Benjamin Whang
William E. Gilbert
Stephen Zilliacus
Carderock Division
Naval Surface Warfare Center
Bethesda, MD 20084

Two Visually Meaningful Correlation Measures for Comparing Calculated and Measured Response Histories

Two visually meaningful correlation measures are proposed for comparing calculated and measured response histories. One is an error index which is a simplification of RSS (root-sum-square) error factor, and the other is an inequality index that is a simplification of Theil's inequality coefficient. The first compares the difference between the calculated and the measured histories to the measured history. The second compares the difference between the two histories to the sum of the two. The proposed correlation measures are compared to other existing measures, namely, Geers' Error Factors, RSS Error Factor, and Theil's Inequality Coefficient for ease of interpretation and visualization. © *1994 John Wiley & Sons, Inc.*

INTRODUCTION

Often in comparing calculated and measured response histories, the two histories are shown on the same graph, and the goodness of fit is subjectively judged by the author or the presenter. One might hear or read meaningless statements such as "As you can see, the calculated results agree well with the experimental results," or "Good agreement is found between experiments and predictions." Someone looking at the two curves in Fig. 1(a), for example, might say that the correlation is "good." Another might say that the correlation is "bad." Sometimes the same two curves are plotted as shown in Fig. 1(b) to improve the appearance of correlation. Clearly, the "beauty" of correlation must not be left to "the eye of the beholder," and *numerical* correlation measures that are independent of the way one plots the curves are needed for objective comparison. In addition, for a correlation measure to be credible, it must be *visually meaningful.*

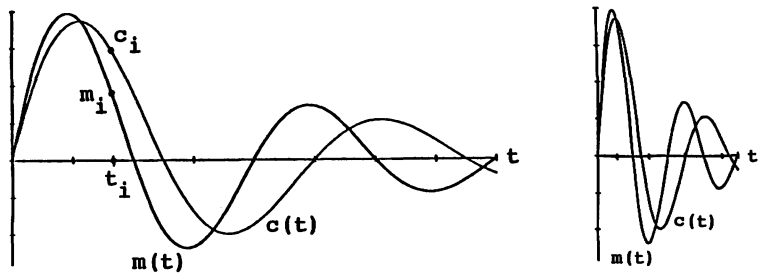
For over a decade, the Navy's shock community has been using Geers' (1984) error factors (magnitude, phase, and comprehensive) as a comparison tool to judge the "goodness" of calculated response histories against the measured. Although the Geers' pioneering work has served its purpose, several pathological problems with the Geers error factors (noticed by the senior author and by Dawson, 1992) led to this work.

In comparing any two response histories (both calculated, both measured, or one calculated and the other measured), a common practice is to assume one of the two to be true (exact or accurate) and the other approximate, and any discrepancy or deviation from the true is associated with the term *error.* For example, when calcu-

Received November 22, 1993; Accepted January 13, 1994

Shock and Vibration, Vol. 1, No. 4, pp. 303–316 (1994)

CCC 1070-9622/94/040303-14



(a) (b)

FIGURE 1

lated values are compared to measured values, it may be scientifically correct to assume the measured to be true. Very often, however, the measured values can be as uncertain as the calculated values, and thus there would be no justification for favoring the measured over the calculated, in which case any difference between the two is associated with the term *inequality* as opposed to error.

To accommodate both error and inequality, two correlation measures are proposed herein: Zilliacus' *error index* and Whang's *inequality index*. The proposed error index compares the difference between two histories to the one assumed to be true, and for convenience the assumed true values will be called the measured, $m(t)$, and the other the calculated, $c(t)$. The error index is a simplification of the well-known RSS (root-sum-square) error factor. The proposed inequality index, on the other hand, compares the difference between two histories to the *sum* of the two, without assuming one of the two to be true. The two histories can be any combination of calculated and measured; however, in this article, one is called the calculated, $c(t)$, and the other the measured, $m(t)$, for convenience. The inequality index is a simplification of Theil's (1975) inequality coefficient.

CORRELATION MEASURES

In what follows, c_i are the calculated values, and m_i are the measured values [Fig. 1(a)].

Geers' Error Factors (M, P, C)

Magnitude Error Factor (M).

$$M = \frac{\sqrt{\sum c_i^2}}{\sqrt{\sum m_i^2}} - 1. \tag{1}$$

Because the first term can be less than 1, M can be negative, and because the first term can be greater than 2, M can exceed 1.

Phase Error Factor (P).

$$P = 1 - \frac{\sqrt{\left|\sum c_i m_i\right|}}{\sqrt{\sqrt{\sum c_i^2}\,\sqrt{\sum m_i^2}}}. \tag{2}$$

Because the second term cannot exceed 1, P is bounded between 0 and 1.

Comprehensive Error Factor (C).

$$C = \sqrt{M^2 + P^2}. \tag{3}$$

C is the vectorial sum of its orthogonal components M and P, and because M can exceed 1, C can exceed 1.

RSS *Error Factor (R).*

$$R = \frac{\sqrt{\sum (c_i - m_i)^2}}{\sum m_i^2}. \tag{4}$$

R is the RSS of the differences between c_i and m_i divided by the RSS of m_i. Obviously, R can exceed 1.

Zilliacus' Error Index (Z).

$$Z = \frac{\sum |c_i - m_i|}{\sum |m_i|}. \tag{5}$$

Z is the area of the residual $(c_i - m_i)$ divided by the area of the measured, and it can exceed 1.

Theil's Inequality Coefficient (T).

$$T = \frac{\sqrt{\sum (c_i - m_i)^2}}{\sqrt{\sum c_i^2} + \sqrt{\sum m_i^2}}. \tag{6}$$

T is the RSS of the differences between c_i and m_i divided by the *sum* of the RSS of c_i and the RSS of m_i. T is bounded between 0 and 1.

Whang's Inequality Index (W).

$$W = \frac{\sum |c_i - m_i|}{\sum |c_i| + \sum |m_i|}. \tag{7}$$

W is the area of the residual $(c_i - m_i)$ divided by the *sum* of the areas of the calculated and measured. W is bounded between 0 and 1.

DISCUSSION

Figures 2(a), 3(a), and 4(a) are directly from Geers (1984), and they indicate Geers' magnitude (e_m), phase (e_p), and comprehensive (e_c) error factors along with RSS (e_r) error factor for three different sets of comparisons. The starting times for the calculated $c(t)$ and the measured $m(t)$ are different because Geers' error measures require $m(t)$ [or $c(t)$] to be adjusted horizontally until e_c is minimized. This adjustment was deemed necessary in fairness to the analysts because the starting time ($t = 0$) for $m(t)$ is usually not well known. In the future, before any comparison is made, a common reference time, such as the time from detonation, should be clearly indicated on each history (calculated *and* measured) so that the adjustment of starting times would not be necessary.

Figures 2(b), 3(b), and 4(b) correspond to their respective figures *without* time adjustments, and in each figure, M, P, C, and R are shown. Although Cs are greater than e_cs, the differences are not drastic. Geers (1984) points out that the e_rs in Figs. 2(a) and 3(a) seem to be unacceptably large. However, because the upper bound of e_r is limitless (not bounded by 100%) there is no basis for stating that these values appear too large. In fact, e_c is not bounded either. At least $e_r = 100\%$ can be understood to be the case when the RSS of $(c_i - m_i)$ is equal to the RSS of m_i, although not easily visualized. The case where $e_c = 100\%$ is beyond visualization.

The most troublesome of these figures is Fig. 3 that shows $e_m = 0\%$ (or $M = 0.009$) and $e_c = 4\%$ (or $C = 0.139$). That magnitude errors of these sets are not near zero.

In Fig. 5, Figs. 2(b), 3(b), and 4(b) are repeated, and Z, T, and W are added for comparison. Z and R are consistently similar to each other as expected, but the advantage of Z over R is that Z is easier to visualize than R. Both Z and R can be greater than 1. As pointed out earlier, Z can be visualized as the ratio of the area of the residual $(c_i - m_i)$ to the area of the measured.

The difference between R and T is that the denominator of T has an additional term, the square root of the sum of the squares of c_i. Conceptually, T compares the RSS of the residuals to the sum of the RSSs of the measured and the calculated; therefore, when $(c_i - m_i)$ are small, T tends to be about half of R. T can never be greater than 1; R is unbounded.

The difference between Z and W is that the denominator of W has an additional term, the sum of the absolute values of c_i; therefore, when $(c_i - m_i)$ are small, W tends to be about half of Z. W can never be greater than 1; Z is unbounded.

T and W are consistently similar to each other as one might expect, but the advantage of W over T is that W is easier to visualize than T. As pointed out earlier, W can be visualized as the ratio of the area of the residual $(c_i - m_i)$ to the sum of the areas of the calculated and the measured.

Figures 6 and 7 show what happens when the sign of one of the two curves is reversed. As pointed out by Dawson (1992), M, P, and C are insensitive to the sign reversal. (Also, note that in these cases R and Z are greater than 1, T and W are not.) Recently, to remedy this particular problem, Geers (1993) proposed a revision to P that in turn affects C. The new Geers' error factors are as follows (M is unchanged):

$$P_{\text{new}} = 1 - \frac{\sum c_i m_i}{\sqrt{\sum c_i^2}\sqrt{\sum m_i^2}} \tag{8}$$

$$C_{\text{new}} = \sqrt{M^2 + (P_{\text{new}})^2}. \tag{9}$$

Because the second term in P_{new} is bounded between +1 and −1, P_{new} is now bounded between 0 and 2. C_{new} can still exceed 1. As shown in Figs. 6(b) and 7(b), P_{new} and therefore C_{new} are now sensitive to the sign reversal. In fact, $P_{\text{new}} = 2.0$ is an indication of sign reversal or of being completely out of phase. However, there still remain

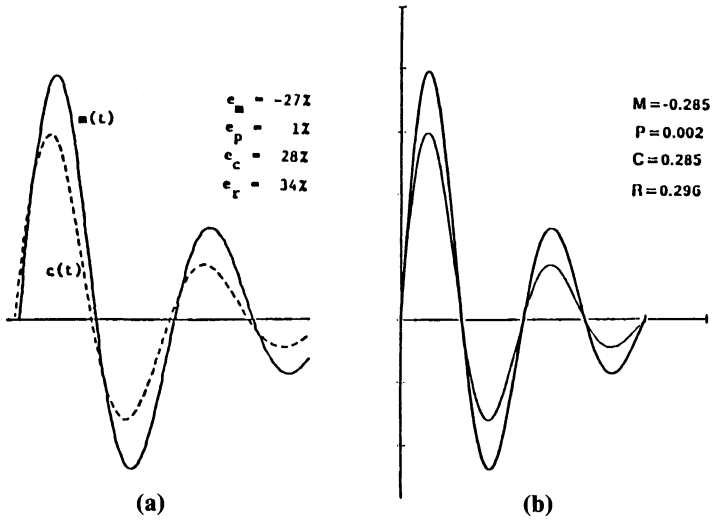


FIGURE 2

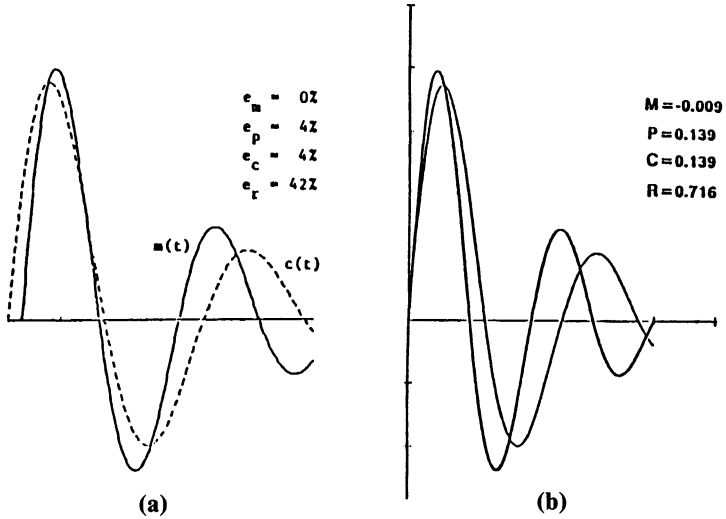


FIGURE 3

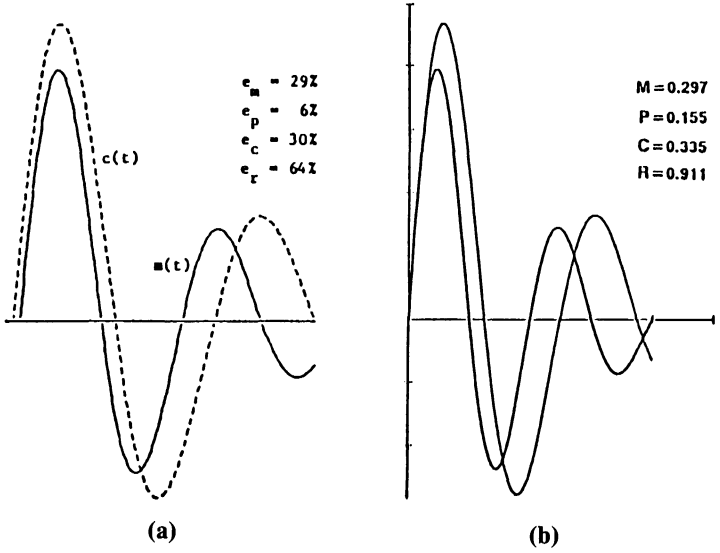


FIGURE 4

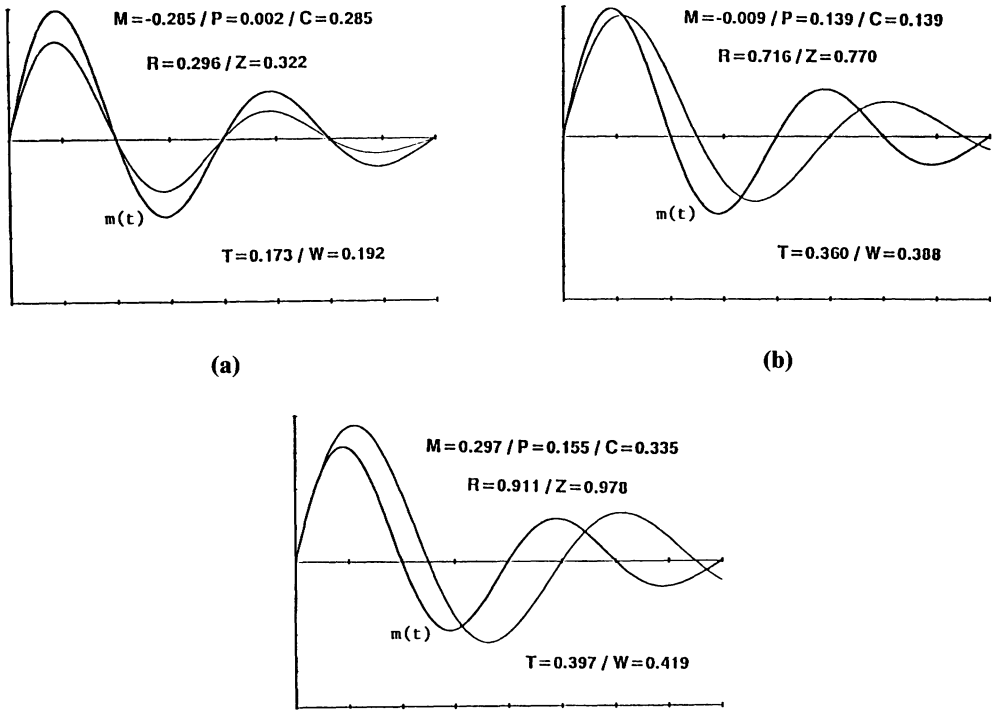


FIGURE 5

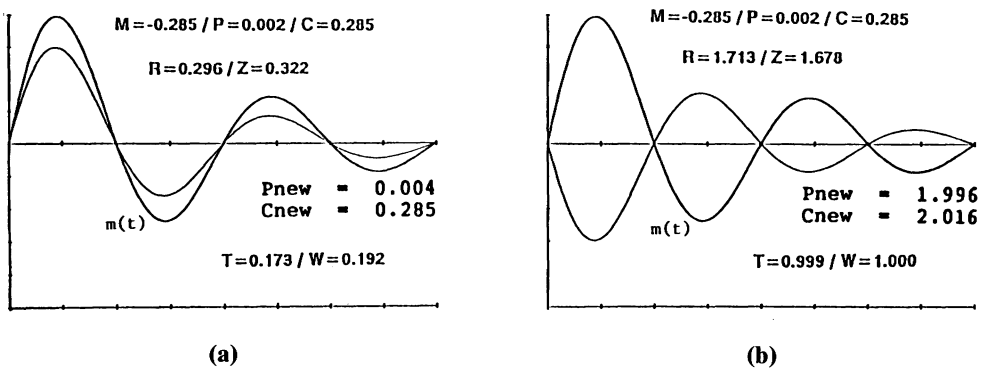


FIGURE 6

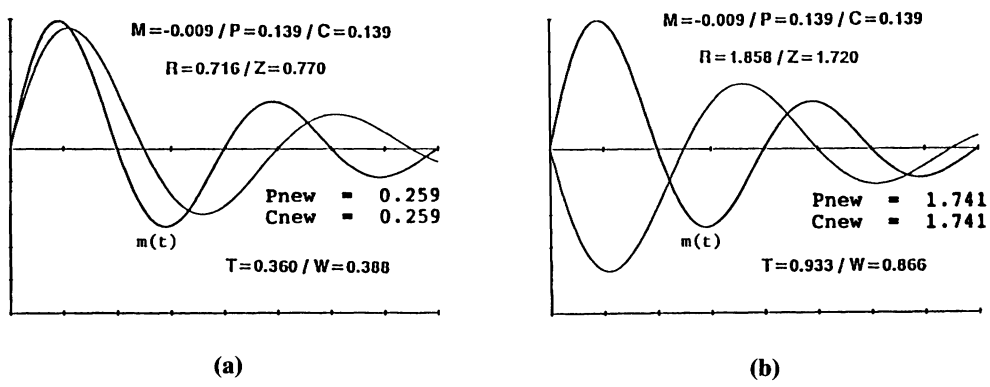


FIGURE 7

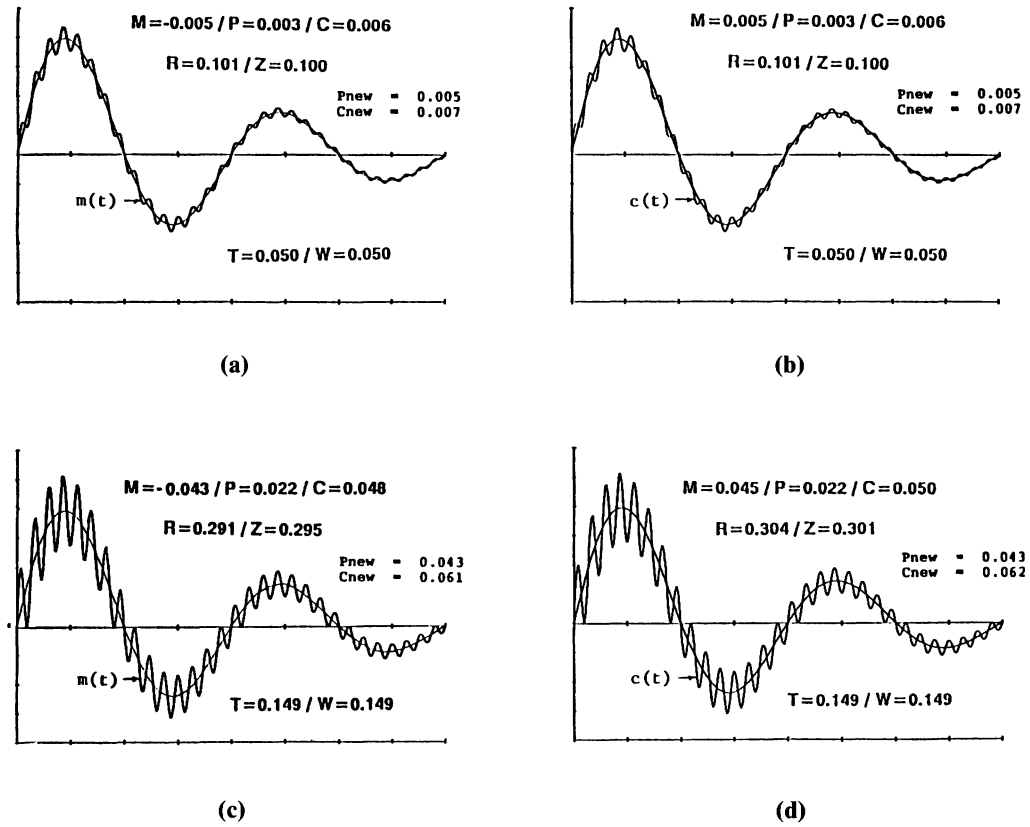


FIGURE 8

a few more pathological problems with P_{new} and C_{new} that will be discussed.

The fact that Z can exceed 1 is meaningful because the residual can be greater than the measured. Similarly, the case of $W = 1.0$ is meaningful because that happens when the residual is equal to the sum. However, with C (or C_{new}) the vectorial sum of M and P (or P_{new}), the meaning of C (or C_{new}) $= 1.0$ is not clear.

Figure 8 shows the effect of reversing $c(t)$ and $m(t)$. M is sensitive to the reversal as pointed out by Dawson (1992), but P and P_{new} are not. Also, as expected, R and Z are sensitive to the reversal and T and W are not.

Figure 9 is not an example of common occurrence but is presented here to show the need to distinguish between "early" time comparison and "late" time comparison. The integration limits have been arbitrarily selected as 0–1.0, 1.0–2.0, and 0–2.0.

Figures 10–12 are related to a common case involving strain records showing a permanent set. Figure 10(a) shows that when $c(t)$ and $m(t)$ are very close to each other, all of the measures discussed here are reasonable. In Fig. 10(b), R, Z, T, and W appear to be reasonable, C and C_{new} appear to be low because P and P_{new} are low. P and P_{new} tend to be low when the curves do not cross the time axis [See Eqs. (2) and (8)]. Figures 11 and 12 are examples of the cases where M, C, C_{new}, R, and Z are close to each other.

Figure 13 is presented here to show that when $W = 1.0$, when the difference is equal to the sum,

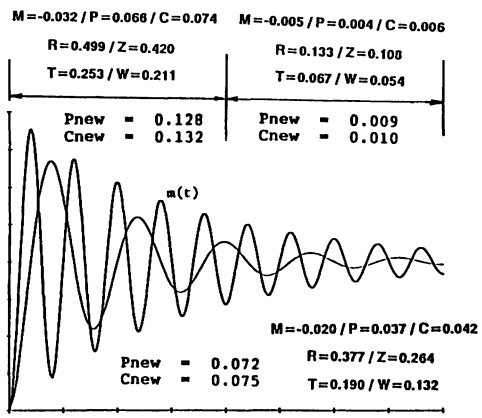
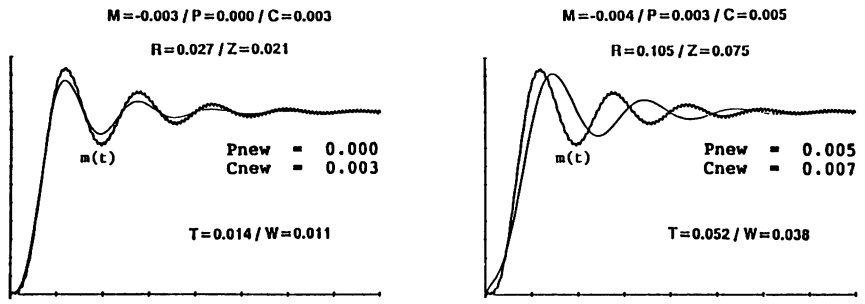
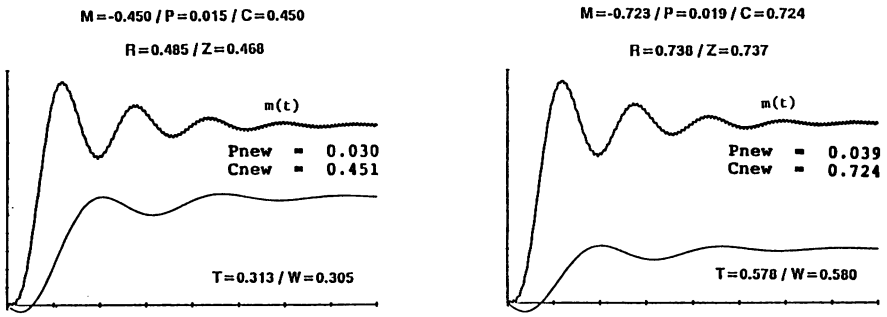


FIGURE 9



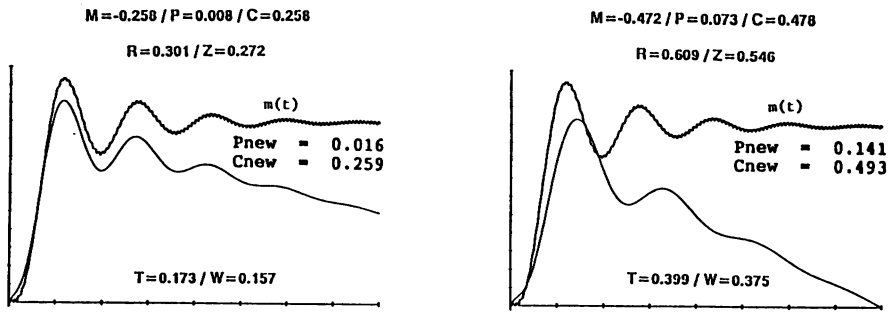
(a) (b)

FIGURE 10



(a) (b)

FIGURE 11



(a) (b)

FIGURE 12

the measure does not distinguish the degree of inequality, that is, W indicates that Fig. 13(a,b) are "equally unequal." On the other hand, C and C_{new} values show Fig. 13(b) to be in greater error than Fig. 13(a), because of the nature of M [see Eq. (1)], but R and Z values show the reverse. P_{new} does not distinguish the degree of error when two curves are completely out of phase.

VISUALIZATION

For a correlation measure to be visually meaningful, it must be simple and consistent with eyeballing, consistent with the mental process of human eyes to compare two response histories. One possible process is to pick off a vertical distance (absolute value) of the difference between

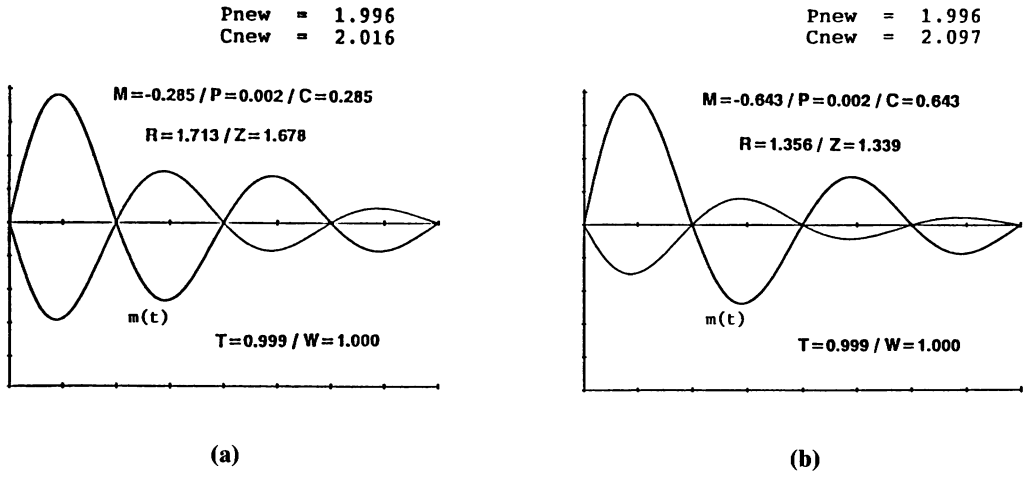


FIGURE 13

two curves at a particular time and divide that distance by another vertical distance at that particular time. These ratios can be obtained at various times, and with relative ease, they can be averaged mentally in an approximate way.

The above process is expressed mathematically as follows:

$$A = \frac{1}{n}\sum_{i=1}^{n}\frac{|c_i - m_i|}{|m_i|} \quad \text{or} \quad B = \frac{1}{n}\sum_{i=1}^{n}\frac{|c_i - m_i|}{|c_i| + |m_i|}. \tag{10}$$

A resembles Z, and B resembles W. However, the problem with A and B is that when $c(t)$ and $m(t)$ intersect simultaneously on the time axis, that is when $|c_i - m_i| = 0$ and $|m_i| = 0$ or when $|c_i - m_i| = 0$ and $|c_i| + |m_i| = 0$, the condition of indeterminacy occurs [see Fig. 5(a), for example). This problem has been avoided in Z and W by summing in the numerator and in the denominator separately. This enables one to associate Z and W as the ratios of areas. It should be pointed out here that visualizing or mentally estimating

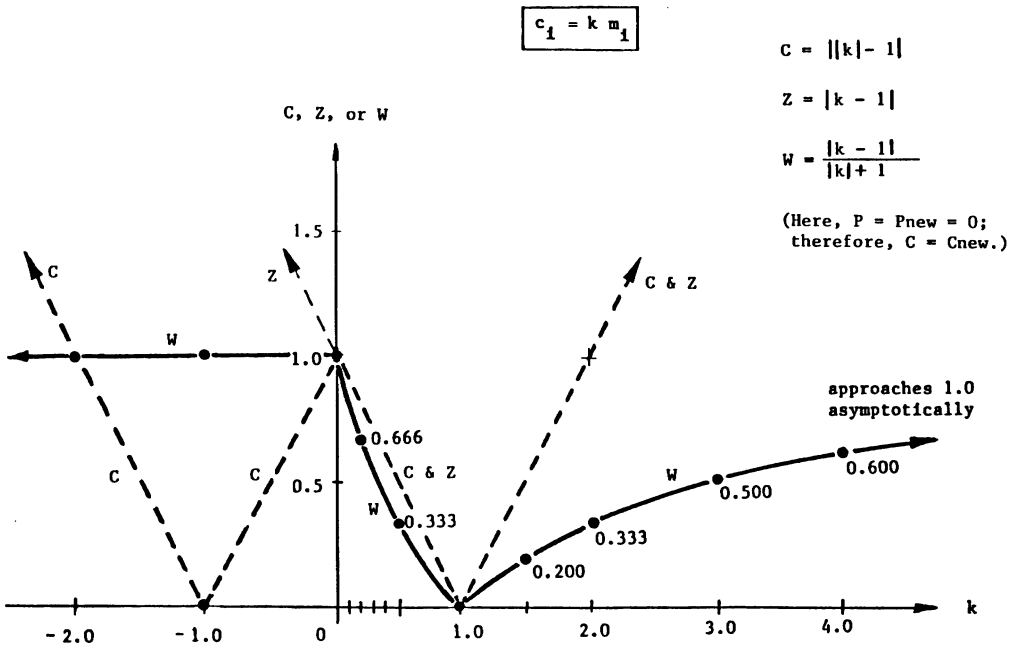


FIGURE 14

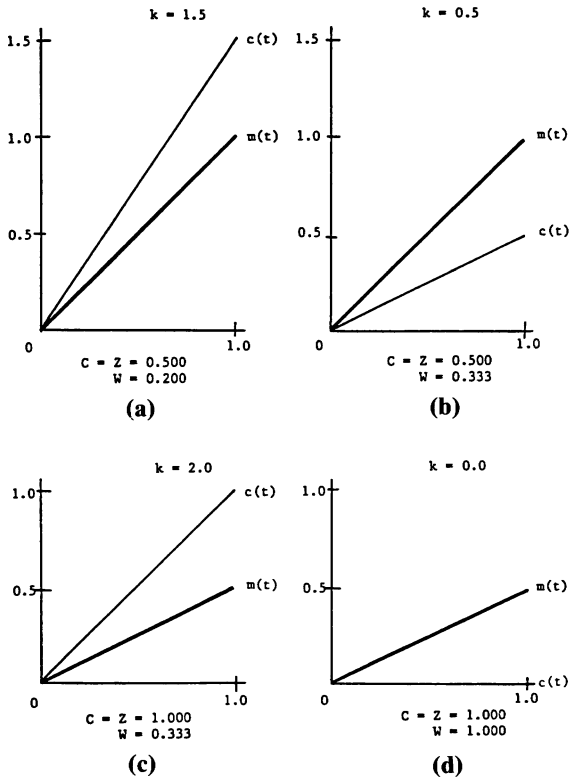


FIGURE 15

M, P, C, P_{new}, C_{new}, R, and T is not natural for most people because it involves taking the square root of the sum of squares of many numbers.

Geers (1993) used a simple case of $c_i = km_i$, where k is a constant, to point out correctly that W is not symmetric about $k = 1$. For example, if one analysis produced $c_i = 0.5m_i$, W would be 0.333 (=0.5/1.5); if another analysis produced $c_i = 1.5m_i$, its W would be 0.200 (=0.5/2.5). And because each differed from the measured by $0.5m_i$, Geers states that associating two different values of W for these two analyses is "not proper." It is true that W is not symmetric about $k = 1$ as shown in Fig. 14 for this simple case. (C is *locally* symmetric about $k = 1$, and *globally* symmetric about $k = 0$. Z is symmetric about $k = 1$.) However, the way human eyes compare two curves is *not* symmetric about $k = 1$. Using the Geers' example of $c_i = km_i$, Fig. 15 shows the asymmetric nature of eyeballing. Figure 15(a) has the appearance of better correlation than Fig. 15(b), even though both $c(t)$s differ from $m(t)$ by 50%. Figure 15(c,d) make the point more strikingly.

Figure 16 further shows the asymmetric nature of eyeballing for exponentially decaying sine

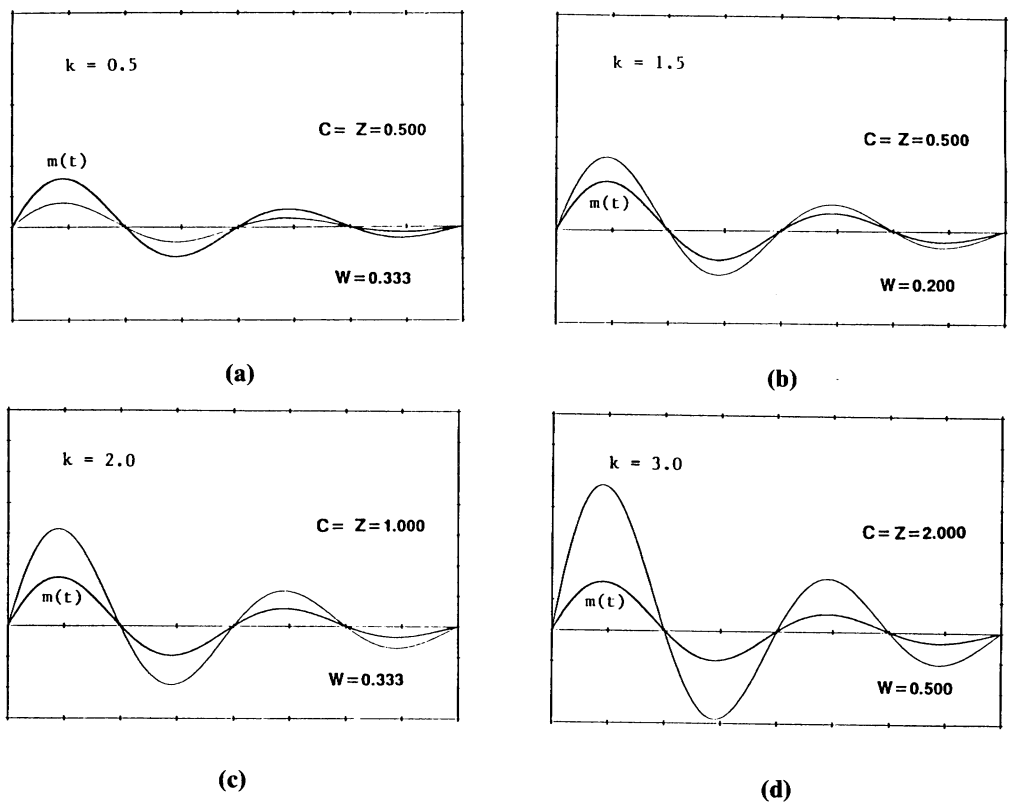


FIGURE 16

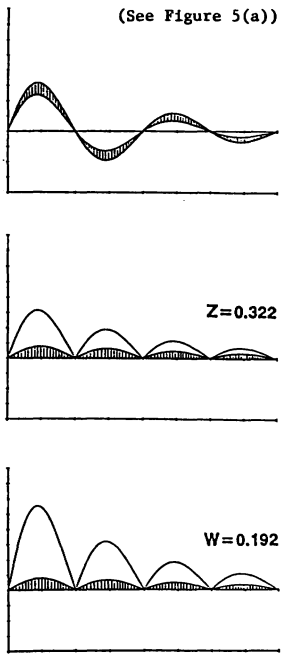


FIGURE 17

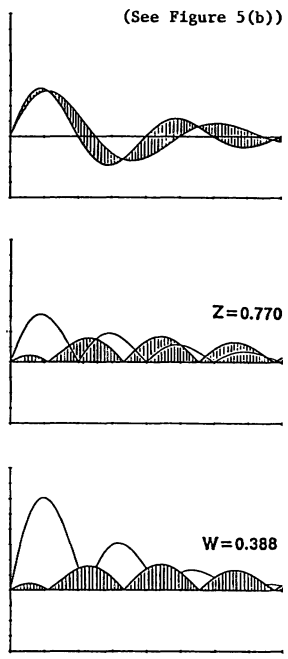


FIGURE 18

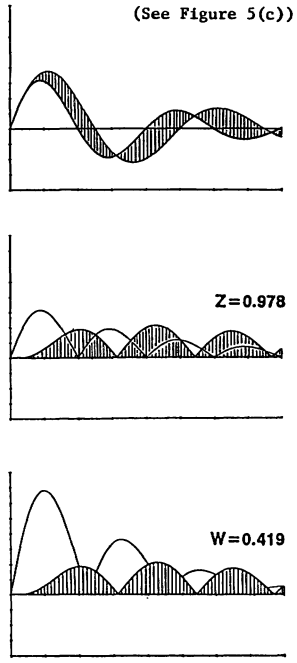


FIGURE 19

waves for k values of 0.5, 1.5, 2.0, and 3.0. Figure 16(b) appears to have better correlation than Fig. 16(a), even though they both differ from the measured by 50% as shown by $C = Z = 0.500$ in both cases. (In this example, $P = P_{new} = 0$; therefore $C = C_{new}$.) Figure 16(a,c) have the same W, but they have different C and Z. Figures 16(c) and 16(d) show that $Z = 1.0$ in general means the

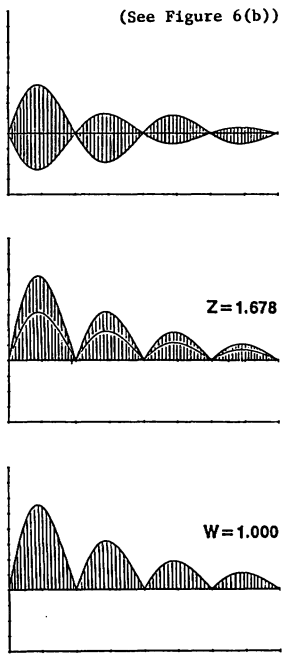


FIGURE 20

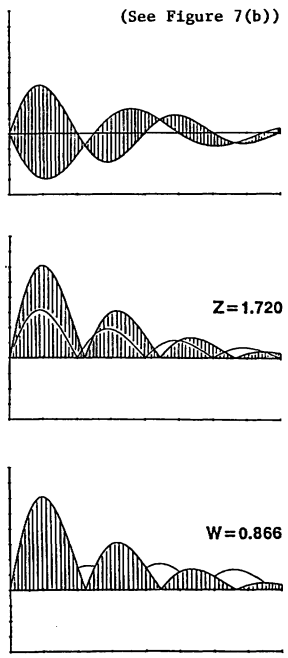


FIGURE 21

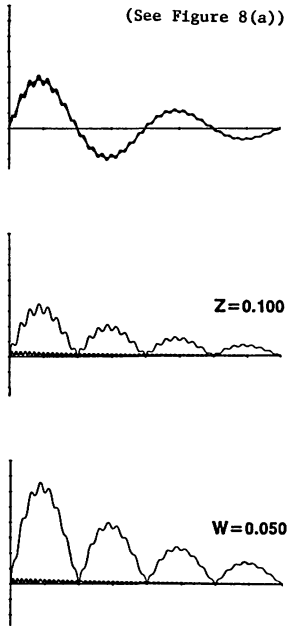


FIGURE 22

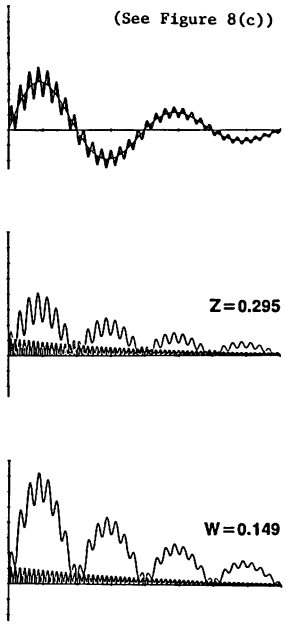


FIGURE 23

"calculated" on the whole is twice the "measured" and that $Z = 2.0$ means the "calculated" on the whole is three times the "measured", etc.

To complete the discussion on visualization, Figures 17–34 are provided to show how Z and W can be interpreted. In each figure, under the original $c(t)$ and $m(t)$, the rectified (absolute-valued) residual $[c(t) - m(t)]$ is compared to the rectified $m(t)$ for Z, and below that, the rectified residual is compared to the sum of the rectified $c(t)$ and $m(t)$ for W. In almost all of the cases shown, the Z-curves intersect, but in all cases, the W-curves never intersect, although at some points they are coincident, which happens when the difference is equal to the sum.

One might recognize that both Z and W are the ratios of means: Z is the ratio of the mean of the rectified residual (shaded area) to the mean of the rectified measured, and W is the ratio of the mean of the rectified residual (shaded area) to the mean of the sum of the rectified calculated and the rectified measured. However, visually comparing the areas is more appealing than averaging. (After all, visually comparing the areas is the whole idea behind pie charts.)

As experience is gained in associating the values of Z and W with their corresponding plots, a consensus may be reached on qualitative words to go with certain ranges of these values. For example, *excellent* may be assigned for W values less than 0.1, *good* for W values between 0.1 and 0.3, etc., and similar words for Z. These words are clearly subjective, and the range of values for each word changes as both the computational and experimental technologies improve.

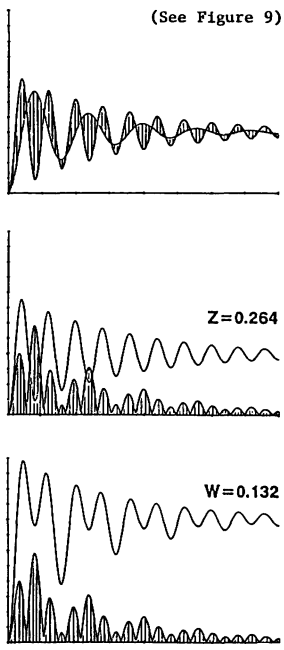


FIGURE 24

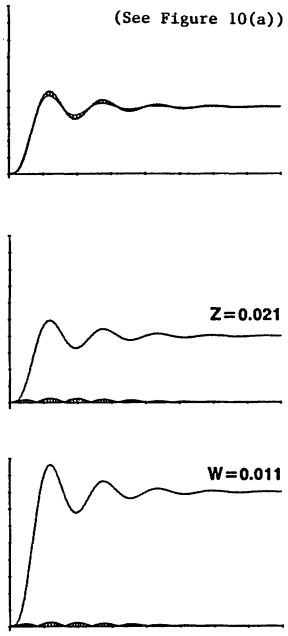


FIGURE 25

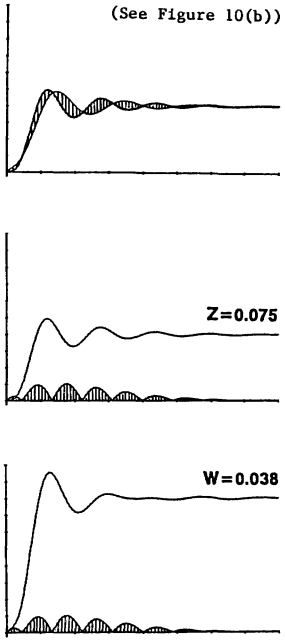


FIGURE 26

An even better way might be to let the values of Z and W speak for themselves instead of associating subjective words to certain arbitrary values. For example, $W = 0.2$ would mean 20% unequal or 80% equal, or $Z = 0.3$ would simply mean 30% error.

CONCLUSION

Two visually meaningful correlation measures have been proposed for comparing calculated and measured response histories: Zilliacus' error index and Whang's inequality index. The error

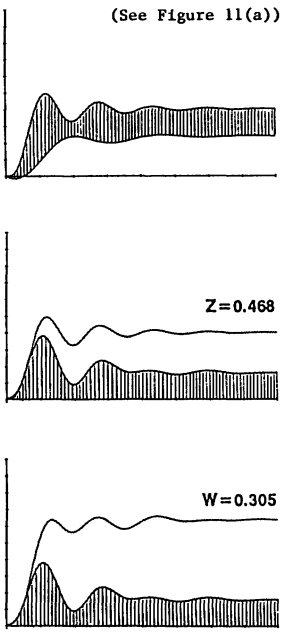


FIGURE 27

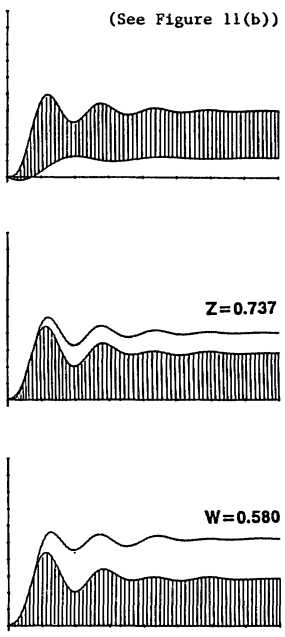


FIGURE 28

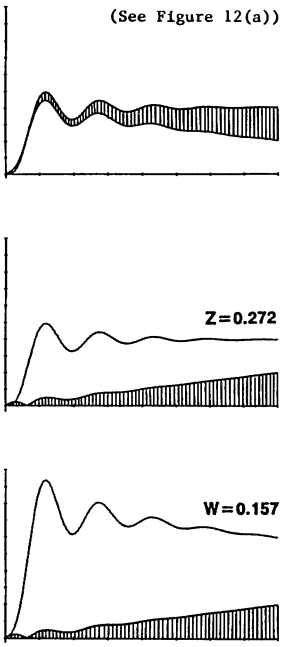


FIGURE 29

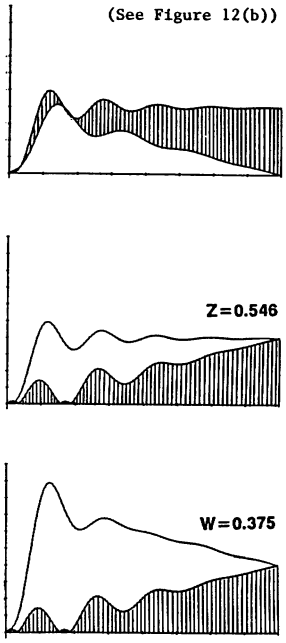


FIGURE 30

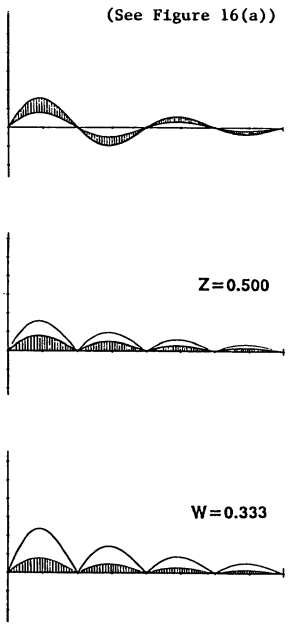


FIGURE 31

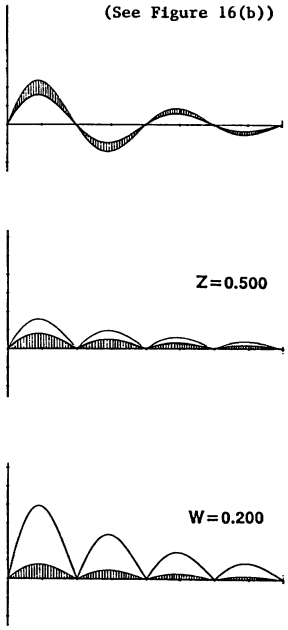


FIGURE 32

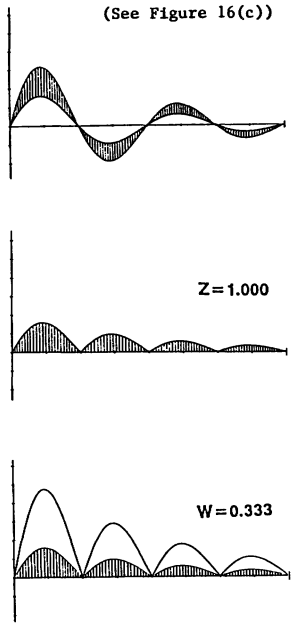


FIGURE 33

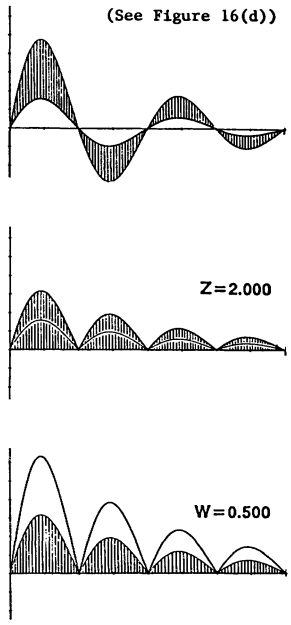


FIGURE 34

index is appropriate when there is justification for favoring the measured over the calculated. The inequality index is appropriate when there is no justification for favoring one over the other. However, whether there is justification or not, both Z and W may be used without adjusting starting times as long as what they are comparing and *what they are being compared to* are kept in mind.

The authors wish to thank Dr. T. L. Geers of the University of Colorado for stimulating and insightful discussions during the draft stage of this article.

REFERENCES

Dawson, R., October 1992, "A Comparison of Internal Fluid Model Experimental Data and Predicted Responses by Applying the Geers Error Function," CDNSWC Report SSPD-93-177-5.

Geers, T. L., 1984, "An Objective Error Measure for the Comparison of Calculated and Measured Transient Response Histories," *Shock and Vibration Bulletin,* Vol. 54, Part 2, pp. 99–107.

Geers, T. L., May 1993, personal communication, letter to B. Whang.

Theil, H., 1975, *Economic Forecasts and Policy,* pp. 31–42, North–Holland Publishing Co., Amsterdam.