

Mathematics Notes
Note 75

November 1981

Tolerance of Spectral Estimation

F. I. Tseng, T. Sarkar and W. F. Walker
Department of Electrical Engineering
Rochester Institute of Technology
Rochester, New York 14623

Abstract

A statistical expression for the mean square error of a spectrum estimation has been derived in terms of the variances and covariances of the amplitude and phase errors of a complex data sequence. No restrictions need be imposed on the magnitude of these variances and covariances. Numerical results have been systematically presented in graphs, which illustrate the dependence of the spectrum error on the standard deviation and correlation distance of the amplitude and phase errors. It is shown that large phase error tends to dominate the spectrum error, and that large correlation distances worsen the spectrum error and sharpen its dependency on the frequency index. An expression to estimate the variance of frequency error has also been derived under the assumption of small phase errors. Numerical results are given, which demonstrate the linear dependency of the frequency error on the phase error and shows that a large correlation distance worsens the frequency error while a large number of samples reduces it.

I. INTRODUCTION

With the advent of the fast Fourier Transform (FFT) algorithm, the discrete Fourier Transform (DFT) has been the most widely used tool in time series analysis, filtering, oil exploration, earthquake analysis, and many other areas. One of the important applications is the estimation of the spectrum in terms of given data. Owing to various uncontrollable factors there are errors in the collected data. In order to determine the tolerances required of the measuring instruments, it is necessary to determine the confidence limits for the estimators of the spectrum. This paper derives a statistical expression for the expected power spectrum of a complex data sequence when the amplitudes and phases of the sequence are subject to random errors. Another related problem of determining the variance of the frequency estimation is also studied.

Various possible sources of errors in digital data acquisition will first be discussed. Because of these sources of errors the probability distribution of the amplitude and phase errors is assumed to be jointly Gaussian. From this assumed probability distribution it is possible to derive confidence limits on estimates of the variances of the spectrum errors. Computer programs are developed to calculate and plot the data to relate the tolerance of the spectrum to the errors of the data sequence.

II SOURCES OF ERROR

In this section a brief review of the various sources of error is presented.

(1) Quantization Errors.

In digital processing of sampled analog signals, the quantization

error is dependent on the number of digits in the digitizer word. Welch [1] has shown that rounding after each two-point transform in evaluating a FFT leads to a relative rms output error ϵ which is bounded above by

$$\epsilon = (0.3)2^{(M+3)/2}2^{-B} \quad (1)$$

for a transform of 2^M samples using B-bit arithmetic.

(2) Aperture error in the sampling device.

The aperture error is caused by the fact that the sampling of the continuous function is not done with a delta function, but rather with finite-width functions. This results in attenuation of the high-frequency information.

(3) Jitter in the sampling device.

Jitter is the process whereby sampling does not take place at the precise instant it should. In general, it can be shown that the error is larger at higher frequencies and that there is a flat, white noise component added in the process [2].

(4) Other sources of error.

- a) Noise, nonlinearities and dropout in the digitizer.
- b) Noise added by amplifiers and extraneous electrical noise picked up from the environment.
- c) Distortion caused by signal levels being too high or excessive noise level owing to the signal being too low.
- d) Folded high-frequency information owing to the antialiasing filters.

III TOLERANCE OF POWER SPECTRUM [3]

If $\{x_n\}$, $n=0,1,\dots,N-1$, is a sequence of complex numbers, where

the real parts of $\{x_n\}$ represent the in-phase time samples and the imaginary parts of $\{x_n\}$ represents the quadrature time samples, then the spectrum of the signal $x(t)$ is given by

$$X(j\omega) = \sum_{n=0}^{N-1} x_n e^{-j\omega nT} \quad (2)$$

where T is the sampling interval, The DFT can be efficiently evaluated by using a FFT algorithm provided N is a highly composite number. In practice, the sequence $\{x_n\}$ is often padded with $L-N$ zeros to make L a highly composite number and also to suppress the picket-fence effect.

Thus let

$$\hat{x}_n = \begin{cases} x_n & \text{for } 0 \leq n \leq N-1 \\ 0 & N \leq n \leq L \end{cases} \quad (3)$$

then

$$X_k = X(j\omega_k) = \sum_{n=0}^{L-1} \hat{x}_n e^{-j2\pi kn/L} \quad k=0,1,2,\dots,L-1. \quad (4)$$

where $\omega_k = 2\pi k/LT$. If $L=2^M$ the spectrum $\{X_k\}$ can be calculated by using a radix-2 FFT. Substituting (3) into (4) yields

$$X_k = \sum_{n=0}^{N-1} x_n W^{nk} \quad k=0,1,2,\dots,L-1 \quad (5)$$

where $W = e^{-j2\pi/L}$.

Due to various sources of errors, as discussed in section II, the complex sequence $\{x_n\}$ is subject to random variations. Let $\{A_n\}$ and $\{\phi_n\}$ be the amplitudes and phases, respectively, of $\{x_n\}$ with mean values A_n^0 and ϕ_n^0 , respectively. We write

$$A_n = A_n^0 (1 + \delta_n) \quad (6)$$

and

$$\phi_n = \phi_n^0 + \zeta_n \quad (7)$$

where $\{\delta_n\}$ and $\{\zeta_n\}$ are random variables with zero mean values. In view of (5) - (7), the expected value of the spectrum is

$$\langle X_k \rangle = \sum_{n=0}^{N-1} A_n^0 e^{j\phi_n^0} \langle (1+\delta_n) e^{j\zeta_n} \rangle W^{nk} \quad (8)$$

Similarly, the expected value of the power spectrum can be written as

$$\langle |X_k|^2 \rangle = \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} A_m^0 A_n^0 e^{j(\phi_m^0 - \phi_n^0)} \eta_{mn} W^{k(m-n)} \quad (9)$$

where

$$\eta_{mn} = \langle (1+\delta_m)(1+\delta_n) e^{j(\zeta_m - \zeta_n)} \rangle \quad (10)$$

It is convenient to express η_{mn} in terms of the joint characteristic function $f(v_1, v_2, v_3, v_4)$ of the four random variables $\delta_m, \delta_n, \zeta_m,$ and ζ_n . By definition,

$$f(v_1, v_2, v_3, v_4) = \langle e^{j(v_1 \delta_m + v_2 \delta_n + v_3 \zeta_m + v_4 \zeta_n)} \rangle \quad (11)$$

From (11), we find that η_{mn} in (10) can be obtained as follows

$$\eta_{mn} = \left[f - j \frac{\partial f}{\partial v_1} - j \frac{\partial f}{\partial v_2} - \frac{\partial^2 f}{\partial v_1 \partial v_2} \right]_{\substack{v_1=v_2=0 \\ v_3=1, v_4=-1}} \quad (12)$$

In general, a characteristic function is expressible in terms of the correlation functions of the random variables. If the probability distribution of the random variables is Gaussian, we may write f in a particularly simple form

$$f(v_1, v_2, v_3, v_4) = \exp\left\{-\frac{1}{2} \bar{v}' [c] \bar{v}\right\} \quad (13)$$

where $\bar{v}' = [v_1, v_2, v_3, v_4]$ is a 1×4 row matrix and is the transpose of the column matrix \bar{v} , and $[c]$ is a 4×4 covariance matrix for the amplitude and phase fluctuation of the time samples. To simplify the analysis we make the following assumptions:

- (1) $\{\delta_n, \zeta_n\}$ are jointly Gaussian.
- (2) The errors in the time samples are stationary, that is, the variances are independent of n :

$$\langle \delta_n^2 \rangle = \Delta^2 \quad (14)$$

and

$$\langle \zeta_n^2 \rangle = \sigma^2 \quad \text{for all } n \quad (15)$$

(3) The covariances depend only on $|m-n|$:

$$\langle \delta_m \delta_n \rangle = \Delta^2 e^{-(m-n)^2/\alpha^2} \quad (16)$$

$$\langle \zeta_m \zeta_n \rangle = \sigma^2 e^{-(m-n)^2/\beta^2} \quad (17)$$

$$\langle \delta_m \zeta_n \rangle = \Delta \sigma e^{-(m-n)^2/\gamma^2} \quad (18)$$

$$\langle \delta_n \zeta_n \rangle = \gamma_0 \Delta \sigma \quad (19)$$

In Eqs. (16)-(18), α , β , and γ are, respectively, the normalized correlation distances of the amplitude and phase errors. The covariance matrix $[c]$ between the m^{th} and the n^{th} elements is then

$$[c] = \begin{bmatrix} \Delta^2 & \Delta^2 \alpha_{mn} & \Delta \sigma \gamma_0 & \Delta \sigma \gamma_{mn} \\ \Delta^2 \alpha_{mn} & \Delta^2 & \Delta \sigma \gamma_{mn} & \Delta \sigma \gamma_0 \\ \Delta \sigma \gamma_0 & \Delta \sigma \gamma_{mn} & \sigma^2 & \sigma^2 \beta_{mn} \\ \Delta \sigma \gamma_{mn} & \Delta \sigma \gamma_0 & \sigma^2 \beta_{mn} & \sigma^2 \end{bmatrix} \quad (20)$$

Substitution of (20) in (13) and the result, in turn, in (12) yields

$$\eta_{mn} = \{1 + \Delta^2 [\alpha_{mn}^2 + \sigma^2 (\gamma_0 - \gamma_{mn})^2]\} e^{-\sigma^2 (1 - \beta_{mn})} \quad (21)$$

Following a similar procedure, we have

$$\langle (1 + \delta_n) e^{j\zeta_n} \rangle = [f - j \frac{\partial f}{\partial v}] \quad v_1 = v_2 = v_3 = 0, v_4 = 1$$

which reduces to

$$= (1 + j\gamma_0 \Delta \sigma) e^{-\sigma^2/2} \quad (22)$$

Combining (5), (8), (9) and (22) we can write the mean square error as

$$\begin{aligned} \langle |\epsilon_k|^2 \rangle &= \langle |X_k|^2 \rangle - |\langle X_k \rangle|^2 \\ &= \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} A_m^0 A_n^0 e^{j(\phi_m^0 - \phi_n^0)} [\eta_{mn} - (1 + \gamma_0^2 \Delta^2 \sigma^2) e^{-\sigma^2}] w^{k(m-n)} \end{aligned} \quad (23)$$

Eq. (23) shows that the mean square error $\langle |\epsilon_k|^2 \rangle$ depends on the data sequence, the variances and covariances of the data sequence errors, and on the frequency index. In order to concentrate the investigation on the effect of measurement errors, we assume that the data samples are uniform, that is, $A_n^0 e^{j\phi_n^0} = 1$ for all n . Then Eq. (23) is simplified to

$$\langle |\epsilon_k|^2 \rangle = \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} [\eta_{mn} - (1 + \gamma_0^2 \Delta^2 \sigma^2) e^{-\sigma^2}] W^{k(m-n)} \quad (24)$$

Since $W = e^{-j2\pi/L}$, Eq. (24) indicates that the maximum error occurs at $k=0$,

$$\langle |\epsilon_0|^2 \rangle = \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} [\eta_{mn} - (1 + \gamma_0^2 \Delta^2 \sigma^2) e^{-\sigma^2}] \quad (25)$$

For small errors, Eq. (25) reduces to

$$\langle |\epsilon_0|^2 \rangle \approx \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} (\Delta^2 \alpha_{mn} + \sigma^2 \beta_{mn}) \quad (26)$$

which shows that the mean square error of the spectrum is proportional to the sum of the covariances of the amplitude and phase errors in the time samples.

IV TOLERANCE OF FREQUENCY ESTIMATION [4]

In estimating the frequency of a single tone signal (e.g. estimation of a carrier frequency), the error in the estimation is strongly dependent on the phase errors in the time samples. Hence, to simplify the analysis we shall assume that the amplitude errors are negligible and the phase errors are small. Thus let $x_n = A_n e^{j\zeta_n} = A_n (1 + j\zeta_n)$. Then the spectrum is given by

$$X(f) \approx \sum_{n=0}^{N-1} A_n (1 + j\zeta_n) e^{-j2\pi f n T} \quad (27)$$

Since we are concerned only with the deviation of the frequency, we shall assume that the signal frequency has been translated down so that the estimated frequency, f_{\max} , for which $|X(f)|$ is a maximum, is small. Thus we can write

$$e^{-j2\pi f n T} \approx 1 - j2\pi f n T - 2\pi^2 f^2 n^2 T^2 \quad (28)$$

Substituting (28) into (27) yields

$$X(f) \approx \sum_{n=0}^{N-1} A_n (1 + 2\pi f T n \zeta_n - 2\pi^2 f^2 T^2 n^2) + j \sum_{n=0}^{N-1} A_n (\zeta_n - 2\pi f n T - 2\pi^2 f^2 T^2 n^2 \zeta_n) \quad (29)$$

The power spectrum is then given by

$$\begin{aligned} |X(f)|^2 \approx & \left(\sum_{n=0}^{N-1} A_n \right)^2 + 4\pi f T \left(\sum_{n=0}^{N-1} A_n \right) \left(\sum_{n=0}^{N-1} n \zeta_n A_n \right) + 4\pi^2 f^2 T^2 \left(\sum_{n=0}^{N-1} n \zeta_n A_n \right)^2 \\ & - 4\pi^2 f^2 T^2 \left(\sum_{n=0}^{N-1} A_n \right) \left(\sum_{n=0}^{N-1} n^2 A_n \right) + \left(\sum_{n=0}^{N-1} A_n \zeta_n \right)^2 + 4\pi^2 f^2 T^2 \left(\sum_{n=0}^{N-1} n A_n \right)^2 \\ & - 4\pi f T \left(\sum_{n=0}^{N-1} n A_n \right) \left(\sum_{n=0}^{N-1} A_n \zeta_n \right) \end{aligned} \quad (30)$$

To find f_{\max} we let $\frac{\partial |X(f)|^2}{\partial f} = 0$ which yields

$$f_{\max} = \frac{\left(\sum_{n=0}^{N-1} A_n \right) \left(\sum_{n=0}^{N-1} n \zeta_n A_n \right)}{2\pi T \left[\left(\sum_{n=0}^{N-1} A_n \right) \left(\sum_{n=0}^{N-1} n^2 A_n \right) - \left(\sum_{n=0}^{N-1} n A_n \right)^2 \right]} \quad (31)$$

Since the mean value of the random phase errors $\langle \zeta_n \rangle = 0$, Eq. (31) shows that the mean value of the frequency deviation $\langle f_{\max} \rangle$ is zero. Thus the standard deviation of the frequency error can be written as

$$\sigma_f = \sqrt{\langle f_{\max}^2 \rangle} \quad (32)$$

To simplify the analysis we let $A_n = 1$ for all n . Then Eq. (31) can be

simplified as

$$f_{\max} = \frac{6 \sum_{n=1}^{N-1} n \zeta_n}{\pi N T (N-1) (N-2)} \quad (33)$$

Substitution of (33) into (32) then yields

$$\sigma_f = \frac{6\sigma}{\pi N T (N-1) (N-2)} \sqrt{\sum_{n=1}^{N-1} \sum_{m=1}^{N-1} n m e^{-(m-n)^2 / \beta^2}} \quad (34)$$

If we normalize the frequency deviation with respect to the resolution bandwidth $F=1/NT$, we can write the normalized frequency error as

$$\frac{\sigma_f}{F} = \frac{6\sigma}{\pi (N-1) (N-2)} \sqrt{\sum_{n=1}^{N-1} \sum_{m=1}^{N-1} n m e^{-(m-n)^2 / \beta^2}} \quad (35)$$

Eq. (35) shows that the frequency error is directly proportional to the phase error. It also indicates that as the correlation distance β increases the frequency error increases also.

V. NUMERICAL RESULTS AND DISCUSSION

(A) Errors of Spectral Estimation

From Eq. (8) it is seen that the noise-free signal level at frequency index $k=0$ is N ; thus we can write the normalized root-mean-square error of the spectrum as

$$\sqrt{\langle |\epsilon_k|^2 \rangle} / N \quad (36)$$

and for the maximum error, which occurs at the frequency index $k=0$, as

$$\sqrt{\langle |\epsilon_0|^2 \rangle} / N \quad (37)$$

Based on Eqs. (24), (25), (36) and (37), a large amount of numerical data have been obtained, which can be used to estimate the confidence limits of the spectrum estimation in terms of the tolerances of the digital data acquisition.

The spectrum error is a function of the amplitude error Δ , the phase error σ , the correlation distances α , β , and γ , the number of time

samples N , the cross-correlation factor γ_0 , as well as the frequency index k . We choose $N=64$ and let $\gamma_0=1$ in all cases. We first computed the maximum spectrum errors [Eqs. (25) and (37)] as functions of amplitude error Δ . The results are presented systematically in Figures 1(a) through 1(e) and Figures 2(a) through 2(e). Next, we computed the spectrum errors [Eqs. (24) and (36)] as a function of the frequency index k by keeping the correlation distances constant ($\alpha=\beta=\gamma=1, 3$ and 5 , respectively, for Figures 3, 4 and 5). In each graph the phase error σ is kept constant and the amplitude error is increased from 0% to 25%. Examination of Figure 1(a) shows that the spectrum error is linearly proportional to the amplitude error for the case of zero phase error. As expected, the spectrum error approaches zero as both the amplitude and phase errors approach zero. The effect of correlation distances is clearly seen in Figures 1(a) through 1(e), namely as correlation distances increase the spectrum error increases in direct proportion. Figures 2(a) through 2(e) clearly indicate the dominant effect of the phase error as relative to the amplitude error. This implies that as the phase error gets larger the amplitude error has less influence on the spectrum error. Figures 3-5 illustrate the frequency dependency of the spectrum error. In all cases the maximum error occurs at frequency index $k=0$, and the spectrum error tapers off as k increases. The frequency dependency of the spectrum error increases with the increase of the amplitude and phase error. The frequency dependency becomes even more pronounced as the correlation distance increases.

(B) Errors of Frequency Estimation

The dependency of the frequency error on the phase error σ and the correlation distance β is clearly seen from Eq. (35) and is noted in Section V. Again we choose $N=64$ and plot the normalized frequency error as a function of the phase error σ for different values of the correlation distance.

It is seen that the frequency error is linearly proportional to the phase error. It should be noted here that this linear dependency is the result of assuming a small phase error in the derivation. To show the effect of the correlation distance we plot in Figure 7 the frequency error versus the correlation distance β for different values of phase errors. It is seen that the spectrum error increases as β increases and the effect becomes more pronounced as the phase error increases.

Finally, Figure 8 demonstrates the dependency of the normalized frequency error upon the number of time samples N . It should be noted that the frequency error is normalized with respect to the resolution bandwidth $F = 1/NT$, which is inversely proportional to N . Examination of Figure 8 indicates that as N increases the frequency error decreases rapidly for small N but it tapers off for large N .

VI CONCLUSIONS

This paper derives a statistical expression for the mean-square error of a spectrum estimator in terms of the amplitude and phase error of a complex time sequence and their correlation distances. No restrictions need be imposed on the magnitude of these errors. Numerical results have been systematically presented in graphs. Examination of these graphs indicates that, for zero phase error, the spectrum error is

linearly proportional to the amplitude error, while for a large phase error the amplitude error has a diminishing influence. They also show that a large correlation distance tends to worsen the spectrum error and sharpen the dependency of the spectrum error on the frequency index.

We also derived an expression to estimate the variance of frequency estimation under the assumption of small phase errors. The results demonstrate the linear dependency of the frequency error on the phase error. They also show that a large correlation distance worsens the frequency error while a large number of samples tends to reduce the frequency error.

REFERENCES

- [1] P.D. Welch, "A Fixed-Point Fast Fourier Transform Error Analysis," IEEE Trans. Audio Electroacoustics, AU-17(2), pp. 151-157, 1969.
- [2] A.V. Balakrishnan, "On the Problem of Time Jitter in Sampling," IRE Trans. Information Theory, Vol. IT-8, No. 3, pp. 226-236, April 1962.
- [3] F.I. Tseng and D.K. Cheng, "Gain Optimization for Arbitrary Antenna Arrays Subject to Random Fluctuations," IEEE Trans. Antennas Propagat. Vol. AP-15, No. 3, pp. 356-366, May 1967.
- [4] Y.S. Shifrin, "Statistical Antenna Theory", the Golem Press, Boulder, Colorado, 1971.

Figure Captions

Figure 1. RMS Spectrum Errors $\sqrt{\langle |\epsilon_0|^2 \rangle} / N$ as a function of amplitude error Δ for different values of correlation distances α , β , and γ . ($N=64$, $\gamma_0=1$)

- (a) Phase error $\sigma=0^\circ$
- (b) Phase error $\sigma=5^\circ$
- (c) Phase error $\sigma=10^\circ$
- (d) Phase error $\sigma=15^\circ$
- (e) Phase error $\sigma=20^\circ$

Figure 2. RMS Spectrum Errors $\sqrt{\langle |\epsilon_0|^2 \rangle} / N$ as a function of amplitude error Δ for different values of phase error σ . ($N=64$, $\gamma_0=1$)

- (a) Correlation distances $\alpha=\beta=\gamma=1$
- (b) Correlation distances $\alpha=\beta=\gamma=3$
- (c) Correlation distances $\alpha=\beta=\gamma=5$
- (d) Correlation distances $\alpha=\beta=\gamma=7$
- (e) Correlation distances $\alpha=\beta=\gamma=9$

Figure 3. RMS spectrum errors $\sqrt{\langle |\epsilon_k|^2 \rangle} / N$ as a function of the frequency index k for different values of amplitude error Δ . ($\alpha=\beta=\gamma=1$)

- (a) Phase error $\sigma=0^\circ$
- (b) Phase error $\sigma=5^\circ$
- (c) Phase error $\sigma=10^\circ$

Figure 4. RMS spectrum errors $\sqrt{\langle |\epsilon_k|^2 \rangle} / N$ as a function of the frequency index k for different values of amplitude error Δ . ($\alpha=\beta=\gamma=3$)

- (a) Phase error $\sigma=0^\circ$
- (b) Phase error $\sigma=5^\circ$
- (c) Phase error $\sigma=10^\circ$

Figure 5. RMS spectrum errors $\sqrt{\langle |\epsilon_k|^2 \rangle} / N$ as a function of the frequency index k for different values of amplitude error Δ . ($\alpha=\beta=\gamma=5$)

- (a) Phase error $\sigma=0^\circ$
- (b) Phase error $\sigma=5^\circ$
- (c) Phase error $\sigma=10^\circ$

Figure 6. Normalized frequency error σ_f/F as a function of the phase error σ , ($N=64$), for $\beta=1,3,5,7,9$.

Figure 7. Normalized frequency error σ_f/F as a function of the correlation distance β , ($N=64$), for $\sigma=5^\circ, 10^\circ, 15^\circ, 20^\circ$.

Figure 8. Normalized frequency error σ_f/F as a function of number of time samples N for a fixed phase error $\sigma=5^\circ$ but for different values of correlation distances $\beta=1,3,5,7,9$.

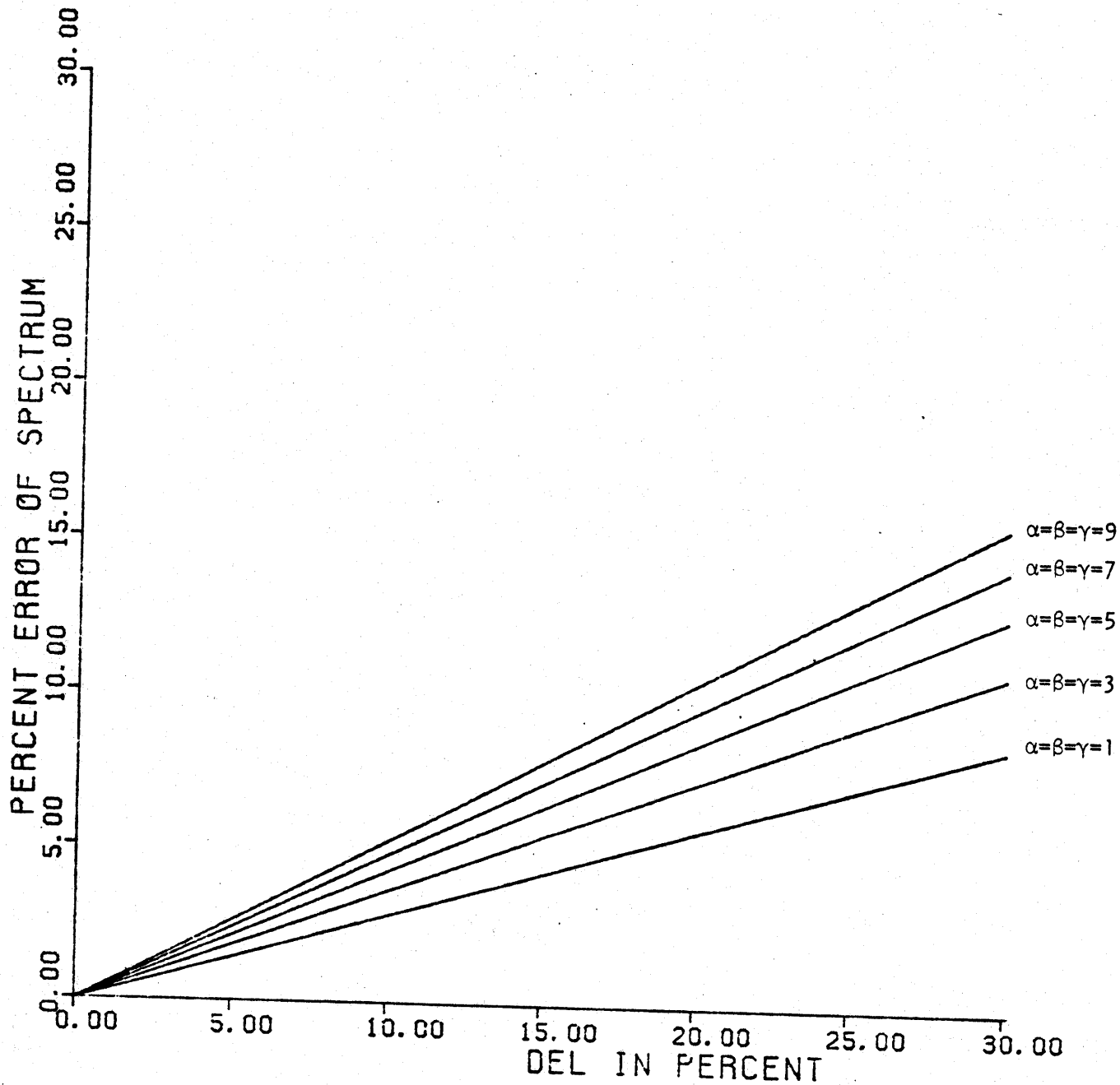


Figure 1(a)

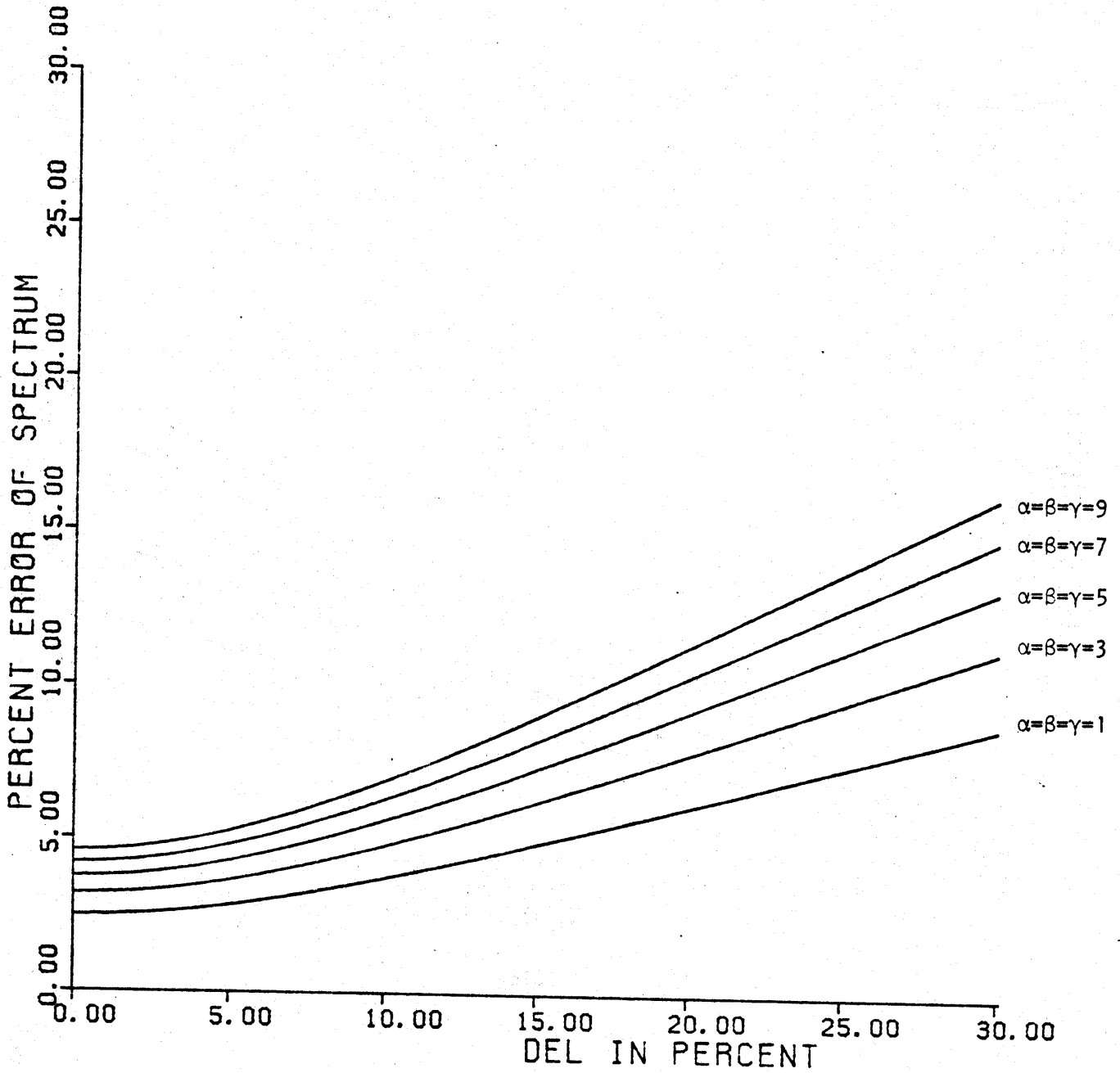


Figure 1(b)

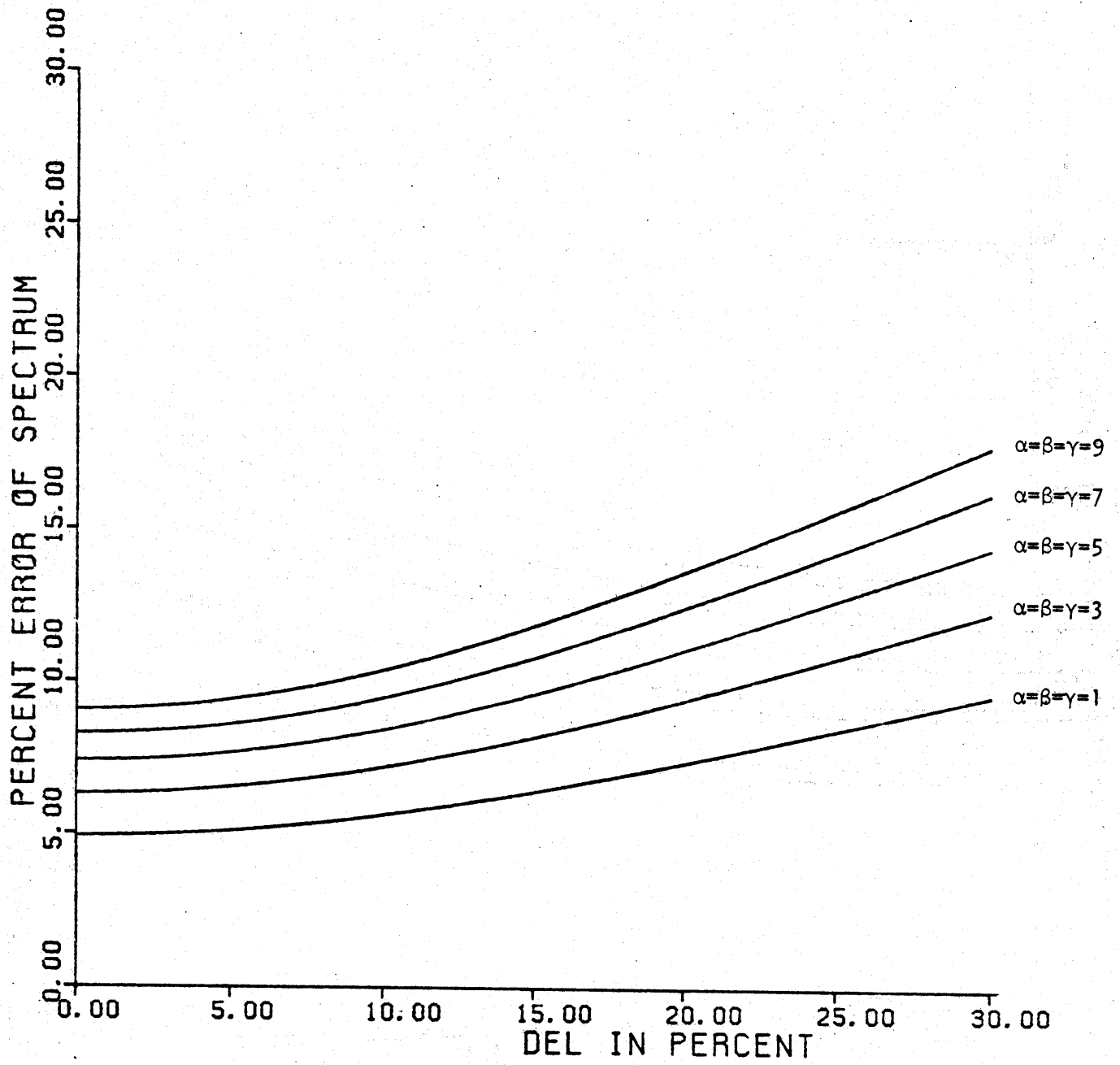


Figure 1(c)

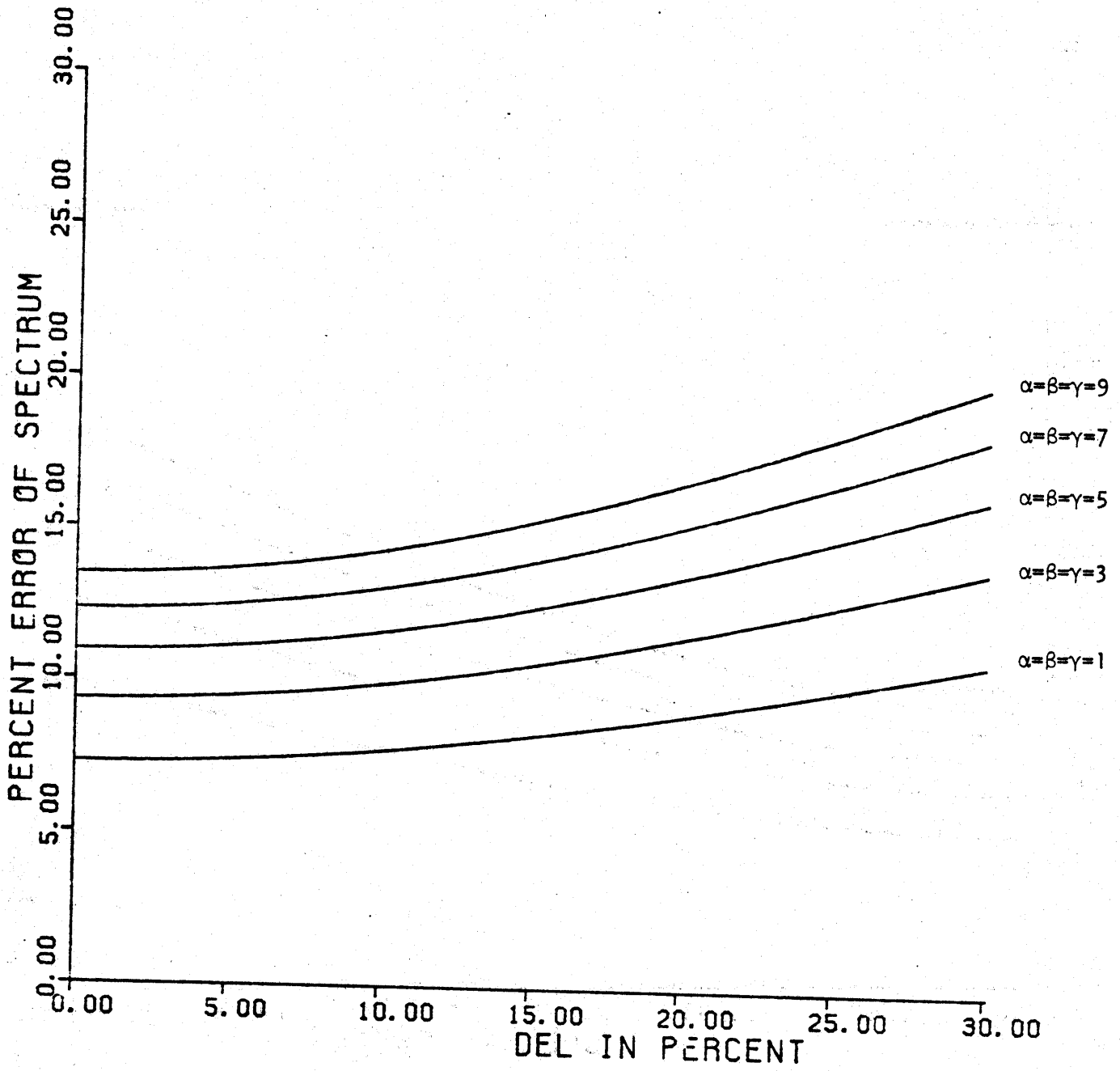


Figure 1(d)

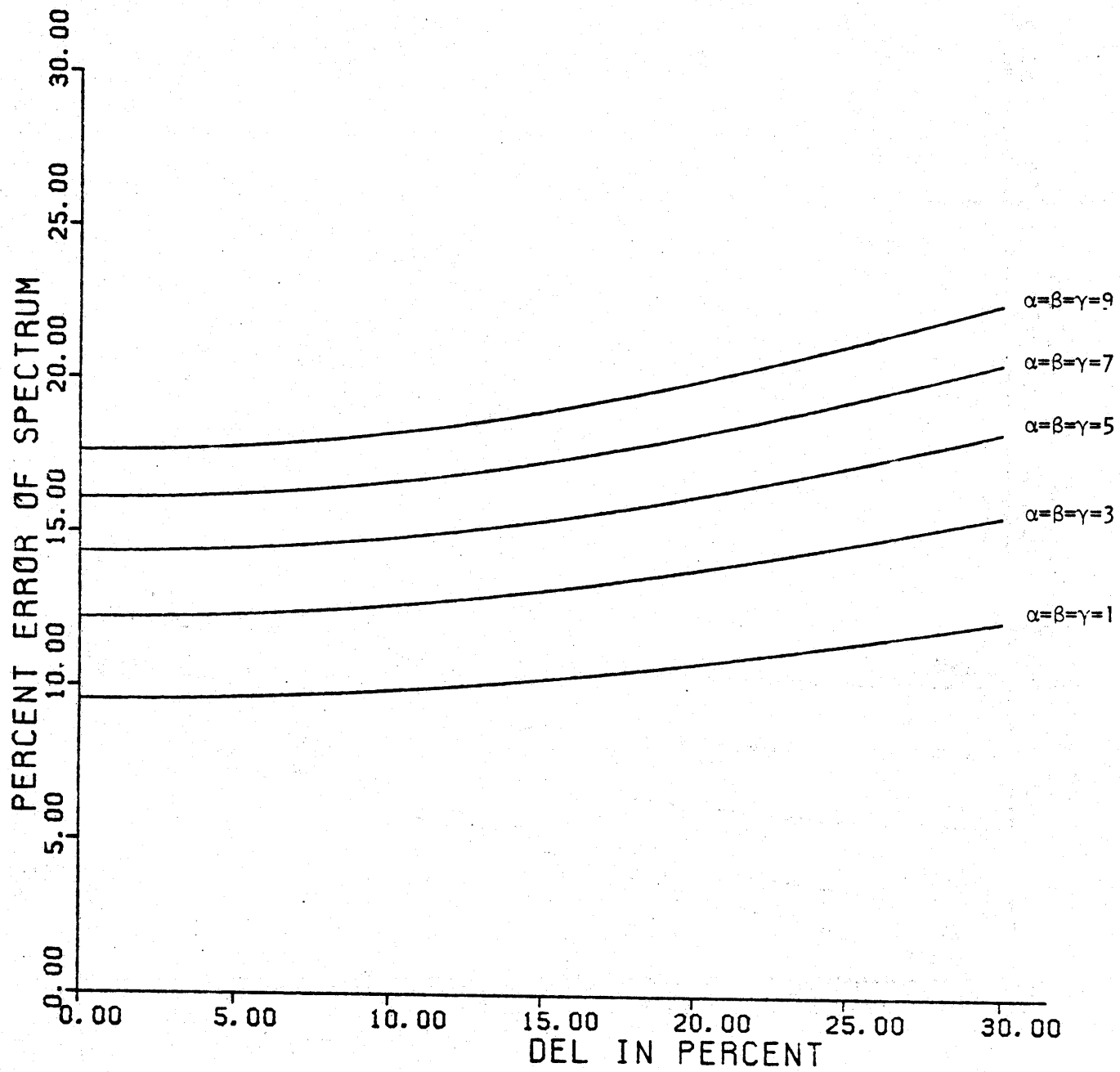


Figure 1(e)

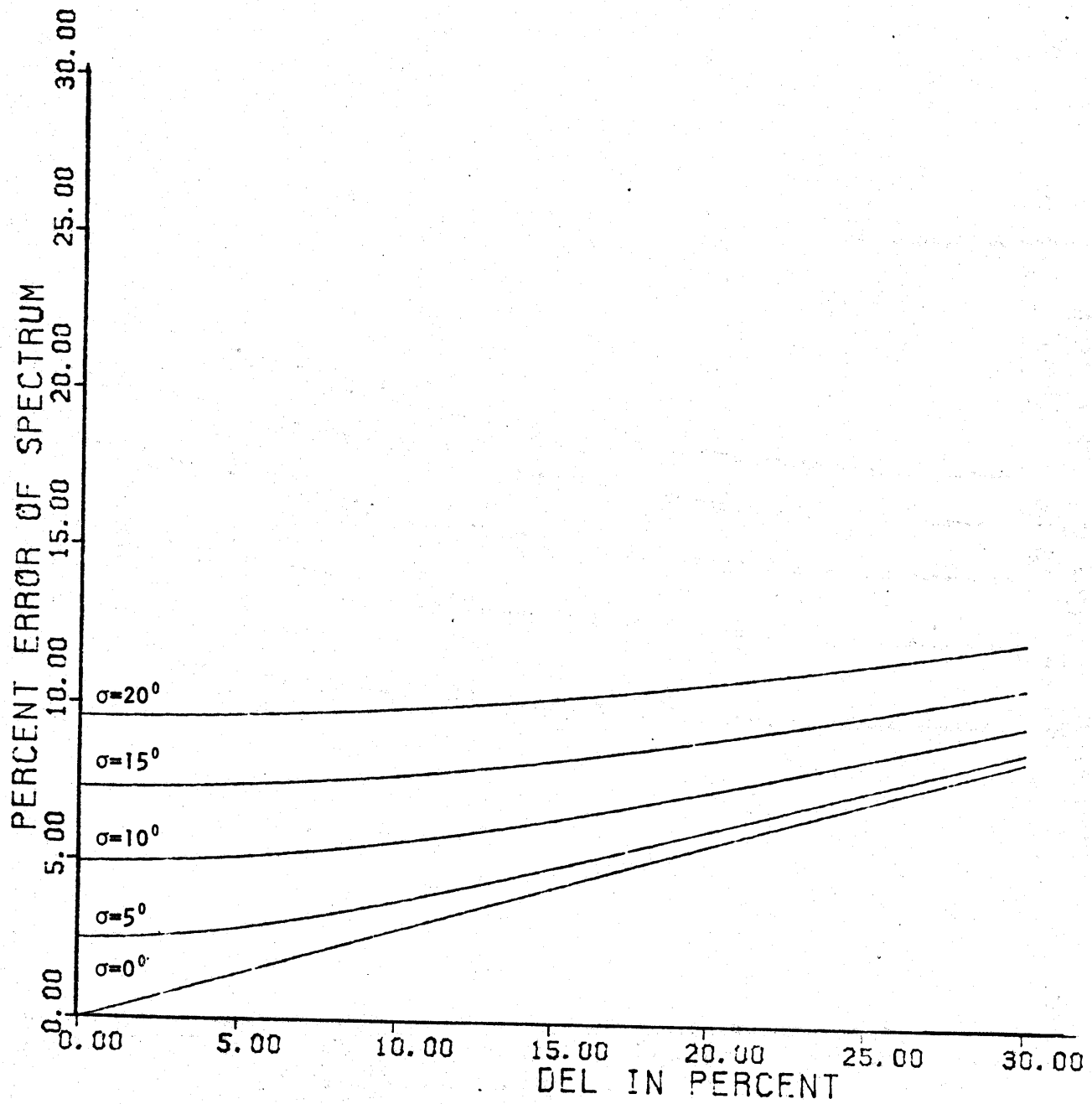


Figure 2(a)

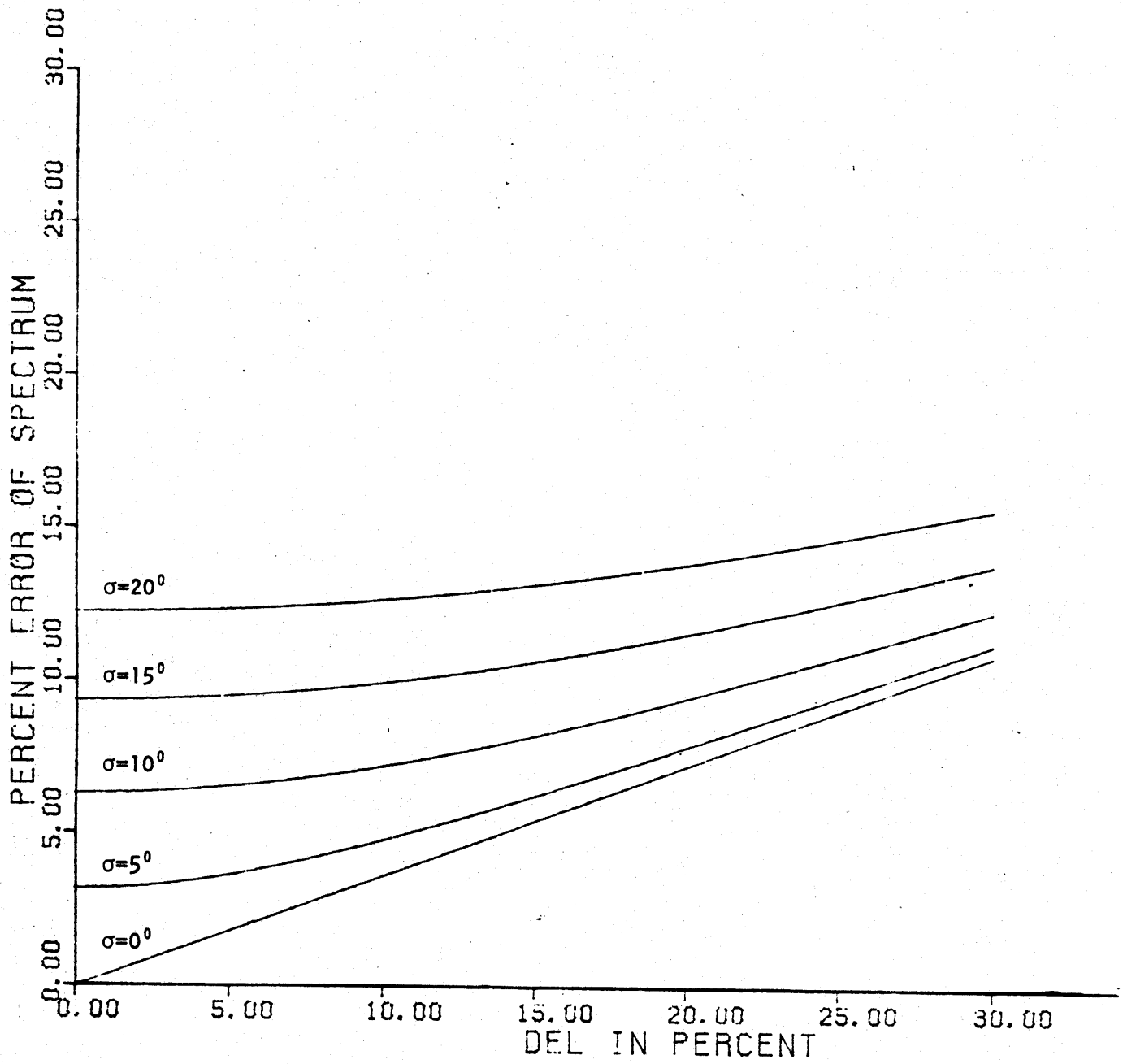


Figure 2(b)

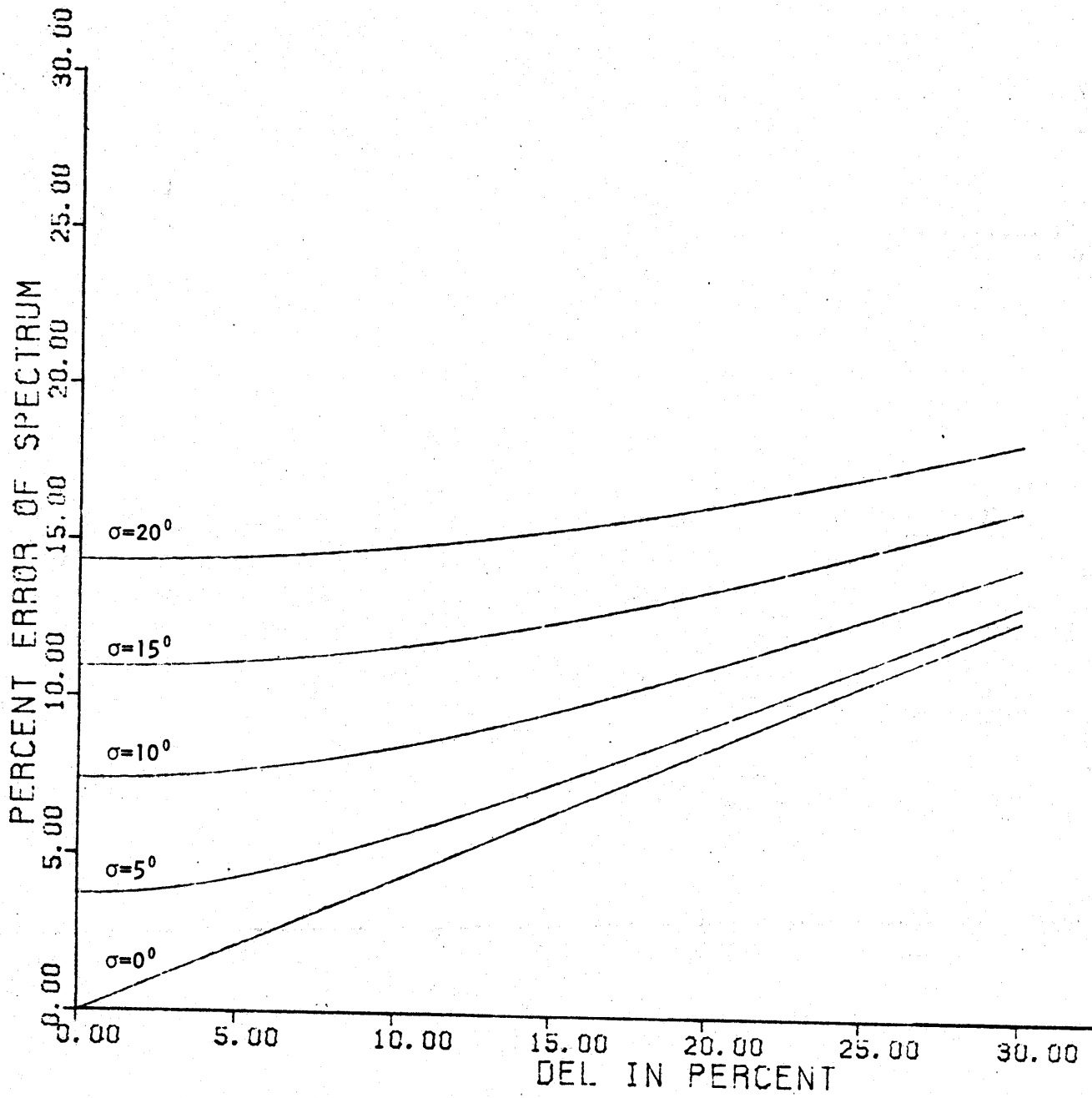


Figure 2(c)

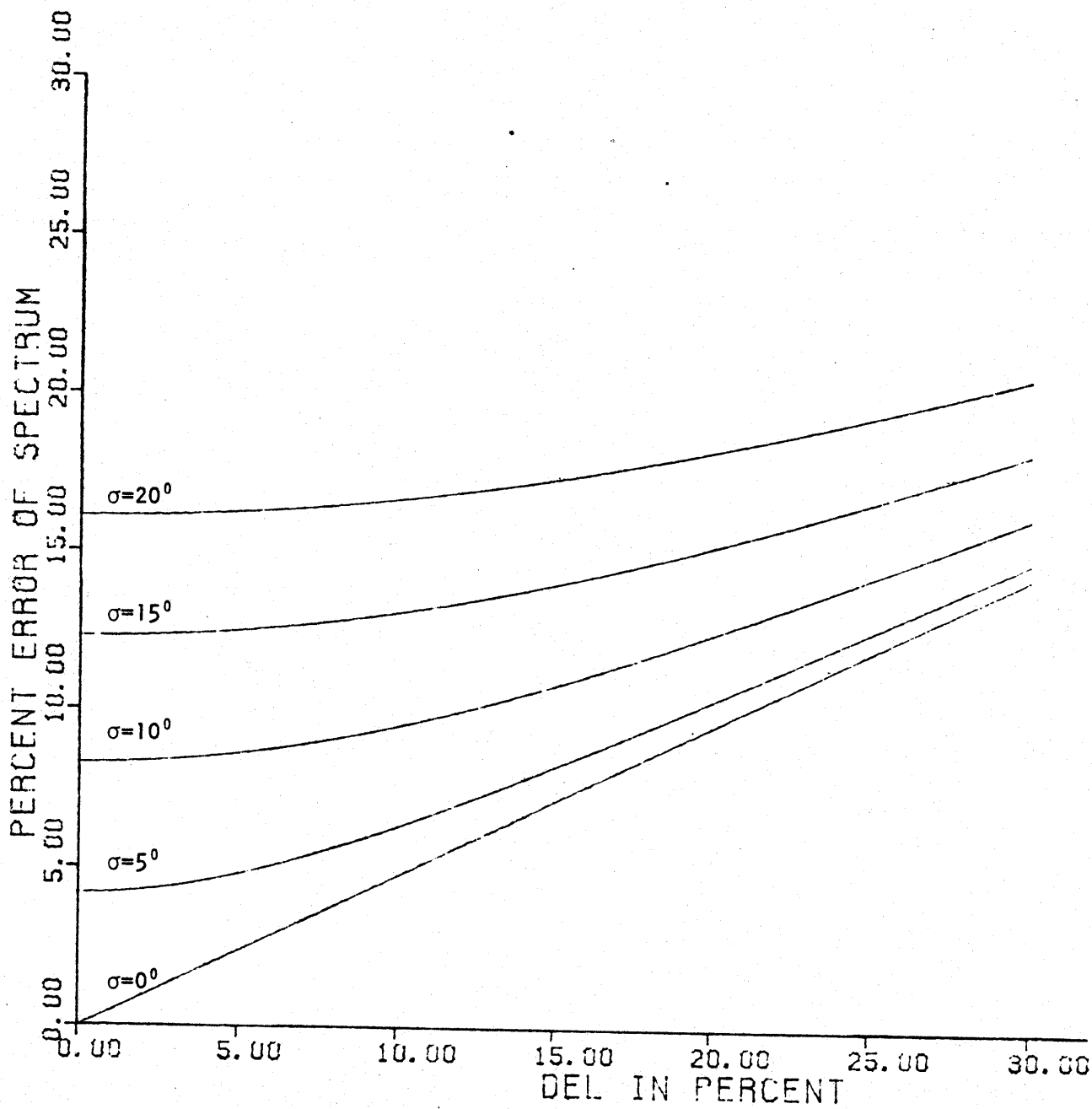


Figure 2(d)

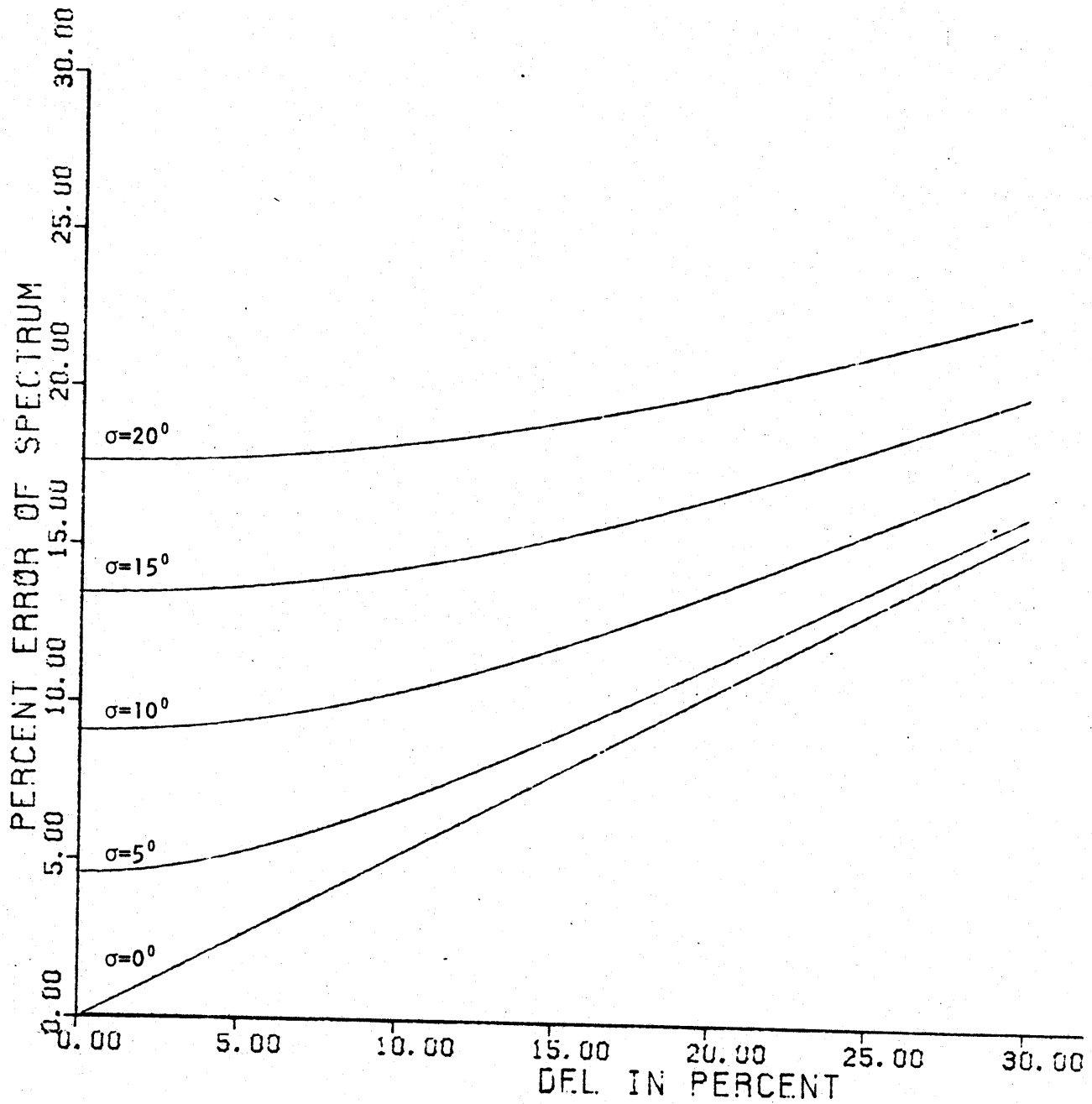


Figure 2(e)

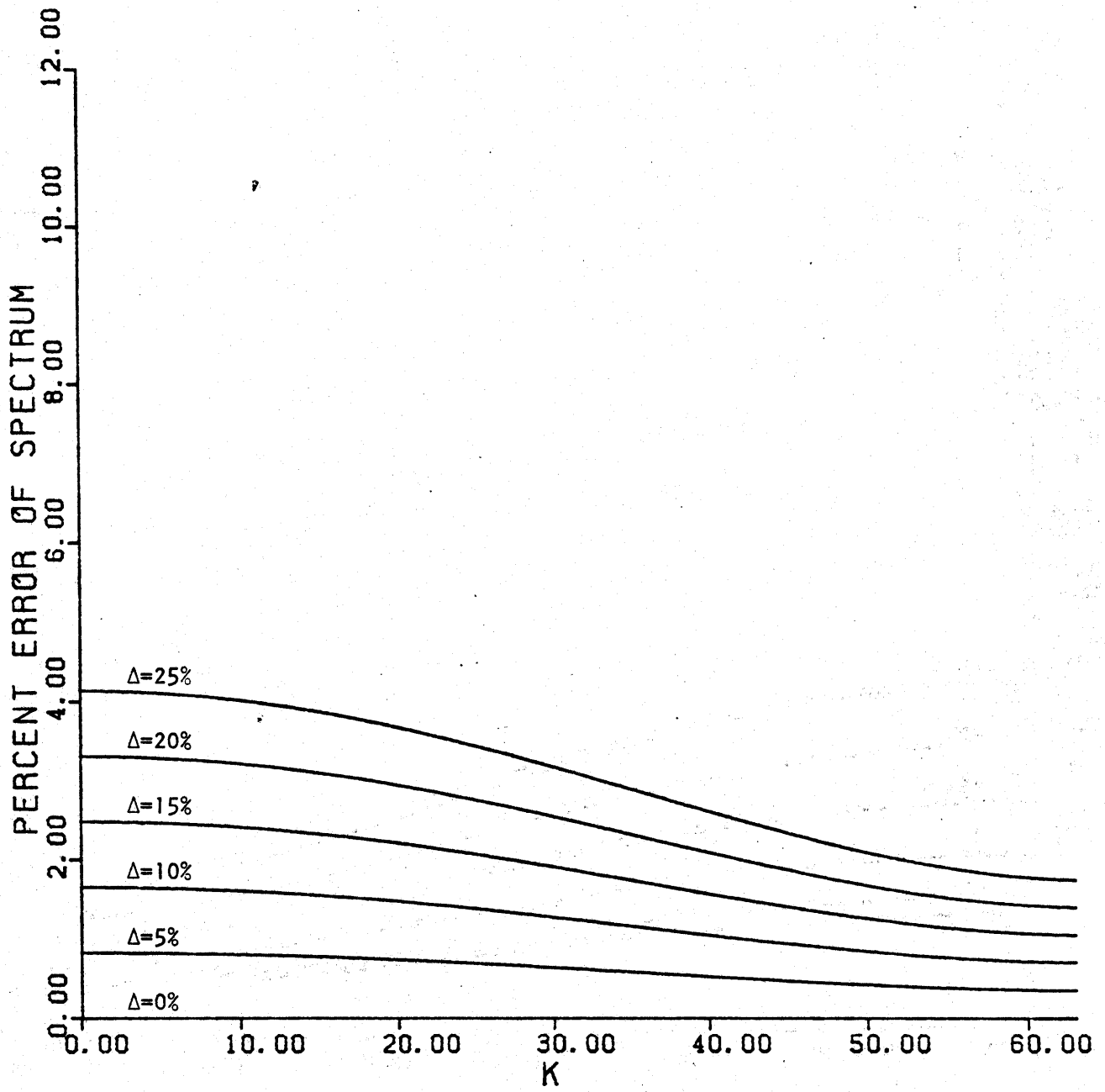


Figure 3(a)

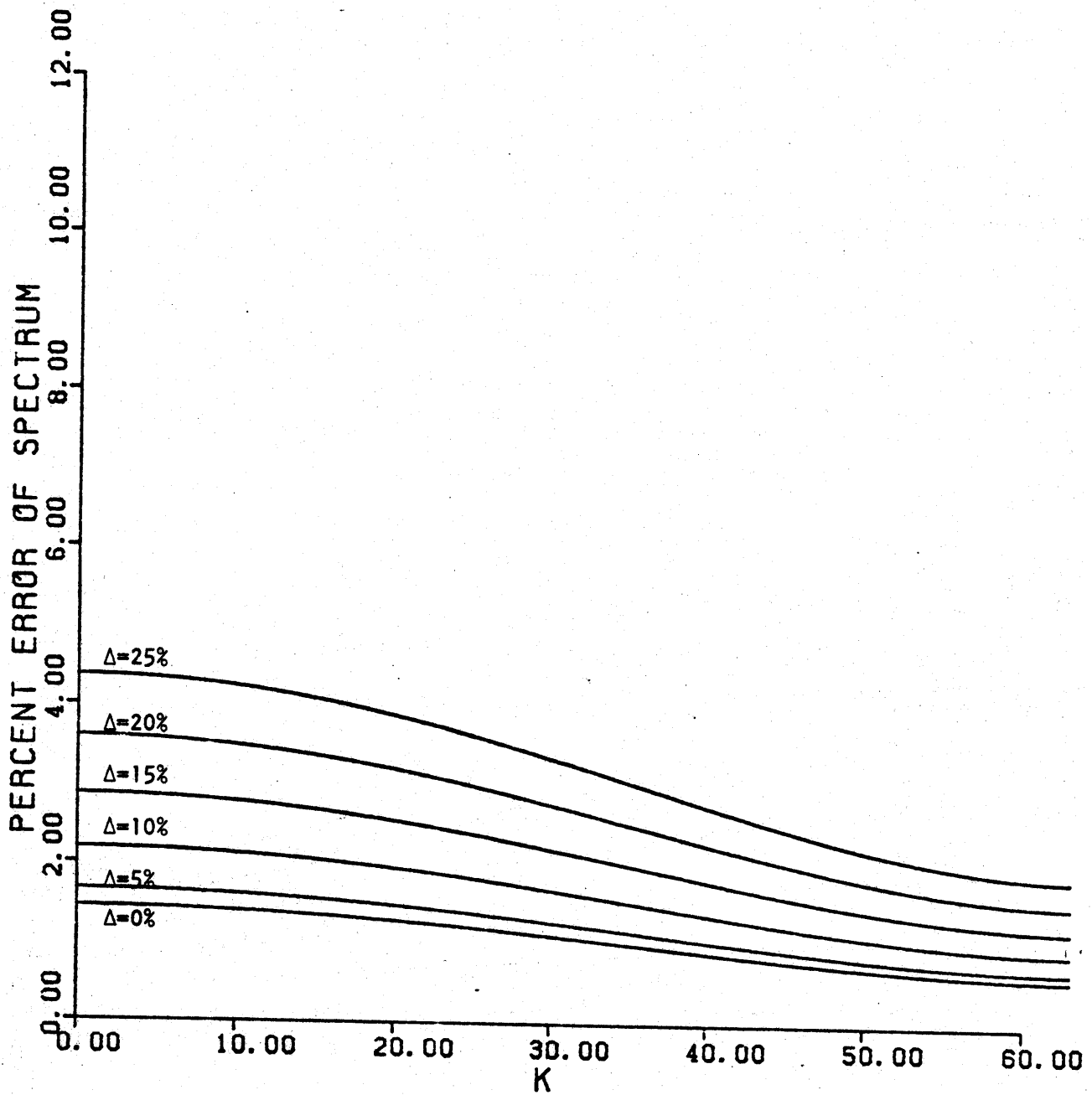


Figure 3(b)

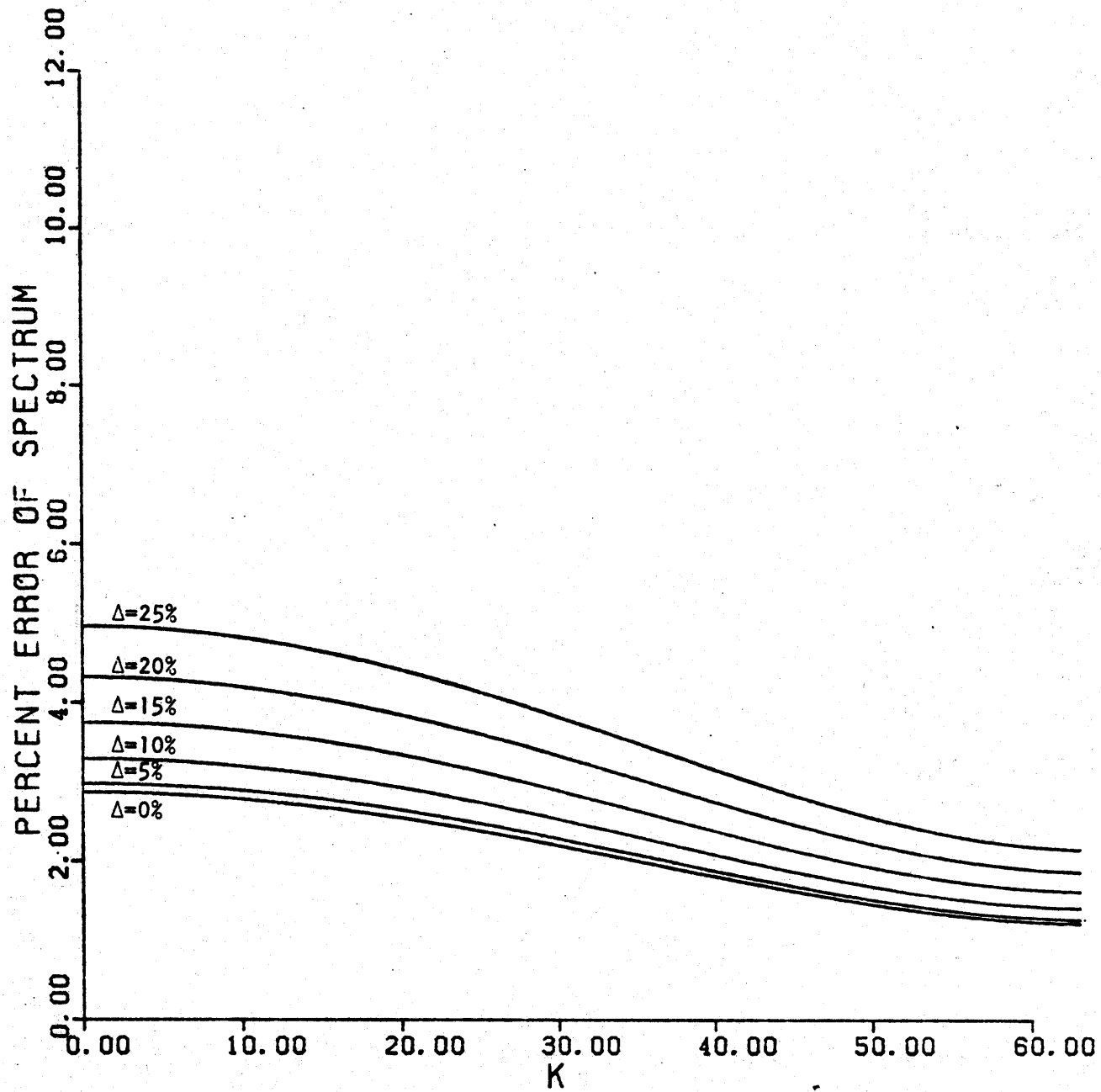


Figure 3(c)

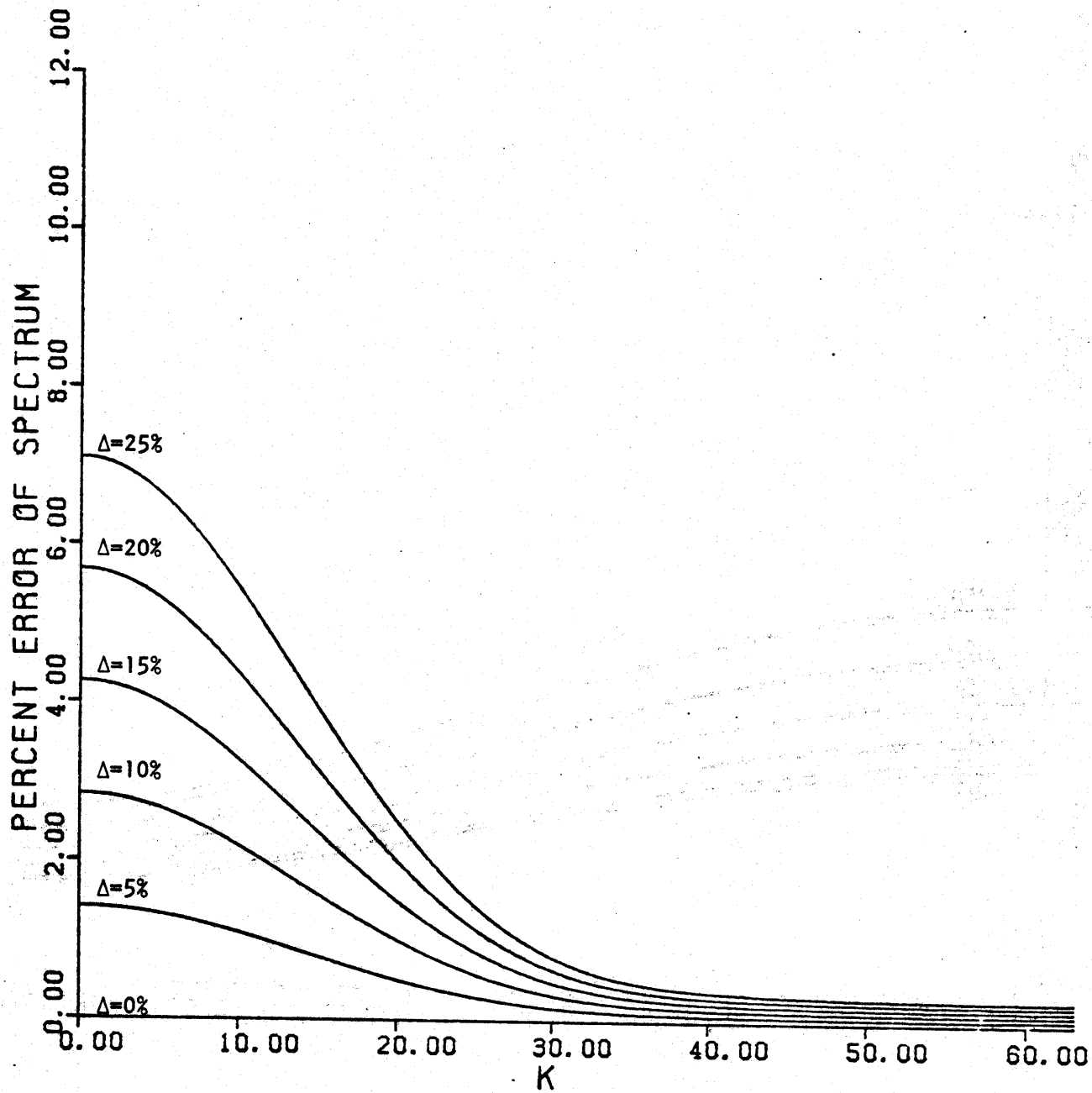


Figure 4(a)

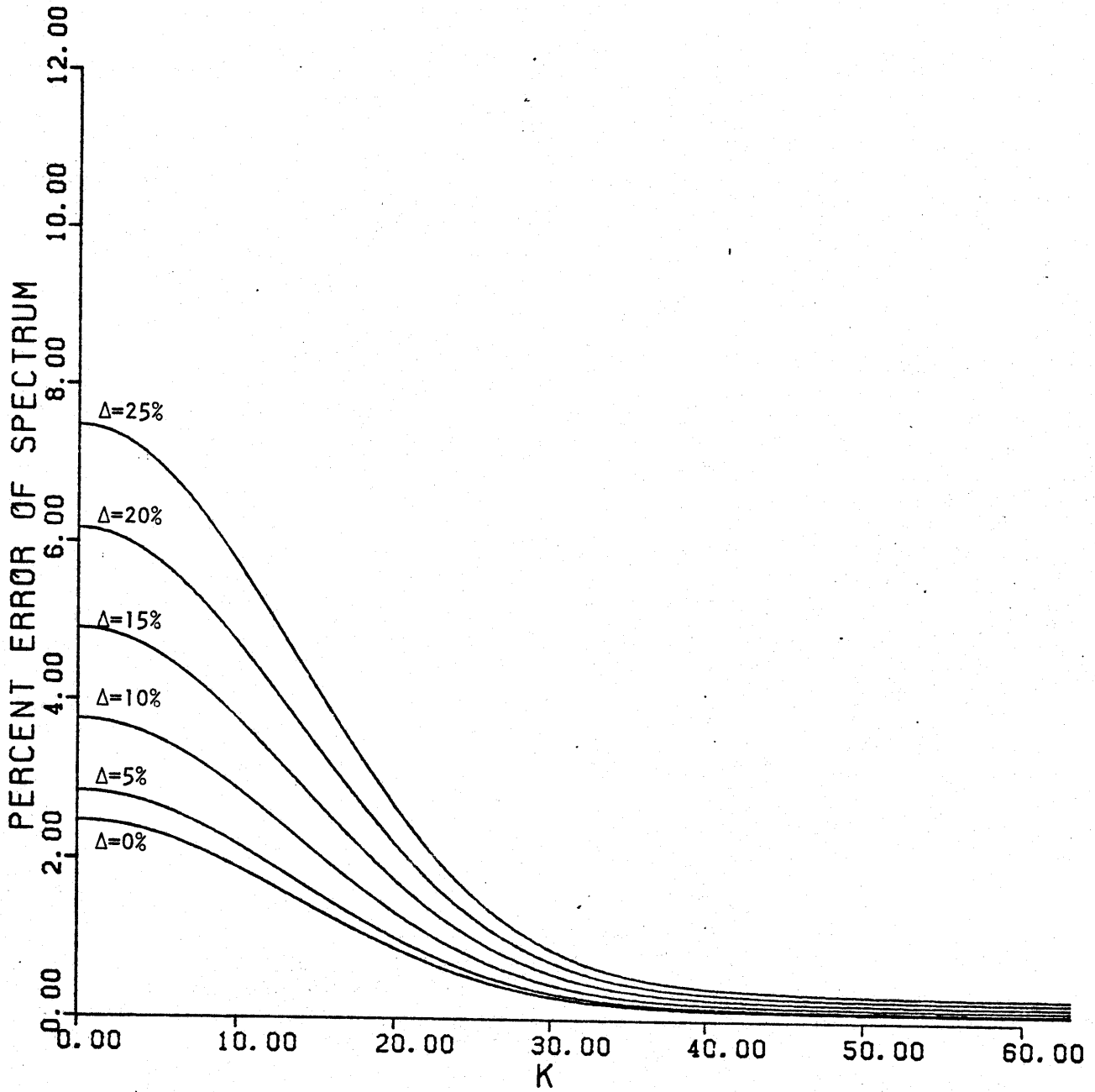


Figure 4(b)

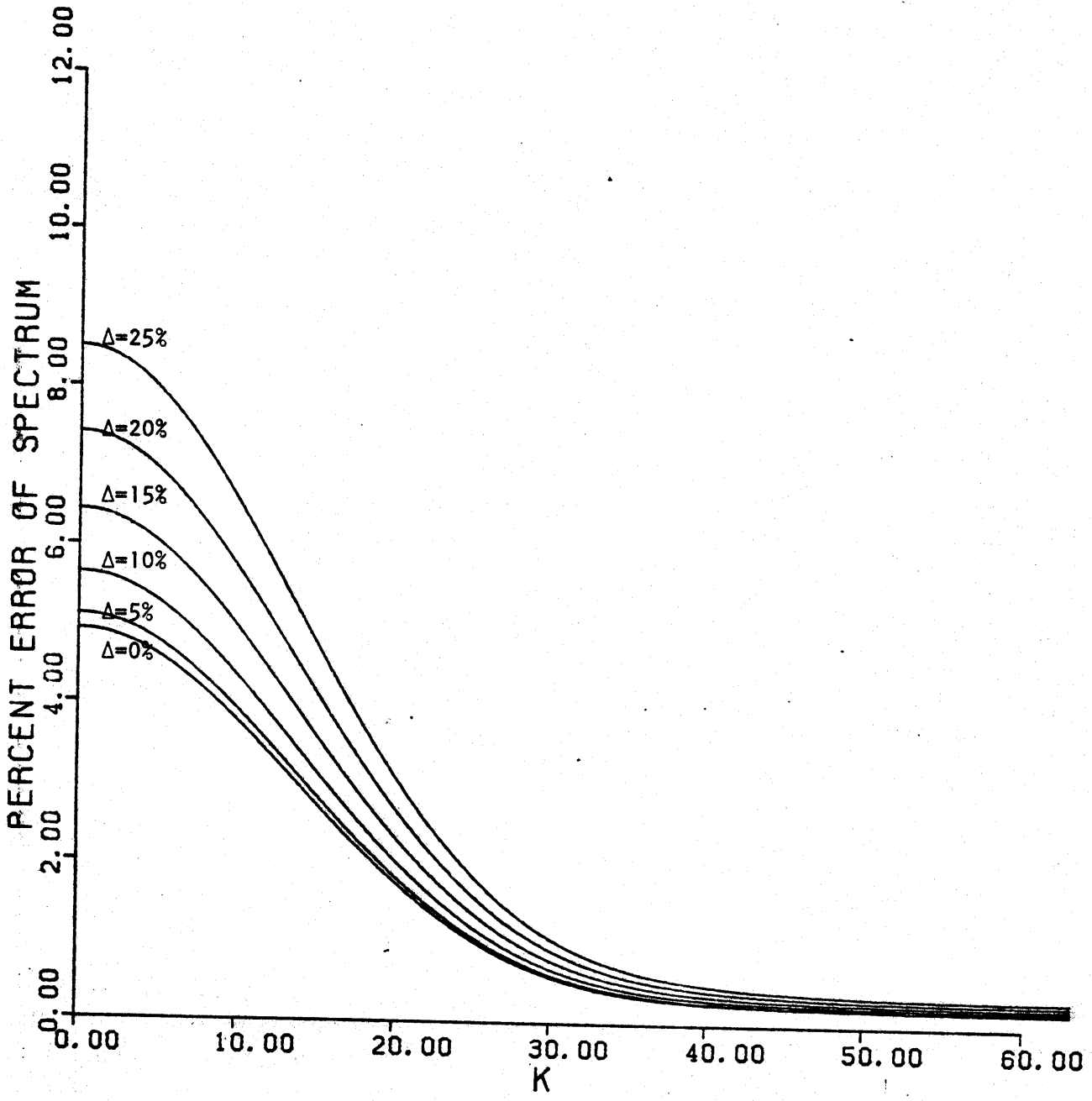


Figure 4(c)

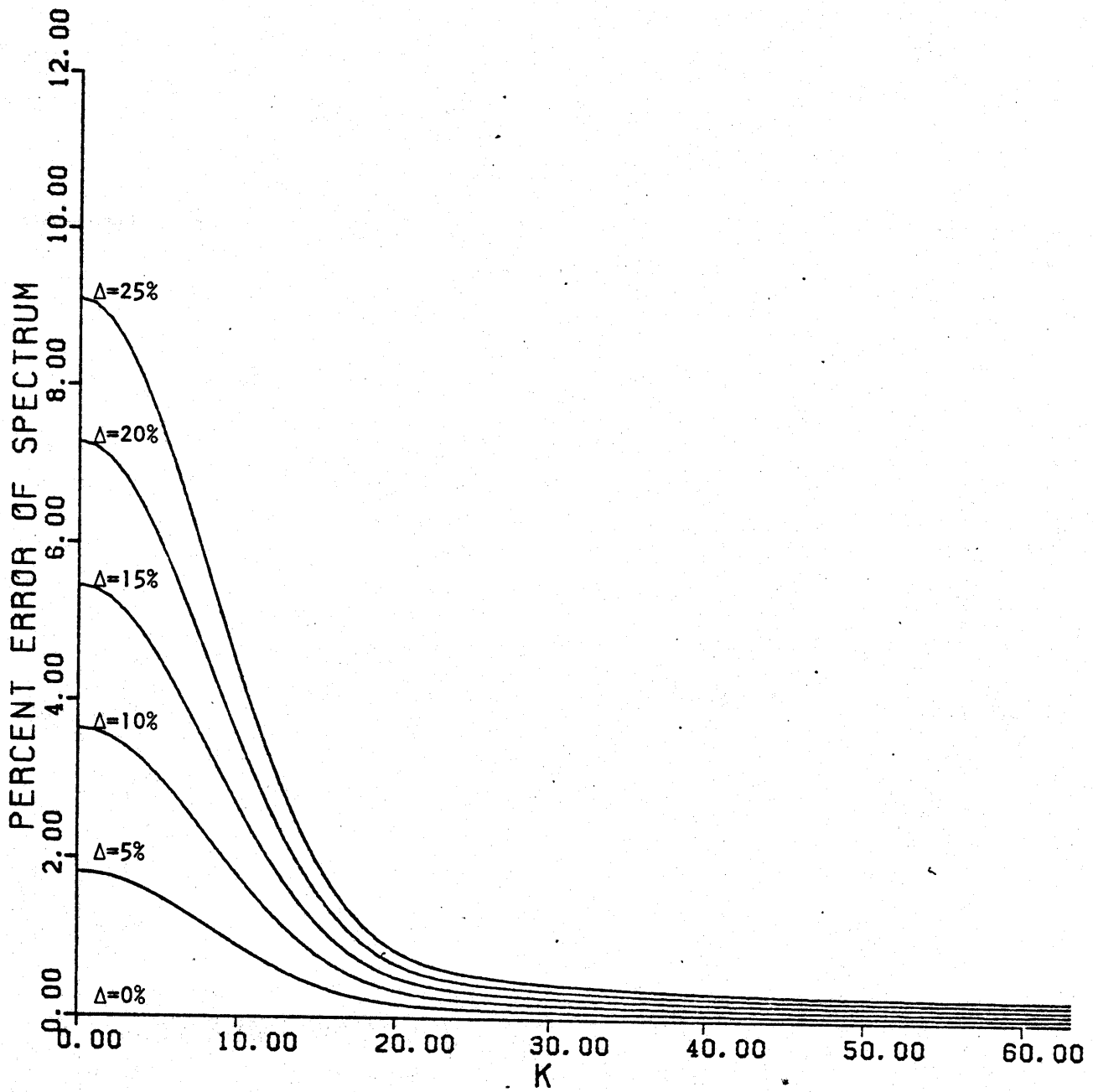


Figure 5(a)

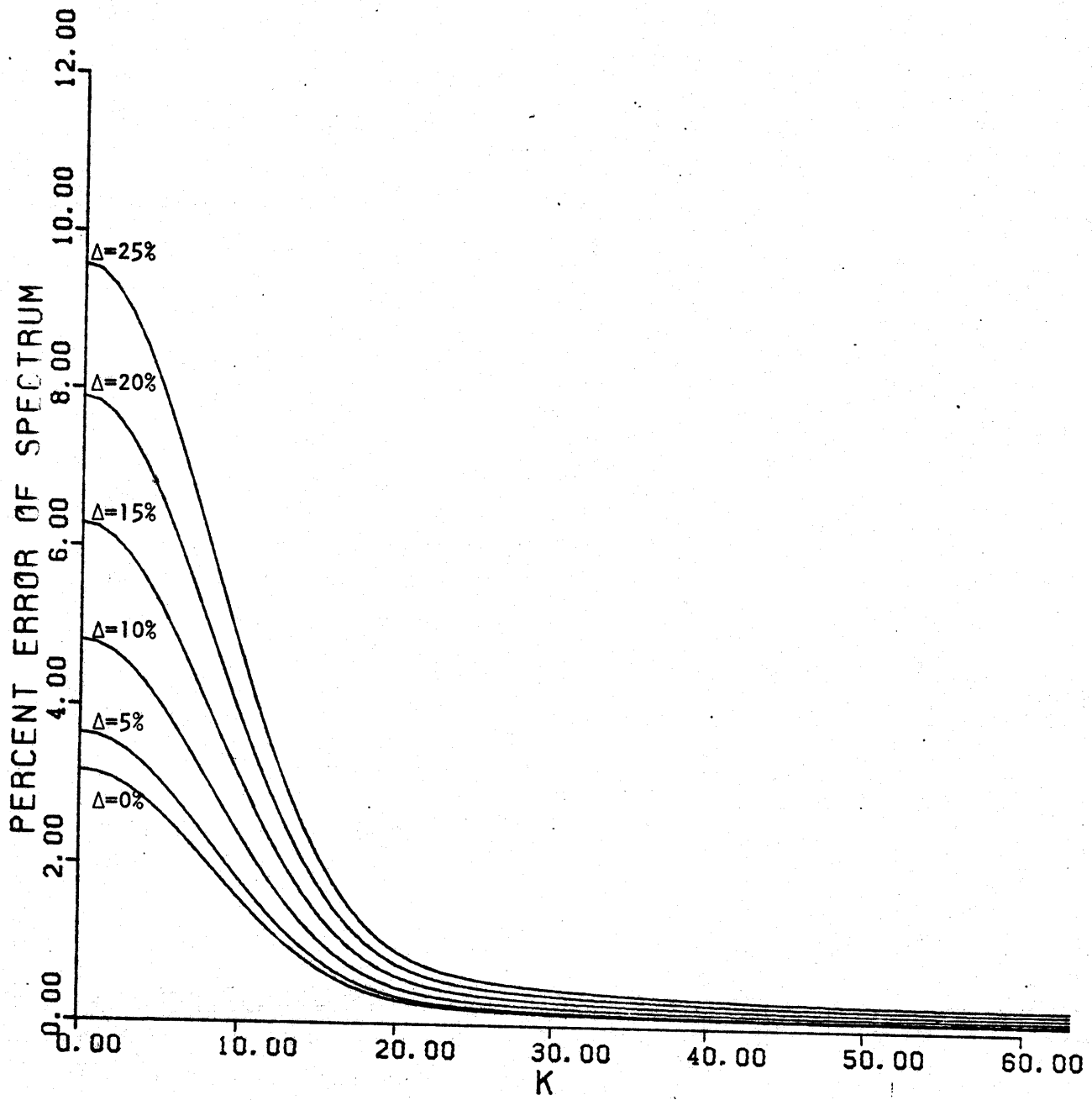


Figure 5(b)

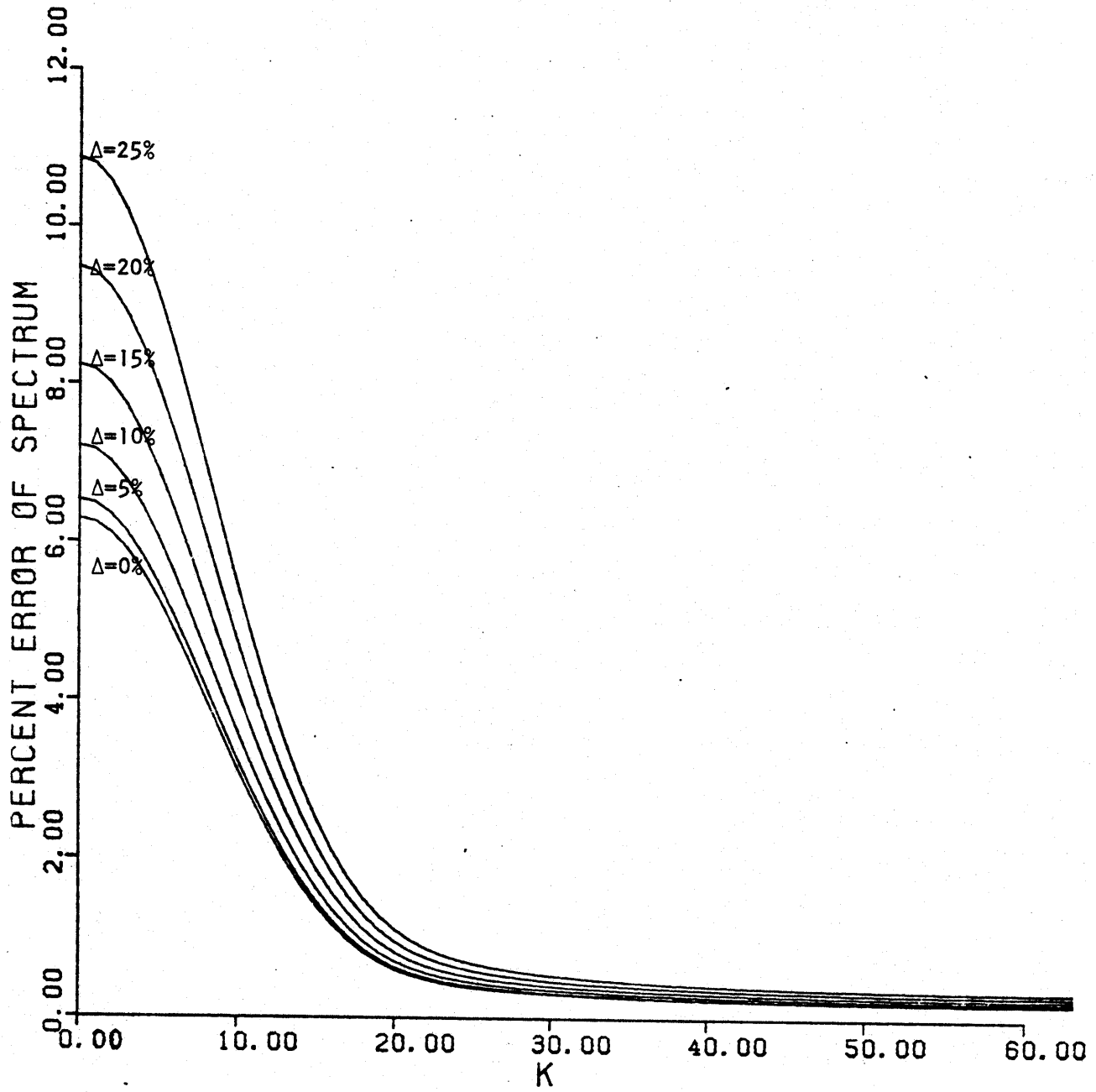


Figure 5(c)

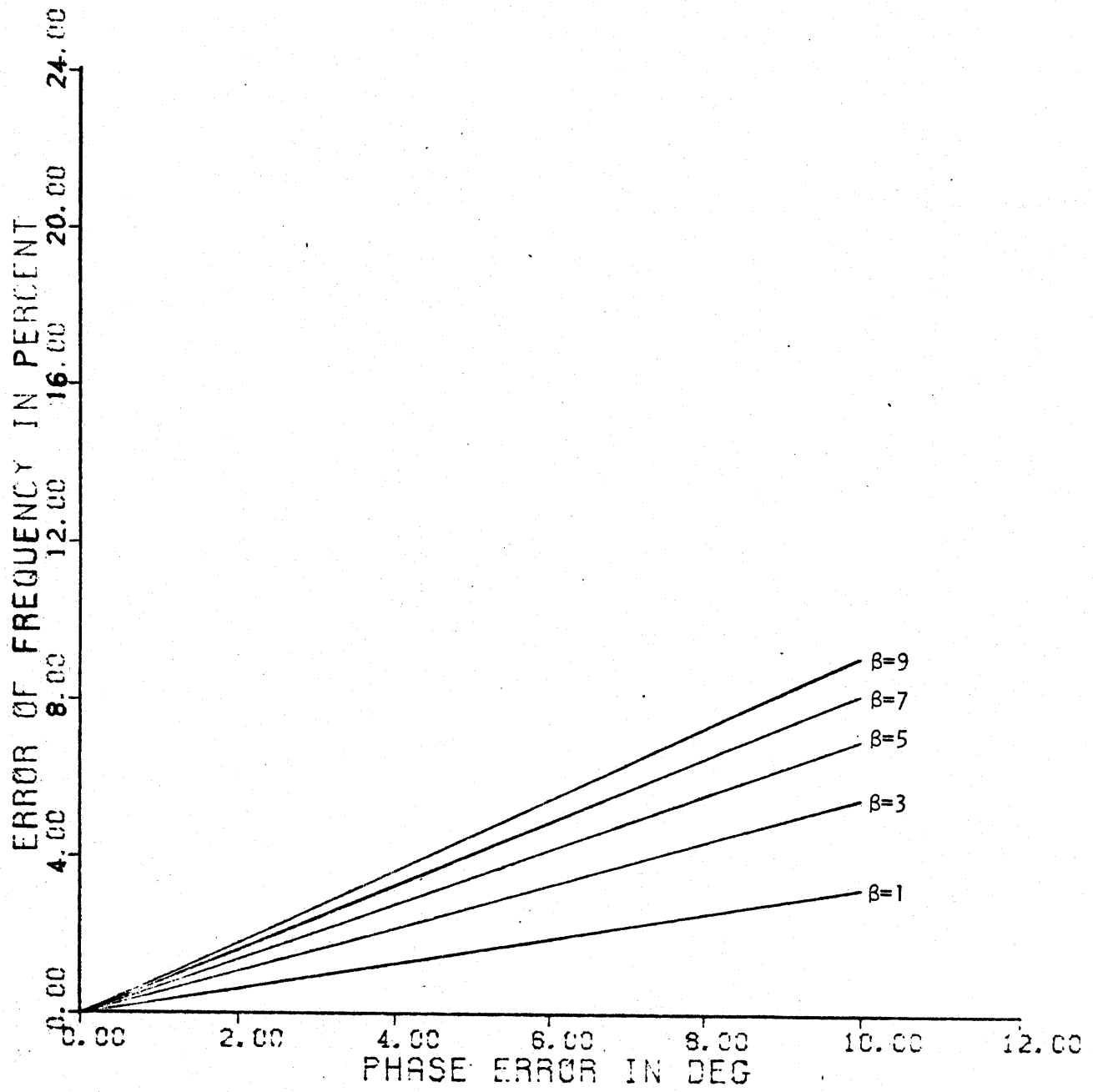


Figure 6

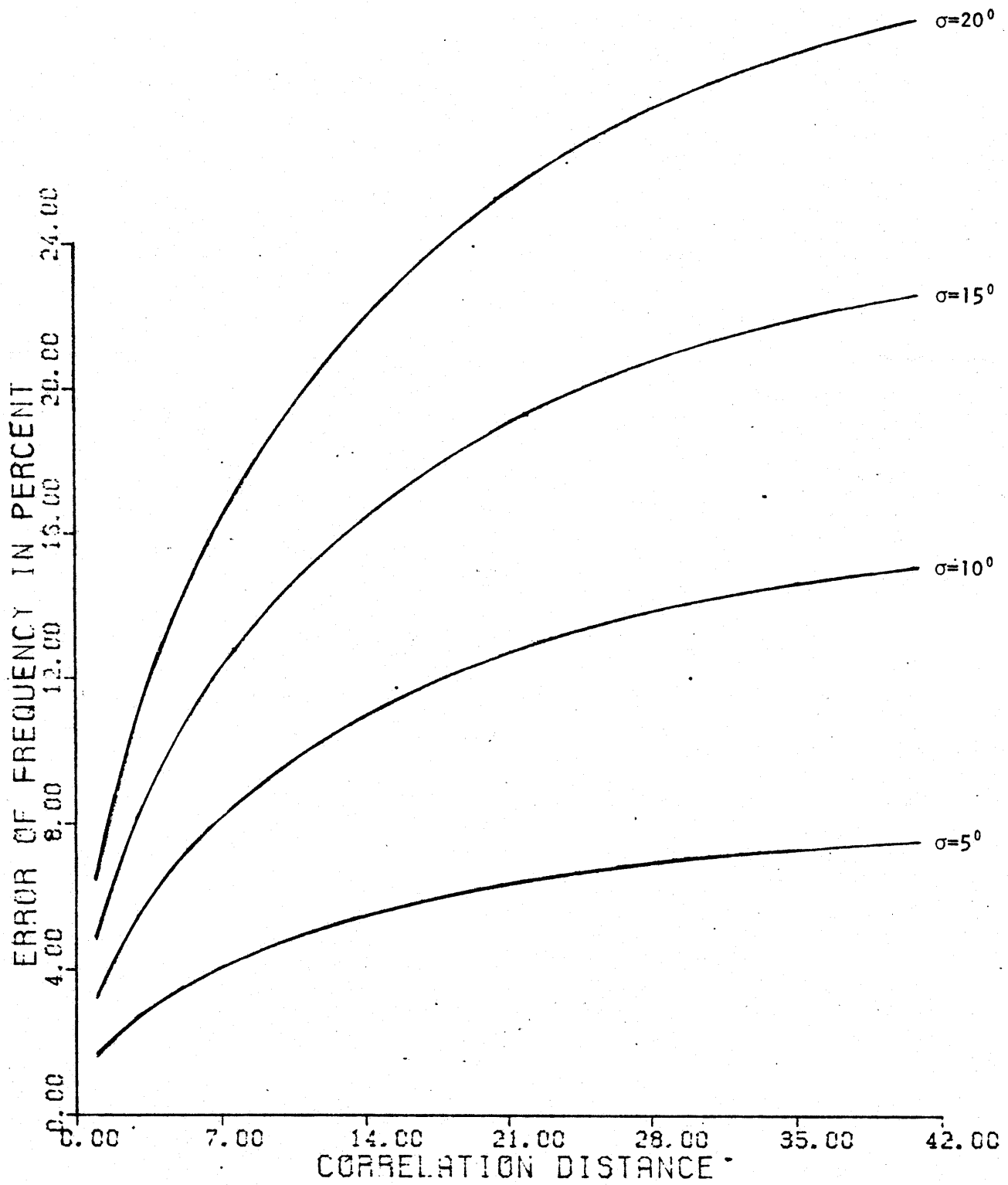


Figure 7

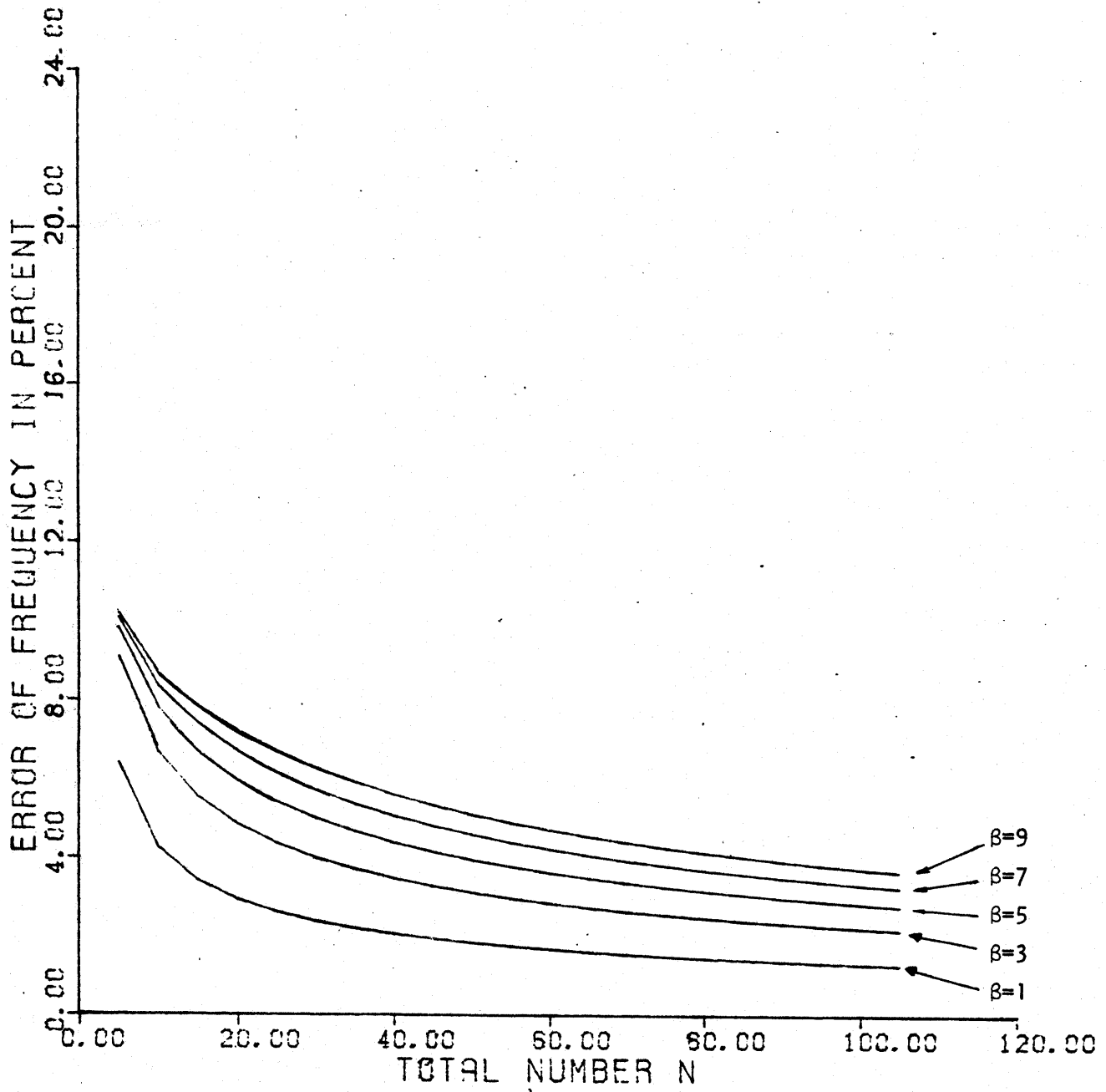


Figure 8