1994

Quantifying Seasonal-Variation In Somatic Tissue - Surfclam Spisula-Solidissima (Dillwyn, 1817) - A Case-Study

Joseph G. Loesch
Virginia Institute of Marine Science

David A. Evans
Virginia Institute of Marine Science

Journal of Shellfish Research, Vol. 13, No. 2, 425–431, 1994.

QUANTIFYING SEASONAL VARIATION IN SOMATIC TISSUE: SURFCLAM *SPISULA SOLIDISSIMA* (DILLWYN, 1817)—A CASE STUDY[1]

JOSEPH G. LOESCH AND DAVID A. EVANS
The College of William and Mary
Virginia Institute of Marine Science
School of Marine Science
Gloucester Point, Virginia 23062

ABSTRACT Condition indexes are commonly derived from bivalve species. Usable meat yields (UMY, in l/bu) from 181 daily landings of Atlantic surfclams, *Spisula solidissima* (Dellwyn, 1817), at a Virginia processing plant in 1974 and 160 landings in 1975 were used as an index in our analysis. The data were fitted to a basic sinusoidal model and a two-compartment sinusoidal model to demonstrate the utility of these models for quantifying cyclic events. The basic model, $x = x_0 + A\cos 2\pi t + B\sin 2\pi t$, is linear in its independent variables and fitted by multiple regression, with x = UMY, t = time in years, where x_0, A, and B are constants determined by the regression procedure (x_0 = mean UMY). Its alternate form is $x = x_0 + r\cos 2\pi(t - t_0)$, with x, x_0, and t as before, r = amplitude of the sinusoidal variation, and t_0 = time when the maximal UMY occurs; r and t_0 are related to A and B as $r = \sqrt{A^2 + B^2}$, and $t_0 = (1/2\pi)\tan^{-1}(B/A)$. The sinusoidal fit to the 1974 data was highly significant ($p < 0.0005$); therefore, the null hypothesis that the data are not a function of time was rejected. The annual mean yield, x_0, was 5.93 l/bu, t_0 was 0.45 (i.e., the maximal UMY occurred about mid-June), and the amplitude r was 0.730; thus, the difference between the lowest and highest yields, $2r$, was almost 1.5 l/bu. Similar estimates were determined from the 1975 data and the combined data. The fit was recalculated for both data sets after excluding *apparent* outliers. As expected, the root-mean-square residual (RMS_{res}) decreased, whereas the coefficient of determination (R^2) increased with the removal of the apparent outliers, but the fitted parameters were inconsequentially affected. A fit of the data to a two-component sinusoidal model, $x = x_0 + A_1\cos 2\pi t + B_1\sin 2\pi t + A_2\cos 4\pi t + B_2\sin 4\pi t$, modeled an annual variation with an asymmetric rise and fall. As a demonstration, the data were also fitted to a parabolic model, $x = a_0 + a_1t + a_2t^2$. Although this model produced fits comparably as close as the sinusoidal models, the coefficients are not interpretable in a simple manner, as is the case with the sinusoidal fits, and it does not allow asymmetric behavior.

KEY WORDS: *Spisula solidissima,* condition index, usable meat yields, seasonal variation, maximum, minimum, sinusoidal, parabolic

INTRODUCTION

Condition indexes are commonly derived for bivalve species. Various index models have been used; in general, the condition indexes reflect a relationship between soft tissue weight and the size of the cavity formed by the two valves. The indexes are used primarily to estimate seasonal meat quality or the effects of disease and pollution on meat quality. It has been suggested that a condition index for oysters be used to monitor pollution. Lawrence and Scott (1982) and Crosley and Gale (1990) reviewed and evaluated bivalve condition index methodologies; in each study, the authors recommended that a standardized index be used, although their models were somewhat different. The presentations and literature cited by those authors and references in the index of papers published in the *Journal of Shellfish Resource* (Castagna et al. 1992, Mann et al. 1993) provide an ample introduction to bivalve condition indexes.

Herein, we present methodologies for estimating seasonal indexes, the maximal and minimal annual values, associated confidence intervals, and tests of significance, regardless of the condition index used.

METHODS

Condition Index

To demonstrate the model, we use a condition index defined as usable meat yields (UMY) in liters per bushel of the Atlantic surfclam, *Spisula solidissima* (Dillwyn, 1817).

Source of Data

The UMYs were determined from daily landings of surfclams at the C&D Seafood Co. in Oyster, Virginia—181 landings totaling 167,564 bushels in 1974 and 160 landings totaling 270,170 bushels in 1975. In both years, the surfclams were harvested in an area approximately between 8.5 to 17.5 nautical miles offshore of Cape Henry and south to the North Carolina state line.

Sinusoidal Model

Loesch (1977) reported the relationship between mean monthly water temperature and mean monthly usable meat yield per bushel (mean UMY) for surfclams. The data in terms of *daily* UMYs are resurrected herein to assess parameters not previously considered in order to demonstrate the utility of sinusoidal functions for quantifying cyclic events exhibited in the life history of many marine species.

The basic sinusoidal model used was

$$x = x_0 + A\cos 2\pi t + B\sin 2\pi t$$

where

$$x = \text{UMY in l/bu}$$

$$t = \text{time of the year (in years)}$$

and the model parameters determined by regression procedure are

$$x_0 \text{ (annual mean UMY in l/bu), and } A \text{ and } B.$$

The sample data were fitted to the model by regressing x on $\cos 2\pi t$ and $\sin 2\pi t$. Although $\cos 2\pi t$ and $\sin 2\pi t$ both depend on t, they are *linearly* independent of each other and therefore can be used

[1]Contribution 1896 of The College of William and Mary, Virginia Institute of Marine Science, School of Marine Science, Gloucester Point, Virginia, USA.

as independent variables in a multiple linear regression procedure.

The model is alternatively expressed

$$x = x_0 + r \cos 2\pi(t - t_0)$$

where x_0 is the mean UMY, r is the amplitude of the sinusoidal variation, and t_0 is time when the maximal UMY occurs; r and t_0 are related to A and B as follows:

$$r = (A^2 + B^2)^{1/2}$$

and

$$t_0 = (1/2\pi) \tan^{-1}(B/A). \text{ [see footnote 2]}$$

Two-Component Sinusoidal Model

A feature of the basic sinusoidal model is that the rise and fall on either side of the maximum (or minimum) are symmetrical. This could be regarded as an unrealistic constraint to put upon the model. The problem is addressed by including additional terms to account for the additional feature. The appropriate extension of the sinusoidal model is to include *two* additional terms that constitute an additional sinusoidal component with a period of 6 months, i.e., one-half of the period of the basic sinusoid:

$$x = x_0 + A_1 \cos 2\pi t + B_1 \sin 2\pi t + A_2 \cos 4\pi t + B_2 \sin 4\pi t$$

The function is still linear in the parameters, and the fit can again be performed using a standard regression procedure. As with the one-component model, an alternative expression is:

$$x = x_0 + r_1 \cos 2\pi(t - t_0^{(1)}) + r_2 \cos 4\pi(t - t_0^{(2)}),$$

where r_2 is the amplitude of the second component. The interpretation of $t_0^{(1)}$ and $t_0^{(2)}$ in terms of time of maximum is, however, now more complex.

Alternative Quadratic (Parabolic) Models

For method comparison purposes, in addition to fitting the data to a sinusoid, we consider the quadratic function:

$$x = a_0 + a_1 t + a_2 t^2$$

[2]There are two angles in the range 0–2π radiants whose tangent is B/A. The appropriate one lies in the quadrant where its *cosine* has the same sign as A and its *sine* has the same sign as B.

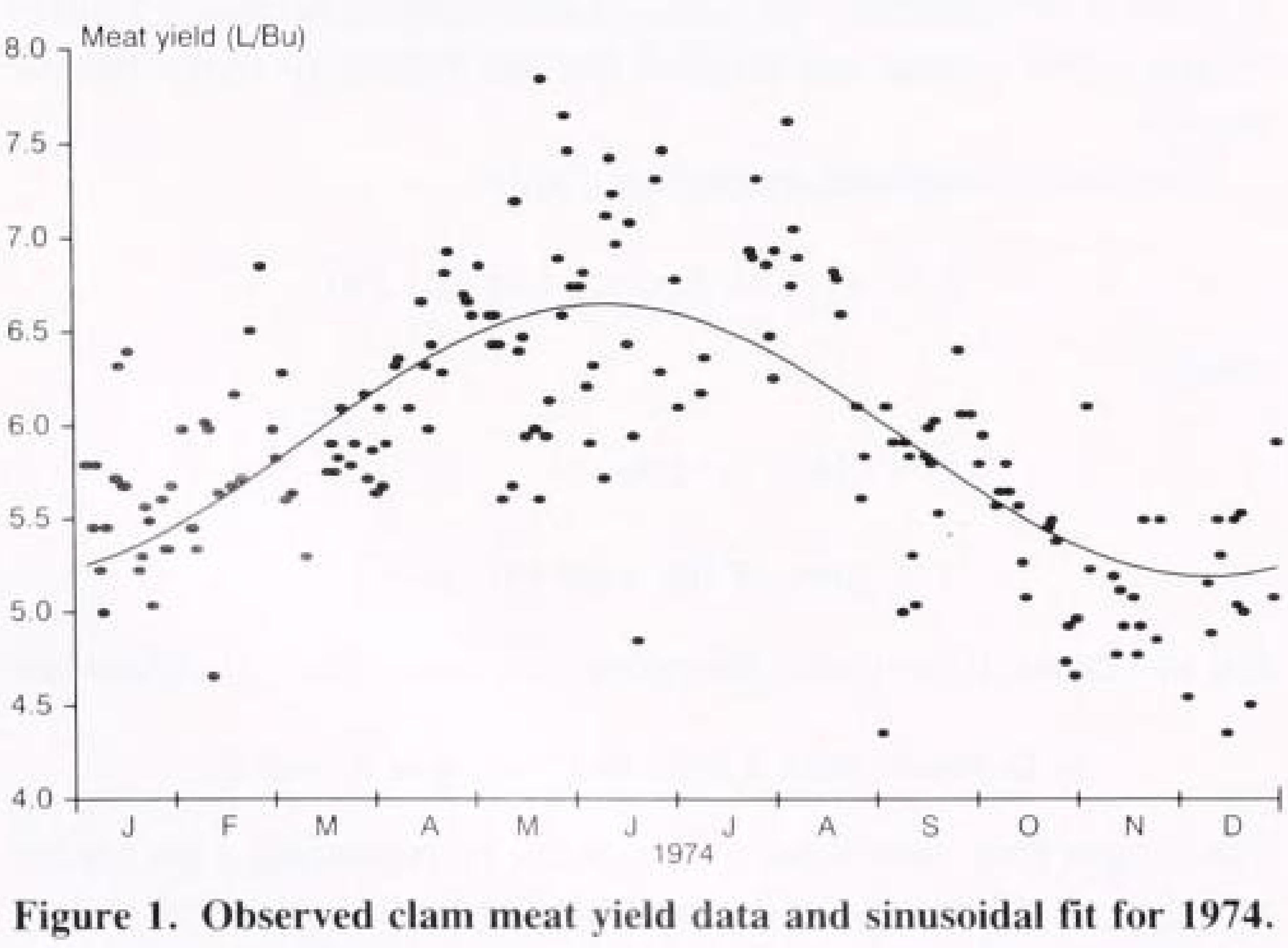

Figure 1. Observed clam meat yield data and sinusoidal fit for 1974.

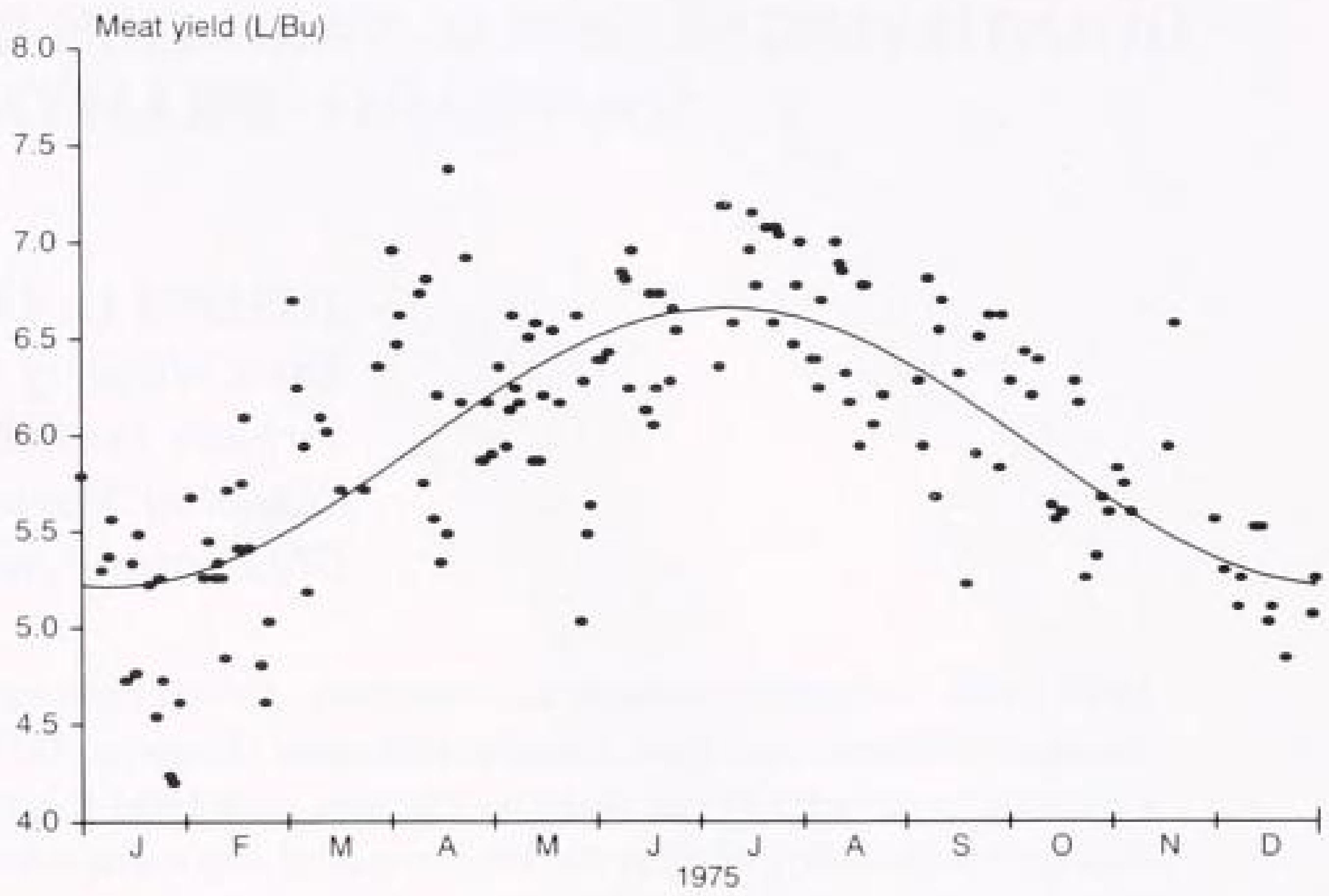

Figure 2. Observed clam meat yield data and sinusoidal fit for 1975.

as an alternative model. This function describes a parabola, containing a single maximum (when $a_2 < 0$) or minimum (when $a_2 > 0$). The position of the maximum (or minimum) is given in terms of the model parameters by the following expressions:

$$t_{max} = -\frac{a_1}{2a_2};\ x_{max} = a_0 - \frac{a_1^2}{4a_2}$$

In order to treat the feature of asymmetry, a term in t^3 can be added to the quadratic model to give a cubic model:

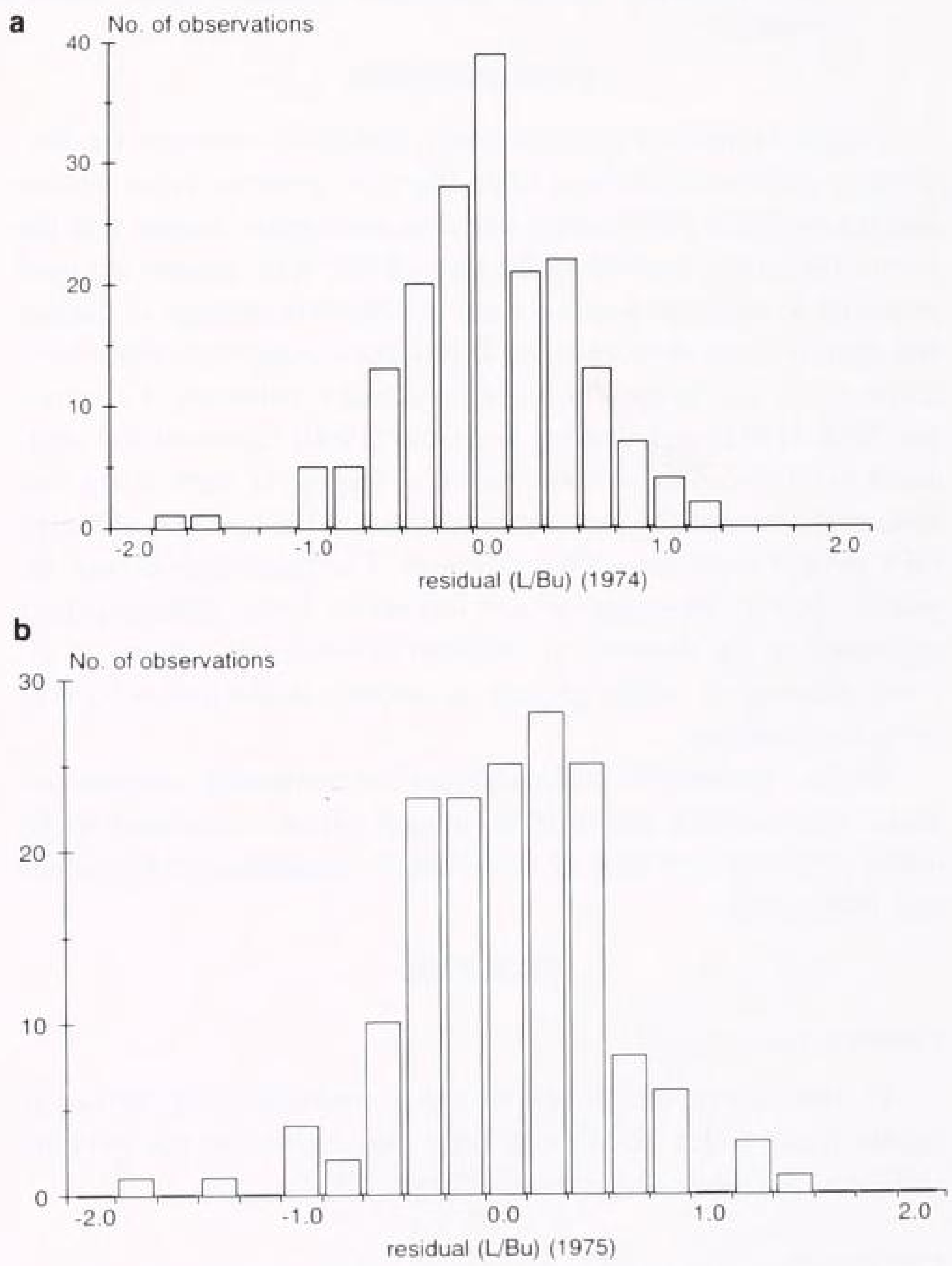

Figure 3. Distribution of residuals from the fit to data from (a) 1974 and (b) 1975.

TABLE 1.

Results from fitting a sinusoidal model $x = x_0 + A\cos 2\pi t + B\sin 2\pi t$ to the clam meat yield data. The model is alternatively expressed as $x = x_0 + r\cos 2\pi(t - t_0)$.

Year/Cut	N	x_0	A	B	r	t_0	R^2	RMS_{res}	p
1974, all	182	5.93	−0.682	0.261	0.730	0.45	0.53	0.496	<0.0005
1.96σ cut ("5%")	172	5.91	−0.720	0.232	0.756	0.45	0.63	0.416	
1.65σ cut ("10%")	162	5.93	−0.709	0.233	0.746	0.45	0.68	0.367	
1975, all	159	5.94	−0.718	−0.094	0.724	0.52	0.54	0.479	<0.0005
1.96σ cut ("5%")	150	5.96	−0.715	−0.134	0.727	0.53	0.66	0.375	
1.65σ ("10%")	146	5.96	−0.712	−0.141	0.726	0.53	0.69	0.354	
Both years	341	5.93	−0.706	0.097	0.713	0.48	0.50	0.502	<0.0005

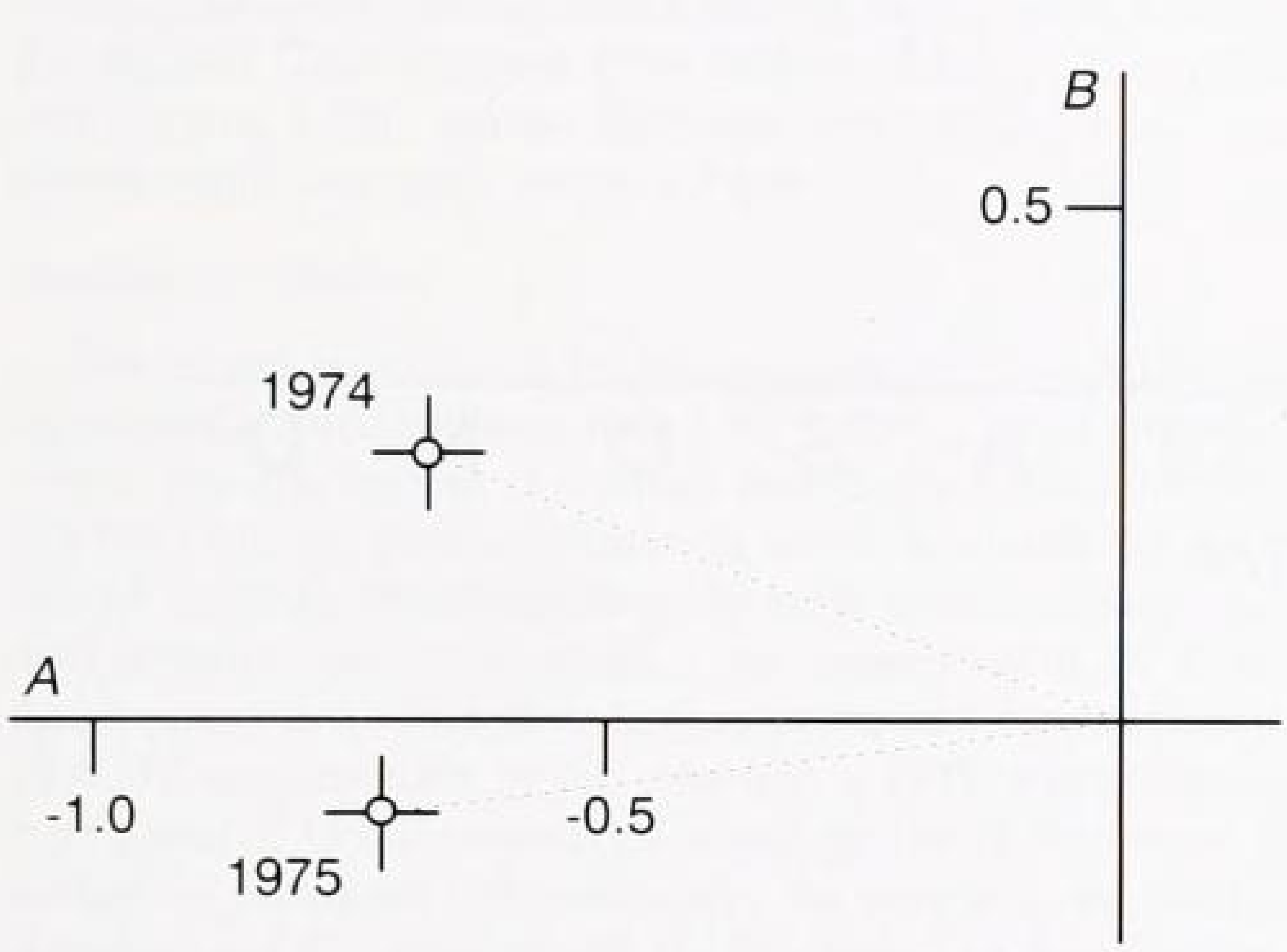

Figure 4. Geometrical representation of the sinusoidal fit parameters for the 2 years. The error bars represent the standard error in the estimation of the parameters.

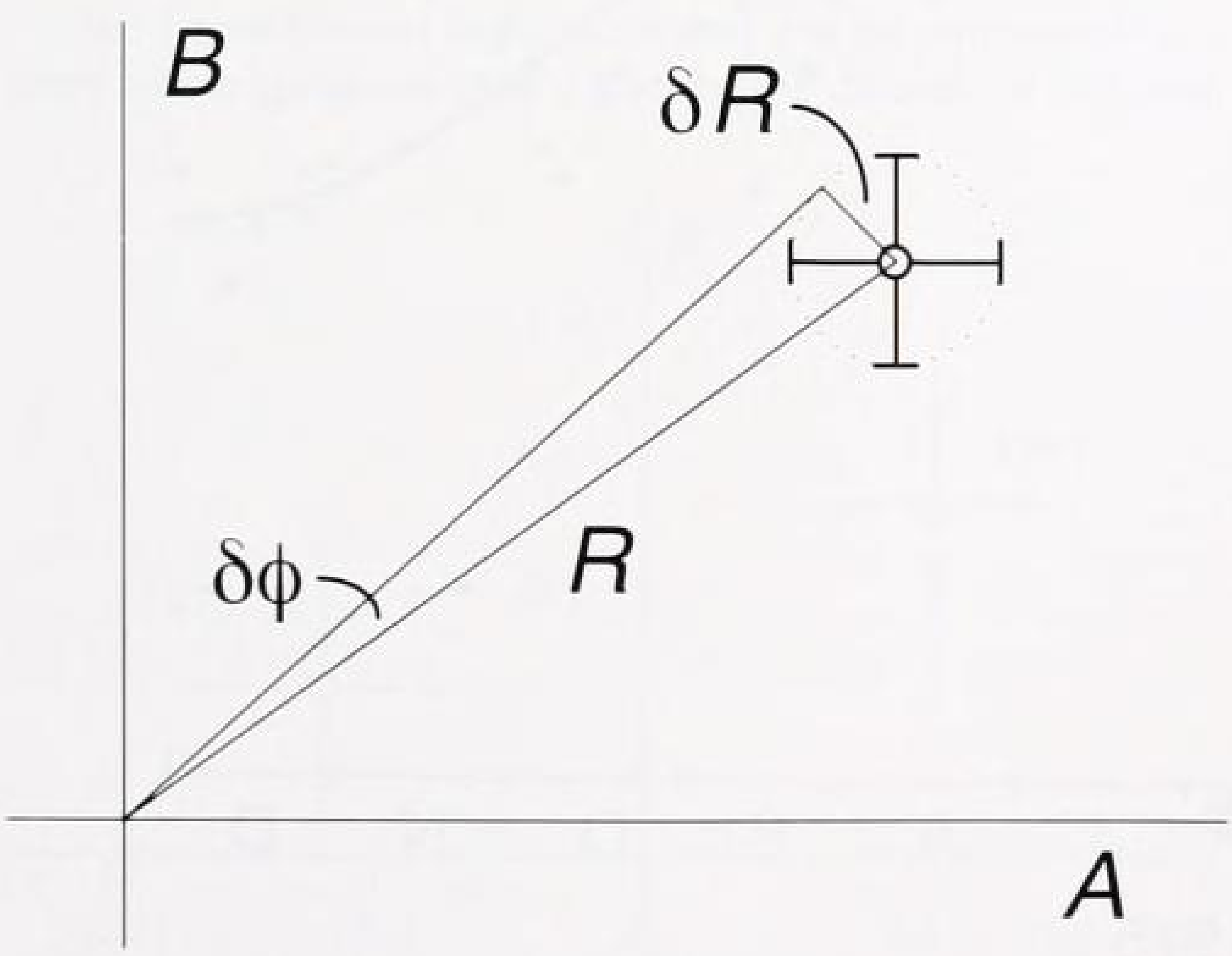

Figure 5. Geometrical representation of the derivation of the approximate error in the phase angle.

$$x = a_0 + a_1 t + a_2 t^2 + a_3 t^3.$$

The addition of this extra term usually results in completely different values for the coefficients obtained from the quadratic fit; furthermore, interpretation of the coefficients becomes obscure. In the data under consideration, another problem occurs. Over the range of t, the value of t^3 is very nearly linearly dependent upon t and t^2, the normal equations are ill-conditioned, and thus, the coefficients are very poorly determined. The cubic expression $x = a_0 + a_1 t + a_3 t^3$ can be fitted to the data and gives an asymmetrical curve about a maximum. It is, however, logically inadmissible because an attempt is being made to characterize an additional feature (the asymmetry of the slopes) *with the same number of parameters as was used in the quadratic fit*. The three parameters in the quadratic can be associated with three defining features of a parabola: the location of the maximum (or minimum) requires two parameters; the "shallowness" of the curve is the third. An additional parameter is therefore required to explain any asymmetry. Consequently, a cubic expression with the term in t^2 suppressed cannot characterize *independently* the asymmetry and location of the maximum. For this set of data, the quadratic fit is as far as one can go with simple polynomial models and this model does not incorporate any asymmetry.

RESULTS

Basic Sinusoidal Model: 1974 and 1975

The seasonal variation in UMY is obvious, as is the variation about the fitted curve (Figs. 1 and 2). The basic sinusoidal fit to

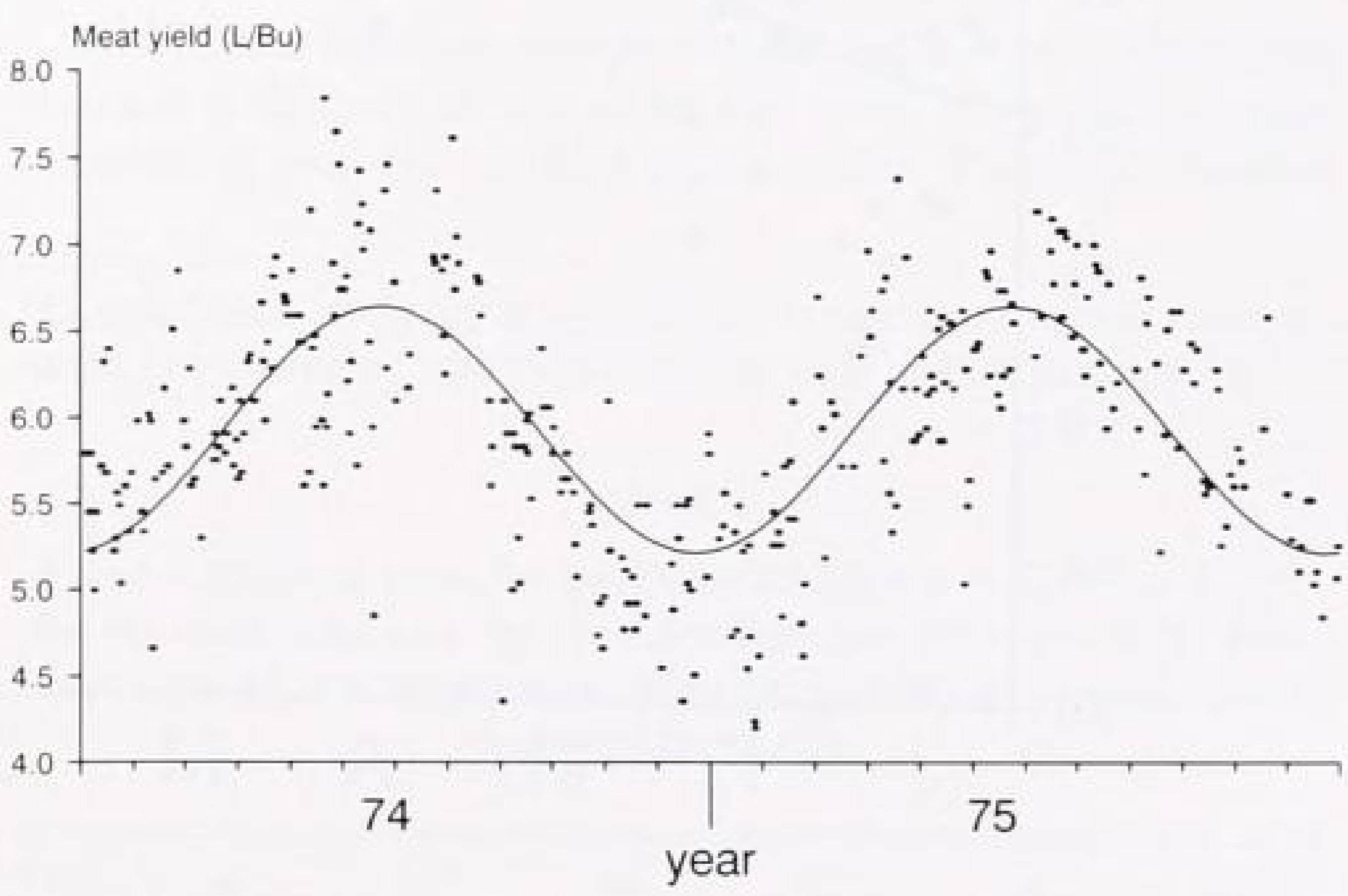

Figure 6. Observed clam meat yield data and sinusoidal fit for the 2-year period 1974 (74) to 1975 (75).

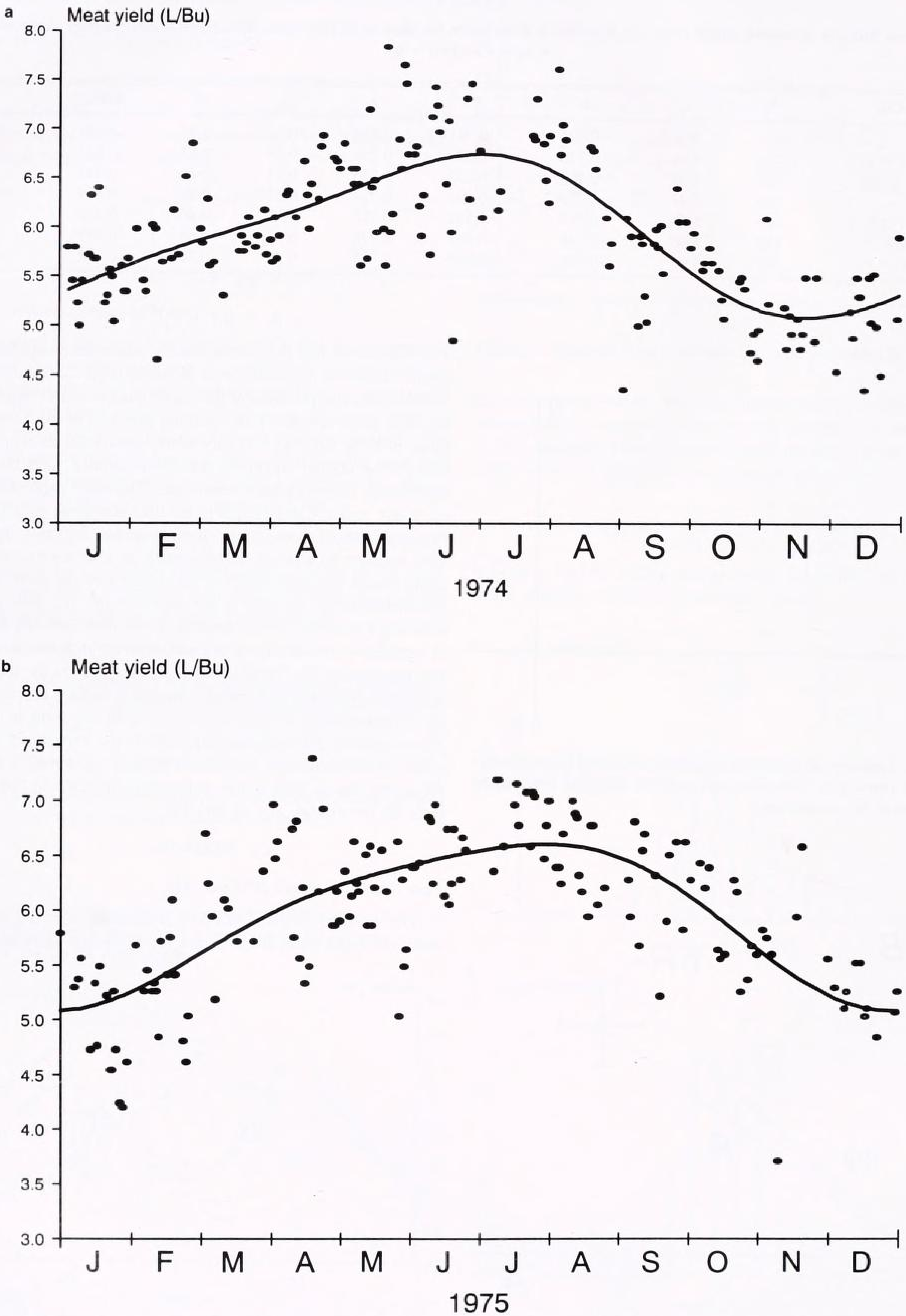

Figure 7. Two-component sinusoidal fit for (a) 1974 data and (b) 1975 data. The additional component (with a 6-month period) allows asymmetry to appear in the fitted function.

TABLE 2.

Coefficients for fit of a two-component sinusoidal function to the clam yield data. The fitted function is: $x = x_0 + A_1 \cos 2\pi t + B_1 \sin 2\pi t + A_2 \cos 4\pi t + B_2 \sin 4\pi t$.

Year	A_0	A_1	B_1	A_2	B_2	R^2	p	RMS_{res}
1974	5.927	−0.712	0.264	0.107	0.141	0.558	<0.0005	0.483
1975	5.937	−0.731	−0.082	−0.119	0.063	0.557	<0.0005	0.473

the 1974 data was highly significant ($p < 0.0005$); therefore, the null hypothesis that the data are not a function of time was rejected. The mean yield, x_0, was 5.93 l/bu, and t_0 was 0.45; thus, the maximal UMY occurred about mid-June. The amplitude r was 0.730, giving a difference between the lowest and highest yields ($2r$) of almost 1.5 l/bu.

For 1975 the sinusoidal fit was also highly significant ($p < 0.0005$); the mean yield, x_0, was 5.94 l/bu, and t_0 was 0.52; thus, the maximal UMY occurred about early to mid-July. The amplitude (r) was 0.724, and the difference between the lowest and highest yields was again almost 1.5 l/bu.

Sensitivity to "Outliers"

The fit was recalculated for both data sets after excluding *apparent* outliers that lay more than $1.96 \times RMS_{res}$ from the initial curves. The distributions of residuals from the fit to all of the 1974 and 1975 data are shown in Figure 3a and b. Assuming the residuals are normally distributed about the curve with a common standard deviation equal to the RMS_{res}, one expects about 5% of the datum points to be excluded. In fact, of the 182 observations in 1974, 10 were excluded by this criterion; in 1975, 9 observations from a total of 159 were excluded. Thus, the rate of occurrence of outliers is consistent with normality. As expected, the RMS_{res} decreased and R^2 increased with the rejection of apparent outliers (Table 1). The point estimates of the parameters, however, were inconsequently affected. The procedure was repeated for a cutoff of $\pm 1.65 \times RMS_{res}$, corresponding to an expected 10% rejection rate. Similar results were obtained and are presented in Table 1. All subsequent analyses use the whole set of observations.

Comparison Between Years

The fit coefficients for a single year can be represented by a point whose coordinates are A and B. The distance of this point from the origin is equal to the amplitude r, and the angular position of the point, measured counterclockwise from the x-axis, is an angle $\phi = 2\pi t_0$. Figure 4 shows the two points corresponding to the two years. The cross arms represent the standard errors in the estimation of A and B. They are all approximately equal to 0.053. As is discussed later, these errors are uncorrelated so that the standard error in r for each year is also approximately 0.053. The difference between the two amplitudes is 0.007; the standard error in this quantity is approximately $\sqrt{2} \times 0.053 = 0.075$. This gives a t-statistic of 0.09. There is, therefore, no evidence of a difference in amplitude between the years.

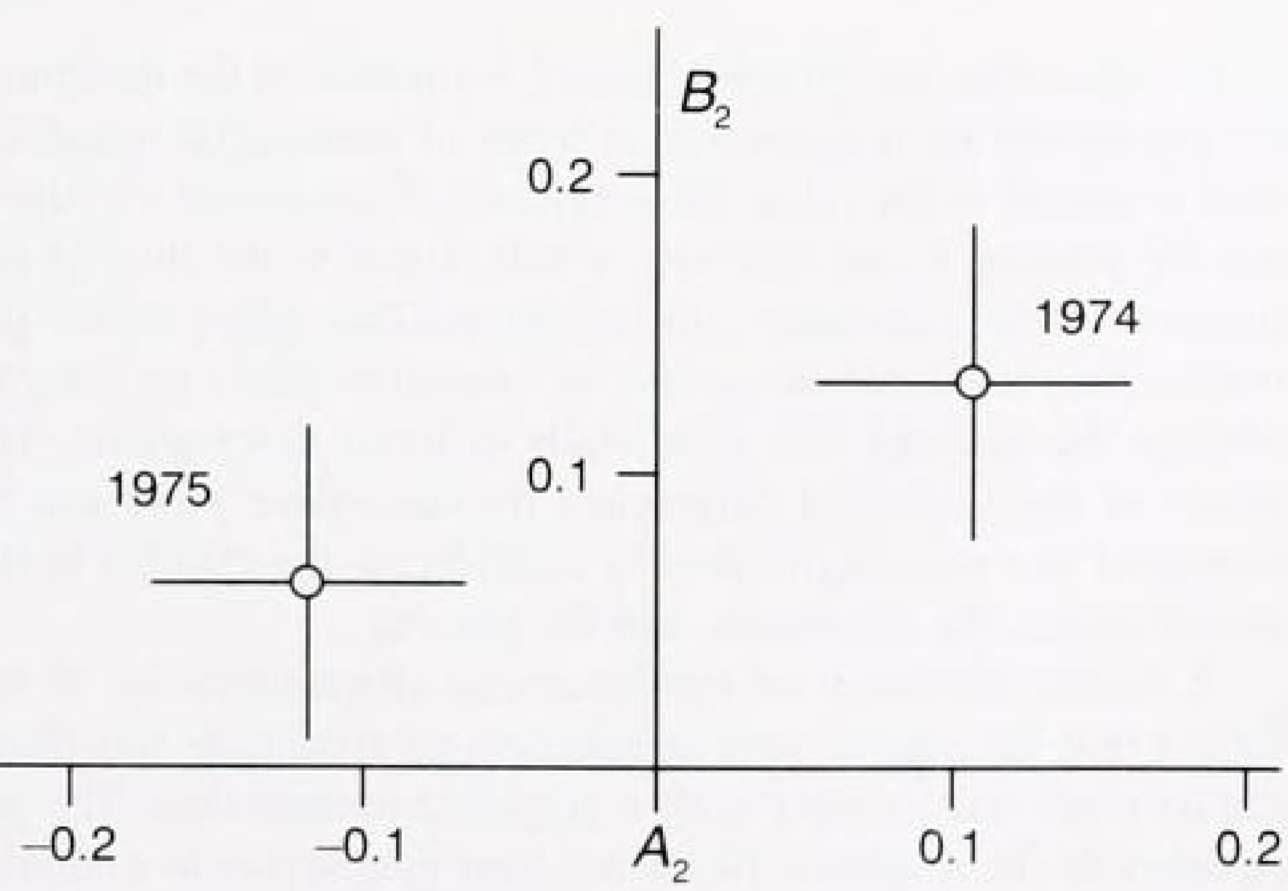

Figure 8. Geometrical representation of the second sinusoidal component for each of the years. The error bars represent the standard error in the estimation of the parameters.

The treatment of the phase angles is different. Because the error in r is much smaller than r itself, one may say that, approximately:

$$\delta\phi = \textit{error in phase (in radians)} \approx \textit{(error in r)/r}$$

This result is demonstrated in Figure 5. For each of the years, the error in the phase is approximately $0.053/0.73 = 0.073$ radians. This, in turn, corresponds to $0.073/2\pi = 0.012$ years for the error in t_0. The error in the difference of the t_0 values is $0.012 \times \sqrt{2} = 0.017$. The observations shows:

$$t_0(1975) - t_0(1974) = 0.52 - 0.45 = 0.07$$

Thus, the t-statistic[3] is $0.07/0.017 = 4.1$. The appropriate degrees of freedom are very large (the total number of observations less two sets of three parameters), so that the Student's t-distribution is very nearly normal. The observed value is therefore highly significant. It can be concluded that 1975 was a year in which the maximum yield occurred later than in 1974. Qualitatively, a glance at Figure 4 indicates that the angle between the two dotted radii is much larger than can be accounted for by the size of the standard error.

Combined Data

Although it has been demonstrated that the best-fit sinusoidal variation is different in each of the two years, the whole two-year sequence of data can be fitted to the model. The fitted function

[3]An unfortunate ambiguity of notation occurs here: the Student's t statistic is not to be confused with the use of t for time and the parameter t_0.

TABLE 3.

Fitted coefficients from the parabolic function $x = a_0 + a_1 t + a_2 t^2$ for the clam yield data for the calendar years 1974 and 1975. Also shown are R^2, the significance of the fit, and the root-mean-square deviation for the fit.

Year	a_0	a_1	a_2	R^2	p	RMS_{res}
1974	5.257	5.420	−6.151	0.475	<0.0005	0.524
1975	4.709	7.005	−6.800	0.532	<0.0005	0.483

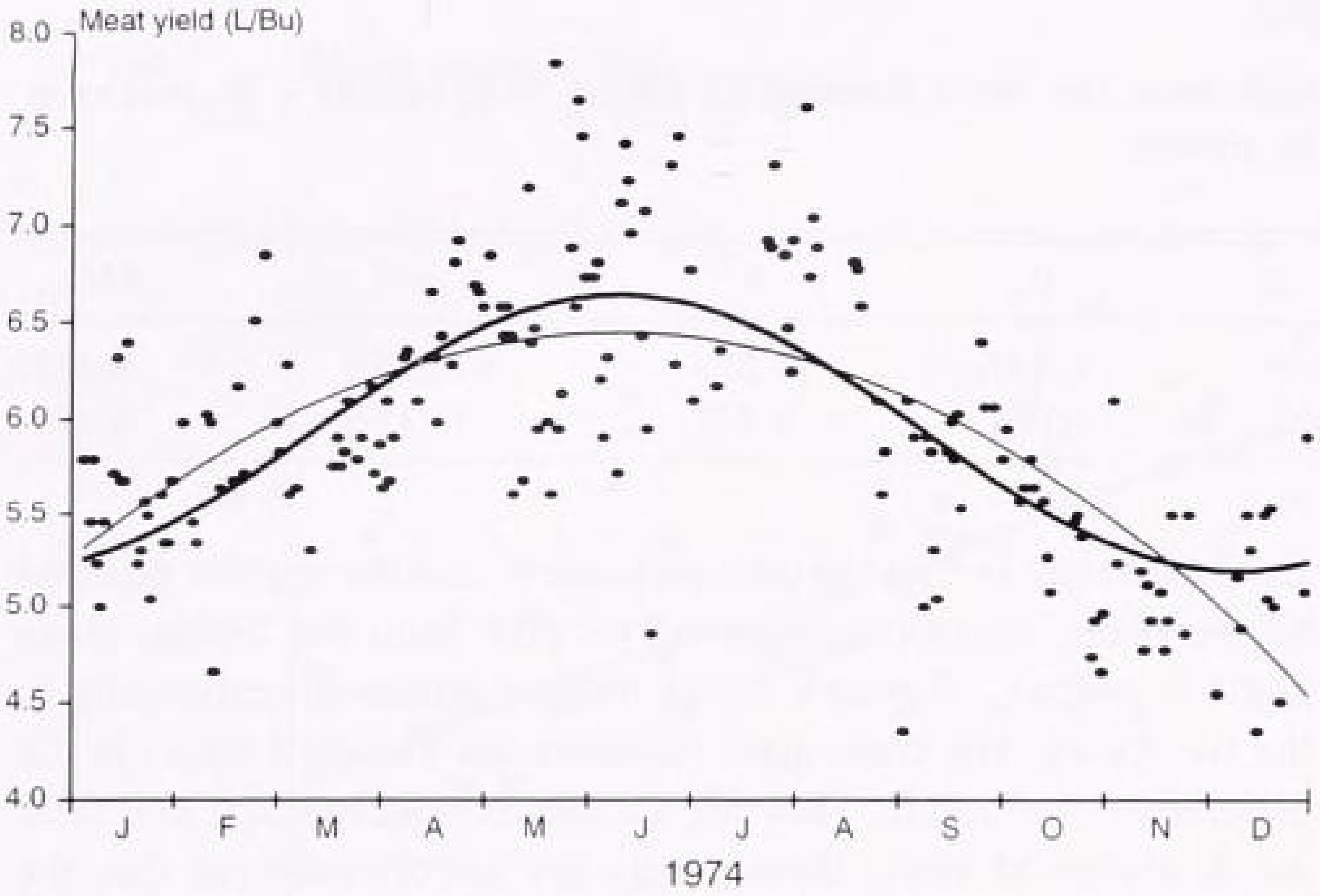

Figure 9. Quadratic and sinusoid fits for 1974 data. The heavy curve is the sinusoid.

together with the data is shown in Figure 6, and the fitted parameters are given in Table 1. The same annual variation is now imposed upon each year and thus represents a compromise between the two differing years. This is reflected in a reduced value of R^2 and an increased value of RMS_{res} compared with the single-year fits. For a data set that spanned several years, the overall fit would be useful in establishing a "typical" annual variation. This would then enable a classification of each year by comparing, in some suitable manner, the fit of one year's data with the "typical" year.

Two-Component Sinusoidal Model

The results of this fit for the 1974 and 1975 data are shown in Figure 7a and b. The coefficients for each year are given in Table 2. For both years, the incorporation of the extra component is statistically significant: in 1974, $p(A_2) = 0.039$ and $p(B_2) = 0.007$, and in 1975, $p(A_2) = 0.032$ and $p(B_2) = 0.237$. Note that the significant presence of the component is indicated by rejecting the null hypothesis, $H_0{:}r_2 = 0$. This is fully justified for only one of the parameters, A_2, B_2 being significantly different from zero. It is seen that, in each year, the fit reproduces an asymmetric rise

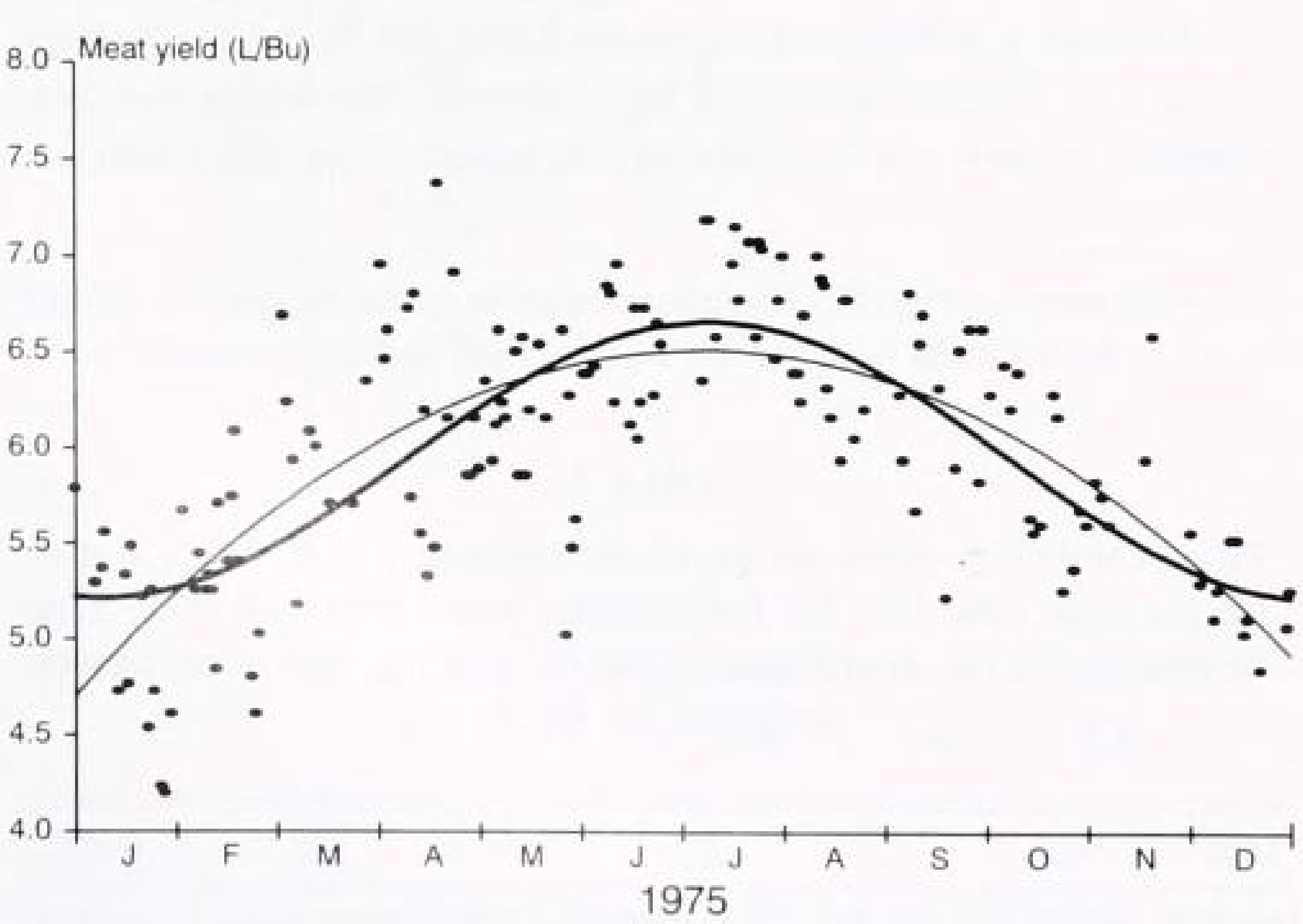

Figure 10. Quadratic and sinusoid fits for 1975 data. The heavy curve is the sinusoid.

TABLE 4.

Comparison of the parabolic and sinusoidal fits for the clam yield data for the 12 months July 1974 to June 1975.

Parabola	a_0	a_1	a_2	R^2	p	RMS_{res}
	5.216	0.058	6.999	0.488	<0.0005	0.520
Sinusoid	x_0	r	t_0	R^2	p	RMS_{res}
	5.787	0.752	5.646	0.546	<0.0005	0.490

and fall. The details of this feature appear to differ markedly between the two years. Qualitatively, the peak occurrence in 1974 is relatively sharp compared with that in 1975, where the persistence of larger values into August and September is quite marked. The difference between the years is statistically significant, as may be judged from Figure 8, which is a plot displaying the coefficients of the second component in a manner similar to that in Figure 4 for the first component.

Alternative Quadratic (Parabolic) Models

Table 3 shows the fitted values for the three parameters a_0, a_1, and a_2 for each of the years. Figures 9 and 10 show the datum points, the fitted parabola, and the fitted sinusoid for each of the years. Although the tightnesses of the fits, characterized by the values of R^2 and RMS_{res}, are of the same order as the sinusoids, the interpretation of the model parameters is less clear. An inverted parabola represents quite well the occurrence of the maximum yield approximately halfway through the year. The estimates of maximal UMY are $t_{max} = 0.441$ for 1974 and $t_{max} = 0.515$ for 1975, in quite good agreement with the sinusoidal fit. If, however, the data for the 12 months from July 1974 through June 1975 are fitted to a quadratic function, a totally different set of coefficients is obtained (Table 4), with in particular, $a_2 > 0$, i.e., a concave-up parabola with a single *minimum*. On the other hand, fitting the sinusoidal function to the July 1974 to June 1975 data gives values for the model parameters r and t_0 (Table 4) that are very similar to those from the two fits for the data from January through December in 1974 and 1975. These two fitted functions and the data are shown in Figure 11.

DISCUSSION

It is seen that the sinusoidal model is superior to the quadratic: the parameters are interpretable in terms of meaningful quantities such as annual mean value, the amplitude of the annual variation, and the phasing of the sinusoid, which relates to the time of occurrence of the maximum and minimum. The values of the parameters are relatively insensitive to where the year's data begin, whereas the quadratic fits give totally different descriptions. The values of the sinusoidal parameters for successive years can be compared in a meaningful way by considering the changes in the overall mean, the amplitude, and the phasing.

A further advantage of the sinusoidal characterization of the data is that the year-to-year comparison of amplitude and phase can be made quite simply with a graphical presentation. The parameters in the quadratic fit do not lend themselves to a similar simple geometric interpretation. Calculation of the errors in functions of the parameters (such as differences) is more complicated because the errors of determination in the parameters are corre-

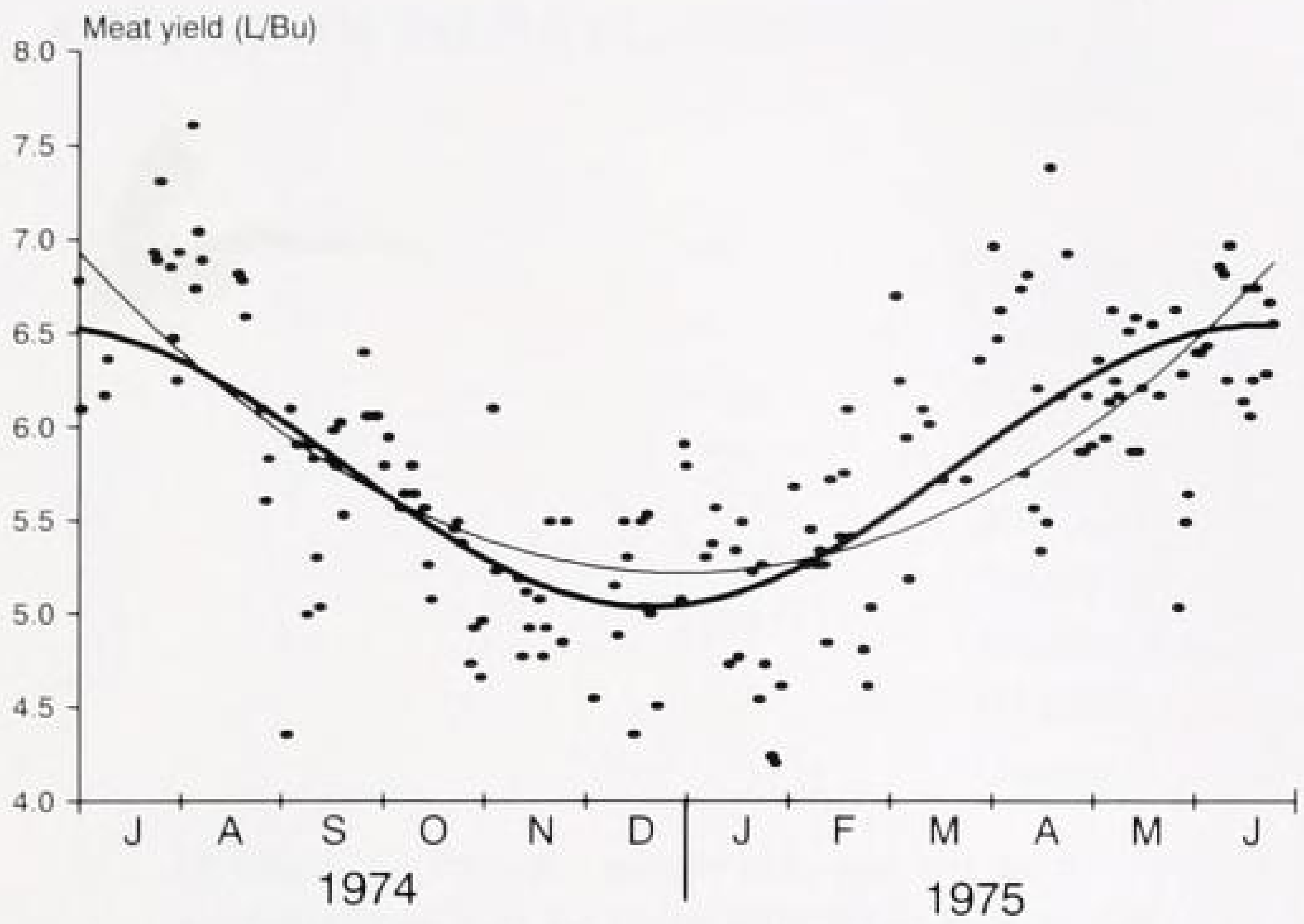

Figure 11. Quadratic and sinusoid fits for the July 1974 to June 1975 sequence. The heavy curve is the sinusoid.

lated. It is a feature of the sinusoidal functions that, if the observations are evenly spaced over a complete period (or multiple of periods), the determinations of the parameters are perfectly uncorrelated. For a relatively large number of points that are distributed approximately uniformly over a complete period, as is the case here, the correlation is negligible. The addition to the model of a second sinusoidal component with a 6-month period allows the characterization of an asymmetric rise and fall. The coefficients of the basic sinusoid are little affected by the addition of these extra terms to the model. This is because the determinations of the parameters are uncorrelated with each other as noted above.

The quadratic fit has no intrinsic merit and is merely an arbitrary parametrization of the observations. The model is suggested because the data for a complete year from January to December show the presence of an apparent maximum and the quadratic function can reproduce this feature. However, 12 months of data from July through June exhibit the opposite appearance, with the presence of a minimum, which gives rise to a totally different fit.

LITERATURE CITED

Castagna, M., N. Lewis, C. M. McFadden & M. Gibbons. 1992. Index of papers published in the Journal of Shellfish Research. *J. Shellfish Res.* 11:561–581.

Crosley, M. P. & L. D. Gale. 1990. A review and evaluation of bivalve index methodologies with a suggested standard method. *J. Shellfish Res.* 9:233–237.

Lawrence, D. R. & G. I. Scott. 1982. The determination and use of condition index of oysters. *Estuaries* 5:23–27.

Loesch, J. G. 1977. Usable meat yields in the Virginia surf clam fishery. *Fish Bull.* 75:640–642.

Mann, R. E., M. Lynch, M. Castagna, B. M. Campos, & N. Lewis. 1993. Index of papers published in the Proceedings of the National Shellfisheries Association. *J. Shellfish Res.* 12:158–182.