GEOMAGNETIC HOURLY AVERAGE AND MINUTE VALUES FROM DIGITAL DATA

C A GREEN

Geomagnetism Research Group
Report Number 85/19
September 1985

ABSTRACT

The production of geomagnetic hourly average values from discrete one minute values,and the production of the one minute values themselves from more rapidly sampled data,are investigated. It is shown that hourly average values obtained from an average of one minute values within the hour should form a homogeneous set with those previously obtained by handscaling of analogue records. Generation of the one minute values by digital filtering of more rapidly sampled data is recommended. A simple cosine weighting applied to 5 or 10 second sampled data (13 or 7 point filter respectively) is suggested.

INTRODUCTION

Hourly average values of geomagnetic field elements have long been the staple output of magnetic observatories and have been used for may scientific and soem commercial applications. When most observatories operated analogue recording instruments,hourly average field values were something that could be produced relatively simply by handscaling methods. A large world-wide databank of these data has been built up and international standards are in use for their format of storage and exchange.

For shorter period field variations,however,no such stable situation has existed up to now. The value of such data was recognised (particularly for the study of external inluences on the geomagnetic field) but few observatories were in a position to produce them easily. For aperiod in the late 1950s to early 1960s attempts were made to accumulate a databank of 2.5 minute values but the tedium of producing these by the hand digitization of analogue records led to this effort not being sustained.

In this decade,however,the introduction into observatories of digital sampling and recording systems has greatly reduced the effort necessary to produce and store short period variation data. At this time about 33% of all observatories have digital recording and there is also an increasing trend towards full automation. One minute values have become the de facto standard for short period observatory digital output and it has recently been suggested at the International Association of Geomagnetism and Aeronomy (IAGA) that these one minute values be used in the generation of the hourly average values.

In this paper we discuss the generation of hourly average field values from digital data and compare them with those produced by traditional methods. We also investigate the generation of the one minute values themselves following another suggestion at IAGA that these be obtained by digital filtering methods.

HOURLY AVERAGE VALUES

The process of taking an average corresponds to a form of low pass filtering of the original data. If the data, $x(t)$, are in continuous form, as on an analogue record, the transfer function of this averaging filter can be derived as

$$G(f) = \sin(\pi fT)/(\pi fT) \quad \ldots (1)$$

where f is frequency and the T is the time interval over which the average is taken (see e.g. Owens, 1978). If, however, the data are discrete samples taken at a time interval Δt (i.e. $x_n = x(n.\Delta t)$) a modified expression is obtained. For an average over M data points

$$G(f) = \sin(M\pi f\Delta t)/M\sin(\pi f\Delta t) \quad \ldots (2)$$

To ease comparison between the continuous and discrete cases we can use $T = M.\Delta t$ and rewrite (2) as

$$G(f) = \sin(\pi fT)/M\sin(\pi fT/M) \quad \ldots (3)$$

Because $N\sin(\theta/N) \leq \theta$ for all θ the magnitude of the transfer function in (3) is greater than or equal to that in (1) for all f but the value obtained from (3) converges to that from (1) for large M and small f. Another difference to note is that the discrete transfer function has a sine function in the denominator and is therefore symmmetric about the Nyquist frequency f_N ($= 1/2.\Delta t$). Kennedy (1980) discusses the difference between mean values obtained from continuous and discrete data and derives an equivalent expression to (2).

Obtaining geomagnetic hourly average values from a continuous analogue record is equivalent to putting T = 60 minutes in equation (1), giving

$$G(f) = \sin(60\pi f)/(60\pi f) \quad \ldots (4)$$

It has been suggested that for discrete data the hourly average be obtained from an average over the hour of one minute sampled data. Putting M = 60 and Δt = 1 minute in equation (2) gives

$$G(f) = \sin(60\pi f)/60\sin(\pi f) \quad \dots (5)$$

It can easily be shown that that the difference between (4) and (5) for f in the range $0 \leq f \leq f_N$ is insignificant,particularly as the current amplitude resolution for published hourly values is 1 nT. Thus it can be said that hourly average values obtained from an average of discrete one minute values should form a homogeneous set with those previously obtained from analogue records.

The hourly average value from continuous records is usually centred on the half-hour. To retain this for discrete data there are two options:

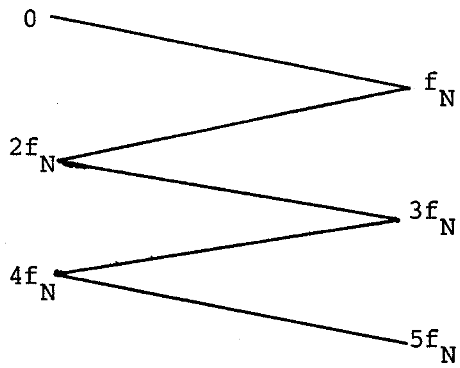
(a) Average 60 one minute values each centred on the half-minute,between t = 0 minutes 30 seconds and t = 59 minutes 30 seconds. This is strictly what has been used in putting M = 60 in (2) to obtain (5).

(b) Average 61 one minute values each centred on the minute,between t = 0 minutes and t = 60 minutes. This corresponds to putting M = 61 in equation (2) giving

$$G(f) = \sin(61\pi f)/61\sin(\pi f) \quad \dots (6)$$

Transfer function values obtained from (6) are still negligibly different from those obtained from (4) for $0 \leq f \leq f_N$. It should be noted,however,that successive hourly average values obtained using the second option are not absolutely independent of each other because the minute values at t = 0 minutes and t = 60 minutes are each used in the calculation of two hourly averages.

Up to this point it has not been defined what a one minute value is. Ideally it should be a simple spot value that in sampling terms tells us everything about the frequency content of the data below the Nyquist frequency (= 1/120 Hz) and nothing above this frequency. Unfortunately this cannot be,because the spectrum of geomagnetic variations just does not cease for frequencies $>1/120$ Hz, so if one minute spot values were to be used the spectrum for frequencies $0 \leq f \leq f_N$ would be contaminated by contributions from $f > f_N$. The mechanism by which this happens is known as aliasing, and can be thought of as a folding operation thus



etc.

The power spectrum for $f > f_N$ is folded back about f_N such that for $0 \leq f \leq f_N$ there is a contribution at f from power at frequencies $(2nf_N - f)$ and $(2nf_N + f)$ where n is an integer. If $S(f)$ is the true power spectrum at f and $S_a(f)$ the aliased spectrum, we can write

$$S_a(f) = S(f) + \sum_{n=1}^{\infty} S(2nf_N - f) + \sum_{n=1}^{\infty} S(2nf_N + f)$$

To estimate the effects on aliasing on our hourly average values obtained from discrete one minute values, aliased spectra were calculated for true geomagnetic spectra assumed to behave as $S(f) = (f/f_N)^{-\lambda}$ where $\lambda = 1, 2$ and 3. The summations above were terminated at $n = 5$ which corresponds to a cut-off period of about 11 seconds for a sampling period of one minute. This limit incorporates most of the geomagnetic pulsation band into the aliased spectrum. Defining the normalized frequency p as f/f_N we have

$$S_a(p) = S(p) + \sum_{n=1}^{5} (2n-p)^{-\lambda} + \sum_{n=1}^{5} (2n+p)^{-\lambda}$$

The quantity finally calculated was the percentage increase in the true spectrum at p due to aliasing and this quantity is plotted in figure 1 for the three values of λ; p is on a log scale. With a sampling interval of one minute, the value of p corresponding to to a period of one hour is $p = 1/30$. A vertical dashed line has been drawn on figure 1 at this frequency. The precentage increase in the spectrum at this frequency due to aliasing is 7.6% for $\lambda = 1$, 0.08% for $\lambda = 2$, and 0.00001% for $\lambda = 3$. Values of λ between 2 and 3 are usual for the background geomagnetic spectrum and the effect of aliasing would be negligible for these values of λ. The presence of large amplitude magnetic pulsations (usually with periods between 20 seconds and 600 seconds) could effectively flatten the spectrum and it can be seen that aliasing becomes more significant as λ decreases. It is recommended, therefore, that the one minute values used for hourly value calculations be 'average' or filtered values from which the frequency content of the data above f = 1/120 Hz has been reduced as much as is practical.

A proposal has already been made at IAGA that the one minute values should be derived from digital filtering methods and possible simple options for this are discussed in the next section.

ONE MINUTE VALUES

As a starting point we make the assumption that successive one minute values should be as independent from each other as possible, thus approximating spot values. This means that each value should be derived from other discrete spot values sampled within that minute period.

We also require that the final minute value be centred on the minute or half-minute and that the filter has zero phase shift at all frequencies. The filtering operation in the time domain is a convolution with a series of filter weights and is thus defined by

$$y_n = \sum_{k=-m}^{+m} w_k x_{n+k}$$

where y_n and x_n are the discrete values of the filtered and raw data series respectively at $t = n.\Delta t$ and w_k the filter weights. For zero phase shift the filter has to be symmetric (i.e. $w_{-k} = w_k$) and for the gain of the filter to be unity at zero frequency for a low pass filter

$$\sum_{k=-m}^{+m} w_k = 1$$

The operation as defined above has the total number of filter weights (M) = 2m+1, which is an odd number. This means that if all possible samples taken in a minute are used to obtain the filtered value, successive filtered values will have one of the original data values in common; the first and last values in each minute are used in two filtering operations. This does not meet the criterion of full data independence but the use of an odd number gives more filtering options and the significance of the overlap decreases with increasing M. The transfer function for the filter as defined above is

$$G(p) = w_0 + 2 \sum_{k=1}^{m} w_k \cos(\pi k p)$$

where p is the normalized frequency as before.

For our purposes the ideal transfer function is one that has unit gain for $f < 1/120$ Hz and zero gain for $f > 1/120$ Hz. This ideal 'step function' low pass filter can only be achieved with an infinite number of filter weights. To retain independence of filtered values the number of filter weights can only be increased by increasing the data sampling rate. Practical considerations such as data storage, instrumental response times and reliability suggest that sampling rates should not be pushed to extremes. Also, the amount of computation necessary to generate the filtered data values is proportional to the number of filter weights. Sampling intervals of 30,15,10,5,2 and 1 second will be considered. If all the values in the minute are used in the filter and M is odd, these intervals correspond to values of M =3,5,7,13,31, and 61 respectively.

There are several types of filter weighting scheme that could be used but this discussion will be confined to four that are simple to implement and have a transfer function that is smooth in the pass band.

(a) Rectangular weighting

For this scheme $w_k = 1/M$ for all k,which corresponds to the taking of a simple mean of the data over M points. The transfer function of this filter is

$$G(p) = \sin(M\pi p/2)/M\sin(\pi p/2)$$

(b) Triangular weighting

For this scheme $w_k = (1 - |k|/L)/L$ where $L = m+1 = (M+1)/2$,and the transfer function is

$$G(p) = (\sin(L\pi p/2)/L\sin(\pi p/2))^2$$

(c) Cosine weighting

For this scheme $w_k = (1 + \cos(\pi k/L)/2L$ where L is as above,and the transfer function is

$$G(p) = \frac{1}{L} + \frac{\cos(L\pi p/2)\sin(m\pi p/2)}{L\sin(\pi p/2)} - \frac{\sin(L\pi p/2)\sin(m\pi(p+1/L)/2)}{2L\sin(\pi(p+1/L)/2)} + \frac{\sin(L\pi p/2)\sin(m\pi(p-1/L)/2)}{2L\sin(\pi(p-1/L)/2)}$$

(d) Binomial weighting

For this scheme $w_k = \frac{(2m)!}{2^{2m}(m+|k|)!(m-|k|)!}$ where $2m = M-1$,and the transfer function is

$$G(p) = \cos^{2m}(\pi p/2)$$

For illustration,these weighting schemes and their transfer functions are shown plotted in figures 2 and 3 for the case M = 7. The modulus of the transfer function has been plotted and where a lobe is shaded this is an indication that it is really negative. From these figures some general conclusions can be drawn. The rectangular function has the most most rapid attenuation with frequency but large lobes (alternately negative and positive) outside the main pass band. The traingular function gives less rapid attenuation for the same value of M but smaller side lobes which are always positive because the transfer function is a squared sinc function. Cosine weighting gives slightly less rapid attenuation than triangular but smaller (alternating sign) side lobes;these result from the transfer function being the sum of three sinc functions and the side lobes tend to cancel each other out. Binomial weighting has a transfer function that has the least rapid attenuation for a given M but no side lobes at all.

A filter is required that attenuates frequencies < 1/120 Hz as little as possible and passes the minimum above this frequency up to the Nyquist frequency. An integral transfer function can be defined that is

$$H(p) = \int_0^p |G(p)|.dp \bigg/ \int_0^1 |G(p)|.dp$$

This is the fraction of the total response of the filter which is between 0 and p. Suitable criteria for judging the effectiveness of the filter might be that $G(p_c)$ and $H(p_c)$ be greater than some chosen values,where p_c is the normalized frequency for a period of 120 seconds,given by $p_c = (1/120)/f_N = (1/120)/(2.\Delta t) = 1/(M-1)$, if all the values in the minute are used.

Figures 4 and 5 show $G(p_c)$ and $H(p_c)$ respectively plotted as a function of M for the four weighting schemes. The limiting values chosen were $G(p_c) > 1/2^{\frac{1}{2}}$ (equivalent to the transfer function being 3 dB down in power at this point) and $H(p_c) > 0.5$. These limits are represented by the horizontal dashed lines on the figures. Both criteria must be satisfied. Rectangular weighting fails to achieve this for any M. For triangular weighting only M = 13 is satisfactory, for cosine weighting M = 7 or 13,and for binomial weighting only M = 5. There seems no advantage in choosing triangular weighting over cosine weighting. Their values of $H(p_c)$ are almost identical but cosine weighting has $G(p_c)$ higher. In terms of the criteria we have adopted there is little difference between the remaining options. Binomial weighting with M = 5 has a slightly better integral response and a slightly worse differential response than cosine weighting with M = 7; the difference is accentuated for cosine weighting with M = 13.

Possible aliasing effects must also be considered however. Choosing M = 5 implies

Δt = 15 seconds, and M = 7 and M = 13 imply Δt = 10 seconds and Δt = 5 seconds respectively. The Nyquist frequencies for Δt = 15,10 and 5 seconds are 1/30 Hz, 1/20 Hz and 1/10 Hz. Frequencies above these Nyquist frequencies are aliased. We require the aliasing to be small for $f \leq 1/120$ Hz. Figure 6 shows the same aliasing curves as in figure 1. The vertical dashed lines are drawn at the values of p corresponding to f = 1/120 Hz for the three different sampling intervals. The percentage increases in the spectrum at f = 1/120 Hz due to aliasing are summarized in Table 1 for the different values Δt (and M) and for the different values of the spectral index λ .

λ between 2 and 3 is typical for the background geomagnetic spectrum and aliasing is small for all Δt for these values. With λ = 1, aliasing becomes much more significant and increases with increasing Δt. A value of λ = 1 is not usually encountered but the presence of large amplitude magnetic pulsations can effectively flatten the spectrum so the use of a sampling interval of 10 seconds or 5 seconds (implying cosine weighting filters of M = 7 or M = 13 respectively) is recommended. Also, because the Nyquist periods for Δt = 10 seconds and Δt = 5 seconds are 20 seconds and 10 seconds, and most significant pulsations have periods > 20 seconds, the majority of the pulsation activity will not be aliased if either of these shorter sampling intervals is used.

CONCLUSIONS

It has been shown that geomagnetic hourly average values derived from an average of discrete one minute values will form a data set that is homogeneous with that previously obtained from continuous records. 60 one minute values centred on the half-minute give a slightly closer approximation to the continuous case than the use of 61 one minute values centred on the minute but the difference would not be detectable at the 1 nT resolution used for hourly values.

An investigation into the generation of the one minute values themselves has suggested that they be derived from 10 or 5 second sampled discrete values, low pass filtered using a symmetrical cosine weighting function (7 or 13 point filters respectively). Our results do not give sufficient evidence to choose conclusively between 10 and 5 second sampling; either would seem satisfactory. In this situation it might be preferable to opt for 10 second sampling because this would require less data storage and/or processing and prove less of a problem for the instrumentation.

ACKNOWLEDGEMENTS

Discussions with J C Riddick and A J Forbes were very helpful in preparing this report, which is published with the permission of the Director, British Geological Survey (NERC).

REFERENCES

Kennedy, J S (1980) Comments on "On the detrending and smoothing of random data" by A J Owens. J. Geophys. Res., 85, 219.

Owens, A J (1978) On the detrending and smoothing of random data. J. Geophys. Res., 83, 221.

TABLE 1

	Δt = 5 seconds M = 13	10 7	15 5
λ = 1	19.0%	38.2%	57.6%
2	0.5%	2.1%	4.7%
3	≪ 0.1%	0.1%	0.5%

Percentage increase due to aliasing in the power spectrum at f = 1/120 Hz for different sampling intervals (Δt) and spectral indices (λ).

FIGURE CAPTIONS

Figure 1 : Percentage increase in the power spectrum due to aliasing for spectra assumed to vary as $p^{-\lambda}$ where p is the normalized frequency f/f_N. The dashed line is for f = 1/120 Hz when Δt = 1 minute.

Figure 2 : Filter weights and transfer function of a 7 point filter for (a) rectangular weighting and (b) triangular weighting.

Figure 3 : Filter weights and transfer function of a 7 point filter for (a) cosine weighting and (b) binomial weighting.

Figure 4 : Magnitude of the filter transfer function,G,at f =1/120 Hz vs. number of filter weights,M,for different filter weighting schemes: (a) rectangular (b) triangular (c) cosine (d) binomial.

Figure 5 : Integral filter transfer function,H,at f = 1/120 Hz vs. number of filter weights,M,for different filter weighting schemes: (a) rectangular (b) triangular (c) cosine (d) binomial.

Figure 6 : Percentage increase in the power spectrum due to aliasing for spectra assumed to vary as $p^{-\lambda}$ where p is the normalized frequency f/f_N. The dashed lines correspond to f = 1/120 Hz when Δt =15,10 and 5 seconds.

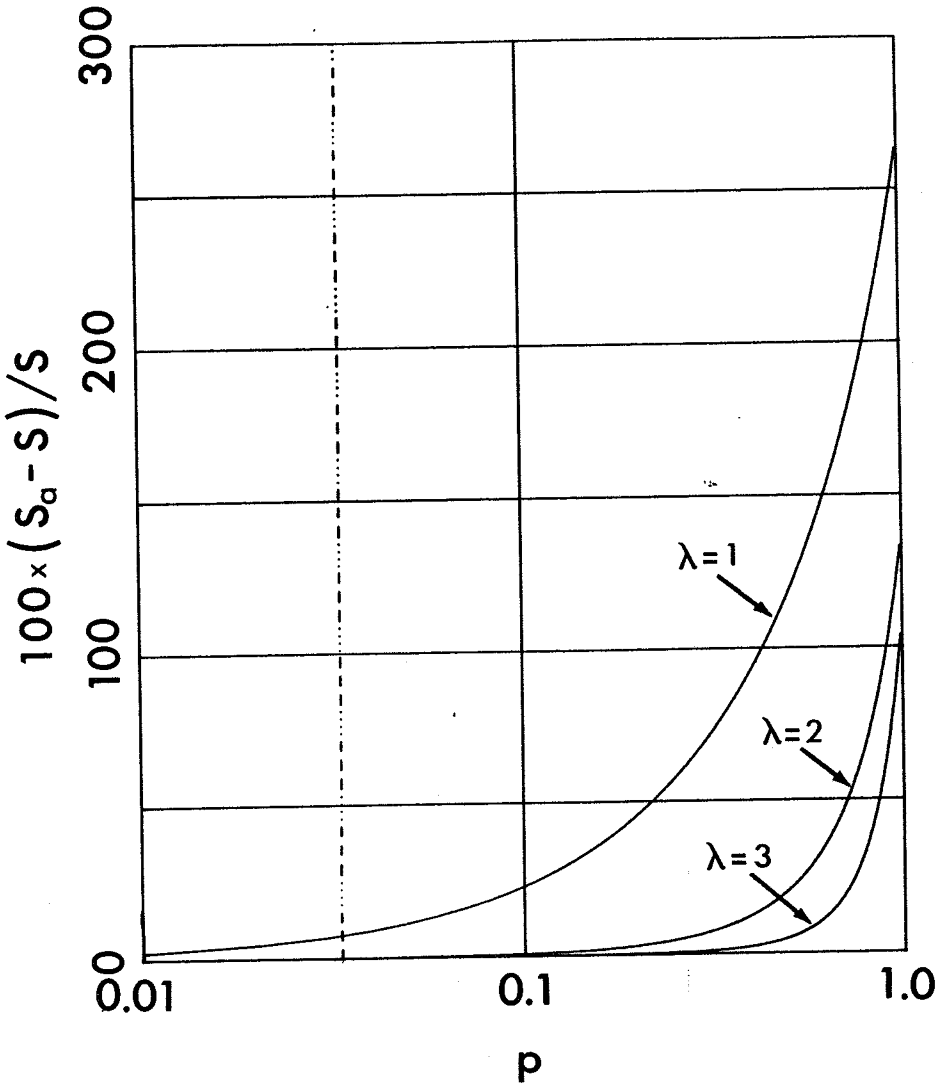


FIGURE 1

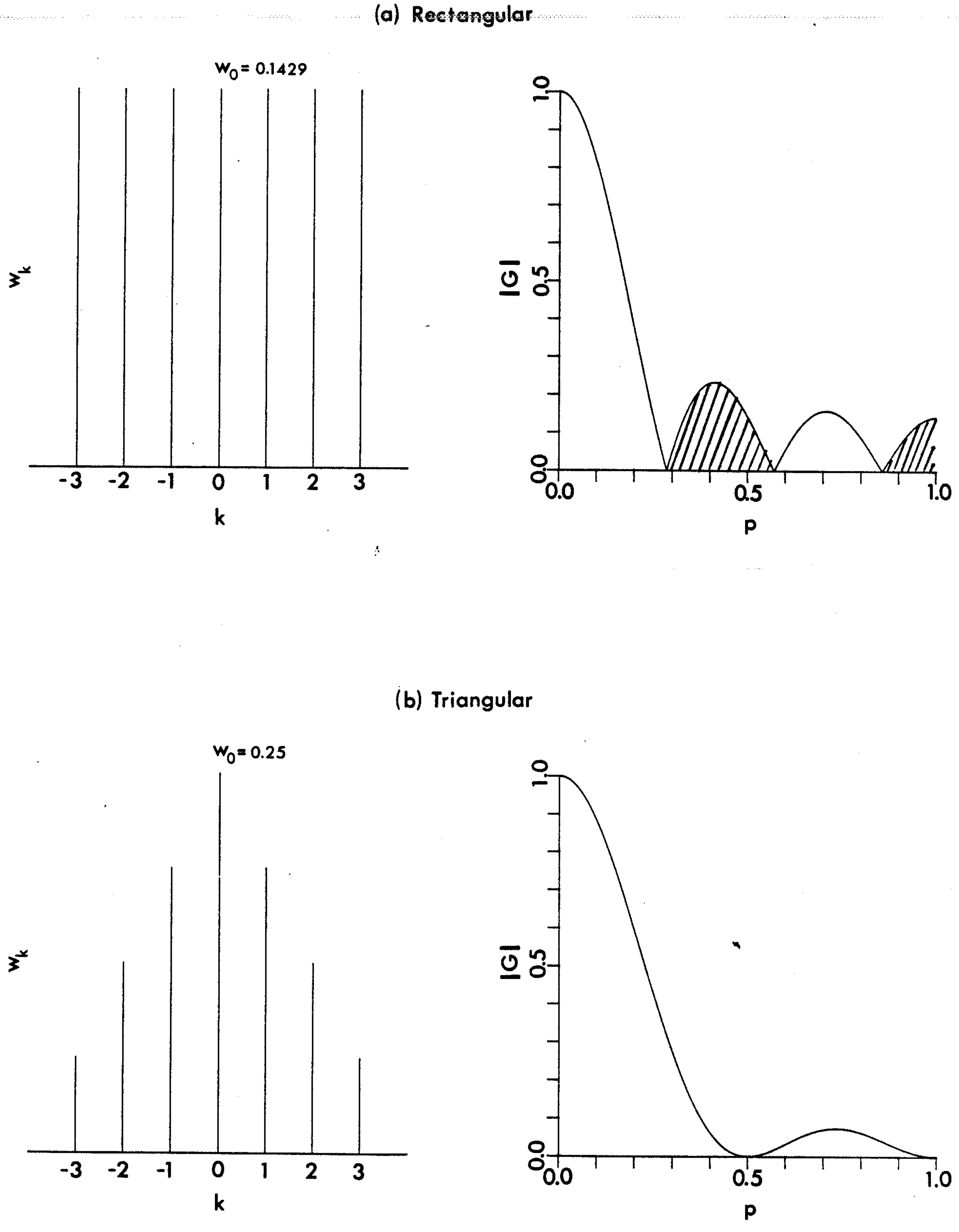
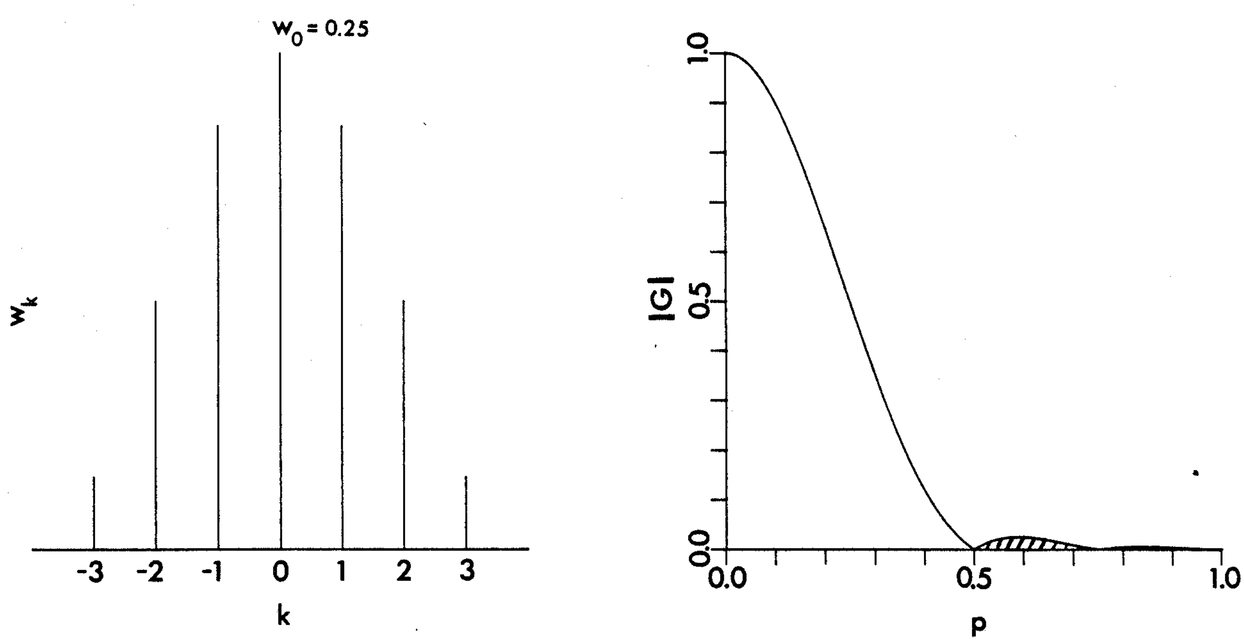

(a) Rectangular
$w_0 = 0.1429$
w_k
-3 -2 -1 0 1 2 3
k
|G|
1.0
0.5
0.0
0.0 0.5 1.0
p
(b) Triangular
$w_0 = 0.25$
w_k
-3 -2 -1 0 1 2 3
k
|G|
1.0
0.5
0.0
0.0 0.5 1.0
p

FIGURE 2

(a) Cosine



(b) Binomial

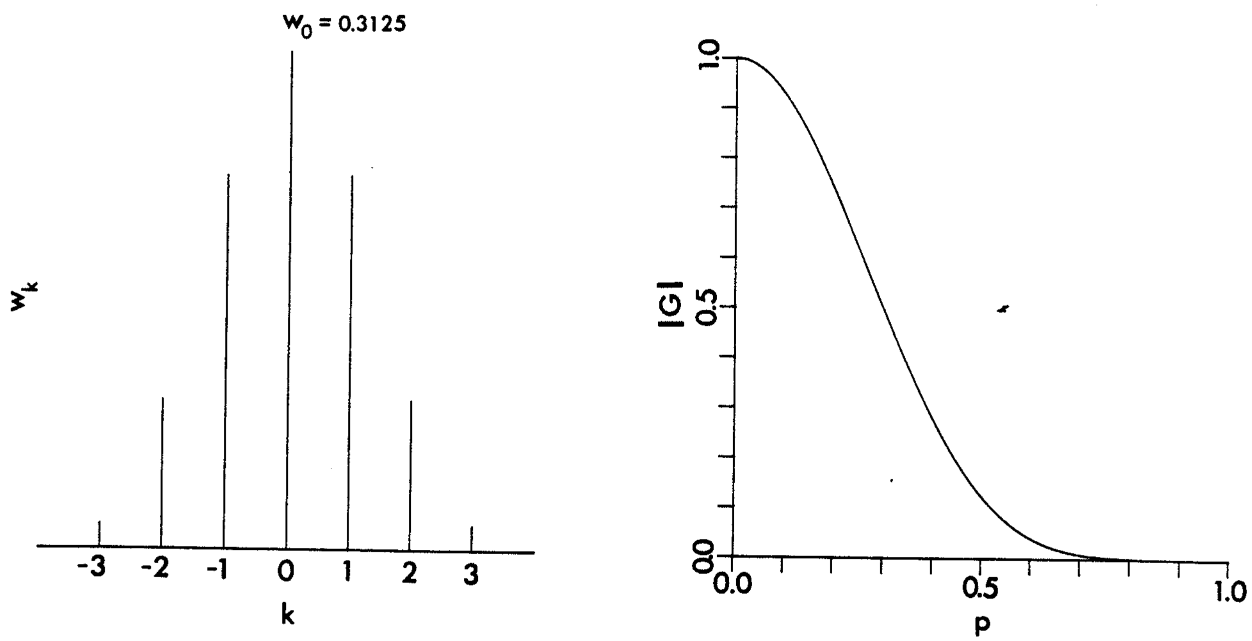


FIGURE 3

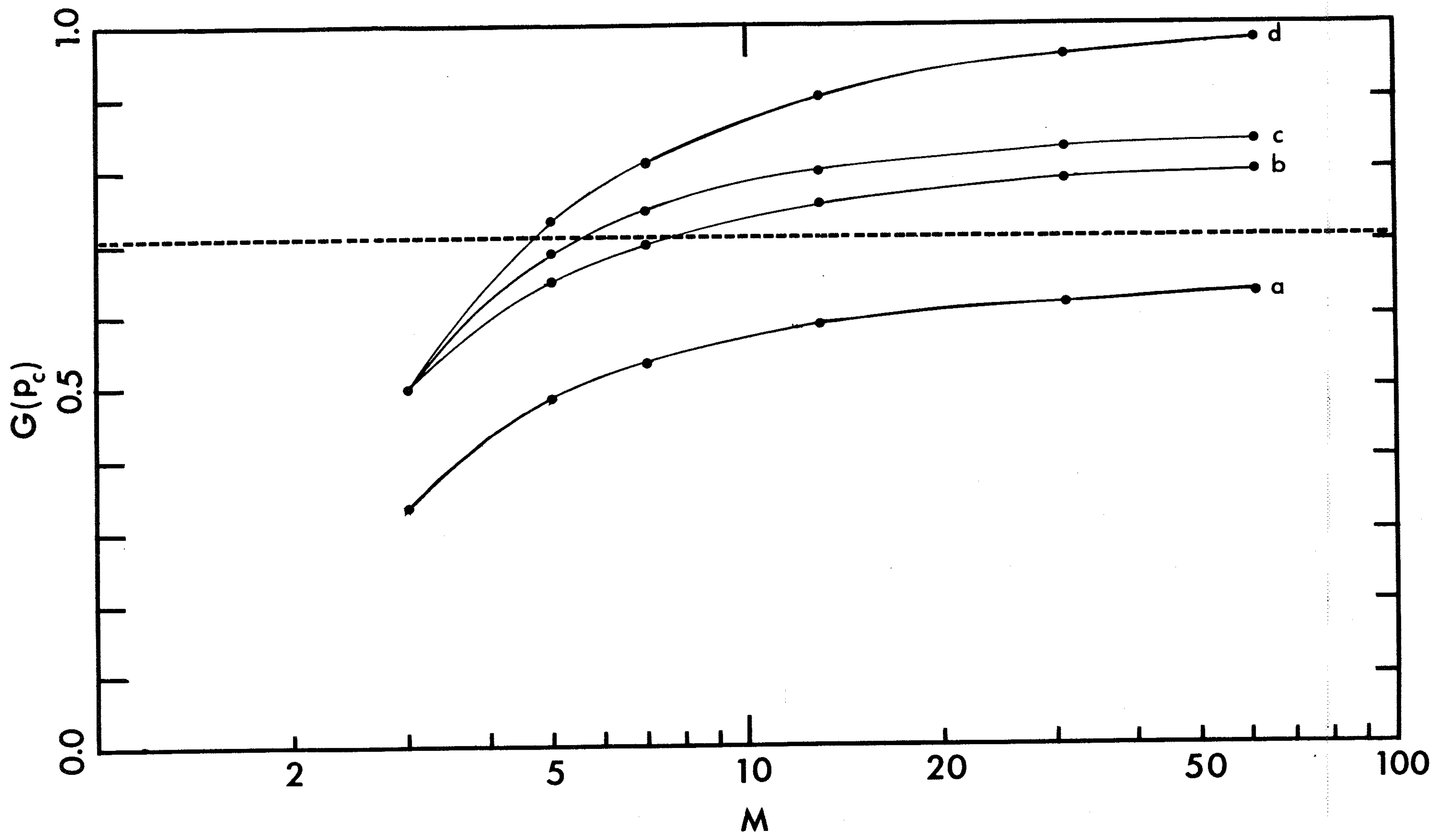


FIGURE 4

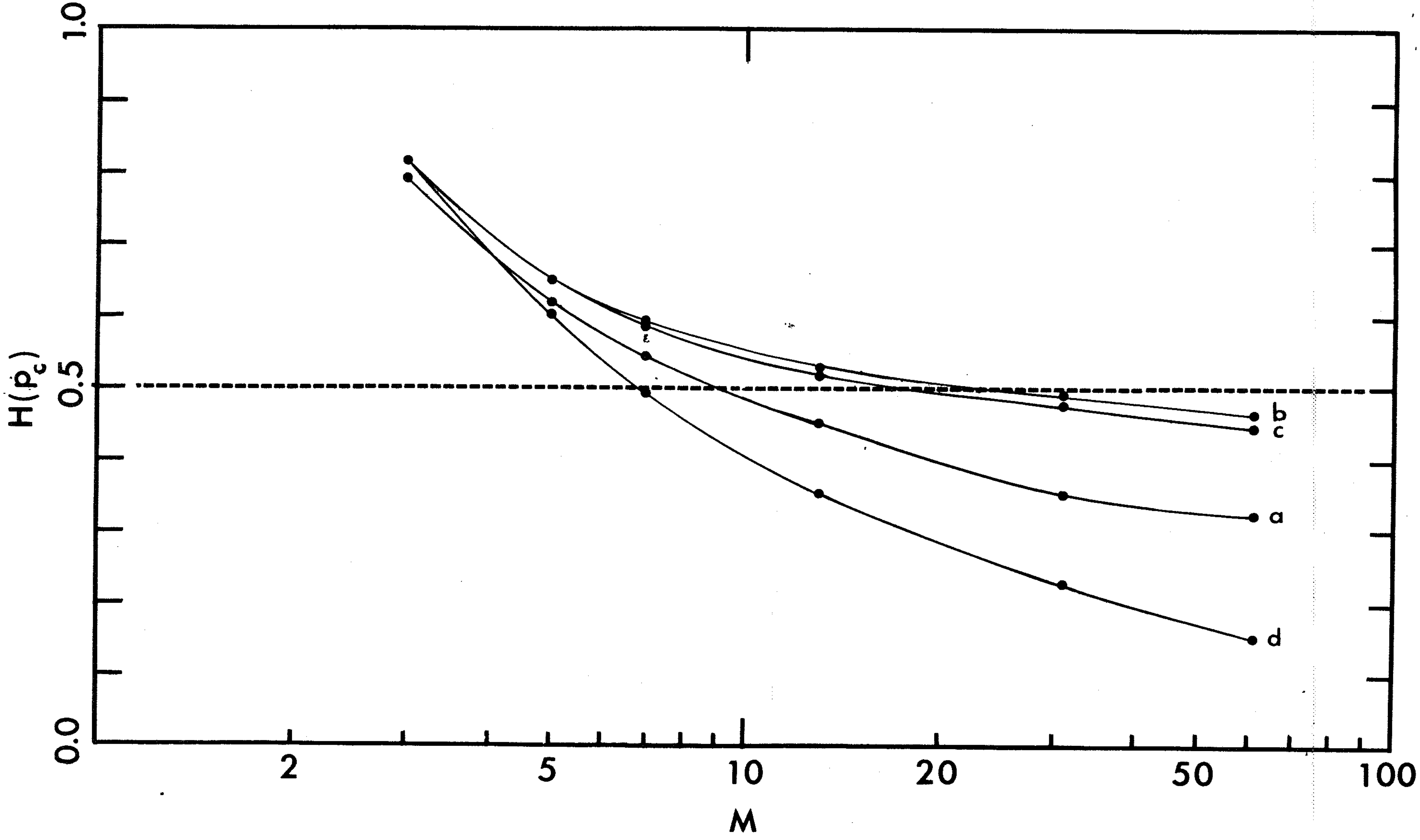


FIGURE 5

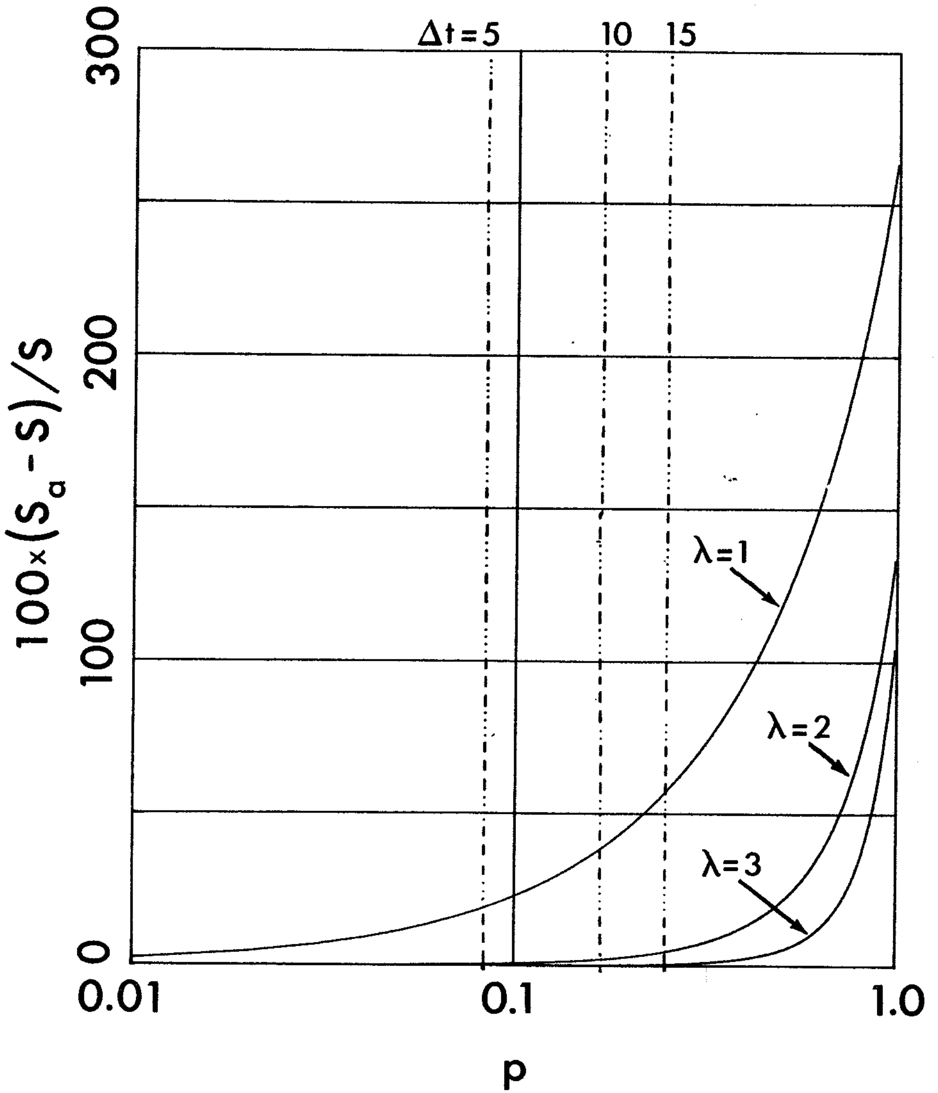


FIGURE 6