K. L. Bocock, [illegible] K. Lindley, Christine A. [illegible], J. K. Adamson and J. A. Webster

R & D 74/54
April 1974

1. Introduction

In an earlier paper (BOCOCK, 1973), one of us described the collection of soil temperature data in a deciduous woodland, Meathop Wood, in the southern Lake District (Merlewood project 301/12). In a separate study (Merlewood project 108), various soil characteristics, including soil temperature, were examined in 48 woodlands scattered over the Lake District and around the northern shores of Morecambe Bay. HOWARD and BENEFIELD (1970) provided an introduction to this study.

Temperature data were collected in different ways in the two projects. In 301/12 they were recorded hourly for over four years using a strip-chart recorder and thermistor probes, whereas, in project 108, a modification of the sucrose inversion technique of BERTHET (1960) was used to determine successive four-weekly mean temperatures over one year on each of the 48 sites.

In both projects, data analysis problems arose, for example, because of the number of data involved and the way in which data were collected, and because some values were missing. In addition, we required a satisfactory method of describing and comparing temperature patterns at different times and soil depths. Harmonic analysis solved many of these problems and we suggest that this method could be used profitably by many more workers if they understood its rudiments and gained some appreciation of its potential. These are the basic reasons for producing this paper, in which we give an outline of the principles of harmonic analysis and describe some of the ways in which we have used the method with temperature data.

2. Basic principles

2.1 Analysis

Harmonic analysis is concerned with the representation of a waveform as a series of mathematical terms and the calculation of estimates for the parameters associated with each term. A waveform is a line representing the change in one variable, such as temperature, with another variable, such as time. It may be a smooth curve such as a sine (Figure 1) or cosine curve, or it may follow the general trend of such a curve, but with corrugations imposed by variation in the basic data used (Figure 2). The analysis is usually applied to data which clearly change or are expected to change periodically, but it may also be used to describe curves which show little evidence of regular oscillations, for example that given in Figure 3.

A sine curve, such as $y = \sin x$ (Figure 1), is a simple example of the type of data which one may wish to examine. The dependent variable, y, could be temperature, but here it is expressed as sine values ranging from the + 1 to -1. The amplitude of the curve, a, is equivalent to the half-range. The independent variable, x, could be time, but here it is expressed in radians. The latter are easily converted to angular degrees, because 2π radians are equivalent to 360° angular. In the example, the period of wavelength of the curve, that is the units of x required to include one complete cycle, from $y = 0$ to $y = 1$ back through $y = 0$ to $y = -1$ and back to $y = 0$, is 2π radians.

Any periodic function, f (x), whose period is 2π, can be expressed in the form

$$f(x) = a_o + c_1 \sin(x + A_1) + c_2 \sin(2x + A_2)$$

$$+ c_3 \sin(3x + A_3) \ldots\ldots\ldots\ldots \qquad (1)$$

$$\ldots\ldots c_n \sin(nx + A_n)$$

a_o is the mean value around which the curve oscillates and is equivalent to y = 0 in Figures 1 and 4. The first term after a_o, $c_1 \sin(x + A_1)$, is called the fundamental or first harmonic, $c_2 \sin(x + A_2)$, the second harmonic and $c_n \sin(nx + A_n)$, the nth harmonic. Fundamental and first harmonic are used here interchangeably for the first term but note that some authors may follow the practice of the physicists and refer to term one as the fundamental and terms two to n as harmonics.

Each term may be expressed graphically as a sine curve, the period of the curve being equivalent to the fundamental period divided by the number of the harmonic. A set of data may thus be modelled by a series of superimposed sine curves whose periods decrease and whose frequency of oscillation increases with the harmonic number. In a plot of the nth harmonic, the abscissa always retains the length of the fundamental as this covers the span of the basic data set, but it is divided into 360^on or $2\pi n$ radians (Figure 4). In the specially selected example illustrated in Figure 4, the observed data (Table 1) are modelled by two superimposed curves. If there are N basic data, a complete model which accounts for 100% of the variability in the data, is constructed by addition of N/2 sine terms or curves.

So far in this paper, we have arbitrarily chosen to use sine terms and curves in the discussion, but equation 1 can also be written in a cosine form

$$f(x) = a_o + s_1 \cos(x + \alpha_1) \ldots\ldots\ldots\ldots s_n \cos(nx + \alpha_n) \qquad (2)$$

or a cosine and sine version

$$f(x) = a_o + c_1 \sin x \cos A_1 + c_1 \cos x \sin A_1 \ldots\ldots\ldots\ldots$$

$$\ldots\ldots c_n \sin nx \cos A_n + c_n \cos nx \sin A_n \qquad (3)$$

or

$$f(x) = a_o + a_1 \cos x + b_1 \sin x \ldots\ldots\ldots\ldots \qquad (4)$$

where $a_n = c_n \sin A_n$

and $b_n = c_n \cos A_n$

Equation 4 can be re-written as

$$y = a_o + \sum_{n=1}^{\infty} (a_n \cos x + b_n \sin x) \qquad (5)$$

which indicates that the dependent variable y may be estimated by adding the mean a_o and the sum of up to an infinite number of composite terms each of which includes the independent variable x and coefficients a_n and b_n. Equation 5 is an expression of the Fourier series on which harmonic analysis is frequently but not invariably based. Further information on this series may be obtained from standard texts such as CALUS and FAIRLEY (1970), DAVIS (1973) and SMITH (1966).

A Fourier series and, in the strictest sense, Fourier analysis use an infinite number of terms to describe an infinite number of closely adjacent data points, that is a continuous curve. Harmonic analysis, a special case of Fourier analysis is concerned with a particular number of data points (N) which are usually equally spaced along, or more strictly around, a particular portion of the hypothetical continuous curve. It is possible to use Fourier-based techniques for analysis of irregularly sampling data, but such data require a more sophisticated approach than that adopted in this paper.

From the previous paragraph, it follows that the basic equation for Fourier-based analysis of temperature data (t) can be written

$$t = t_o + \sum_{n=1}^{N/2} (p_n \cos x \, q \sin x) \quad (6)$$

with a change of notation from equation 5 because we are now dealing with a specific type of data. Note that n is now limited to the theoretical maximum of N/2 terms, N being the number of data points equally spaced in time.

Some authors (CONRAD and POLLAK, 1950; FUGGLE, 1971) use equation 6 in their analysis but, for convenience of interpretation and presentation of the results of the analysis, this equation is more usually expressed solely in cosine terms, as in equation 2 (TURNER, 1971), or, in sine terms as in equation 1 (BALLARD, 1972; BROOKS and CARRUTHERS, 1953; JAGER, 1972; KRISHNAN and KUSHWAHA, 1972; MOTTERSHEAD, 1971; PEARCE and GOLD, 1959; SIEGENTHALER, 1933; VAN WIJK and DE VRIES, 1963; WEST, 1952). We use equation 6 to allow calculation of the coefficients p_n and q_n and the following version as the final model

$$t_i = t_o + \sum_{n=1}^{N/2} a_n \sin (nx + A_n) \quad (7)$$

This indicates that temperature t_i, referring to some particular point in time i, may be estimated from t_o, the mean of N data, plus the sum of up to N/2 terms. Each term may be represented by a sine curve as in Figure 4. Each curve has an amplitude or half-range a_n which is algebraically equivalent to $\sqrt{p_n^2 + q_n^2}$ where p_n and q_n are the coefficients from equation 6.

x, which is sometimes called the time angle (CONRAD and POLLAK, 1950), equals iz where i = 0, 1, 2 .... N-1 and $z = \frac{360^o}{P}$ in angular degrees or $\frac{2}{P}\pi$ in radians. The latter is equivalent to w, the radial frequency, of VAN WIJK and DE VRIES, (1963). P denotes the length of the fundamental period in the time units which are of particular interest. For example, if 52 weekly mean temperatures are being analysed, one complete period or cycle of the fundamental curve extends over 52 weeks so P = 52. In the 27th week, therefore, i = 26 and

$z = \frac{360^o}{52}$ so $x = 26 \left(\frac{360}{52}\right)$ or 180^o angular. P and N are often numerically equal, as in this example, but P may be higher than N if the object of the analysis is to obtain estimates of y for x values lying between the sampling points, as in section 3.3 below. Note that, for all harmonics, P retains the same value, but x changes from 0 to N - 1 $\left(\frac{360}{P}\right)$, and n, the frequency of oscillation of the sine curve increases from one for the fundamental to N/2 for the N/2th curve (Figure 4).

A_n in equation 7, the phase angle, offset angle or angle of lag, is equivalent to arctan (p_n/q_n) ($= 1/\tan (p_n/q_n)$) where p_n and q_n are the coefficients from equation 6. In Figure 1, the sine curve is drawn so that it begins at y = 0, reaches a peak at 90° and a trough at 270° and, because we have based this curve on published sine values, the fundamental curve accounts for virtually all the variability in the data. For the point at the beginning of this curve, equation 7 would read as follows

$$0 = 0 + 1 \sin (0 (\frac{1\ 360}{P})) + A_n)$$

so $\sin (A_n) = 0$ and $A_n = 0^\circ$

Where $y = 0.7071$, $\sin (A_n) = 0.7071$ and $A_n = 45^\circ$. At the peak of the curve, $y = 1$, $\sin (A_n) = 1$ and $A_n = 90^\circ$.

From this example, it is clear that the phase angle indicates the starting point of the curve on the angular scale being used for that curve. It therefore also determines the values of x at which the extremes of y occur, that is where $nx + A_n$ is 90° (0.5π) for the maximum or 270° (1.5π) for the minimum. Other desired values on the curve may be calculated by setting the equation 7 variables appropriately as in section 3.3 below.

In the second example of harmonic analysis (Figure 4) as in the first (Figure 1) the fundamental curve begins at $\bar{y} = 0$ so the phase angle is 0. The angle for the second harmonic has to be 180° to complete the model of the data. Third and subsequent harmonics are of negligible importance and may be ignored. Table 2 gives the complete analysis of the data.

Some computer programs produce phase angle values in a form uncorrected for the quadrant in which the angle lies. This is indicated by the negative sign attached to some angles and by all the angles being less than 90°. Correction should be made as follows by reference to the signs taken by p_n and q_n in equation 6 (BROOKS and CARRUTHERS, 1953; CONRAD and POLLAK, 1950).

p_n +	q_n +	A_n no correction
p_n +	q_n -	$180^\circ - A_n$
p_n +	q_n -	$180^\circ + A_n$
p_n +	q_n +	$360^\circ - A_n$

The above correction is done automatically in the current Merlewood programs (HASCV, HAST and HAFIT).

Equation 7 holds for both odd and even numbers of data but the calculations of the last harmonic vary slightly with type of number. CONRAD and POLLAK (1950) p. 131 give relevant details. The Merlewood programs are constructed to cope with odd and even-numbered data sets representing less than or more than one fundamental cycle or period. Examples of the sort of differences one can expect between the outputs of harmonic analyses of regularly varying data where the data sets contain less than one and up to three periods are given in Table 3.

Note particularly that the mean around which the harmonic curves oscillate varies with the number of data and that, if the data represent x complete periods, the xth term and terms which are multiples of x account for most of the variability. Other terms have a negligible influence and reflect differences between periods, and the corresponding curves are virtually straight lines. Where the data set does not cover exactly one or more periods, then more than one term will frequently be required to account for a high percentage ($> 90\%$) of the variability in the data. Clearly, inspection of the output from harmonic analysis can reveal much about the pattern of the data which are being analysed.

2.2 Synthesis

Harmonic synthesis involves assembling the fundamental and a selected number of harmonic terms and then predicting values from the mathematical model which has been created. If N equally-spaced observations were used in the analysis and if N/2 harmonics are used in the synthesis and N equally-spaced data are predicted, then the difference between the observed and predicted values should be negligible. The total variance for the observed data is given by $\sum (t - t_o)^2/N$ using the notation of equation 7. PANOFSKY and BRIER (1958) and many other authors state that the variance of the fundamental or any harmonic is given by a $a_n^2/2$ except for the last harmonic where the value is a_n^2, a_n being the amplitude of curve n. However, this applies only to even-numbered data sets. For odd-numbered sets the variance is $a_n^2/2$ for all curves. The total of the variances for n harmonics can be expressed as a proportion of the total variance and the synthesis stopped when the proportion reaches some arbitrarily selected limit, or, alternatively, when the harmonic variance itself drops to a selected value. A Merlewood computer programme, HASCV, which calculates the total variance and the percentage of the total variance accounted for by n harmonics, is available (Table 2).

Harmonic analysis and synthesis programs may be tested a) using worked examples from the literature (CARSON 1961; CONRAD and POLLAK 1950) or b) by checking that synthesis with N/2 terms (i) gives predicted values which are virtually the same as the original observed data (ii) accounts for exactly 100% of the total variance. Another valuable test is to use a set of sine values extracted from tables as test data (Figure 1). If values covering only one period are used, then the output for the fundamental should be as follows: mean 0.0, amplitude 1.0, phase angle 0, and much higher than 99.9% of the total variability accounted for, the particular percentage depending on the number of decimal places in the test data.

3. Application of harmonic analysis and synthesis

3.1 Introduction

There appear to be two main types of approach to the use of this method. The first, of which we have had no experience but which is used in meteorology, attempts to answer the questions "Are there any identifiable periodic components in a given waveform?". "How many?" and "What are their statistics?". Relevant information and examples of this approach are given by BROOKS and CARRUTHERS (1953), HARTLEY (1949), HORN and BRYSON (1960), PANOFSKY and BRIER (1958).

The second approach, which we illustrate below, assumes that there are periodic fluctuations over defined periods, for example one day or one year, and is concerned with describing these fluctuations objectively. The mathematical models which are constructed are then used in various ways, for example, as predictive models, or in the comparison of changes [illegible] between two variables with changes in a third variable (3.2) in the estimation of missing data values (3.3) and in the adjustment of data (3.4).

3.2 Description and comparison of patterns of change in variables

This application is illustrated using the results of our examination of weekly mean soil temperatures calculated from hourly data for a three year period and for six soil depths (Project 301/12). Because 365 days is not exactly equal to 52 weeks it was necessary to begin each 'year' on the same date to ensure synchrony of data. The one or two residual days at the end of the year were ignored. A typical data set is shown in Figure 2 together with the fitted fundamental curve and points based on the complete model. From Figures 2, 3 and 12 it is clear that in accordance with theory our observed temperature data are modelled very well indeed if all the harmonics are used. An indication of the gradual build-up of the complete model is obtained from the cumulative percentage variance curves (2.2 above the Figure 5). Note that the shape of these curves varies with soil depth.

The fundamental term normally accounts for more than 90% of the total variability in our temperature data but two or more harmonics are usually required to account for more than about 95% of the variability (compare CARSON 1961; FUGGLE 1971; KRISHNAN and KUSHWAHA, 1972; TURNER 1971). The fundamental curve indicates the main trend in the data (Figure 2) and the harmonic curves indicate various unspecified types of minor variation. As PANOFSKY and BRIER (1958) point out, each harmonic does not necessarily have any physical or biological meaning. Nevertheless, careful inspection and consideration of the harmonics is always advisable because occasionally, as in HORN and BRYSON'S (1958) study of precipitation patterns over North America, they have a very definite significance in relation to the variable being studied. Where strong periodicity does not occur in the observed data or in the residual data after extraction of one or more terms then the remaining terms provide merely a mathematical description of the changes in the data.

Where the fundamental accounts for most of the variability, its statistics, the mean, the amplitude and the phase angle, are frequently used in comparisons, but caution is required in their use (see Section 3.3 below). With our data, we compared the temperature curves for different soil depths and years (Figures 6, 7(a) and 8). A full account of the background to this study will be given elsewhere (BOCOCK and WHITE, in preparation). Here we wish to give only an outline of the analysis to illustrate how this approach can form a useful basis for other analyses.

The study involved using regression analysis to test the linearity of the relationships between the dependent variable (mean, amplitude or angle) and the independent variable, soil depth, for each year and the use of a covariance analysis program to allow comparison of the parallelism and elevation of the curves for the three years. All the nine regressions fitted by the method of least squares, were linear or only slightly curved (quadratic term not significant or only just significant $P = 0.05$) (Figures 6, 7(a) and 8).

The annual mean temperature, unlike the amplitude and phase angle of the fundamental curve does not have to be calculated using harmonic analysis but it is an essential statistic of each harmonic term. The slopes of the first and third mean temperature curves (Figure 6) where not significantly different from 0($P > 0.20$) but the slope of the second curve was significant ($0.001 < P < 0.02$) but only very slightly positive so in general there is little change in annual mean temperature with soil depth. The difference between the three regression coefficients were not significant ($0.05 < P < 0.10$) so the lines must be regarded as parallel. The difference in elevation of the lines was highly significant ($P < 0.001$) so at least two of the lines had a signficantly different elevation. The elevations of the two lines which had the most similar elevation, lines two and three, were only just signficantly different from each other ($0.01 < P < 0.05$) and were highly significantly different ($P < 0.001$) from the elevation of line one.

Amplitudes of the fundamental curves were transformed to their natural logarithms to reduce curvature. The slopes of the three lines (Figure 7(a)) were all negative and significant ($0.001 < P < 0.01$). The difference between the slopes was not significant ($P > 0.20$), but there was a highly significant difference between the elevations of the lines ($P < 0.001$). Further tests indicated that the elevations of lines two and three were not significantly different ($0.05 < P < 0.20$), but both these values were highly significantly different from the elevation of line one ($P < 0.001$).

The third statistic of the fundamental term, the phase angle, decreased significantly ($P < 0.001$) with soil depth (Figure 8) and the slopes in the three years were not significantly different ($P > 0.20$). Assuming linearity, the elevations of lines one and two were not significantly different ($0.05 < P < 0.20$) but were different from the elevation of line three ($P < 0.001$).

Having examined the temperature variable soil/depth/time relationships individually as outlined above it is then possible to compare these relationships. For example, it is worth noting that, for each temperature variable, the regression for one year differs significantly from the other two regressions, but that the odd year is 1966-67 for mean and amplitude (Figures 6 and 7(a)) and 1968-69 for phase angle (Figure 8). Or, again, according to the theory of heat conduction in a semi-infinite homogeneous medium, the lag extreme values should increase, that is the phase angles should decrease, and the logarithm of the temperature amplitude should decrease linearly with soil depth (CARSON, 1961). GLOYNE (1971) states that the annual mean temperature is the same at all soil depths in a homogeneous soil. Clearly, analysis of our temperature data suggests that the Meathop soil is almost, but not quite, entirely homogeneous with depth (Figures 6, 7(a) and 8), but in some aspects, organic matter, root, stone and moisture content, the Meathop soil is known to be very heterogeneous (SATCHELL, in preparation). This anomaly can be at least partly explained by the small number of data in each data set and the relatively high variability which together make demonstration of curvature more difficult.

An alternative approach to the comparison of temperature curve statistics for different years and soil depths is to use harmonic analysis combined with analysis of variance as described in detail by BLISS (1970). If required for comparisons, variances can be calculated for individual amplitudes or phase angles (GUEST, 1961; BLISS, 1970).

As there are defined relationships between mean temperatures, time and soil depth in Meathop Wood (Figures 6-8) it would be possible to construct a model, similar to that used by FLUKER (1958), which would allow prediction of weekly mean temperatures at any soil depth in the range 0-50 cm or at any point in time within the three years studied. Estimation for times or soil depths outside these limits would be a dubious procedure, but might be possible if air temperature was included in the model as a continuously measured and analysed variable. The latter possibility is being examined currently as one way of predicting soil temperature from air temperature and other climatic variables (BOCOCK et al., in preparation).

3.3 Interpolation of values from a harmonic model curve

Interpolation is possible using harmonic models ranging in complexity from the fundamental curve to one built of N/2 harmonics. However, it must be stressed that the predictions from the models apply only to the same sub-divisions of the time-scale as those used for the observed data. For example, below, we have estimated mean temperatures centred on each day of the year from a model based on four-weekly means so these estimates can only be four-weekly means.

We have used interpolation with models incorporating N/2 harmonics in several ways. For the Meathop data, we estimated maximum, minimum and range for each model by creating a computer disk file of predicted weekly means, one centred on each day of the year, and by scanning this for maximum and minimum values. Amplitude (half-range) derived from these extreme values (Figure 7(b)) changes with soil depth and year in a similar way to the amplitudes of the corresponding fundamental curves (Figure 7(a)), but the absolute values of the data and the slopes of the regressions differ slightly with the complexity of the model used. This suggests that in some respects the fundamental may not be a satisfactory model of changes in our data.

This point is clarified by consideration of the mean lag in temperature with soil depth (0 to 50 cm) calculated ~~for~~ from either the phase angle of the fundamental (Figure 8) or the times of extremes in the complete model (Figure 9). These sources give estimated mean lags of respectively about 14 and 3 days.

The former figure appears to be the more realistic of the two judging by published temperature lag figures of 6 days (MOCHLINSKI, 1970) or 8 days (ANONYMOUS, 1968) per 30 cm soil depth, and 12.6 days for 10-50 cm depth (SIEGENTHALER, 1933). An explanation of the difference in our two estimates of lag is obvious if we consider the weekly mean data for one year, say 1967-68. First, the fundamental expresses the general seasonal trend in the data (Figure 2) regardless of the percentage of variability for which it accounts (Figure 5). The maxima, ~~26~~ 23 July for 0 cm (Figure 2) and ~~9~~ 6 August for 50 cm are therefore the trend maxima. Second, the complete model accounts for virtually all the variability in the data, so the lag is virtually estimated from the times of maxima of the observed means, 24 August for 0 cm (Figure 2) and 25 August for 50 cm. The shorter the period on which each of the observed means is based, the lower the lag is likely to be. On the basis of MOCHLINSKI'S (1970) value of two hours per four inches (10 cm) for diurnal lag, we would have expected a lag of only 10 hours if we had used hourly data in our analyses.

In project 108, a different type of data-handling problem arose. Each site was sampled every four weeks but only a quarter of the 48 sites could be sampled every week. As a result, the mean temperatures for the sites visited in week n were not comparable to those for sites visited in weeks n + 1 to week n + 3, nor to data other than temperatures, which were collected as spot readings each time a site was visited. To surmount these difficulties we carried out a harmonic analysis and synthesis of each set of 13 data (Table 4) and then tracked along the model curve a standard number of days from each sampling date and read off the data (Figure 10). In this way, we were able to produce for all sites, sets of temperature data which were strictly comparable with each other and with other data.

It appears to be advisable to check the shape of the model curve between sampling points before treating data as described above because a curve that fits the observed data very well may show some corrugations between the points (Figure 11) and these may reduce the accuracy of interpolated values. It is anticipated that this phenomenon will be of importance only with very small data sets where the number of harmonics in the model is insufficient to produce a well-smoothed curve. The corrugations are expected to be greater if the data are spot readings rather than means derived from integrated temperature such as those used in the current project, 108.

3.4 Calculation of missing value estimates using iterative harmonic analysis

Harmonic analysis has proved to be useful in the calculation of missing value estimates of mean temperature in project 108 where each complete data set contained only 13 four-weekly means (Figure 12). The procedure, which may be carried out using a widely applicable program HAFIT (Table 5), involves the following steps in the first series of iterations:-

a) Insertion in the computer data file of a missing value estimate derived by linear interpolation from adjacent values. Where the missing value occurs at the beginning or end of a data set, an observed data from the other end of the set is used as one of the "adjacent" values.

b) A pre-selected value is added to or subtracted from the linear interpolate. The value can be determined only by experience with a particular type of data. For the data sets from project 108, which all produced approximately the same shape of curve (Figure 12) but which had amplitudes ranging from 4° to 10°C, an initial value of about -1.0°C was suitable for the first series of iterations (Table 5).

c) Harmonic analysis is carried out on the revised data set and estimates of mean temperatures corresponding to the observed values including the initial missing value estimate are calculated together with sums of squares of differences between the observed and predicted values.

d) The missing values estimate is incremented by a small amount and c) is repeated on the revised data set.

e) The sums of squares of differences calculated in d) is compared with that calulated previously in c). If the second sum is less than the first, the missing data estimate is printed out otherwise the program prints INCREASE (Table 5).

f) Incrementing, harmonic analysis and synthesis continue until the program is stopped manually.

A second series of iterations with a much higher base setting and smaller steps (Table 5) follows the first. The data file is automatically edited to include the data estimate associated [illegible]
of squares, and run as a normal data set with harmonic analysis programs

It is advisable to allow the program HAFIT to run for a few cycles after the first 'minimum' sums of squares has been reached because sometimes the sum increases again slightly before dropping to a second and final minimum value (Table 5).

If there are two missing data, one (A), it does not matter which, is set to a linear interpolated value, then a)-f) (above) should be followed, B being treated as the missing value. A is then treated as the missing value and a)-f) is followed using the revised data set containing the first missing data estimate for B. A is again treated as the missing value, then B, then A and so on until estimates for A and B remain unchanged during two cycles of analysis/synthesis.

Calculation of missing data estimates in this way is surprisingly rapid. Single estimates can be obtained in a few minutes with sets containing only 13 data if a suitable initial setting is chosen for the missing value. In one test, double estimates were produced for one set of 13 data in about 15 minutes, but this procedure usually takes rather longer.

4. Acknowledgements

We are particularly grateful to the following present and past colleagues who assisted in the collection and analysis of the data used as illustrative material in this paper: Mr. J. Heath, Mrs. J. Parrington and Mr. T. Booth (Project 301/12), Mr. A. D. Bailey and Mr. C. Benefield (Project 108).

References

Anonymous (1968). Averages of earth temperature at depths of 30 cm and 122 cm for the United Kingdom. London: Meteorological Office. 232 pp.

Berthet, P. (1960). La mesure ecologique de la temperature par determination de la vitesse d'inversion du saccharose. Vegetatio, 9, 197-207.

Ballard, T. M. (1972). Subalpine soil temperature regimes in south-western British Columbia. Arctic Alpine Res., 4, 139-146.

Bliss, C. I. (1970). Statistics in biology: statistical methods for research in the natural sciences. New York-London: McGraw-Hill. 2, 219-287.

Bocock, K. L. (1973). The collection of soil temperature data in intensive studies of ecosystems. Merlewood Research and Development Paper No. 47, 42 pp.

Bocock, K. L. and White, E. J. (In preparation). Soil moisture tension and soil temperature. In Satchell, J. E. (Editor). The ecology of Meathop Woods (Provisional title). New York-Heidelberg-Berlin: Springer.

Bocock, K. L., Lindley, D. K., Gill, Christine A. and Adamson, J. K. (In preparation). A comparison of methods of estimating woodland soil temperature from air temperature and other climatic variables.

Brooks, C. E. P. and Carruthers, N. (1953). Handbook of statistical methods in climatology. London: H.M.S.O. 330-345.

Calus, I. M. and Fairley, J. A. (1970). Fourier series and partial differential equations. A programmed course for students of science and technology. Chichester: John Wiley-Interscience. 112 pp.

Carson, J. E. (1961). Soil temperature and weather conditions. Argonne, Illinois: Argonne National Laboratory. ANL-6470. 244 pp.

Carson, J. E. (1963). Analysis of soil and air temperatures by Fourier techniques. J. geophys. Res., 68, 2217-2232.

Conrad, V. and Pollak, L. W. (1950). Methods in climatology. Cambridge, Massachusetts: Harvard University Press. 119-133.

Davis, J. C. (1973). Statistics and data analysis in geology. New York-London-Sydney-Toronto: John Wiley. 256-272 and 358-372.

Fluker, B. J. (1958). Soil temperatures. Soil Sci., 86, 35-46.

Fuggle, R. F. (1971). A note on the harmonic analysis of the diurnal temperature waves at three locations on a mountain slope. S. Afr. geog. J., 53, 93-95.

Gloyne, R. W. (1971). A note on the average annual mean of daily earth temperatur in the United Kingdom. Met. Mag., Lond., 100, 1-6.

Guest, P. G. (1961). Numerical methods of curve fitting. Cambridge: University Press. 341-349.

Hartley, H. O. (1949). Tests of significance in harmonic analysis. Biometrika, 36, 194-201.

Horn, L. and Bryson, R. A. (1960). A harmonic analysis of the annual march of precipitation over the U.S. Ann. Ass. Am. Geogr., 50, 157-171.

Howard, P. J. A. and Benefield, C. B. (1970). Selection of characters for the classification of soil humus types. Merlewood Research and Development Paper No. 21, 9 pp.

Jager, T. (1972). The temperature regime in some soils near Tire, Turkey. Versl. landbouwk. Onderz. Ned., 785 180-194.

Krishnan, A. and Kushwaha, R. S. (1972). Analysis of soil temperature in the arid zone of India by Fourier techniques. Agric. Met., 10, 55-64.

Mochlinski, K. (1970). Soil temperatures in the United Kingdom. Weather, Lond., 25, 192-200.

Mottershead, B. E. (1971). Soil temperature at various depths under pasture and bare soil at Armidale, New South Wales. Fld. Stn. Rec. Div. Pl. Ind. CSIRO (Aust.), 10, 5-12.

Panofsky, H. A. and Brier, G. W. (1958). Some applications of statistics to meteorology. University Park, Pennsylvania: Pennsylvania University Press. 254 pp.

Pearce, D. C. and Gold, L. W. (1959). Observations of group temperature and heat flow at Ottawa, Canada. J. geophys. Res., 64, 1293-1298.

Satchell, J. E. (Editor) (In preparation). The ecology of Meathop Woods (Provisional title). New York-Heidelberg-Berlin: Springer.

Siegenthaler, J. (1933). Bodentemperaturen in Abhängigkeit von äusseren meteorologischen Faktoren. Beitr. Geophys., 40, 305-322.

Smith, C. A. B. (1966). Biomathematics. London: Charles Griffin. 1, 420-426.

Turner, J. C. (1971). The harmonic analysis of soil temperature data obtained beneath desert vegetation in central Australia. Met. Res. Bull., 3, 169-183.

Van Wijk, W. R. and De Vries, D. A. (1963). Periodic temperature variations in a homogeneous soil. In Van Wijk, W. R. (Ed.) Physics of plant environment. Amsterdam: North Holland. 102-143.

West, E. S. (1952). A study of the annual soil temperature wave. Aust. J. scient. Res., Series A, Physical Sciences. 5, 303-314.

Table 1. Data analysed and corresponding fundamental and second harmonic values used in construction of Fig. 4.

Basic data	Fundamental values	Second harmonic values
0	0	0
0.0206	0.3420	-0.3214
0.1504	0.6428	-0.4924
0.4330	0.8660	-0.4330
0.8138	0.9848	-0.1710
1.1558	0.9848	0.1710
1.2990	0.8660	0.4330
1.1352	0.6428	0.4924
0.6634	0.3420	0.3214
0	0	0
-0.6634	-0.3420	-0.3214
-1.1352	-0.6428	-0.4924
-1.2990	-0.8660	-0.4330
-1.1558	-0.9848	-0.1710
-0.8138	-0.9848	-0.1710
-0.4330	-0.8660	0.4330
-0.1504	-0.6428	0.4924
-0.0206	-0.3420	0.3214

Table 2. Analysis of the basic data given in Table 1 using Merlewood computer program HASCV.

```
NO. OF SAMPLE UNITS ? 18
NO. OF SINE CURVES ? 9
1 FOR PRINTOUT OF FINAL PREDICTIONS OR 0 ? 0

MEAN = 2.716357E-9 THEN FOR EACH SINE CURVE
NUMBER      AMPLITUDE      ANGLE      CORRECTED ANGLE
VARIANCE OF DEVIATIONS     VARIANCE OF AN OBSERVATION
TOTAL VARIANCE    CUMULATIVE VARIANCE    % VARIANCE
AMP. S. ERROR    ANGLE S. ERROR    S. OF S. OF DEVIATIONS

 1                 .9999873         1.492641E-6     1.492641E-6
                   .1489389         .749981
                   .6249841         .4999873        80.00001
                   .2886715         16.53987        2.234083

 2                -.4999937         4.171093E-6     180
                   4.577383E-5      .8653626
                   .6249841         .6249842        100
                   .3100829         35.53333        5.950598E-4

 3                 4.075874E-8     -76.11811        283.8819
                   5.409634E-5      1.022701
                   .6249841         .6249842        100
                   .3370957         4.738654E+8     5.950598E-4

 4                -9.893856E-6     -.1891228        179.8109
                   6.611399E-5      1.249968
                   .6249841         .6249842        100
                   .3726733         2.158168E+6     5.950260E-4

 5                -3.122760E-6     -.2744725        179.7255
                   8.500305E-5      1.607102
                   .6249841         .6249842        100
                   .4225718         7.753262E+6     5.950213E-4

 6                 4.262186E-8     -88.53922        271.4608
                   1.190042E-4      2.249943
                   .6249841         .6249842        100
                   .4999937         6.721323E+8     5.950211E-4

 7                -1.976756E-5      .07955779       180.0796
                   1.983359E-4      3.749905
                   .6249841         .6249842        100
                   .645489          1.870934E+6     5.950076E-4

 8                -1.545008E-6      .9124067        180.9124
                   5.950071E-4      11.24971
                   .6249841         .6249842        100
                   1.11802          4.146116E+7     5.950071E-4

 9                -2.342751E-8      .6643442        180.6643
                   5.950071E-4      11.24971
                   .6249841         .6249842        100
                   [illegible]      2.734299E+9     5.950071E-4
```

Table 3. Variation in the output of harmonic analysis with the number of periods analysed. The basic data set comprised 13 four-weekly mean soil temperatures collected over one year (Merlewood project 108).

Data examined	Mean	Amplitude	Corrected phase angle	Cumulative % variation accounted for
		Statistics of the first three or nine harmonics		
First 10 of 13	9.78	7.61	16.22	82.60
		1.88	19.08	87.63
		2.81	39.29	98.91
13 basic data	9.68	7.27	52.33	94.84
		0.37	7.80	95.09
		0.19	198.13	95.16
19, basic 13 + first 6 repeated	11.05	5.52	122.08	56.44
		3.94	329.39	85.14
		1.66	352.25	90.27
26, basic 13 + all 13 repeated once	9.68	0	81.00	0
		7.27	52.33	94.84
		0	323.45	94.84
39, basic 13 + all 13 repeated twice	9.68	0	78.31	0
		0	67.00	0
		7.27	52.33	94.84
		0	292.29	94.84
		0	339.60	94.84
		0.37	7.80	95.09
		0	252.16	95.09
		0	341.00	95.09
		0.19	198.13	95.16

Table 4. Output of the computer program HAST after harmonic analysis of 13 four-weekly mean soil temperatures from one of 48 Lake District woodland sites.

```
NO. OF VALUES (INCLUDING MISSING) ? 13
NO. OF MISSING VALUE ? 3
NO. OF SINE CURVES ? 6
LINEAR INTERPOLATE = 14.6041
BASE ADJUSTMENT & STEP SIZE ? -1.0,0.2
```

	ESTIMATE	S. OF S.	MEAN
1	13.6041	.01488866	9.625229
2	13.8041	.0148316	9.640614
3	14.0041	.01479302	9.655998
4	14.2041	.0147725	9.671383
5	14.4041	.01476978	9.686768
6		INCREASE	
7		INCREASE	
8		INCREASE	
9		INCREASE	
10		INCREASE	

```
NO. OF VALUES (INCLUDING MISSING) ? 13
NO. OF MISSING VALUE ? 3
NO. OF SINE CURVES ? 6
LINEAR INTERPOLATE = 14.6041
BASE ADJUSTMENT & STEP SIZE ? -0.4,0.01
```

	ESTIMATE	S. OF S.	MEAN
1	14.2041	.01477251	9.671383
2	14.2141	.01477203	9.67215 2
3	14.2241	.01477147	9.672921
4	14.2341	.01477115	9.673691
5	14.2441	.01477065	9.67446
6	14.2541	.0147704	9.675229
7	14.2641	.01476982	9.675998
8	14.2741	.01476952	9.676768
9	14.2841	.0147693	9.677537
10	14.2941	.01476915	9.678306
11	14.3041	.01476893	9.679075
12	14.3141	.01476882	9.679844
13	14.3241	.01476876	9.680614
14		INCREASE	
15	14.3441	.01476875	9.682152
16		INCREASE	
17		INCREASE	
18		INCREASE	
19		INCREASE	
20		INCREASE	

Table 5. Outputs of computer program HAFIT. This program provides an estimate for a missing data using iterative harmonic analysis. Note that the first series of iterations has a much lower base setting and a larger step than the second series. Harmonic analysis of the completed data set is given in Table 4.

```
NO. OF VALUES (INCLUDING MISSING) ? 13
NO. OF MISSING VALUE ? 3
NO. OF SINE CURVES ? 6
LINEAR INTERPOLATE = 14.6041
BASE ADJUSTMENT & STEP SIZE ? -1.0,0.2
```

	ESTIMATE	S. OF S.	MEAN
1	13.6041	.01488866	9.625229
2	13.8041	.0148316	9.640614
3	14.0041	.01479302	9.655998
4	14.2041	.0147725	9.671383
5	14.4041	.01476978	9.686768
6		INCREASE	
7		INCREASE	
8		INCREASE	
9		INCREASE	
10		INCREASE	

```
NO. OF VALUES (INCLUDING MISSING) ? 13
NO. OF MISSING VALUE ? 3
NO. OF SINE CURVES ? 6
LINEAR INTERPOLATE = 14.6041
BASE ADJUSTMENT & STEP SIZE ? -0.4,0.01
```

	ESTIMATE	S. OF S.	MEAN
1	14.2041	.01477251	9.671383
2	14.2141	.01477203	9.672152
3	14.2241	.01477147	9.672921
4	14.2341	.01477115	9.673691
5	14.2441	.01477065	9.67446
6	14.2541	.0147704	9.675229
7	14.2641	.01476982	9.675998
8	14.2741	.01476952	9.676768
9	14.2841	.0147693	9.677537
10	14.2941	.01476915	9.678306
11	14.3041	.01476893	9.679075
12	14.3141	.01476882	9.679844
13	14.3241	.01476876	9.680614
14		INCREASE	
15	14.3441	.01476875	9.682152
16		INCREASE	
17		INCREASE	
18		INCREASE	
19		INCREASE	
20		INCREASE	

Figure 1. A sine curve which is the basis of the analyses described in this paper. The data used in the construction of the curve are those given in Table 1 for the fundamental.

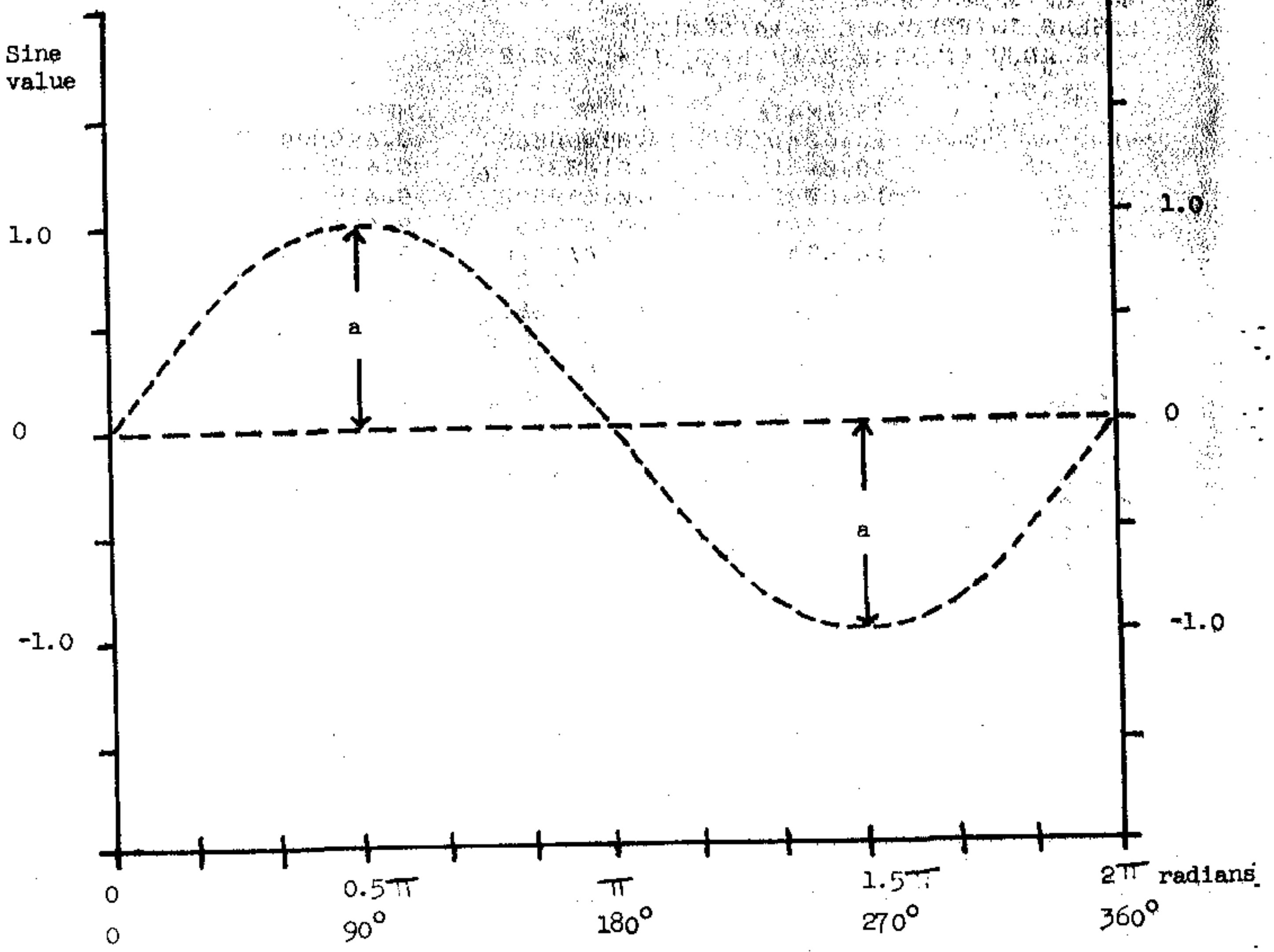


Figure 2. Weekly mean temperature at the soil surface in Meathop Wood in 1967-68, X observed data, O values predicted from a model incorporating 26 harmonics. The smooth curve running through the data represents the fundamental curve. - - indicates the observed data maximum and ---- the fundamental maximum.

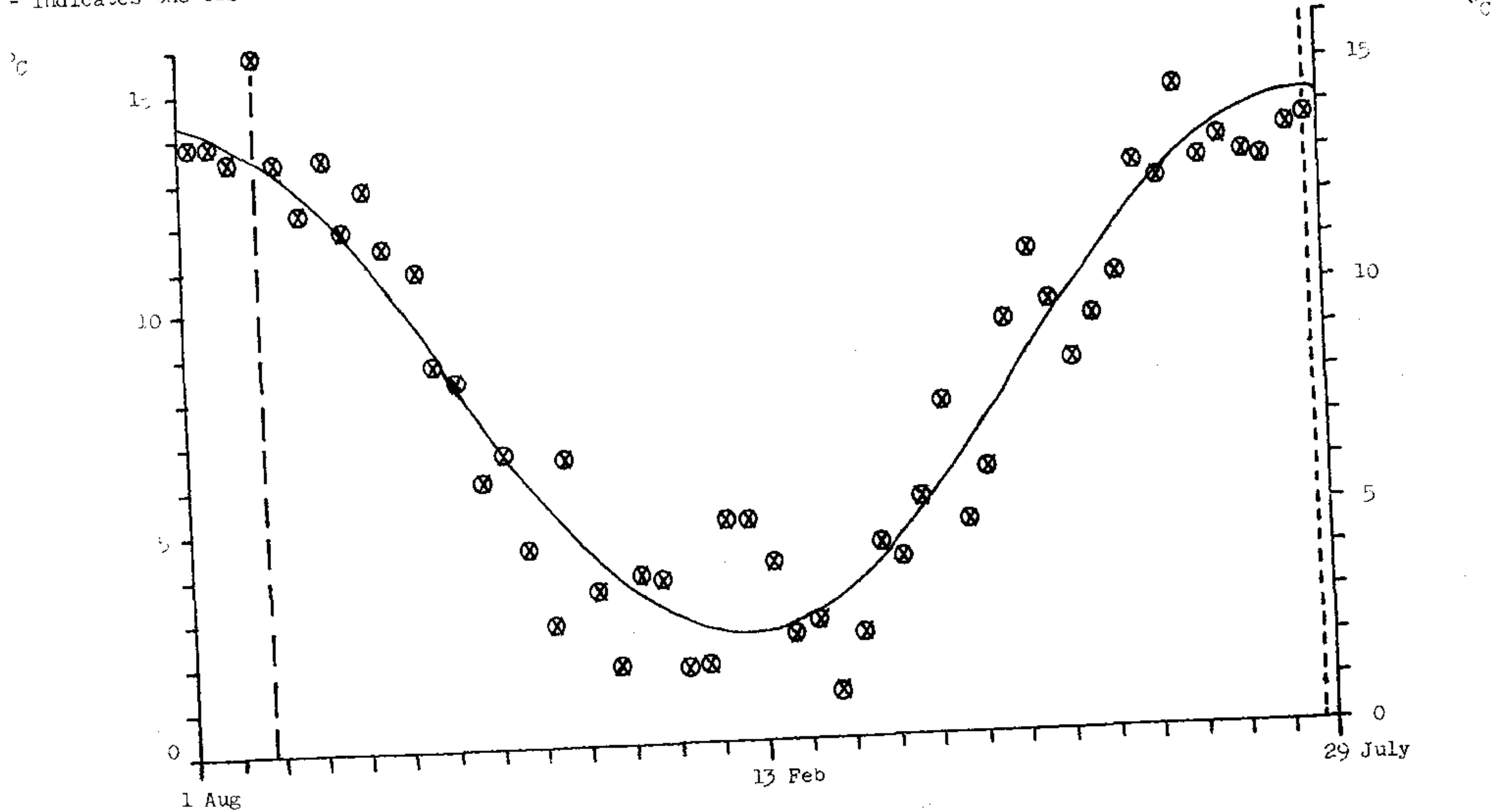


Figure 3. Daily mean temperature at the soil surface in Meathop Wood from 27 February to 26 March 1967 inclusive, X observed values, O values predicted from a model incorporating 14 harmonics.

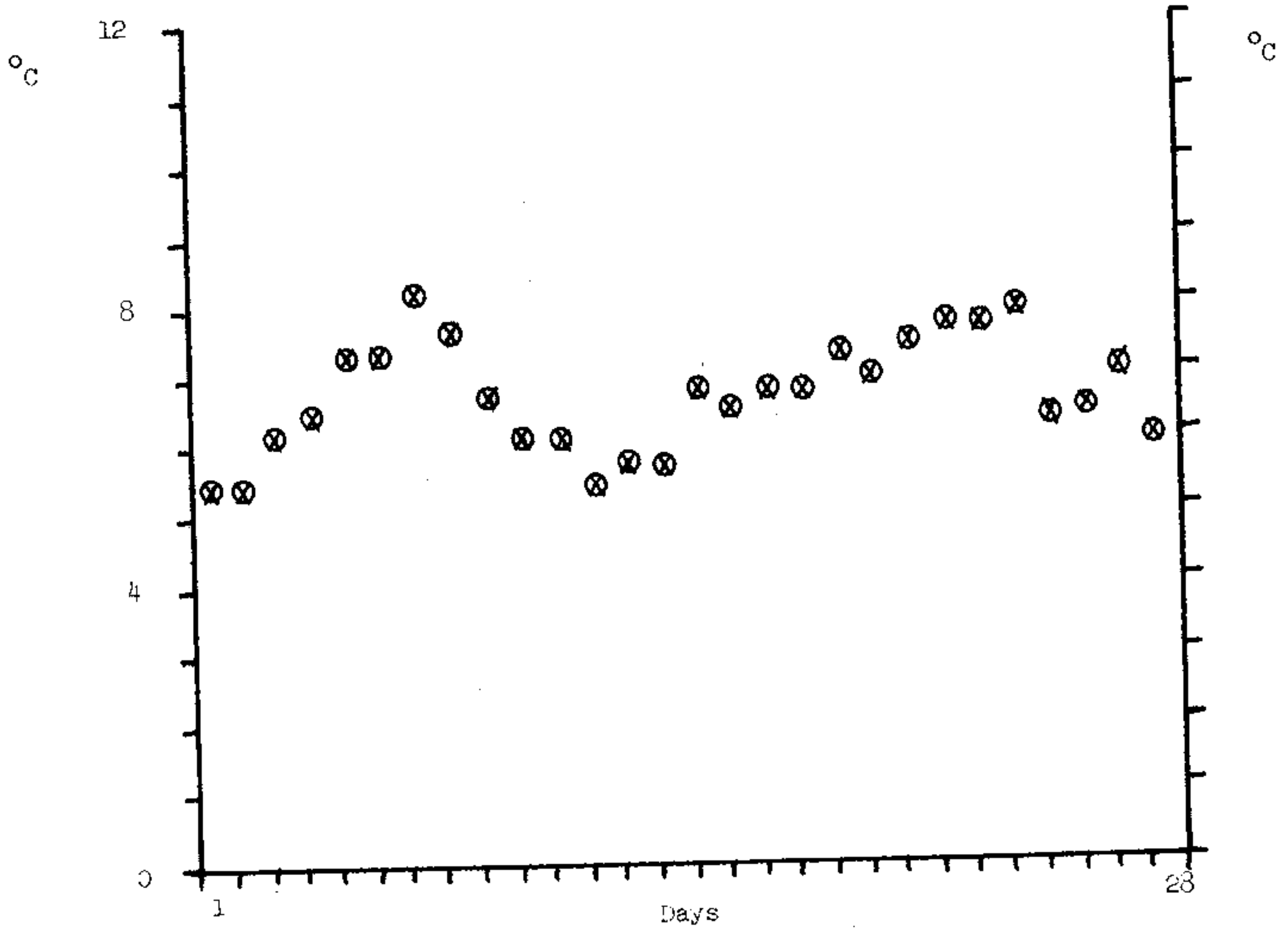
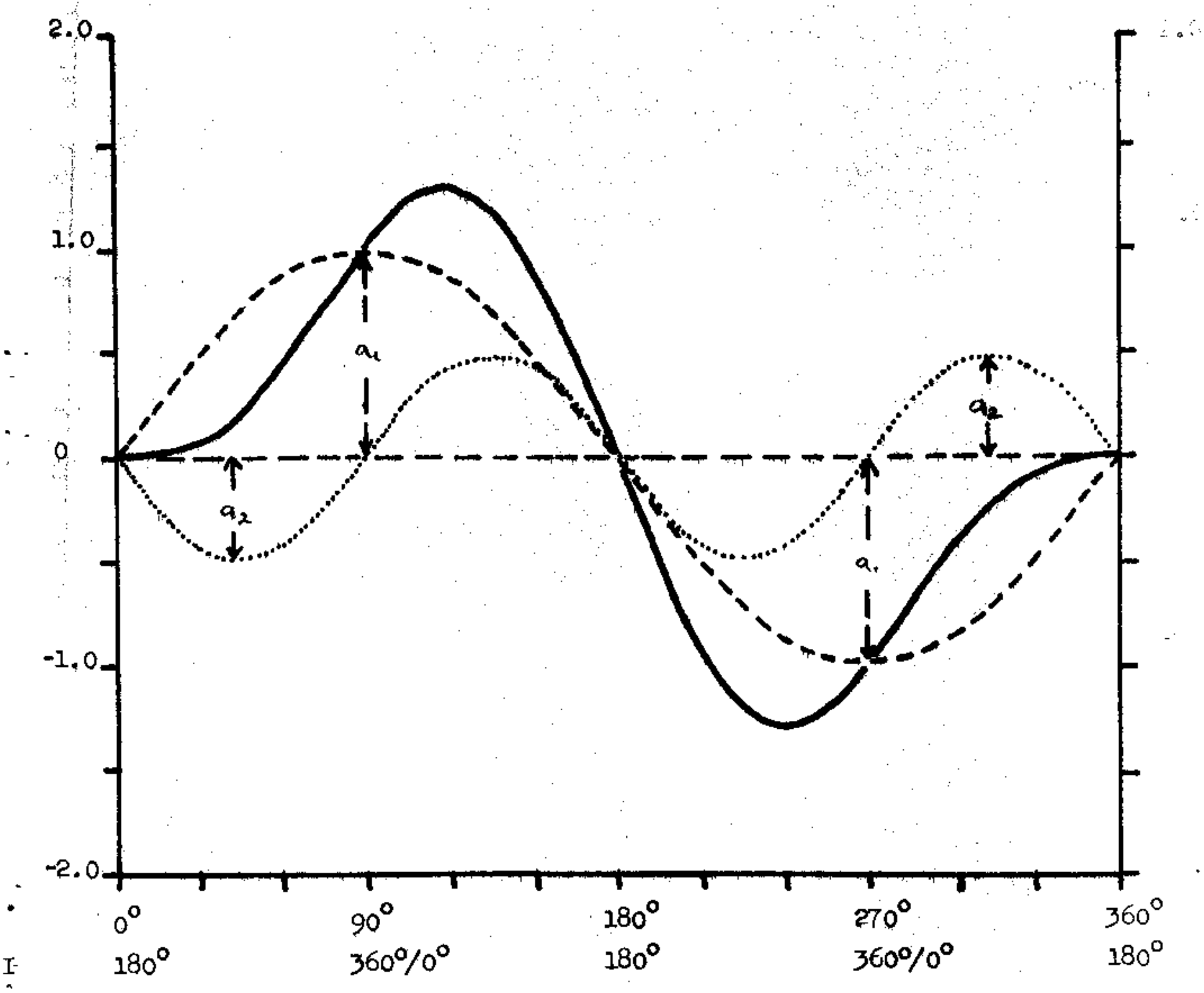


Figure 4. Harmonic analysis of a wave form (——) into two components, a) a fundamental curve (---) with amplitude a_1 and a frequency of one and b) a second harmonic (....) with amplitude a_2 and a frequency of two. See Table 1 for the basic data and Table 2 for a computer print-out of the analysis.



Phase angle of the fundamental (I) and of the second harmonic (II).

Figure 5. Cumulative percentage variance accounted for by successive harmonics derived by harmonic analysis of weekly mean temperatures for 0 cm (X) and 50 cm (O) soil depth in Meathop Wood in 1967-68 (1 August-29 July).

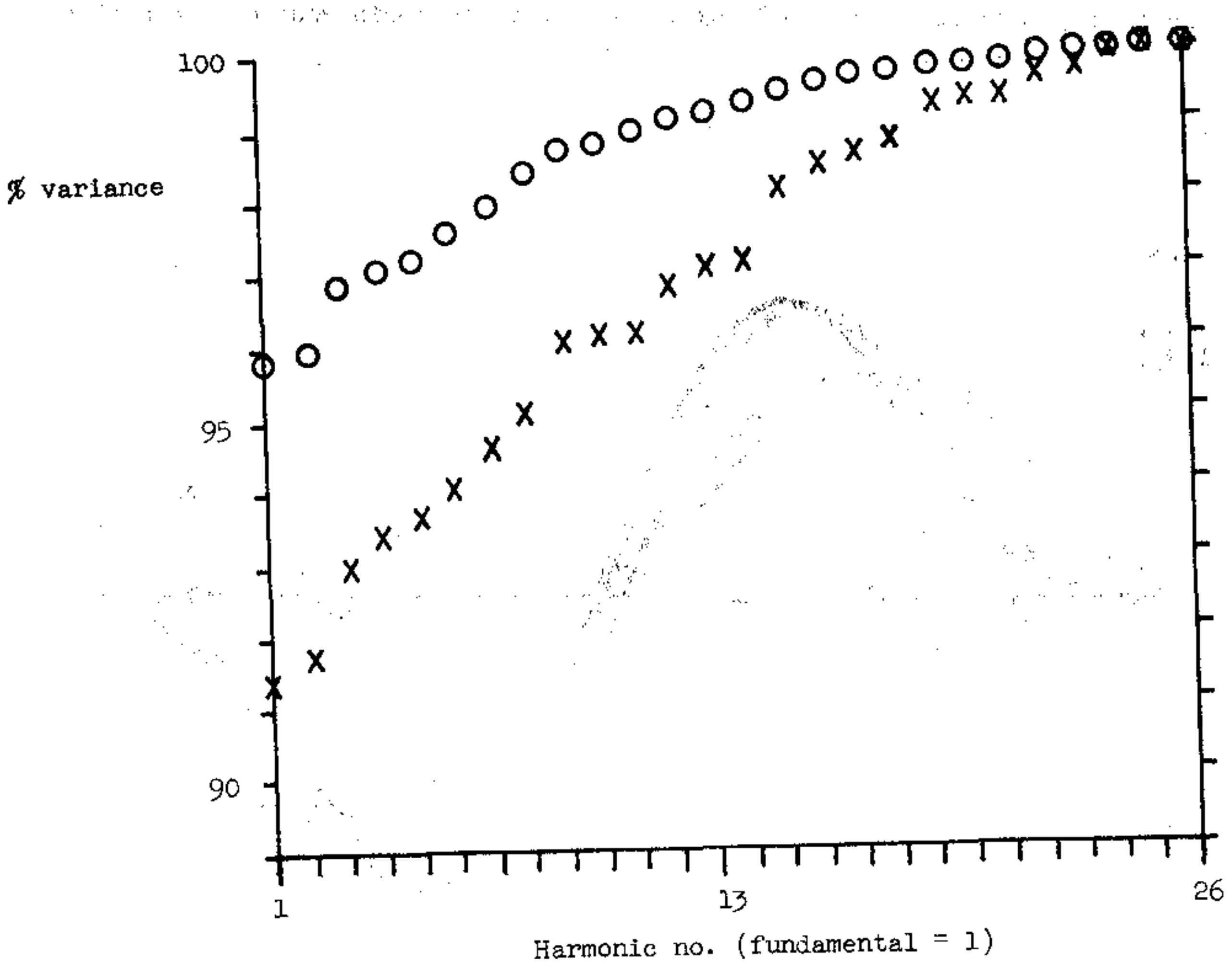


Figure 6. Change in annual mean temperature with soil depth in Meathop Wood, X 1966-67, ⊙ 1967-68, ● 1968-69.

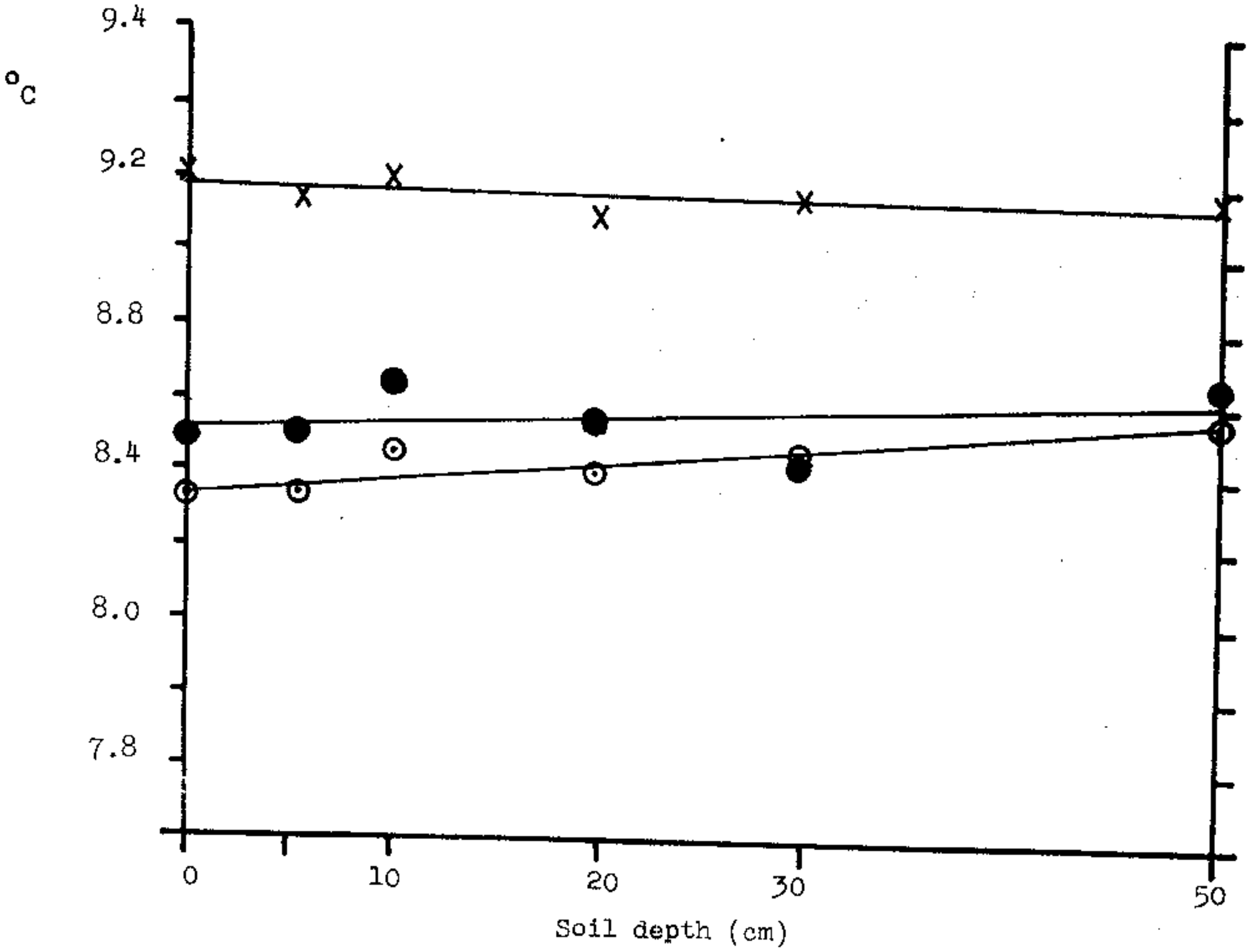


Figure 7. Change in the natural logarithm of amplitude of the weekly mean temperature curves with soil depth in Meathop Wood, (a) for the fundamental sine curve, (b) for the complete model of 26 harmonics. X 1966-67, ⊙ 1967-68, ● 1968-69. ——— line of best fit, ---- linear regression where best line is curvilinear.

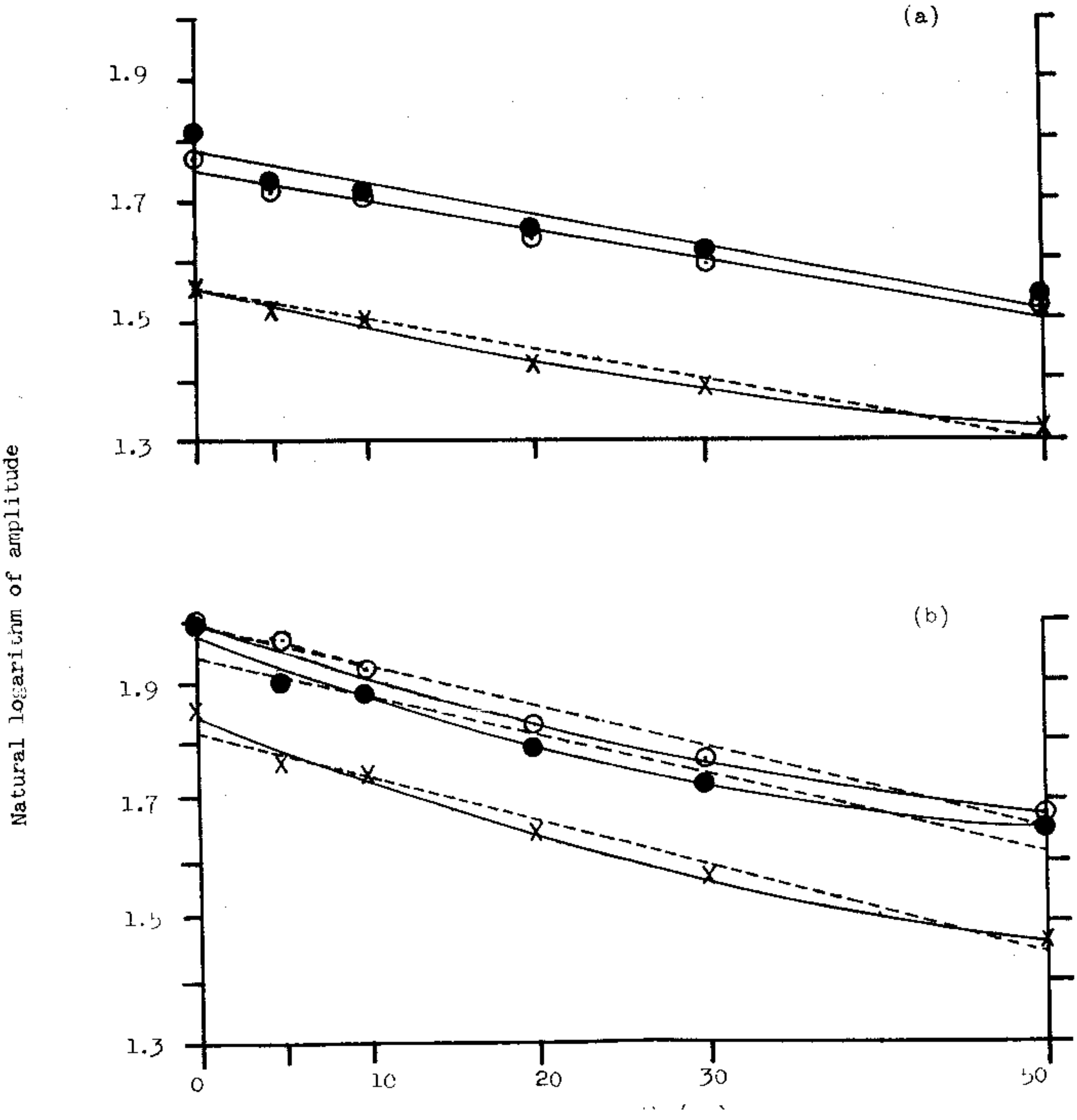


Figure 8. Change in corrected phase angle of the fundamental weekly mean temperature curve with soil depth in Meathop Wood. X 1966-67, ⊙ 1967-68, ● 1968-69. —— line of best fit, ---- linear regression line where best line is curvilinear. Note that 1^o is equivalent to 1.0138^o days (normal year) or 1.016^o days (leap year)

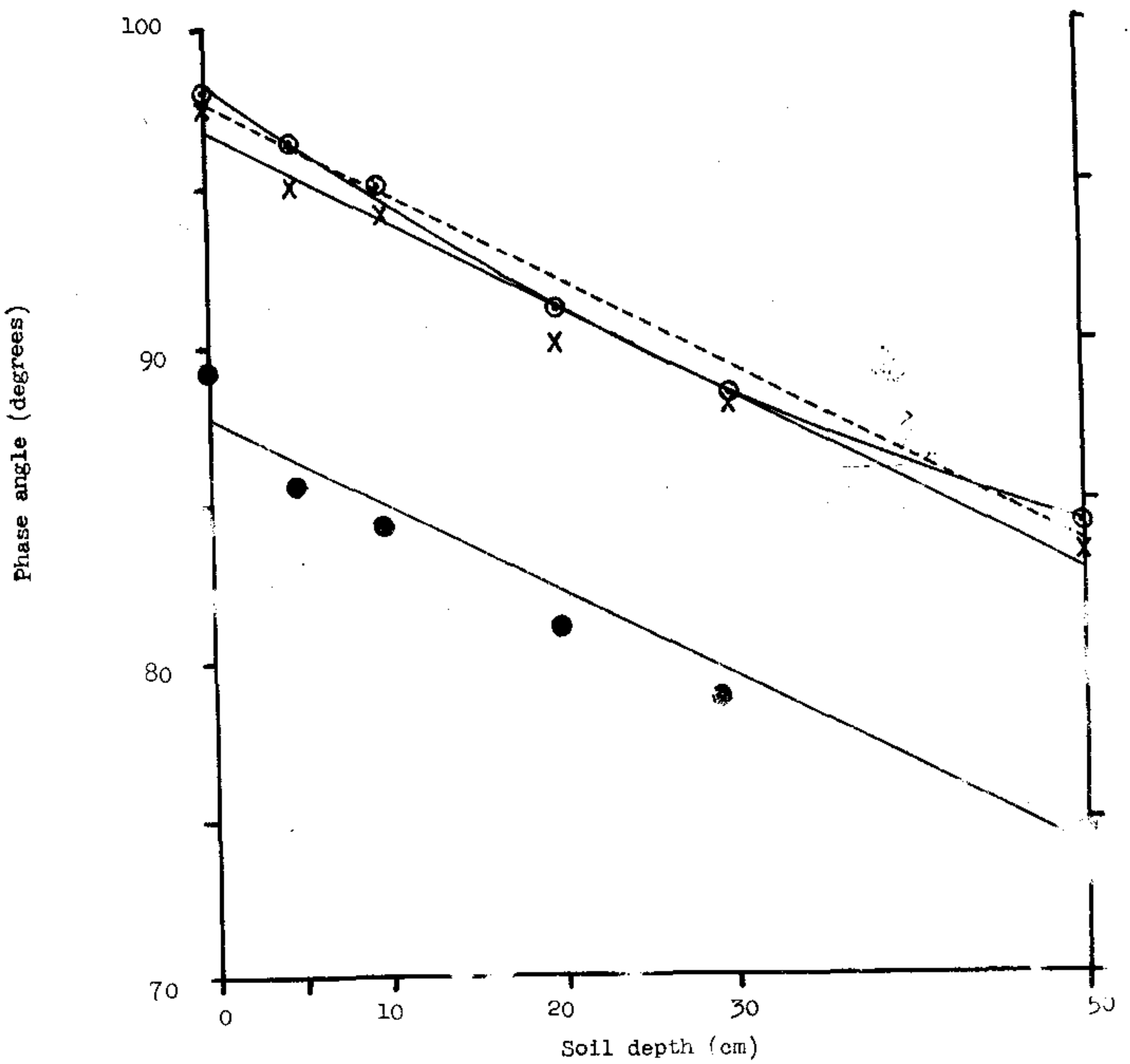


Figure 9. Change in average cumulative lag of the weekly mean temperature curve with soil depth in Meathop Wood. This lag was calculated as (lag_{max} + lag_{min})/2 where the extreme lags were (time of extreme at soil depth d_1) - (time of extreme at depth d_2), d_2 always being the upper depth. X 1966-67, ⊙ 1967-68, ● 1968-69.

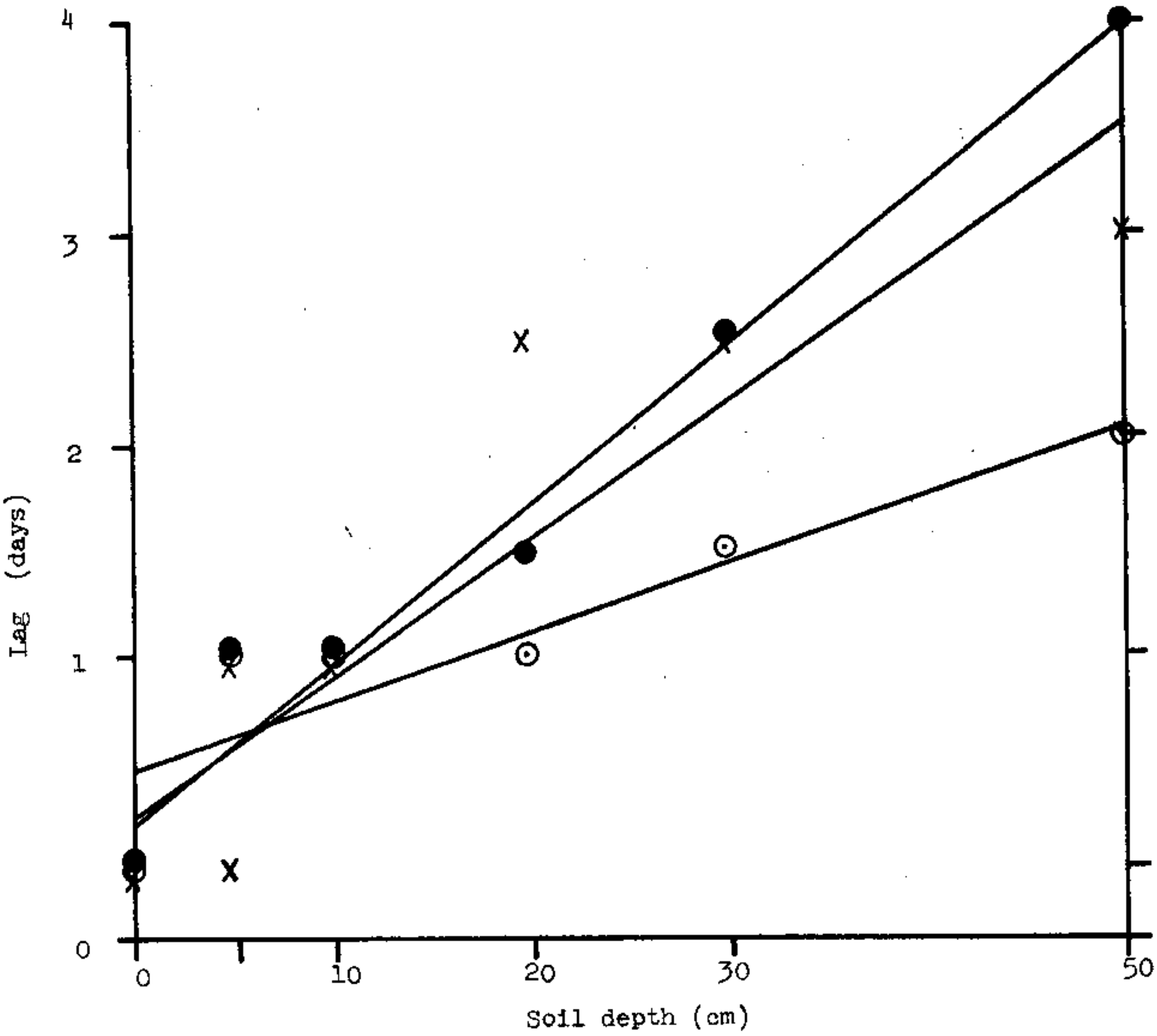


Figure 10. Adjustment to standard points in time of four-weekly mean temperatures collected from 48 sites using staggered sampling. Each of the four curves represents a sampling group of approximately 12 sites.

O Observed data, four weeks apart for successive samplings on the same site, one week apart for data from different sites in adjacent week groups
O Corrected data estimated by tracking along the complete harmonic model curve.

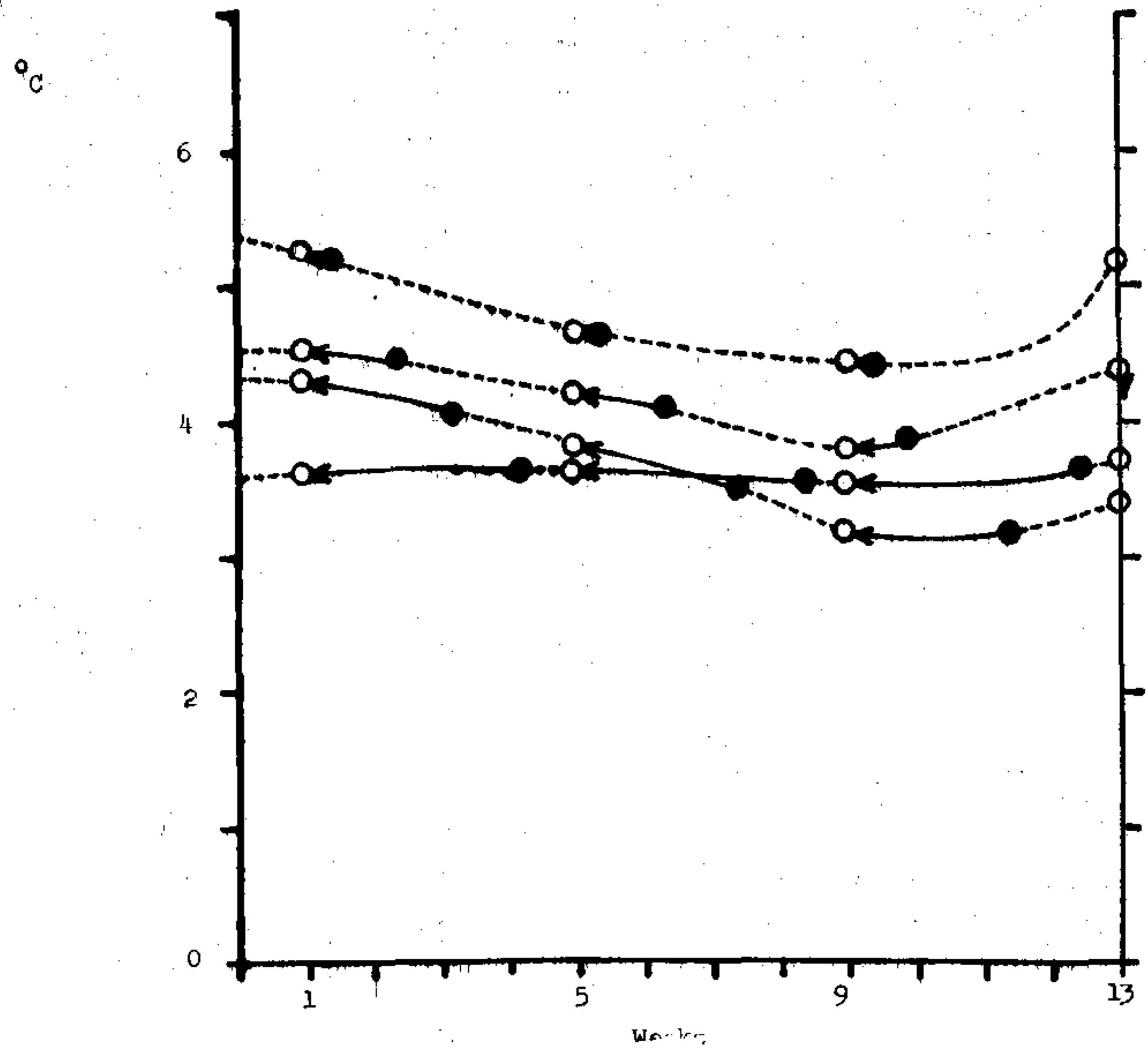


Figure 11. Corrugation of a complete harmonic model curve (O) between observed data points (●). The data are taken from the region of the sixth and seventh points in Figure 12. A straight line has been drawn between the observed data points to emphasize the corrugation.

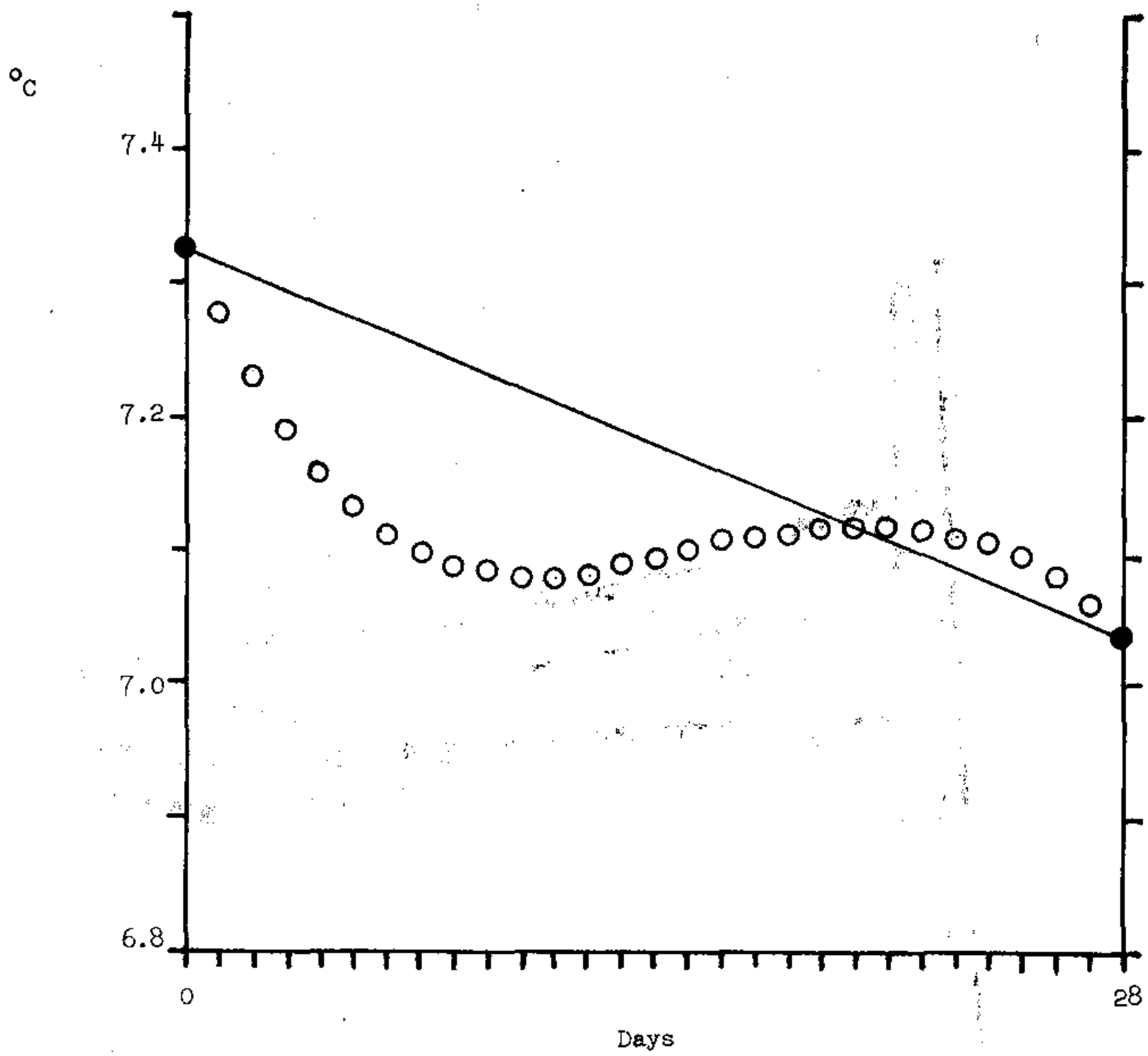


Figure 12. Four-weekly mean temperature at the soil surface of one of 48 Lake District woodland soils examined in 1971-72. X observed data, 0 values predicted from a model incorporating six harmonics, ● missing data estimate, ▲ initial setting for iterative estimation of the missing value.

set