumed for the standard deviation σ of the error component were 1/6, 2/6, 3/6, . . . , 10/6. The main statistics of the phase-diagram results which were examined were the sample mean indicated period $M(\hat{T})$ and the sample variance and standard deviation of the indicated periods, $V(\hat{T})$ and $S(\hat{T})$ respectively. In addition, the *residual mean square* [6, p. 27], denoted by s^2 and defined by

$$s^2 = \frac{\displaystyle\sum_{q=0}^{r-p} (\varphi(n_s; q; p) - \hat{\varphi}(n_s; q; p))^2}{r - p - 1}$$

was used in evaluating the degree to which the regression line $\hat{\varphi}$ fit the phase-diagram data points φ.

EFFECTS OF INCREASED ERROR COMPONENT

In the basic model (1), if the standard deviation σ is arbitrarily small, the observed data become increasingly representative of data obtained from just a sine curve. The presence of a periodic component is obvious, as is the period length. Theoretical considerations then dictate that for any data set of the type (1), selection of an analysis interval $p = n_s T$ and examination of the values $\varphi(n_s; q; p)$ should result in an indicated period $\hat{T}$ arbitrarily close to T. Moreover, if the experiment were to be repeated for a number of sample data sets, it should be expected that the resulting mean value of $\hat{T}$, $M(\hat{T})$, is close to T, and the standard deviation $S(\hat{T})$ should be arbitrarily small.

If the situation were reversed and σ were permitted to increase, then the effects should also be reversed. Sample data sets will bear little similarity to data from sine curves. Moreover, it is quite likely that $M(\hat{T})$ values will be obtained which are not necessarily close to T and that $S(\hat{T})$ values will be obtained which are not arbitrarily small.

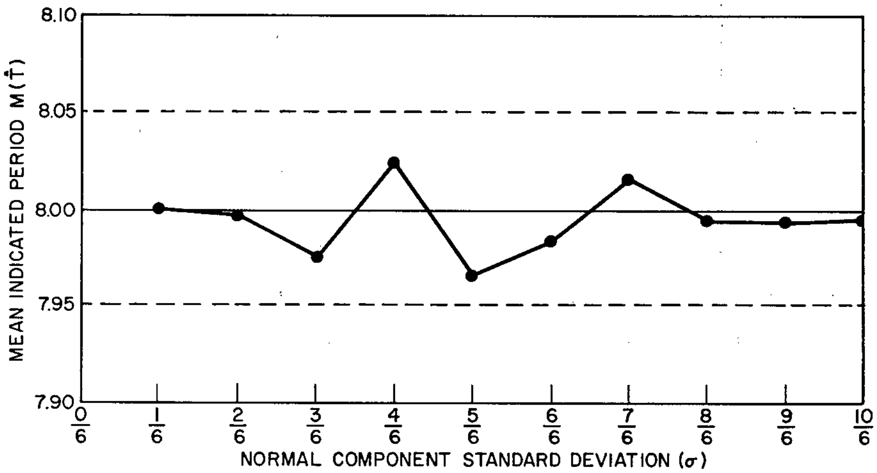

Fig. 2 – Mean indicated period as a function of the error component; eight cycles of the sine curve; actual period = 8; p = 32

Mean Estimated Period

Figure 2 indicates the mean estimated period $M(\hat{T})$ for observed data sets of 65 points as defined in the model (1), analyzed using an analysis interval of $p = 32$ and using $n_s = 4$. For small values of σ, the mean was calculated on the basis of 10 observed data sets; for relatively large values of σ, 15 sets were used.

In general, the observed values of $M(\hat{T})$ differ from $T = 8.0$ by no more than 0.05. Nearly all values are within 0.025 of T. It is not intuitively clear that such differences would be always sufficiently small for confirming the presence of the period $T = 8.0$; one can, however, easily conceive of situations in which such a difference would be acceptable. From the results shown in Fig. 2, it is not clear that increases in σ must necessarily result in changes in $M(\hat{T})$. Although values of σ of 1/3 or less are related to $M(\hat{T})$ values extremely close to 8.0, all the σ values between 1/6 and 10/6 appear to be related to $M(\hat{T})$ values no more than 0.05 from 8.0. No trend of $M(\hat{T})$ as a function of σ is evident from these results. It is appealing to propose that $\hat{T}$ is an unbiased estimator of T; however, additional work will be necessary to either prove or at least provide further support for this hypothesis.

It can be speculated that in cases where σ is known to be large and where numerous sample data sets can be obtained, confirmations (or initial estimates of T, for that matter) can be obtained by applying phase-diagram methods to relatively large numbers of samples; in this way, the mean $M(\hat{T})$ might be expected to converge to a relatively consistent estimate of T. The analytical method could then be employed using phase values only, with no attention being paid to the values in the amplitude diagram. One cannot say, however, that the amplitude diagram can be completely neglected and that the phase diagram can be accepted as the sole source of all information on the estimated period. Increases in σ can be expected to cause significant dispersions among the calculated values of $\varphi(n_s; q; p)$. Then, even though the calculated straight line $\hat{\varphi}$ may generate an estimate $\hat{T}$ reasonably close to T, goodness-of-fit considerations between the φ values and the estimated line may preclude accepting $\hat{T}$ as a reasonable result. This point will be discussed further later in the report.

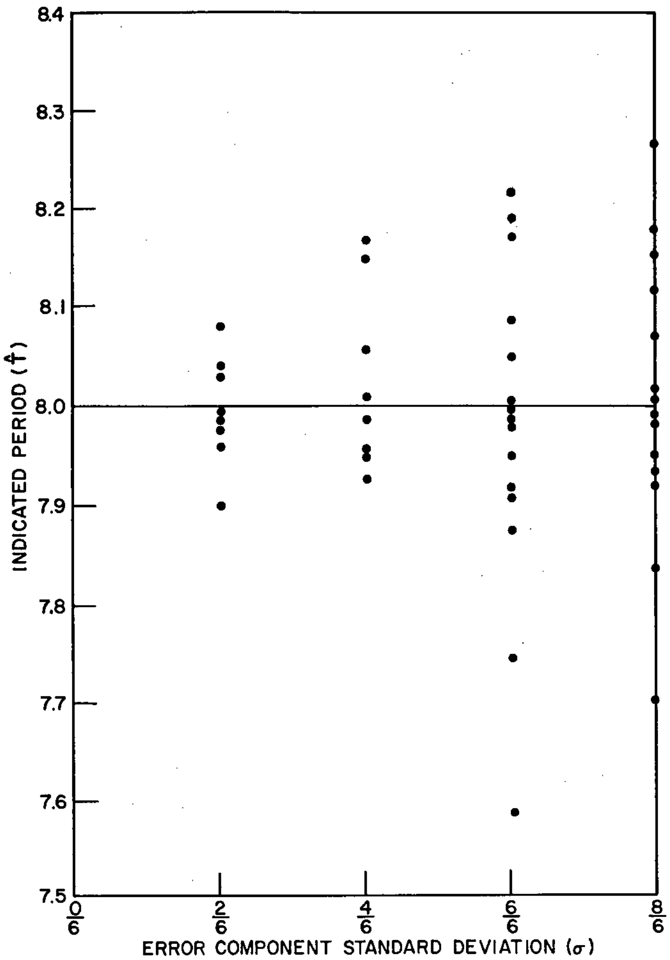

Fig. 3 – Estimated period $\hat{T}$ as a function of the error component; actual period = 8; p = 32

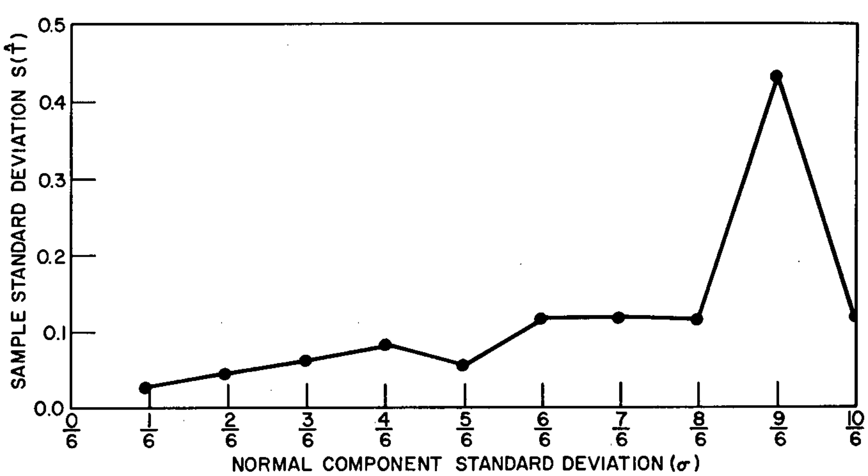

Fig. 4 – Sample standard deviation as a function of the error component; eight cycles of the sine curve; actual period = 8; p = 32

Standard Deviation of the Estimated Period

Although the sample mean $M(\hat{T})$ gives evidence of being unaffected by changes in σ, the sample standard deviation $S(\hat{T})$ is evidently strongly affected by such changes. Since σ does have an influence on the observed data values and seemingly little influence on the values of $M(\hat{T})$, it should be expected that changes in σ should produce noticeable changes in $S(\hat{T})$.

Figure 3 illustrates representative observed values of $\hat{T}$ obtained for various values of σ. As σ increases, the spread in the observed values of $\hat{T}$ increases correspondingly. For $\sigma = 3/6$, about 70 percent of all observed samples were found to have an indicated period between 7.95 and 8.05. As σ increased to 6/6, this proportion decreased to 60 percent, and for $\sigma = 8/6$ the proportion was about 33 percent. Figure 4 shows $S(\hat{T})$ for values of σ up to 10/6. If one takes into account the possibility of sampling fluctuations, especially for the values obtained for $\sigma = 5/6$ and 9/6, it may be assumed that $S(\hat{T})$ increases monotonically with σ. For values of σ below 4/6, $S(\hat{T})$ seems to increase linearly with σ; however, the linear trend is not continued into the larger values.

These results indicate some of the difficulties to be encountered for fairly noisy data if confirmations are performed no more than once on a given set of observed data. The increased variance in $\hat{T}$ will result in increased probabilities that a given value of $\hat{T}$ will be significantly different from T. Thus the presence of a periodic component may be completely overlooked (unconfirmed), or the period length may be incorrectly identified. The ability of the method to permit additional confirmations can be used to lower the probability of such errors.

VALUE OF REPEATED CONFIRMATIONS

In concept, repeated confirming analyses, performed using different analysis interval lengths, should permit the generation of additional evidence when a periodic component is present and, by lack of sustaining evidence, permit the identification of those data sets in which there is no periodic component. For a given size of the data set and an assumed period T, since repeated confirmations are characterized by the choice of $p = n_s T$, it is of interest to investigate whether there is a reasonable order in which the p's might be selected. This section indicates that for the model being investigated, confirming analyses are best performed when the sequence of p values used is increasing.

To illustrate, Figs. 5 and 6 present the results of analyzing observed data samples using analysis intervals of $p = 8, 16, 24, \ldots, 48$ and sample data sets of 15 samples. Figure 5 indicates the sample means $M(\hat{T})$ for underlying $\sigma = 6/6$ and $\sigma = 8/6$. This figure indicates that as p increases, the means $M(\hat{T})$ tend to cluster closer to the actual value of T than when p is small. For the case of $\sigma = 6/6$, $M(\hat{T})$ appears to be within 0.05 of T for all values of p no less than 32; the same result holds for the case $\sigma = 8/6$ for all p which are at least 16. Regardless of the actual values, it seems prudent to avoid using $p = 8$ and the associated value $n_s = 1$. Figure 6 indicates values of the sample standard deviation $S(\hat{T})$. For $\sigma = 6/6$, as p increases, $S(\hat{T})$ has a tendency to decrease, although not in a strictly monotonic fashion; the same is also true of the case $\sigma = 8/6$. For the case of $\sigma = 6/6$, $S(\hat{T})$ is generally less than 0.20; for the larger p values, $S(\hat{T})$ becomes less than 0.10. For the case of $\sigma = 8/6$, $S(\hat{T})$ is generally less than 0.22; as p increases, $S(\hat{T})$ rises also but descends to a value of about 0.15 at $p = 48$.

As mentioned previously, these results tend to support the judgement that for any family of observed data sets, the confirmation process is best carried out by initially using an analysis interval of moderate size and by increasing the size of the analysis interval if additional confirmations are necessary.

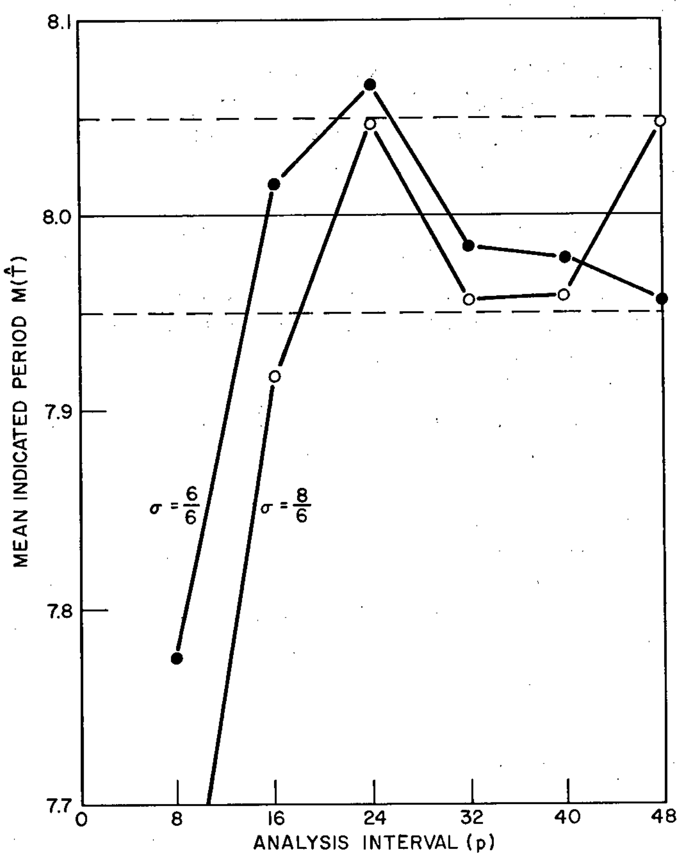

Fig. 5 – Mean indicated period as a function of the analysis interval; eight cycles of the sine curve; actual period = 8

QUALITY OF REGRESSION LINES

The mean residual sum of squares s^2 was used as a measure for examining the degree to which the calculated regression line $\hat{\varphi}$ corresponds to the phase values φ obtained by the analysis. In cases where the calculated values of $\varphi(n_s; q; p)$ lie close to the straight line $\hat{\varphi}(n_s; q; p)$, the differences $\varphi(n_s; q; p) - \hat{\varphi}(n_s; q; p)$ will be small, and thus s^2 will be small; consequently, small values of s^2 can be taken as indicating reasonably good fits between the computed phase values φ and the estimated line $\hat{\varphi}$. Conversely, large values of s^2 indicate the absence of a reasonable fit. This section discusses some facets of the behavior of s^2 as a function of the underlying σ and the analysis interval p.

Figure 7 contains approximations of the sample distribution functions of observed values of s^2 for an analysis interval of $p = 32$, with corresponding $n_s = 4$, and for error-component standard deviations $\sigma = 1/6$, $6/6$, and $9/6$. Values of s^2 are smallest for small values of σ, a result which is consistent with the fact that the observed data sets are only slightly disturbed from the theoretical sine-curve values. The median value of s^2 is about 0.7. Phase diagrams thus conform to practical results in the sense that the smaller the value of σ, the easier it should be to detect or confirm the presence of a periodic component.

UNCLASSIFIED

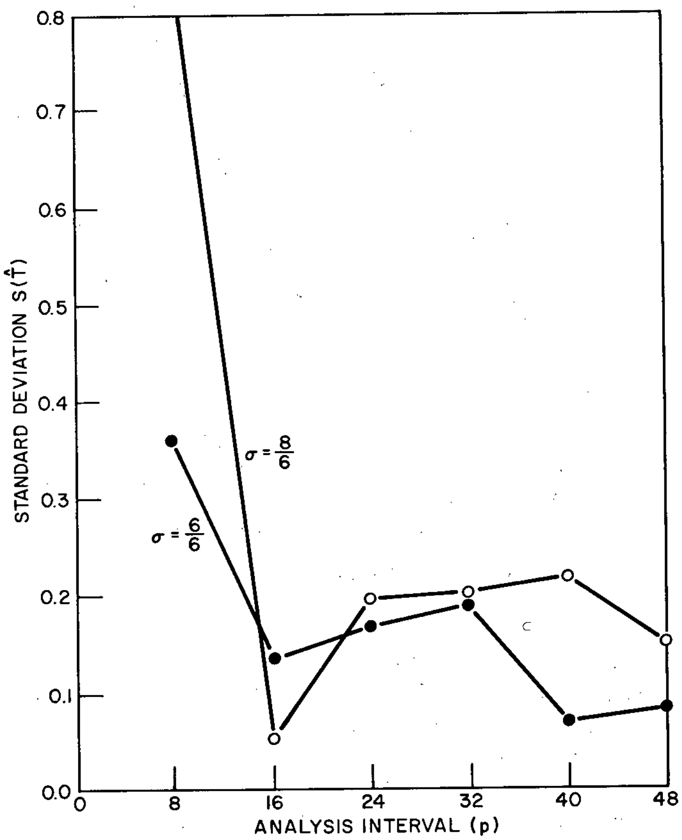

Fig. 6 – Sample standard deviation as a function of the analysis interval; eight cycles of the sine curve; actual period = 8

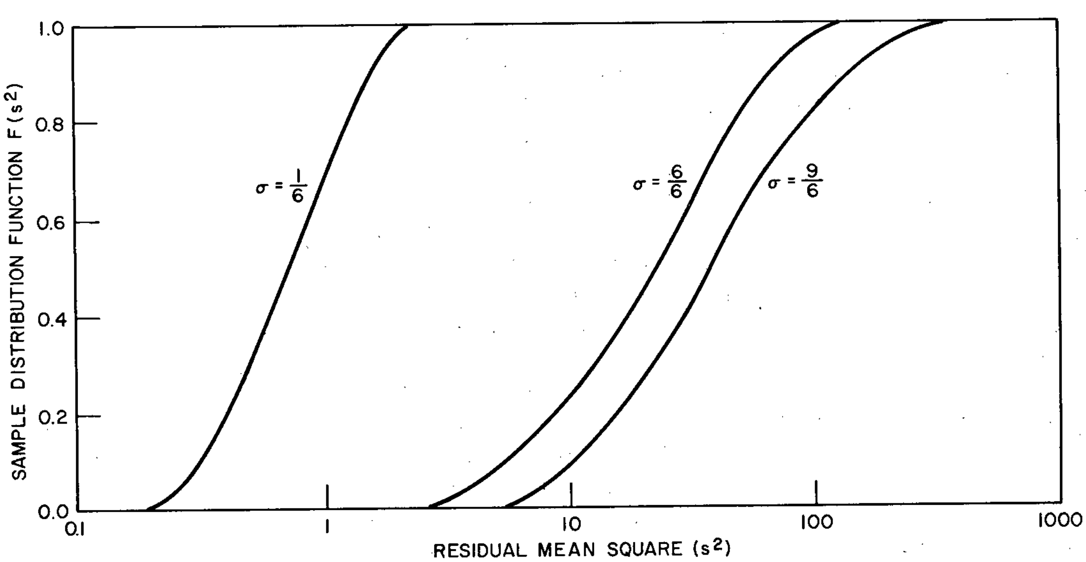

Fig. 7 – Sample distribution function of s^2 as a function of σ; p = 32

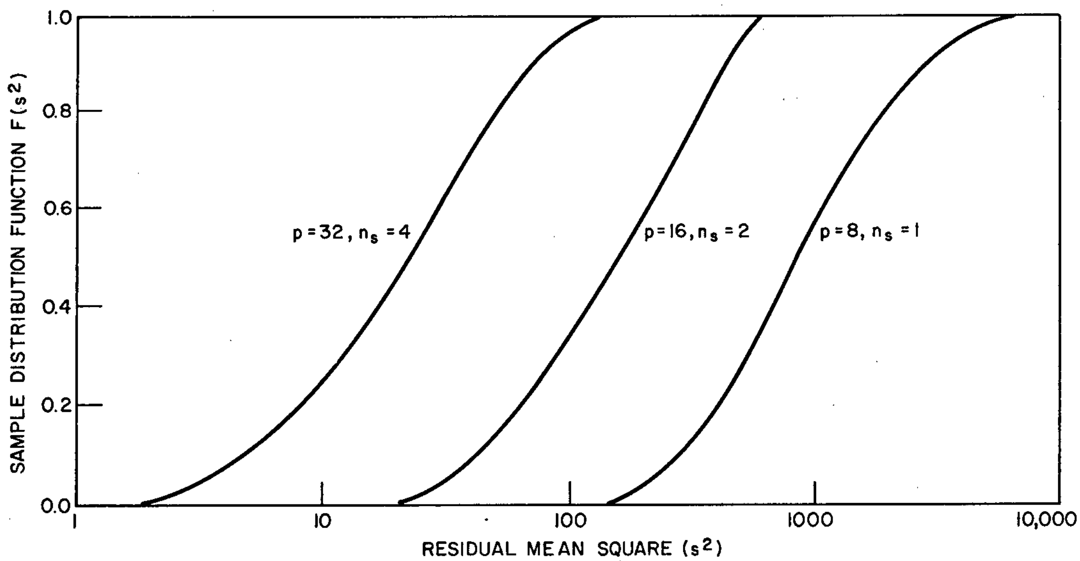

Fig. 8 – Sample distribution function of s^2 as a function of p; σ = 1.0

As σ increases, the values of s^2 also increase. For $\sigma = 6/6$, the median value is about 22; increasing σ to 9/6 increases the median to about 40. Knowledge of this effect may, however, be of slight value in attempting to analyze observed data having unknown underlying structure. If multiple samples of data sets can be obtained under identical conditions, then, as has been stated, the sample mean $M(\hat{T})$ can be used to estimate T. It should be recalled, however, that the method relies on a goodness-of-fit criteria for straight-line estimation as well as on obtaining $M(\hat{T})$ values close to T. With no evidence other than what can be seen in the phase diagram, increases in σ tend to lower the degree of willingness one would have in accepting the straight line $\hat{\varphi}$ as confirming evidence. Additional evidence would be required, for example, by using the data in the amplitude diagram or by using some form of statistical test for the analysis of phase-diagram values.

Statistical tests for use with the phase diagram would be most easily developed on the basis of a set of stochastically independent random variables. From the description of both the model (1) and the methodology, it is clear that for given p and n_s, independence among the various $\varphi(n_s; q; p)$ values occurs only for values of q which are at least p apart. For relatively large p, there would be small numbers of independent values, thus limiting the effectiveness of a test. On the other hand, whereas lower values of p would produce greater numbers of independent variables, the use of such lower values tends to run counter to the result indicated previously that subsequent confirmations are enhanced by increasing p. Additional evidence of the value of taking increasing values of p is shown in Fig. 8, which contains approximations of the observed sample distribution functions of the s^2 values, for $\sigma = 6/6$ and for p = 8.0, 16.0, and 32.0. (One should notice that the horizontal axes in Figs. 7 and 8 have different scales.) The evidence that taking increasing values of p generates lower values of s^2 is consistent with earlier findings and thus increases the acceptance of the estimate $\hat{T}$. The residual mean square s^2 has a relatively large median value of around 900 for $p = 32$, whereas for $p = 32$, it has a smaller median value of roughly 21. Conditions under which the phase-diagram results could be accepted without resorting to the amplitud diagram are still undetermined. As mentioned, the development of a statistical test would be beneficial, but regardless whether such a test procedure is available, further computational experience with the method will still be required.

FINDINGS AND CONCLUSIONS

The following comments relate to the effectiveness of only the phase diagrams when the method of phase and amplitude diagrams is used in the role of confirming the presence of a supposed periodic component.

1. As expected, the effectiveness of phase diagrams is highest when the level of the error component of the observed data is lowest. Increasing the error level decreases the effectiveness of the phase diagram as an independent diagnostic tool and increases the requirement for additional evidence from the amplitude diagram.

2. If repeated samples of observed data can be obtained under identical conditions, then the mean indicated period can be an effective estimator of the actual period. The value of the mean indicated period appears to be independent of the level of the error component. The variance of the indicated period is dependent on the underlying error level, and as this level increases, the chance decreases of properly identifying an actual periodic component on the basis of a phase diagram from a single sample.

3. Since not all the observed data points need be used at any given step of the analysis process, it is recommended that for phase-diagram analysis only a moderate amount of data points be used for the initial confirmation; for subsequent confirmation attempts, increasingly more data points should be used.

SUMMARY COMMENTS

The method of phase and amplitude diagrams may have general use in the analysis of observed time series with a suspected periodic component. Special importance may be found in those areas in which either the number of data points or the number of cycles is small. The tendency mentioned in [3] and [7] for periodogram methods to indicate the presence of periods when no periods are in fact present has not yet been investigated; quite possibly, the dual nature of the clues of the phase-and amplitude-diagram method and the capability to perform repeated confirmations may preclude such false alarms. Further investigations into this method should include some experience with purely autoregressive series, such as those in [7], to determine the false-alarm behavior of the method. Also, some experience using a theoretical model comprising both periodic and autoregressive components should be gained. Finally, in keeping with the recent experience reported in [3], it would seem appropriate to investigate the behavior of the method when applied to Beveridge's classical series of wheat prices.

REFERENCES

1. K. Stumpff, *Grundlagen und Methoden der Periodenforschung*, Julius Springer, Berlin, 1937.
2. J. Blume, *Nachweis von Perioden durch Phasen- und Amplitudendiagramm mit Anwendungen aus der Biologie, Medizin, und Psychologie,* Westdeutscher Verlag, Cologne, 1965.
3. C. W. J. Granger and A. O. Hughes, "A New Look at Some Old Data: The Beveridge Wheat Price Series," J. Roy. Statistical Soc. **134**, Ser. A, Part 3, 413 (1971).
4. A. Sollberger, *Statistical Problems in Biological Rhythms and Biological Rhythm Control*, unpublished, 1968.
5. H. L. Wiener, "A Test of Significance for Periodogram Analysis," Ph.D. thesis, Catholic University of America, 1971.
6. N. R. Draper and H. Smith, *Applied Regression Analysis*, John Wiley and Sons, New York, 1966.
7. M. G. Kendall, *Contributions to the Study of Oscillatory Time-Series*, Cambridge University Press, Cambridge, 1946.

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)* Naval Research Laboratory Washington, D.C. 20390	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

RECENT COMPUTATIONAL EXPERIENCE WITH THE METHOD OF PHASE AND AMPLITUDE DIAGRAMS

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

Final report on one phase of the NRL problem.

5. AUTHOR(S) *(First name, middle initial, last name)*

Howard L. Wiener

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
August 15, 1972	15	7

8a. CONTRACT OR GRANT NO. NRL Problem B01-10 b. PROJECT NO. RR 003-02-41-6152 c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) NRL Report 7435 9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)* OR Report 72-3

10. DISTRIBUTION STATEMENT

Approved for public release; distribution unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Department of the Navy, Office of Naval Research, Arlington, Virginia 22217

13. ABSTRACT

Recent experience with observed time-series data have suggested the possibility of renewed interest in the analysis of periodic behavior. This report summarizes recent computational results using the method of phase and amplitude diagrams, which is an analytic and graphic technique for indicating the presence of periodic components and for determining period length, phase, and amplitude. The method has promise in areas of application involving relatively small numbers of data points or possibly small numbers of cycles (if a period is actually present). As the name implies, two sets of clues are used: phase diagrams and amplitude diagrams. The results reported here discuss the use of phase diagrams only. Findings and conclusions include the influence of an additive error component on the effectiveness of phase diagrams and recommended procedures for performing repeated confirmations of an indicated period length.

DD FORM 1 NOV 65 1473 (PAGE 1)

S/N 0101-807-6801

Security Classification

UNCLASSIFIED

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Statistics						
Statistical computations						
Time series						
Periodicities						
Statistical tests						
Graphical methods						
Computers						

DD FORM 1 NOV 65 1473 (BACK)
(PAGE 2)

Security Classification