Bulletin of the Seismological Society of America, Vol. 70, No. 1, pp. 293-303, February 1980

A NEW ASSESSMENT OF ERRORS FROM DIGITIZATION AND BASE LINE CORRECTIONS OF STRONG-MOTION ACCELEROGRAMS

BY JAFAR SHOJA-TAHERI*

ABSTRACT

Statistical analysis of strong-motion records of the 1966 Parkfield earthquake and the 1952 Taft earthquake indicates that the usable limit of long period of velocity and displacement, calculated by integration of the accelerogram records, is restricted mainly by human reading and base line correction errors. Base line correction errors arise from uncertainties in the coefficients of the fitted straight base line. Calculation by the jackknife method of the variances of the errors in displacement records from linear base line corrections indicates that these errors are more pronounced in the long-period (above 7 sec) range than errors from digitization. The records analyzed in this paper indicate that long-period limits due to the combined errors vary between 7 and 14 sec. Beyond these limits components of displacement spectra from the present analog accelerograms are not reliable measures of ground motion.

INTRODUCTION

The usable frequency band of a strong-motion record is inherently restricted by the combined noise from all the procedures involved in the record processing. The limits of this band may be determined by not only estimating the signal-to-noise ratio at each frequency harmonic, but also considering the fraction of total spectral energy, which is contributed by the processing noise at each harmonic. For a harmonic with frequency ω (within the acceleration response of the instrument) recorded acceleration $\ddot{Y} = A \sin \omega t$ corresponds to ground displacement $Y = -(A/\omega^2) \sin \omega t$. Because the amplitude, A, is contaminated with recording and processing noise, $A = A_{\text{signal}} + A_{\text{noise}}$, at some long period ($\omega \rightarrow 0$), the energy of the errors, $(A_{\text{noise}/\omega 2})^2$, contributed to the total energy of ground displacement, $\sum_{i=1} (A_i/\omega_i 2)^2$ becomes so large that inclusion of this harmonic may significantly distort the actual displacement record. It follows that any study based on analysis of velocity and displacement components requires an estimate of the frequency limit at which the noise energy can be neglected.

The purpose of this paper is to make statistical analysis of the errors due to data processing of strong-motion accelerograms, and also to establish an estimate of the usable frequency band of displacement components.

In an attempt to estimate errors in digitized strong-motion accelerograms, Trifunac *et al.* (1973), made a statistical analysis of errors of a straight thin line, independently digitized by four different operators. They showed that among all errors of recording, record processing, and digitizing, human reading error is the main contributing factor to the variance of the total errors. They concluded that this error limits the usable long periods of double integrated digitized data at about 16 sec.

It is, however, important to note that any estimate of recording errors based on a digitized straight line may not be an unbiased representation of the errors for actual accelerograms. Unlike the straight line which is uniformly thin (see Trifunac

* Present address: Geophysics Department, School of Sciences, Ferdowsi University, Mashad, Iran.

et al., 1973), the trace thickness of an accelerogram varies with the amplitude of the record.

In addition to reading errors due to human eyes, there is another significant and, up to the present, overlooked error which is introduced into the data during base line correction processing. This error originates mainly from the uncertainties in the parameters by which a base line is defined. The base line of a given accelerogram record is usually determined by fitting the "closest" curve (straight line or higher order polynomial) to the data by the least-squares method. In this case, for a given accelerogram, values of the base line parameters strongly depend on the particular selection of digitized data points along the length of record.

The following sections of this paper examine these two prominent errors (reading and base line correction) in some detail and establish a statistical basis for estimating the long-period limits introduced in computed displacements by these errors.

Reading Errors Due to Human Eyes

A direct assessment of reading errors of digitized accelerograms can be made by analyzing a group of records which are obtained from repeatedly digitizing the same record by a number of operators. The meaningful long-period limit of displacement can then be estimated from the variance of the displacement amplitude spectrum at low frequencies. In the absence of such information it is, nevertheless, possible to make a reliable estimate of reading errors because for each accelerogram, the sources that generate these errors, such as the thicknesses of acclerogram trace, the fixed trace, and cross hairs in the typical digitizing machine, are generally well known. The trace thickness of an accelerogram on the Mylar translucent films varies between 0.2 and about 1 mm. In a typical strong-motion record, the high-energy (large-amplitude) episode occupies only a small portion of the total length. It follows that the trace thickness of a significant part of the length of a record stays at 1 mm. The fixed trace is produced by a fixed mirror which is rigidly attached to the accelerograph frame. To eliminate the effect of the systematic errors introduced by the digitizing machine, this trace is digitized and subtracted from the accelerometer trace (Trifunac *et al.*, 1973). This trace is 0.5 to 1 mm. The cross hairs have a finite thickness of 0.1 to 0.2 mm.

It is reasonable to assume that the combined errors from the thicknesses of the accelerogram trace, the fixed trace, and the cross hairs are randomly distributed within bounds of at least ±0.2 mm.

For an accelerogram recorded with sensitivity of about 8 cm/g, for example, extreme bounds of ±0.2 mm are equivalent to values of about ± 2.45 cm/sec^2. In the present paper, a total of 2,048 pseudorandom numbers with acceleration values uniformly distributed between ± 1.225 cm/sec^2, were generated by computer to resemble recording errors with arbitrary bounds of ±0.1 mm. The amplitude spectra of the corresponding displacement were then computed for frequencies between 0.025 and 25 Hz (the Nyquist frequency) and plotted in Figure 1.

The lowest straight-dashed line in this figure is the least-square fitted line to the displacement amplitude spectrum. The two upper dashed lines show the fitted lines to the displacement amplitude spectra corresponding to reading errors of ±0.2 mm and ±0.4 mm, respectively. Conveniently, therefore, the middle one of these lines will be utilized in the remainder of this paper to represent the effect of reading errors of accelerograms in corresponding displacement records.

Errors Due to Base-Line Correction

As mentioned previously, for a given accelerogram the values of the base line parameters depend strongly on the selection of the positions of the data points along the time axis. Variations in the base line parameters have pronounced effects on low-frequency components of the displacement records. During digitization of accelerograms, the time positions of the majority of points (with the possible exception

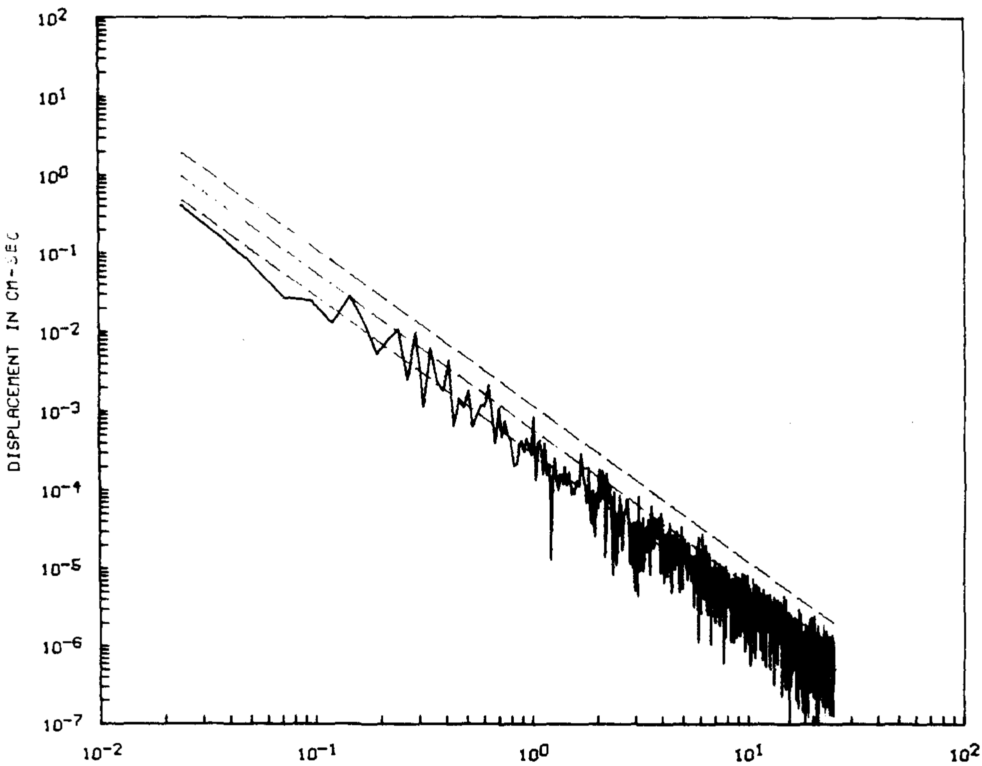

FIG. 1. Displacement amplitude spectrum corresponding to 2,048 pseudorandom numbers with acceleration values uniformly distributed between ±1.225 cm/sec^2. The lowest straight-dashed line is the best fitted line to the displacement amplitude spectrum. The two upper-dashed lines are best fitted lines to the displacement amplitude spectra pertinent to acceleration values distributed between ±2.45 cm/sec^2 and ±4.90 cm/sec^2, respectively.

of points corresponding to peaks and troughs) are selected more or less arbitrarily. Therefore, redigitization of the same record is likely to produce a majority of data points which have different time positions. An important question is to what extent deviation of the time positions affects the base line parameters.

To investigate this question, we make use of the powerful statistical method called the "jackknife technique" (Tukey, 1958; Brillinger, 1964; Brillinger, 1966). The jackknife method entails dividing the sample into a number of different subsamples (chosen in a particular way) and calculating the desired values from all but one of the subgroups, to produce what are known as "pseudovalues." The procedure is repeated over and over, omitting a different subgroup each time. The variance of the population is then estimated directly from these pseudovalues.

Suppose an accelerogram consists of N data points. Let A and B define the base line $A + Bt$ (t is time) that may be calculated from these data points. Separate these points into r groups such that the data points of each group are selected randomly from the total N points. Let A_i and B_i be the parameters of the base line based on all but the ith group. A_i, $B_i (i = 1, \ldots, r)$ are used to determine variances of A and B, respectively.

The unbiased estimates of the variances of A and B can be determined simply from

$$\sigma_A^2 = \sum_{i=1}^{r} (A - A_i)^2/(r - 1)$$

$$\sigma_B^2 = \sum_{i=1}^{r} (B - B_i)^2/(r - 1).$$

Determination of base line parameters and their corresponding variances was carried out for all strong-motion accelerogram records of the 1966 Parkfield earthquake. [The uncorrected accelerogram data reported by the Earthquake Engineering Research Laboratory (Caltech) are utilized in this paper.] These records are selected because they possess relatively high recording quality. For comparison, the parameters and related variances of the Taft 1952 records were also calculated.

For all records, except CH #8 and Temblor, a length of 40 sec (N = 2,000 points with equal intervals, t = 0.02 sec) was selected. For CH #8 and Temblor, which are shorter than 40 sec, lengths of 26 sec (N = 1,300 points) and 30 sec, (N = 1,500 points) were selected, respectively. A straight base-line $A + Bt$ was fitted for each record to the data points by least squares.

For the jackknife procedure, A_i and B_i were calculated based on all but the ith group [the ith group was chosen (sampled) randomly from the N data points]. In this analysis, each record was separated into 10 groups, each of which consisted of $N/10$ data points. Brillinger (1966) indicated that values of r between 10 and 20 are reasonable.

In order to examine how the estimated variance may vary with r, standard errors σ_A and σ_B of the vertical component of CH #2 were calculated. The resulting parameters for these records were $A = 2.398$, $B = -0.1076$. (The values of data are in terms of cm/sec^2.) For $r = 10$, $\sigma_A = 0.61000$ and $\sigma_B = 0.02208$, and for $r = 20$, $\sigma_A = 0.49120$, and $\sigma_B = 0.01767$. It should be noticed that selection of 20 groups gives slightly smaller variances.

Because under normal conditions of digitization the gross effects of the variances do not appear to be overestimated by selection of 10 groups, the data points of all records were therefore separated into 10 groups. The values of A, B, σ_A, and σ_B for CH #2, #5, #8, #12, Temblor, and Taft 1952 are given in Table 1.

At this stage, let us examine the effects of the base line parameters and their variations on the spectra of displacement which are calculated from the accelerograms. Figures 2, 3, and 4 illustrate the analysis made for the N65E component of CH #2, the S25W component of Temblor and the vertical component of Taft, respectively. These components are typical of all the 17 components analyzed in the present paper. The *upper* plots (Figures 2a, 3a, and 4a) illustrate the displacement amplitude spectra calculated from: (1) accelerograms with uncorrected base line (shown by solid line); (2) accelerograms with corrected base line, $A + Bt$ (shown by dotted line); (3) accelerograms with corrected base line, $A + \sigma_A + (B + \sigma_B)t$ (shown by dashed line); (4) accelerograms with corrected base line, $A - \sigma_A + (B - \sigma_B)t$ (shown by dot-dashed line).

The phase variation of displacement spectra, due to variation of the base line parameters, is as important as amplitude variation in controlling the long-period limits of displacement records; the analysis of base line correction errors should, therefore, incorporate this variation with amplitude variation. The middle plots (Figures 2b, 3b, and 4b) show the phase variation introduced by incorporating σ_A and σ_B into the base line corrections.

These plots, as well as the values of σ_A and σ_B shown in Table 1, indicate that for some records (e.g., CH #12 N50E and N40W) variations of the base line parameters are less pronounced than those of others. A possible explanation of this difference

TABLE 1

BASE LINE PARAMETERS AND LONG-PERIOD LIMITS DUE TO BASE LINE CORRECTION, T_b, AND THE COMBINED ERRORS OF BASE LINE AND READING, T_c.

	A	σ_A	B	σ_B	T_b* (sec)	T_c† (sec)
CH #2, Down	2.398	0.610	−0.108	0.022	12.5	11.0
CH #2, N65E	0.475	0.982	−0.023	0.034	13.0	12.5
CH #5, Down	1.529	0.476	−0.070	0.016	12.0	8.0
CH #5, N85E	0.164	0.969	−0.008	0.031	12.0	10.0
CH #5, N05W	−0.739	0.760	0.032	0.024	11.0	10.0
Temblor, Down	0.813	0.290	−0.052	0.014	10.0	7.5
Temblor, S25W	1.514	1.012	−0.096	0.046	8.0	7.0
Temblor, N65W	0.353	0.491	−0.023	0.023	12.0	10.0
CH #8, Down	−0.133	0.444	0.009	0.023	9.0	7.5
CH #8, N50E	0.413	0.890	−0.032	0.044	9.0	7.5
CH #8, N40W	0.487	0.680	−0.037	0.036	9.5	9.0
CH #12, Down	1.634	0.292	−0.076	0.011	12.5	9.5
CH #12, N50E	1.818	0.207	−0.081	0.007	14.5	10.5
CH #12, N40W	3.007	0.138	−0.133	0.005	No limit	13.0
Taft, Down	−0.042	0.286	0.004	0.010	19.0	14.0
Taft, S69E	−0.065	0.623	0.002	0.022	13.0	12.0
Taft, N21E	0.261	0.660	−0.020	0.022	12.5	11.0

* T_b, long-period limit due to base line corrections.

† T_c, long-period limit due to combined errors of base line and reading.

lies in variation of the gross amplitude along the length of an accelerogram; for a record along which the gross amplitude is more uniform, the σ_A and σ_B are expected to be smaller.

As can be seen from the plots, variation of the base line parameters significantly affects the low-frequency components of the amplitude, as well as phase spectra of displacements up to about 0.1 Hz. Displacement spectral energy is mainly contributed by the few lowest frequency components and hence the significant variations observed in amplitude and phase of these components will be expected to mask the characteristics of displacement records.

In order to make an estimation of the long-period limits (longest usable periods), the following method was developed and the results depicted in the lower plots of the figures (Figures 2c, 3c, and 4c).

The method is, simply, to compute at a given frequency ω_i, the absolute value of amplitude variation in the amplitude spectrum $D(\omega_i)$,

$$|\Delta D| = |D_{A+Bt} - D_{A+\sigma_A} + (B + \sigma_B)t|.$$

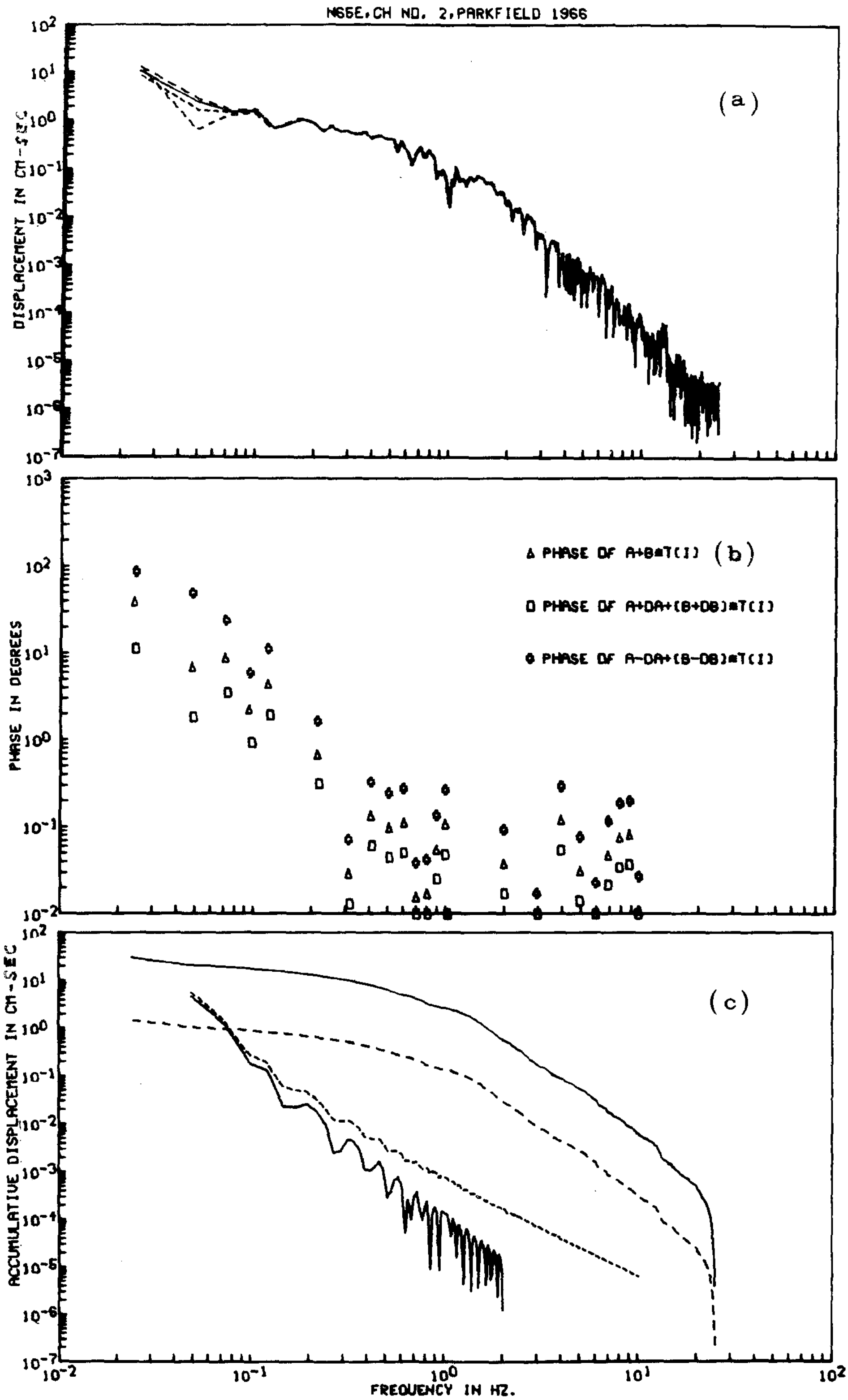

FIG. 2. (a) Displacement amplitude spectra calculated from uncorrected (solid line) and base lines corrected (dotted, dashed, and dot-dashed lines) accelerogram of N65E component of CH #2, Parkfield 1966. (b) Phase variation due to variances of the base line's parameters. (c) Accumulative displacement amplitude spectrum (upper solid line) and its 5 per cent fraction (long dashed line), the amplitude variation due to variances of the base line's parameters (oscillatory solid line), and the combined error (short dashed line).

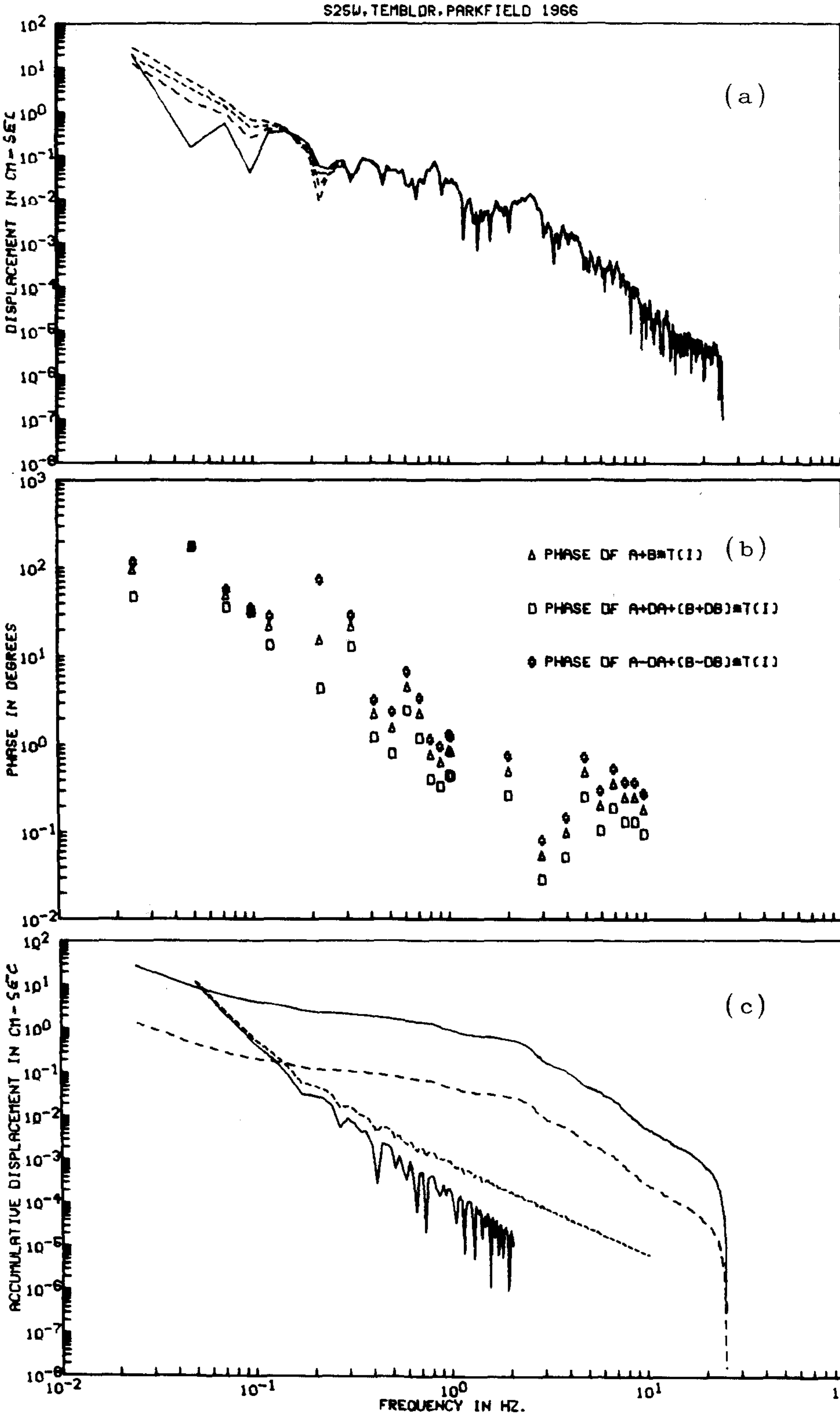

FIG. 3. (a) Displacement amplitude spectra calculated from uncorrected (solid line) and base lines corrected (dotted, dashed, and dot-dashed lines) accelerogram of S25W component of Temblor, Parkfield 1966. (b) Phase variation due to variances of the base line's parameters. (c) Accumulative displacement amplitude spectrum (upper solid line) and its 5 per cent fraction (long dashed line), the amplitude variation due to variances of the base line's parameters (oscillatory solid line), and the combined error (short dashed line).

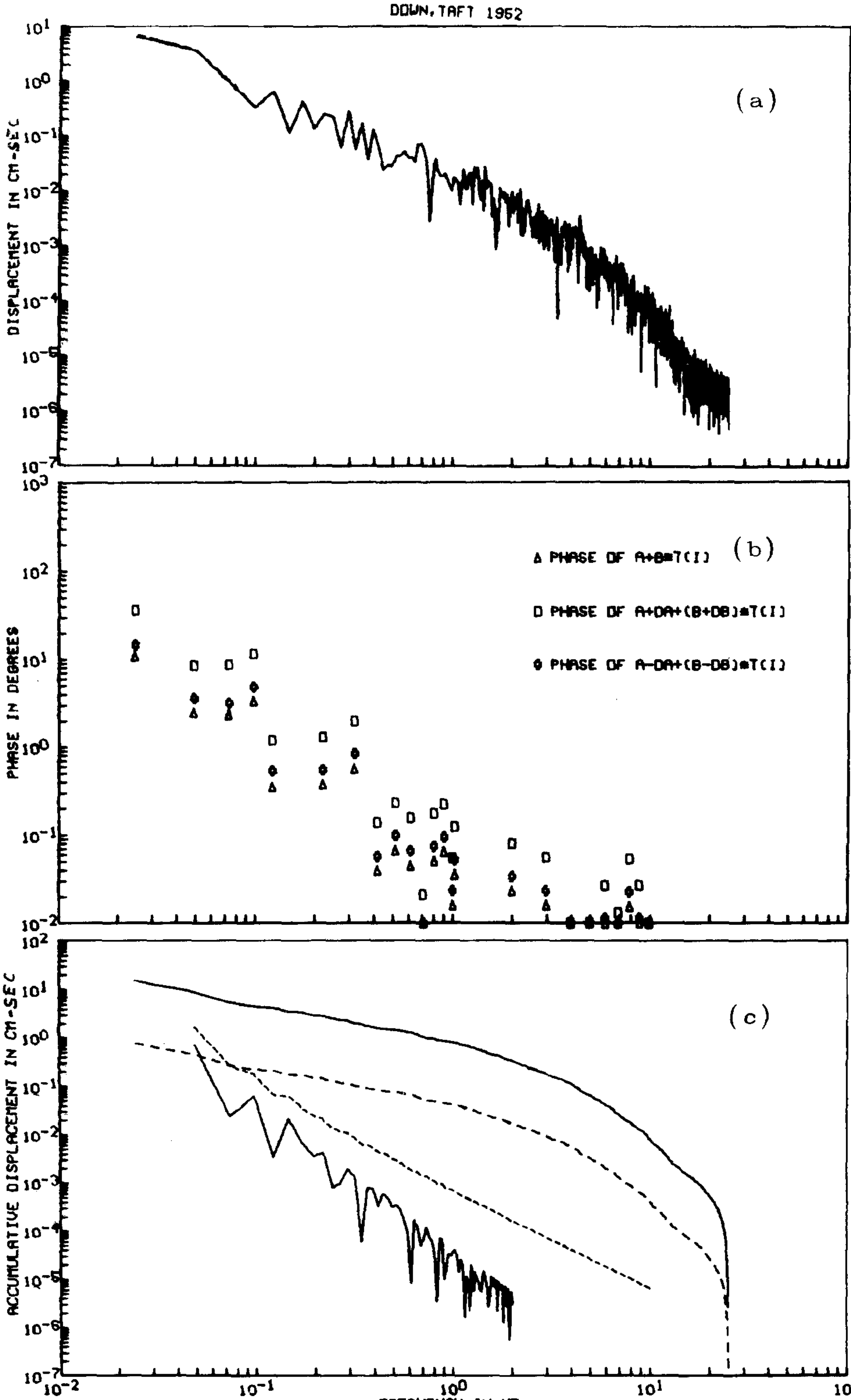

FIG. 4. (a) Displacement amplitude spectra calculated from uncorrected (solid line) and base lines corrected (dotted, dashed, and dot-dashed lines) accelerogram of vertical component of Taft, Kern County 1952. (b) Phase variation due to variances of the base line's parameters. (c) Accumulative displacement amplitude spectrum (upper solid line) and its 5 per cent fraction (long dashed line), the amplitude variation due to variances of the base line's parameters (oscillatory solid line), and the combined error (short dashed line).

D_{A+Bt} is the displacement-amplitude spectrum calculated from the (base line $A + Bt$) corrected accelerograms. This value is compared with some fraction (i.e., 5 per cent) of the accumulative displacement amplitude spectra

$$\sum_{j=i+1}^{N} |D_{A+Bt}|$$

corresponding to the subsequent frequency ω_{i+1} (N corresponds to Nyquist frequency and is equal to 1,024 in this analysis). In the *lower* plots of Figures 2c, 3c, and 4c, the accumulative amplitude spectra are shown by the solid lines. The dashed lines parallel to these lines show the 5 per cent fraction of the accumulative displacement spectra; a fraction of 5 per cent is selected arbitrarily as the basis for estimating the usable long-period limits of displacement records. The amplitude variations, $|\Delta D(\omega_i)|$, are plotted by the oscillatory solid lines. For each record, the crossing point of the $|\Delta D(\omega_i)|$ line with the 5 per cent fraction line defines the long-period limits introduced during base line correction. These periods are listed in Table 1. Although σ_A and σ_B vary with T_b, they tend to decrease with increasing T_b. Least-squares fit to the data gives the following relationships between T_b and σ_A and σ_B,

$$\sigma_A = 0.97 - 0.031T_b$$

$$\sigma_B = 0.042 - 0.0015T_b.$$

Variation of T_b, as shown in Table 1, is between 8 and 19 sec. To obtain a more unbiased relationship between T_b and record length, accelerograms with more variety of record lengths are required. Least-squares fit to 15 of the 17 records listed in Table 1 (components of CH #12, N40W, and Taft, Down, are excluded) gives the following relationship between T_b and record length L

$$T_b = 2.71 + 0.246L.$$

This relation shows that T_b increases with record length L. For L equal to 26, 30, and 40 sec, T_b becomes equal to 9.1, 10.1, and 12.5 sec, respectively. For L equal to 50, and 60 sec (assuming linear extrapolation can be employed from this relation for L longer than 40 sec), T_b is 15.0 and 17.5 sec, respectively.

Combined Errors of Reading and Base Line Correction

The combined errors of reading and base line correction for each record were obtained simply by adding, at each frequency, the value given by the *middle* dashed line in Figure 1 (this line resembles the reading errors bounded by ±0.2 mm) to the value given by the $|\Delta D(\omega_i)|$ curve. The combined errors are shown in Figures 2c through 4c by the short dashed lines. Now the long-period limits due to the combined errors can be readily found from the points where the lines of the combined errors cross the lines corresponding to the 5 per cent fraction of the accumulative displacements. The results are also given in Table 1.

The long-period limits introduced by the combined errors are very much controlled by the degree of fluctuation in both amplitude and phase at long periods. The long-period limits, as shown in the figures and Table 1, vary widely between 7.0 to 14.0 sec.

Least-squares fit to the same 15 records used for the $T_b - L$ relation, gives the following relationship between T_c and record length L

$$T_c = 2.69 + 0.194L.$$

This relation shows that T_c also increases with record length. For L equal to 26, 30, and 40 sec, T_c becomes 7.7, 8.5, and 10.4 sec, respectively. For L equal to 50 and 60 sec (assuming linear extrapolation can be employed to this relation for L longer than 40 sec), T_c becomes 12.4 and 14.3 sec, respectively. This result, therefore, indicates that the long-period limit of 16 sec inferred by Trifunac *et al.*, (1973) is probably too great even for a record length of 60 sec. The large variation of the long-period limits of the 17 strong-motion records analyzed in this paper suggests that the usable long-period limit of each record should be determined independently, case by case. A more representative average value for the long-period limit of actual ground displacement of 40 sec of length appears to be 10 sec.

CONCLUSIONS

Statistical analysis of two major sources of error (human reading and base line correction) in strong-motion records was carried out. These errors limit the usable long period of ground displacements obtained from double integration of accelerogram records. All accelerograms of the Parkfield earthquake of 1966 and also of the Taft earthquake of 1952 were studied to estimate the point at which noise contamination overwhelms the displacement signal.

In the range of the lowest frequency components ($0.025 < \omega < 0.1$ Hz) of displacement amplitude spectra, errors due to uncertainty in base line parameters are more pronounced than are human reading errors. They have, therefore, a significant weight in controlling the usable long-period limits. These errors, however, become considerably smaller than human reading errors in intermediate- and high-frequency components.

The usable long-period limits of the displacement records analyzed in this paper vary widely between 7 to 14 sec. It is indicated that the usable long-period limit, T_c, varies (increases) with record length L. For L equal to 40, 50, and 60 sec, T_c is estimated to be about 10, 12, and 14 sec, respectively.

At a period of about 16 sec, the combined errors in the majority of cases exceed 25 per cent of the accumulative displacement amplitude spectrum. In 35 per cent of the cases, the combined errors at this period exceed 50 per cent of the accumulative displacement spectrum. Therefore, the period of 16 sec suggested by Trifunac *et al.*, (1973) is likely to be an overestimate of the usable long-period limits of displacement records.

It is suggested that the most unbiased estimate of the long-period limit of a strong-motion record must be obtained by analyzing the set of digitized records. Digital records from the newly available digital strong-motion accelerometers should allow the above limit to be significantly extended. Nevertheless, errors of the second type considered here will still arise and will need reanalysis.

ACKNOWLEDGMENTS

I am grateful to Dr. Bruce A. Bolt, who provided the original stimulus for this research as well as valuable criticism and suggestions throughout its course. This work has been supported by the National Science Foundation Grant AEN 74-21548, and the Seismographic Station of the University of California, Berkeley.

During the course of this work, I was on educational leave from the Department of Geophysics, Ferdowsi University, Mashad, Iran. The financial support of Ferdowsi University and Iran's Ministry of Sciences and Higher Education is gratefully recognized.

References

Brillinger, D. R. (1964). The asymptotic behaviour of Tukey's general method of setting approximate confidence limits (the jackknife) when applied to maximum likelihood estimates, *Rev. Inst. Statist.* **32,** 202–206.

Brillinger, D. R. (1966). The application of the jackknife to the analysis of sample surveys, *Commentary* **8,** 74–80.

Trifunac, M. D., F. E. Udwadia, and A. G. Brady (1973). Analysis of errors in digitized strong-motion accelerograms, *Bull. Seism. Soc. Am.* **63,** 157–187.

Tukey, J. W. (1958). Bias and confidence in not-quite large samples (abstract), *Ann. Math. Statist.* **29,** 614.

Seismographic Station
Department of Geology & Geophysics
University of California
Berkeley, California 94720

Manuscript received July 12, 1977