

## A Comparison of Mortality Patterns in Human Populations Residing Under Diverse Ecological Conditions: A Time Series Analysis

By Paul M. Lin<sup>1</sup> and Michael H. Crawford<sup>2</sup>

### ABSTRACT

During the last century, a number of epidemics have swept across the world causing similar mortality peaks in diverse human populations. In particular, the effects of the influenza epidemic of 1918 can be seen in urban and rural human aggregates separated by continents and thousands of miles.

This paper examines mortality periodicity, due to diverse population structures, ecology, and exposure to similar pathogens, through the use of time series analyses. Specifically, raw yearly mortality figures for two Italian alpine communities, Aceglio and Bellino, are compared with those of a Mennonite congregation living in Kansas, United States, for the same time periods. Crosscorrelation, autocorrelation, and power spectrum analyses have been applied in order to identify possible mortality periodicity and to compare these cycles across populations. The mortality cycles occur at approximately 10 years in the Mennonite series, and 13 in Aceglio and Bellino. Explanations are proposed for these data and for the significant correlations exhibited by the three time series. The last century of human existence saw a number of major demographic changes on a world-wide basis resulting from a variety of technological breakthroughs and medical developments. For example, as a result of innovations in transportation, there has been a rapid breakdown of reproductive and geographical isolation of small human populations such as the Mennonites. Due in part to this geographical isolation, communities that were exposed to specific pathogens periodically experienced disease epidemics, and mortality patterns were unique to each population. The incidence and duration of these epidemics depended in part on the demographic structure of the population and the unique historical events that introduced the pathogen into the community.

The purpose of this paper is to explore the mortality patterns of three human populations living under diverse ecological conditions with exposure to various pandemic diseases. In particular, we examine the periodicity of mortality patterns using power spectral, cross-correlation and autocorrelation analyses, and explore some variables which may contribute to this periodicity.

<sup>1</sup>Institute of Social and Behavioral Pathology and Psychiatric Research Unit, Department of Psychiatry, Pritzker School of Medicine, University of Chicago, Chicago, Illinois 60637.

<sup>2</sup>Laboratory of Biological Anthropology, University of Kansas, Lawrence, Kansas 66045.

### *The Populations*

Mortality data from three populations living under diverse ecological conditions are analyzed in this study. These populations are: Alexanderwohl, Kansas (a Mennonite agricultural community), and Acceglio and Bellino (two Alpine agricultural communes of Northern Italy).

*Alexanderwohl.* The Mennonites are a Dutch Anabaptist group that followed the leadership of a Dutch Catholic priest, Menno Simons. These Mennonites were tolerated in the Netherlands during the mid-16th century, however, following the conquest of their homeland by Spain, they experienced extreme persecution. The Mennonites emigrated to East Prussia during the 17th century and remained there until the late 18th and early 19th centuries when most of them immigrated to Russia and established a series of colonies near the Dnieper and Molotschna Rivers in the Ukraine and in the Crimean Peninsula. The final migration of the Mennonites to the New World, established colonies in the United States, Canada, Central America, and Paraguay.

The mortality records utilized for this time series analysis were derived from one Mennonite community, Alexanderwohl, whose origin can be traced to the Netherlands, Prussia, and Russia. Although this community was highly heterogeneous genetically in its origin in the Netherlands, it was a recognizable entity in Prussia (Przechowka) and Russia (Alexanderwohl). Fifteen couples founded the Przechowka Church in 1669, and all but seven families of this congregation immigrated to Russia in 1821 and settled in the Ukraine near the Molotschna River. The congregation adopted the name "Alexanderwohl" in honor of the Russian Czar. Almost all of the Alexanderwohl Mennonites migrated to the United States in 1874. The congregation underwent fission with the majority settling in Goessel, Kansas, while an offshoot community was planted in Henderson, Nebraska. The Alexanderwohl congregation consists of 1122 parishioners residing on the plains of south central Kansas, in Goessel and the surrounding area.

### *Acceglio and Bellino*

In 1968 Acceglio was a community of 482 persons who resided in seven fraziones or hamlets. It is located at the terminus of an Italian Alpine valley, Valle Maira, whose entrance is approximately 100 km south of Turin. Acceglio has been experiencing considerable population diminution, from 911 residents in 1951 to 482 in 1968. This reduction in size is

typical of these Alpine valleys where the young adults migrate to industrial centers such as Turin and Milan for employment.

Bellino was a village of approximately 350 residents in 1970 and is located in the adjoining valley north of Valle Maira, called Valle Varaita. Bellino is a *commune* consisting of ten "fraziones." Lasker and his colleagues from the University of Turin described the degree of reproductive isolation and demography of two of the communities, Bellino and Casteldelfino. The mortality data used in our analysis of Bellino were obtained from Luchetti, et al. (1977).

#### METHODS

A time series model is constructed in order to specify the factors which contribute to movements in the series and to analyze the manner in which these factors interact in influencing the direction and magnitude of the series. The choice of a time series model depends largely upon the investigator's understanding of the phenomenon under study and the objectives of the analysis. In most analyses a general time series is considered to be a mixture of four components—trend, cyclic, seasonality, and irregularity. These four components are variously combined in the construction of multiplicative, additive or mixed multiplicative-additive models. In this study, the primary analytic procedures include a combination of regression and time series analyses. The time series analytical model chosen may be referred to as a sinusoidal model, although this model is not mutually exclusive with the aforementioned models in its multiplicative and additive properties. It may also be assumed that the Mennonite, Acceglio and Bellino sequences were generated by stochastic processes. A probabilistic model, which treats possible stochastic processes, has been adopted in a separate paper (Lin and Crawford, n.d., in preparation).

In time series, trend and cyclicity are relative terms, which depend upon arbitrarily defined unit lengths for the analysis of a series. Each of our three time series is considered to be a mixture of: (a) a trend, or long-run fluctuations; (b) fluctuations about the trend with certain regularity or cyclical pattern; (c) a yearly component; and (d) an irregular or random effect (noise). We attempt to decompose the observed mortality fluctuations into (a) and (b), which constitute the underlying signal or meaningful pattern of variation, upon which a noise or random variation (d), and possibly (c), are superimposed. Patterns of seasonal or monthly fluctua-

tions are subsumed in (c) and are therefore not recoverable in the present study.

It is to be assumed further that a trend is due to relatively permanent forces operating in the same direction, while shorter-term fluctuations about the trend are due to a different set of causes. Given this assumption, residuals, that is, the deviations from the trend, are interpreted as disturbances attributable to the combination of (b), (c), and (d). Customarily, noise in a data set is "filtered" out by smoothing the series. In an exploratory analysis, a number of filters, or estimate equations, were applied to the three sequences. These included three-point and five-point moving averages, Sheppard's five-term equation, Spencer's 21-term equation, and Sheppard-Spencer hybrid seven-term equation. These smoothing operations noticeably increased autocorrelation and cross-correlation. However, these improvements were gained at the cost of reduction of the variances in the original sequences. For this reason, the original data, rather than filtered data, are used for the analysis. Three extreme mortality peaks were observed, 32 in the Mennonite sequence in the year 1950, 37 in the Acceglia and 87 in the Bellino in the year 1918 (Figure 1). These outliers complicate our time series procedures and necessitate some adjustments. These outlying values are substituted by the values of their nearest neighbors (31, 25, and 30, respectively), thus

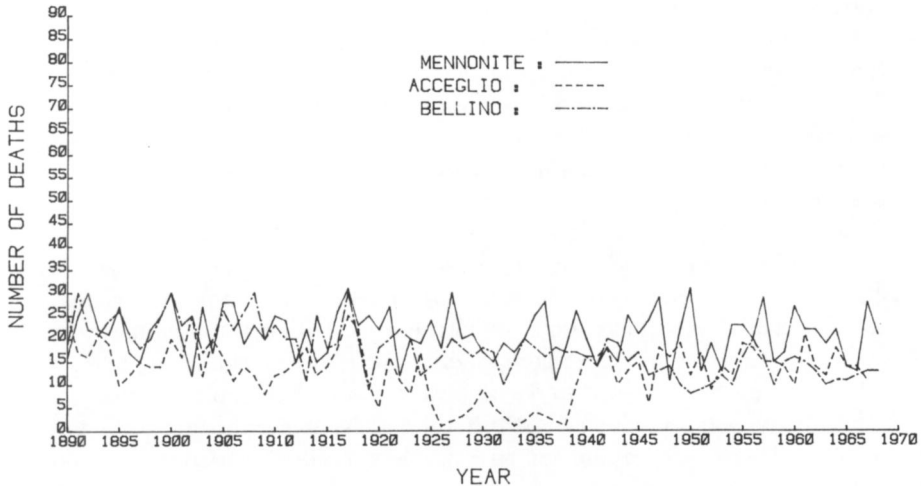


FIG. 1. Raw mortality figures from 1890 to 1967 are plotted for three populations.

minimizing the undesirable effects of these extraneous values in a manner similar to winsorization (Dixon 1960; Dixon and Tukey 1968).

### *Trend*

The essential concept of trend, as described by Kendall (1973, p. 29), requires smoothness and representation by a continuous and differentiable function of the time. The smooth component, commonly referred to as the secular trend, represents the general long-term sweep of the time series data. If the characteristics of the stochastic process remains invariant with respect to time, the process is said to be stationary. Most time series models require a stationary trend, that is, a trend which has no significant linear trend. However, in real-world situations, "probably very few of the time series one meets in practice are stationary" (Pindyck and Rubinfeld 1976, p. 502). Our initial analysis determines whether a trend is stationary or nonstationary (or evolutionary) in order to remove a nonstationary trend if it exists. A statistical procedure employed here consists of a test of a null hypothesis that the slope of the regression on time is zero, or almost zero, with the  $t$  statistic (Davis 1973, pp. 218-219):

$$t = \frac{b_1}{\sqrt{\frac{MS_D}{SS_X}}} \quad (1)$$

where  $MS_D$  is the variation about the regression  $Y_i = b_0 + b_1X_i$  and  $SS_X$  is found by

$$SS_X = \sum_{i=1}^n X_i^2 - \frac{\left(\sum_{i=1}^n X_i\right)^2}{n} \quad (2)$$

For the Mennonite series the null hypothesis  $H_0: \beta_1 = 0$  was tested against its alternative  $H_1: \beta_1 \neq 0$ . In this two-tailed test, either significant positive or negative slopes lead to rejection. The regression equation attained by routine least-squares-fitting POLYNM (Lin, n.d.) is

$$Y = 22.3310 - 0.0311X$$

and  $t = -1.20243$ . There are 79 observations in the Mennonite series, thus, there are 77 degrees of freedom associated with  $SS_D$  ( $n - 2$ ). At a significance level of 5%, the critical value of  $t$  with  $\nu = 75$  degrees of

freedom, by harmonic interpolation, is 1.980. The value of  $t$  at a significance level of 20% is 1.289; thus, the test statistic value,  $-1.202$ , lies outside the critical region. Therefore, we may not reject the null hypothesis that the slope of the regression is zero, or nearly zero. Figure 2.1 clearly indicates that the fitted values for the trend are essentially parallel to the X axis, albeit slightly sloping downward.

The trend for Bellino, in Figure 2.3, appears to be linear and slopes steeply. The slope is not a critical issue here because of the use of residuals in the analysis. It may be assumed that a nonlinear curve might produce a better fit. Therefore, POLYNM is performed on the Bellino series stepwise. The  $F$  value for the first-degree regression equation is 10.339. With 1 and 77 degrees of freedom, this  $F$  statistic is significant at better than the 0.001 level, indicating that a higher degree polynomial may be appropriate. Polynomial regression of the second-degree equation is then fitted to the Bellino series. To test the null hypothesis that the  $p^{\text{th}}$  order term contributes nothing, the  $F$  test, given below, is applied.

$$F = \frac{(R^2 \text{ with } p^{\text{th}}\text{-order term}) - (R^2 \text{ without } p^{\text{th}}\text{-order term})}{\frac{(1 - R^2 \text{ with } p^{\text{th}}\text{-order term})}{(N - p - 1)}} \quad (3)$$

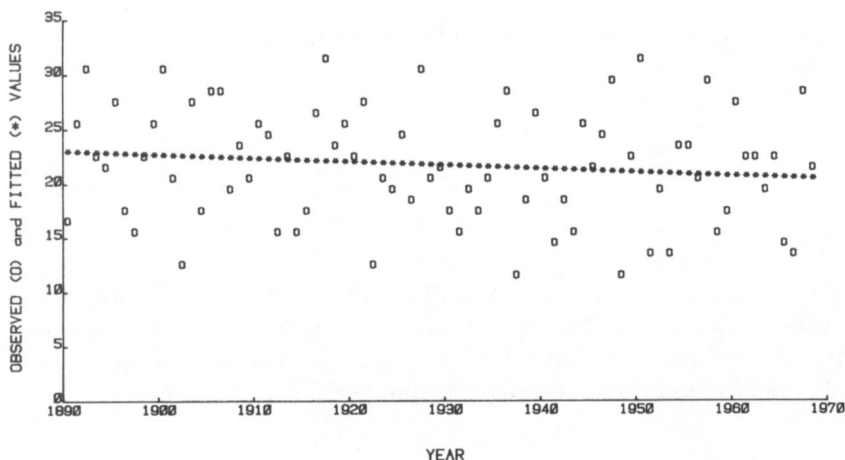


FIG. 2.1. A plot comparing the observed and fitted values for the Mennonite population. O represents the observed data points, \* represents the fitted line.

where  $R^2$  is the square of the multiple correlation, that is, the proportion of variances of  $Y$  explained (Kim and Kohout 1975, p. 372). By substituting the values from second-degree POLYNM we obtain

$$F = \frac{0.53573 - 0.53055}{\frac{1 - 0.53573}{79 - 2 - 1}} = 0.8479$$

With 1 and 76 ( $N - p - 1$ ) degrees of freedom, the  $F$  values, by harmonic interpolation, are 3.966, 1.346, and 0.460 at significance levels of 5%, 25%, and 50%, respectively. These  $F$  values indicate that the null hypothesis cannot be rejected and that higher polynomials are not significant.

The Acceglio series presents a considerable difficulty in regression because of possibly flawed records from 1926 to 1936 wherein the mortality frequencies for the 14-year period were markedly low as contrasted with the other periods. In order to attain a reasonably good fit, it is necessary to raise the polynomial to an unusually high ninth degree. The result of ninth-degree POLYNM is graphically presented in Figure 2.2.

In summary, the plots in Figure 2 facilitate the detection of the presence of long-term trends and shorter-term fluctuations. Least squares regressions were employed to measure the trend components. The Men-

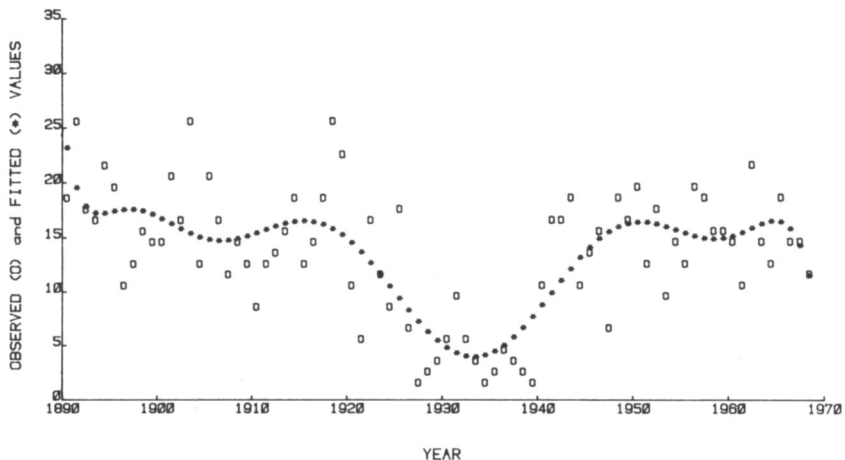


FIG. 2.2. A plot comparing the observed and fitted values for the Acceglio population.

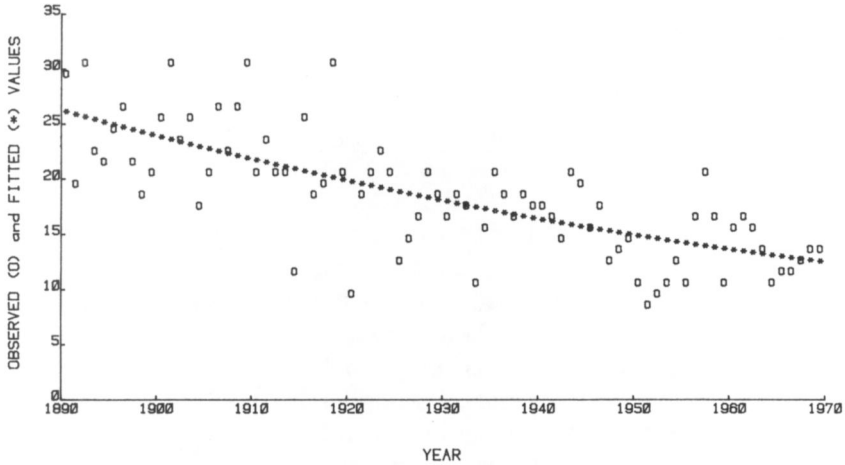


FIG. 2.3. A plot comparing the observed and fitted values for the Bellino population.

nonite time series was decreasing by an approximately constant amount, 0.0311, each period, the straight line or arithmetic trend being represented by  $Y = 22.3310 - 0.0311X$ . Similarly, the Bellino series was decreasing by a much larger constant, 0.2392, the nearly straight line being represented by  $Y = 25.6558 - 0.2392X_1 + 0.0008X_2$ . As to the Acceglio series, its oscillating trend may be better inspected graphically.

### *Residuals*

The secular long-term mortality trends demonstrated above lead to a question of identifying causative factors which contributed to definite changes in the amount of change per period. Obviously, the Mennonite population on one hand and the Acceglio and Bellino populations on the other would not have been subjected to an identical set of causative factors. Nevertheless, there would be at least two factors common to the three populations which contributed to the general decline in mortality over the 80-year period: (1) changes in environments, the foremost being the positive intervention of modern medicine, especially antibiotics, and the improvements in hygiene; and (2) generally decreasing population sizes due to out-migration or fission (or budding). In Acceglio and Bellino, emigration of adults to the urban centers for employment was the primary factor responsible for the decrease in population size. The Mennonite community experienced two fissions during the last century based upon theological disagreements and factionalism.



Our primary interest lies in the relatively shorter, yet possibly systematic, mortality oscillations. These short-term fluctuations, if properly partitioned out, are the components unaffected by a secular trend and are informative as to the pattern or cyclicity of mortality fluctuations. We will attempt to elicit the pertinent information on the short-term mortality fluctuations in the form of residuals.

The mortality frequency variations each year were most likely due to changes in the environment facing the individuals. Unless additional knowledge is available, we may assume that for each observation  $X$  (year), possible observations on  $Y$  (mortality) would differ in a random fashion. This situation may be formally described by adding a random error component to the simplest regression model

$$Y_t = \beta_0 + \beta_1 X_t + e_t \quad (4)$$

where for each observation,  $Y$  is a random variable,  $X$  is fixed at a one-year interval or nonstochastic, and  $e$  is a random error term. Appendix I describes the methods employed in this analysis.

A random disturbance term (hereafter referred to as *residuals*) is a deviation of an observed value,  $Y$ , from a fitted value,  $Y'$ :  $e_t = Y_t - Y'_t$ . Since the residuals are deviations from a fitted line they are independent of the trend, provided that a regression equation applied for curve fitting

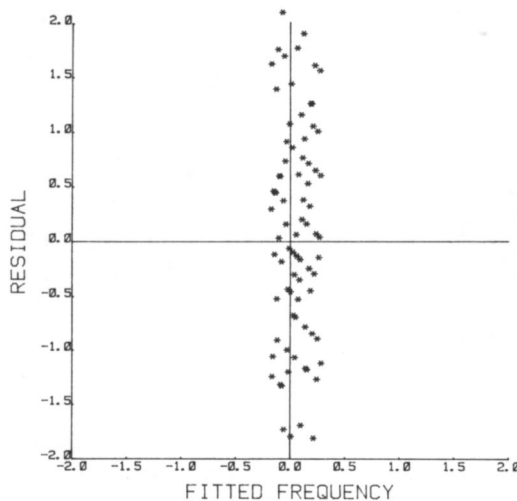


FIG. 3.1. A plot of standardized residuals against standardized fitted values for the Menonite community.

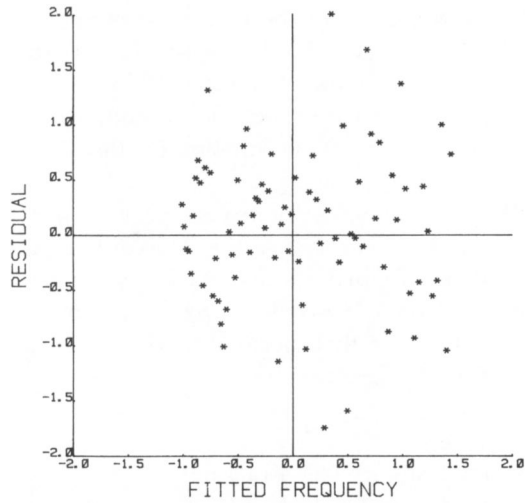


FIG. 3.2. A plot of standardized residuals against standardized fitted values for the Acceglio community.

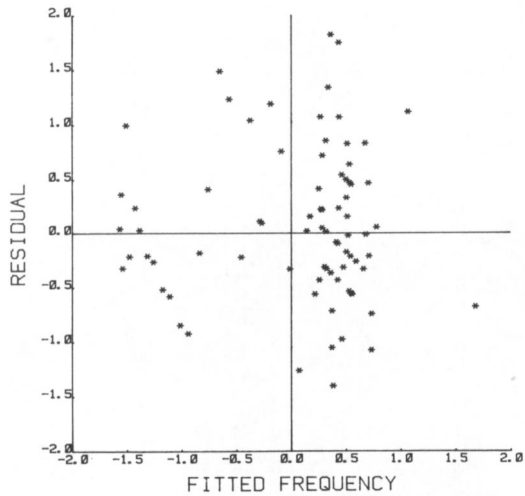


FIG. 3.3. A plot of standardized residuals against standardized fitted values for the Bellino community.

is appropriate. The Mennonite, Acceglio and Bellino series may be considered as appropriately detrended by first-degree, ninth-degree and second-degree polynomial regression, respectively, as shown in the normal probability distributions of the residuals (Figure 4) to be discussed shortly. The residuals (Appendix II thus obtained will be used in Fourier series.

In Figure 3, the standardized residuals are plotted against the standardized predicted values. Since in the scatterplots the residuals are plotted against  $Y'$  values, which also represent a time dimension along the X axis, a direct examination of these residuals permits the identification of a visible pattern.

The patterns of the standardized residuals (Figure 3) indicate the following conditions: (1) there is no need for adding terms to the polynomial regression equations; (2) the generally uniform band patterns suggest homoscedasticity around the regression lines, that is, relative freedom from abnormalities; and (3) outliers or deviant cases are absent, that is, all residuals fall within the  $-2$  to  $+2$  standard deviation unit range, due to the winsorization of the original sequences. The linear regression of the unstandardized residuals on the fitted values yielded zero slopes in

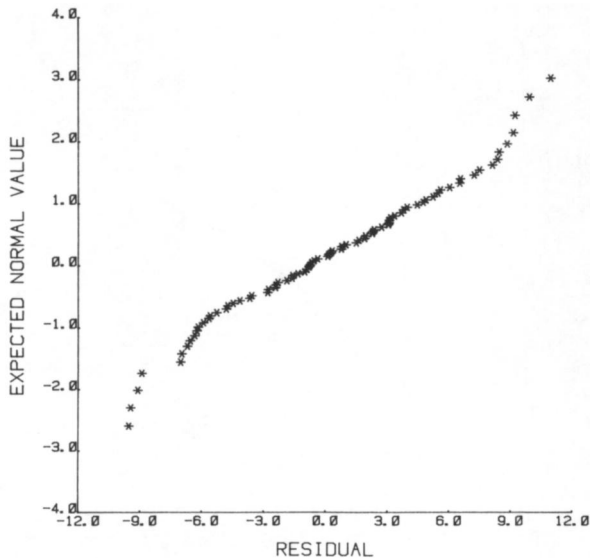


FIG. 4.1. Normal probability values of the residuals plotted for the Mennonite population.

all three populations, indicating that the residuals are free from the secular trend.

To test normality of the residuals, a graphic approach is employed (Figure 4). First, the ordinate of a normal distribution curve is computed in terms of the expected density of  $z$  of observations at any given value of the variable  $y$  (treated as if continuous) by the equation

$$z = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2} \quad (5)$$

where  $\pi = 3.14159$  and  $e = 2.71828$ . The  $z$  values are then transformed to a cumulative form by summing them progressively over the successive intervals, thus producing a sigmoid curve. Since departures from normality are not readily recognizable in a sigmoid form, the cumulative distribution is replotted so that the data points define a straight line if the underlying distribution is normal (Dixon and Massey 1969, pp. 56-66). Instead of using the normal equivalent deviate (NED) (Bliss 1967, pp. 101-107; Finney 1971, pp. 22-25), a residual is plotted in terms of probits in units of  $\sigma$  as an equally spaced scale for a  $y$  ordinate at  $x/\sigma$  (expected

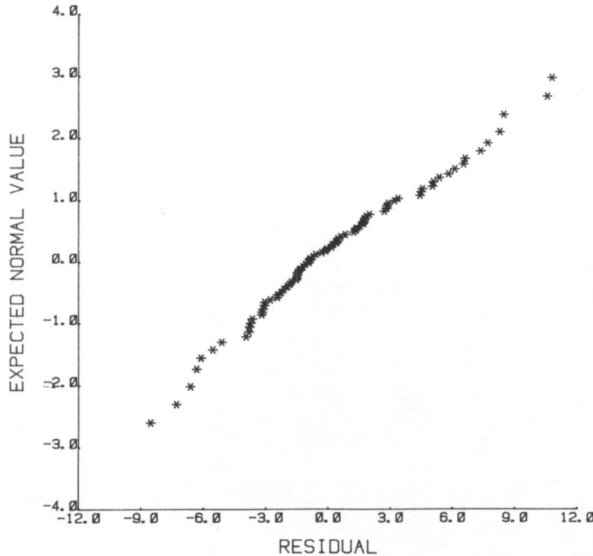


FIG. 4.2. Normal probability values of the residuals plotted for the Acceglio population.

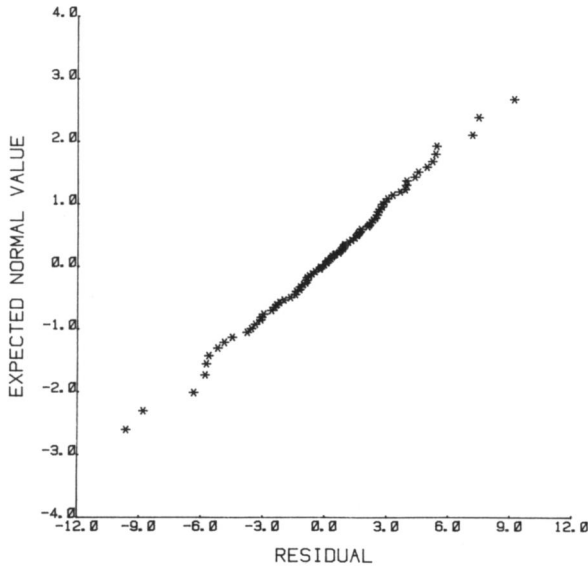


FIG. 4.3. Normal probability values of the residuals plotted for the Bellino population.

normal value) against the corresponding point  $x$  (residual) on the abscissa. Because the mean of the residuals is zero, the plot is dissected into quadrants.

The test of normality with the cumulative probit diagram thus obtained with routine PROBIT (Lin n.d.) is to fit the plotted data points by inspection with a straight line if they are normal. Except for a few percentages at each end of the three sequences, the distributions of the residuals (Figure 4) are satisfactorily normal without major deviations from an overall linear trend (see Appendix III).

#### *Autocorrelation*

Autocorrelation, or serial correlation, occurs when the errors associated with observations in a given time period carry over into the following time period. Ideally, the presence of such autocorrelation is not expected if the regression model is appropriate. However, the assumption that errors corresponding to different observations are uncorrelated, while being an important assumption, often breaks down in time series studies (Pindyck and Rubinfeld, 1976, p. 152).

In Figure 2, the observations are not randomly scattered around the regression line. Instead, these deviations, or residuals, tend to occur on

the same side of the regression line, indicating certain degrees of autocorrelation. However, it is these autocorrelations that may indicate biologically significant phenomena in the mortality trend. It may be inferred that the oscillating serial correlations reveal the effects of mortality causative factors operating in one period and carrying over into the subsequent periods. The residuals are input into the routine AUTOOCR (Davis, 1973, pp. 232-239) by the equation

$$r_L = \frac{[(n-L)(\sum Y_i Y_{i+L}) - \sum Y_i \sum Y_{i+L}]/(n-L)(n-L-1)}{[n\sum_{i=1}^n Y_i^2 - (\sum_{i=1}^n Y_i)^2]/n(n-1)} \quad (6)$$

$$= \frac{\text{cov}(Y_i, Y_{i+L})}{s_Y^2}$$

to compute autocorrelation function,  $r_L$ , defined as the linear correlation between a time series and the same series at a later interval of time, that is, at lag  $L$ . In this routine, the maximum amount of offset between the two series is set not to exceed  $n/4$ .

The linear correlation ( $r_L$ ) of the first 19 and 20 lag intervals of the whole series of 79 observations (Mennonite and Acceglio) and 80 observations (Bellino) are plotted in Figure 5. As in regression, the correlograms indicate that the observation for a current year is somewhat correlated with the observations of the previous or following years. However, the autocorrelation function of this kind can arise because of carry-over effects (Box and Jenkins 1970, pp. 33-34). It suggests that the causative factors which effected particular realizations in the form of residuals were not entirely dissipated from that year. Statistical tests of the hypothesis that an autocorrelation,  $r_L$ , is zero are not discussed in detail here. It would suffice to report the statistics computed by BMDP2T Box-Jenkins Time Series Analysis (Dixon et al. 1981) as follows: (1) the  $t$  values of the means (0.1168 for the Mennonite; 0.1223 for the Acceglio; 0.4066 for the Bellino) are not significant; (2) the standard errors of the autocorrelations range from 0.12 to 0.17 at all lags; and (3) approximately 94% of the autocorrelation values fall within their 95% confidence intervals, indicating again the adequacy of the regression models employed and detrending of the original data. A few comments are to be made on the basis of inspection of the autocorrelograms.

To begin with, it is to be noted that at the first one or two lags the autocorrelation function in all three sequences declines rapidly, suggesting a general stationarity. More specifically, in the Mennonite sequence,

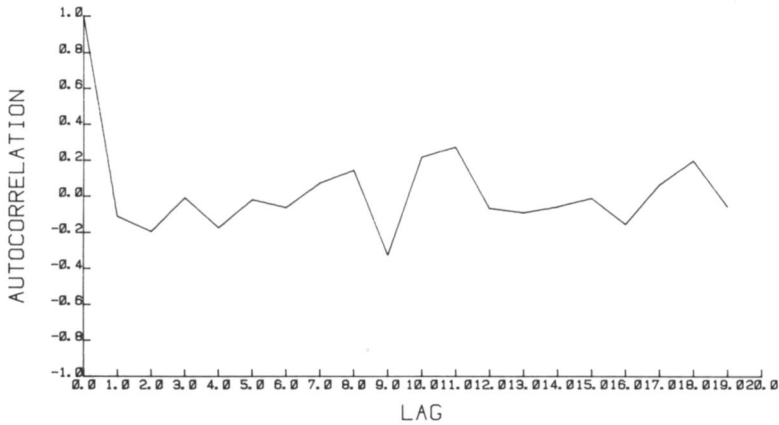


FIG. 5.1. This is a plot of the autocorrelation values, 1.0 to  $-1.0$ , against the lag in the mortality time series for the Mennonites.

the correlation coefficient drops rapidly from 1.0 to  $-0.1957$  at lag 2, drifts upward to 0.1435 at lag 8, drops to  $-0.3280$  at the next lag, and reaches the highest peak of 0.2728 at lag 11. At lag 12 the signal appears to move out of phase itself. The two zero crossings suggest repetitions of similar patterns of residual fluctuations in these segments. No such more or less distinct peaks are present in the Acceglio and Bellino sequences.

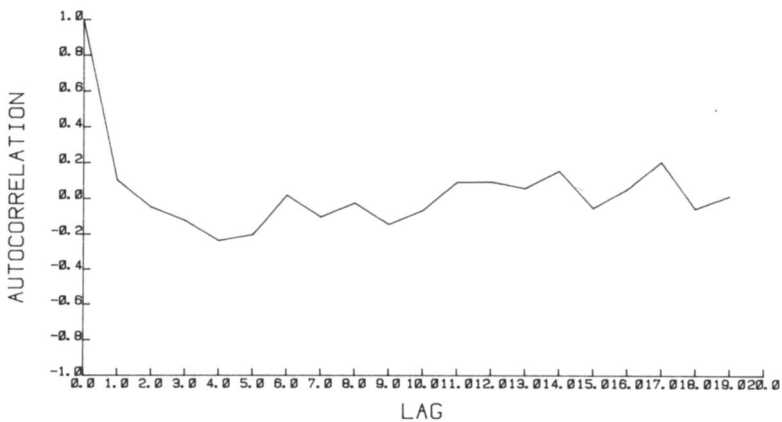


FIG. 5.2. This is a plot of the autocorrelation values, 1.0 to  $-1.0$ , against the lag in the mortality time series for Acceglio.

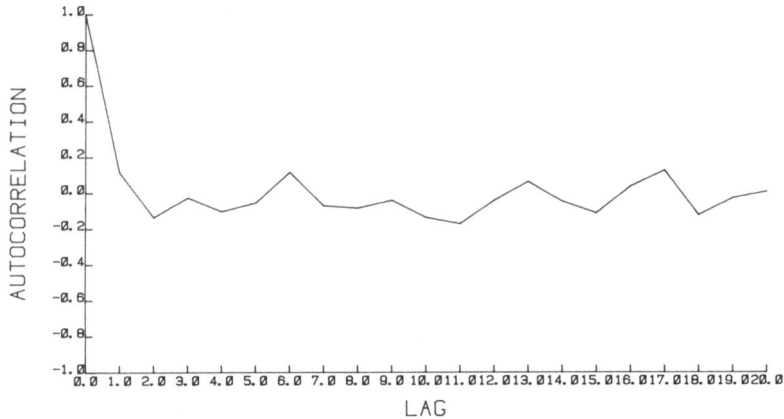


FIG. 5.3. This is a plot of the autocorrelation values, 1.0 to  $-1.0$ , against the lag in the mortality time series for Bellino.

Nevertheless, weak peaks are discernable at lag 14 in the Acceglio and lag 6 in the Bellino sequences. These peaks and troughs observed in a plot of the residual series, with itself lagged in time, are local variations in the series suggesting the periodic components of our time series. This periodic component may be interpreted as representing mortality cycles. Jenkins and Watts (1968, pp. 5-6) caution that “the autocorrelation function is sometimes difficult to interpret because neighboring values can be highly correlated” and consider it as an intermediate step in the estimate of the spectrum. In the following, Fourier series analysis will be employed in the hope that it may better reveal the presence of periodicities, if any, in the data.

#### *Fourier Analysis*

An underlying assumption in this analysis is the existence of a regular, or more or less regular, periodic variation imposed in our time series. In the previous sections a trend, that is, a drift in the average value of the signal, was detected by fitting an appropriate regression. The detrending operation converted the original series containing a trend and a periodic component to the deviations (residuals) from the fitted line. The new time series thus obtained is interpreted as consisting of periodic components (signal) and a random component (noise). In a harmonic model a periodic trend is interpreted as a trigonometric trend. In this section the dominant



periodic components embodied in the new time series by Fourier analysis are extracted and described.

Each of the three new series is a signal specified at discrete, equally spaced times and is expanded in terms of periodic functions. The discrete signal is derived from a continuous signal of duration or length,  $L$  (equal to 79 years for the Mennonite and Acceglio; 80 years for Bellino), by sampling the values of the signal (mortality frequencies) at a spacing of one year, thus a total of 79 and 80 sample values or periodic functions are generated. A periodic function,  $Y_i$ , at a point  $S_i$ , is composed of a sum of the amplitudes of the component sine and cosine waves at a distance  $X_i$  from the origin of the series, whose frequencies are multiples (harmonics) of the first harmonic or fundamental frequency. This function is represented by (Davis 1973, pp. 261-265)

$$Y_i = \sum_{n=1}^{\infty} \left( \alpha_n \cos \frac{2n\pi X_i}{\lambda} + \beta_n \sin \frac{2n\pi X_i}{\lambda} \right) \quad (7)$$

over the limited range  $-\pi$  to  $\pi$ , with the whole sum having a period of  $2\pi$ . The coefficients  $\alpha_n$  and  $\beta_n$  are expressed as the variance (power or intensity)  $s_n^2$  for all  $n^{\text{th}}$  harmonic or frequency component by

$$s_n^2 = \alpha_n^2 + \beta_n^2 \quad (8)$$

Routine FOURER (Davis 1973, pp. 266-269) computes the coefficients and the powers for all harmonics up to the limit  $N/2$ , the Nyquist frequency. The successive values of  $s_n^2$  are plotted against the harmonic number  $n$  (or  $i/\lambda$ ) in the power spectrum, sometimes referred to as a spectrogram. For a possibly better impression of the true periodic structure of the time series, the spectrum is smoothed by the Hanning filter

$$S_n^2 = \frac{S_{n-1}^2}{4} + \frac{S_n^2}{2} + \frac{S_{n+1}^2}{4} \quad (9)$$

and is displayed in Figure 6.

The intensity,  $s_n^2$ , may be considered as a measure of the closeness of fit between the chosen trigonometric function with frequency  $i/\lambda$  and the observed data (Anderson, 1971, p. 109). The observed peaks in the spectrum are expected to determine the harmonic components, although its interpretation may not be straightforward (Kendall 1976:98). An interpretation of the power spectrum is made within the theoretical frame-

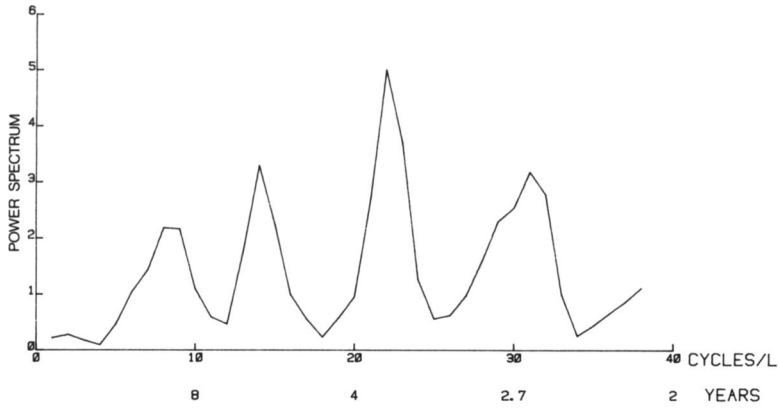


FIG. 6.1. Smoothed power spectrum plot of the Mennonite mortality time series.

work of power spectral analysis, which measures the relative contribution of each harmonic or frequency to the variability or oscillatory nature of the cyclical fluctuations.

### RESULTS AND DISCUSSION

Power spectrum analysis reveals the existence of periodicity in this mortality time series. The plots of the time series in Figures 2 and 5 are suggestive of periodic components. The highest peak in the Mennonite spectrum (Figure 6.1) occurs at a wavelength of about 3.6 years, followed

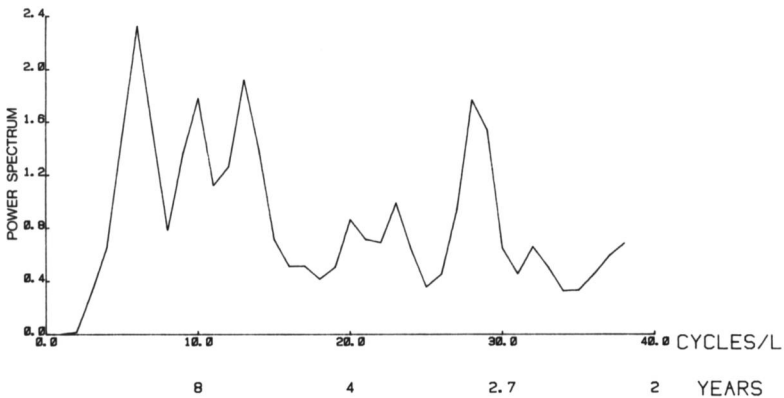


FIG. 6.2. Smoothed power spectrum plot of the Accegljo mortality time series.

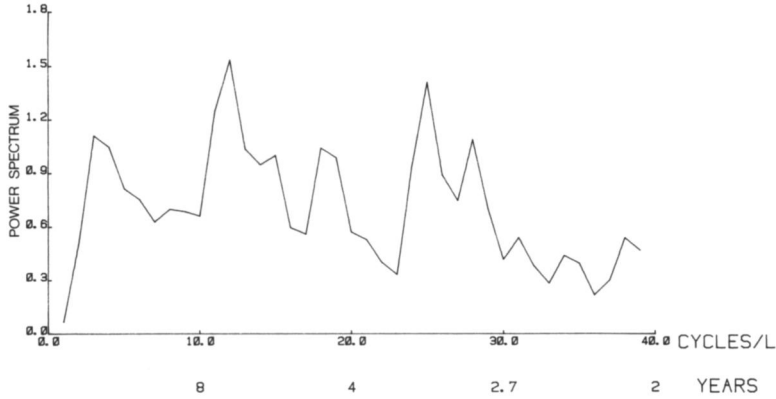


FIG. 6.3. Smoothed power spectrum plot of the Bellino mortality time series.

by a second, lower one at 5.6 years. The third peak occurs at 2.6 years while the fourth is at 10 years. The cyclical components in the Acceglio and Bellino series are examined in the same manner. Information concerning the cyclical components of mortality are extracted from autocorrelation and power spectrum and is summarized in Table 1. The power spectrum cycles are ranked in descending order of wavelengths.

There is a broad concordance between the autocorrelation and the power spectrum in the detection of cyclical components in that 10.5 approximates 9.9 years in the Mennonite series, 14.0 approximates 13.2 years in the Acceglio, and 6.0 is almost equivalent to 6.7 years in the Bellino. The 26.7-year cycle in the Bellino series appears to be a multiple

Table 1

*Mortality Periodicity Interpreted from Autocorrelation and Power Spectrum Analyses*

Cycles	Populations		
	Mennonite	Acceglio	Bellino
Autocorrelation	10.5	14.0	6.0
Power Spectrum Peaks	1	9.9	26.7
	2	5.6	6.7
	3	3.6	4.4
	4	2.6	3.2

of its half wavelength of 13.4 years, comparable to the 13.2-year cycle in the Acceglio series.

In summary, embedded in the Mennonite mortality series are 10-year cycles with 5.6-year subcycles. Similarly, the Acceglio series is characterized by a 13.2-year cycle with 7.9-year subcycles, and the Bellino series contains a 13.4-year cycle with 6.7-year subcycles. The shortest cycles, 2.6 years, 2.8 years, and 3.2 years in the Mennonite, Acceglio and Bellino series, respectively, are the highly localized variations reflecting the serial correlations between the adjacent data points. We propose to term the major periodic components, that is, the 10-year cycle in the Mennonite series and the 13-year cycle in the Acceglio and Bellino series, as the "mortality cycles."

#### CROSSCORRELATION

Following the demonstration of the existence of the cyclicities in the three series, the cycles are then compared. In autocorrelation a time series is compared with itself. This same technique, used to compare one series with another, is crosscorrelation. The equation for crosscorrelation, which is the same as the ordinary linear correlation but is somewhat different from the autocorrelation, is (Davis, 1973, p. 243)

$$r_m = \frac{n \cdot \sum Y_1 Y_2 - \sum Y_1 \sum Y_2}{\sqrt{[n \cdot \sum Y_1^2 - (\sum Y_1)^2] [n \cdot \sum Y_2^2 - (\sum Y_2)^2]}} \quad (10)$$

in which the two series are designated as  $Y_{1i}$  and  $Y_{2i}$  and  $n$  is defined as the number of overlapped positions between the two chains. The equation (10) is equivalent to

$$r_m = \frac{\text{COV}_{1,2}}{s_1 \cdot s_2} \quad (11)$$

where  $\text{COV}_{1,2}$  is the covariance of the overlapped portions of the two sequences, and  $s_1$  and  $s_2$  are the standard deviations of the overlapped segments of the two sequences. The significance of the crosscorrelation coefficient,  $r_m$ , is assessed by the approximate test of the null hypothesis  $H_0: r_m = 0$

$$t = r_m \sqrt{\frac{n^* - 2}{1 - r_m^2}} \quad (12)$$

which has  $(n^* - 2)$  degrees of freedom.

The original raw data, not the winsorized residual series, are input to routine CROSCR (Davis, 1973, pp. 245-246). The plots of Mennonite-Acceglio, Mennonite-Bellino and Acceglio-Bellino crosscorrelations and their partial statistical data are presented in Figure 7 and Table 2.

While the crosscorrelograms (Figure 7) help reveal the characteristics of the time series, they must be interpreted with great caution. The peaks of seemingly high crosscorrelations on both ends of a correlogram result from an increase in the variance of a crosscorrelation because fewer observations are used toward the beginning and the end of "sliding" the two chains past each other. Therefore, little importance should be attached to those high crosscorrelations since a small number of observations are being matched. The most significant crosscorrelations are those based upon a comparison of the entire lengths, or nearly the entire lengths, of both series, lagged from one to a few years in each direction.

With 77  $(79 - 2)$  degrees of freedom, the  $t$  value of 2.306 on the Mennonite-Bellino crosscorrelation at match position 77 is above the 5% value ( $t = 1.991$  by harmonic interpolation). The high crosscorrelation between the Acceglio and Bellino time series at match position 79, given a  $t$  value of 4.235 with 77 degrees of freedom ( $t = 3.422$ ), is to be expected

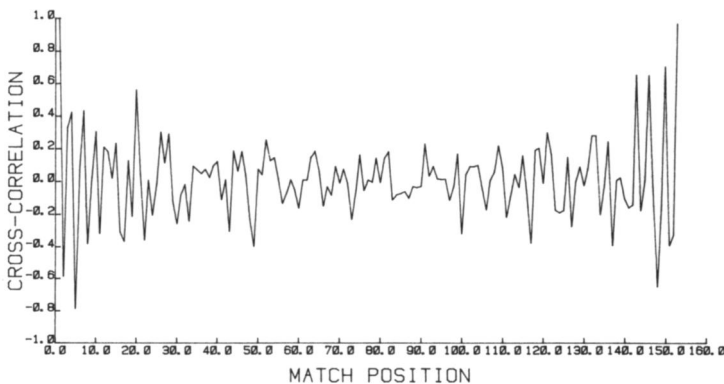


FIG. 7.1. Cross-correlogram matching the mortality time series between Mennonites and Acceglio.

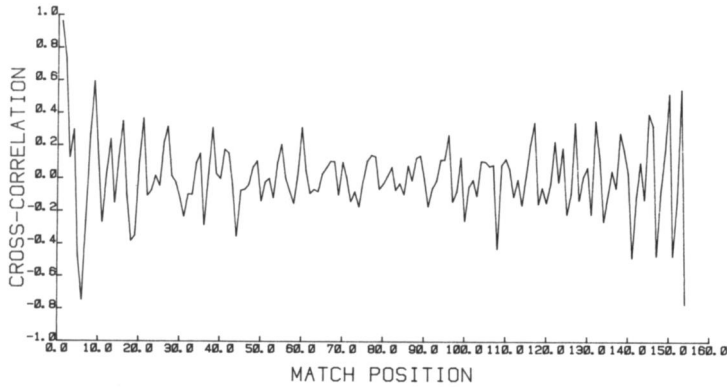


FIG. 7.2. Cross-correlogram matching the mortality time series between Mennonite and Bellino.

in view of their geographical proximity and ecological similarity. The Mennonite-Aceglgio crosscorrelation is significant at the 10% level with a  $t$  value of 1.770 with 75 degrees of freedom ( $t = 1.992, 1.666$  at the 5 and 10% levels, respectively). In summary, each pair of the three time series is significantly, or nearly significantly, crosscorrelated.

In some disciplines, the data utilized deal with multidimensional patterns of observations. However, the mortality fluctuations used in this study represent a unidimensional phenomena observed through time. The analysis of this phenomenon consists, in part, of fitting a theoretical distribution to the observations. The analyses described in this paper

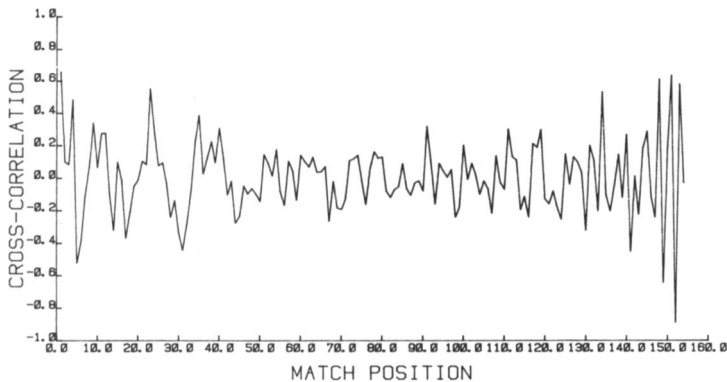


FIG. 7.3. Cross-correlogram matching the time series between Aceglgio and Bellino.

Table 2

*A Comparison of Cross Correlations Between Mennonite, Acceglio and Bellino Time Series*

Match Position	Terms Being Matched		Number of Terms Matched	Correlation Coefficient	t Value
	(1)	(2)			
Mennonite (1)—Acceglio (2)					
75	1 — 77	3 — 79	77	0.20734	1.836
76	1 — 78	2 — 79	78	0.06948	0.607
77	1 — 79	1 — 79	79	0.08341	0.734
78	2 — 79	1 — 78	78	0.11554	1.014
79	3 — 79	1 — 77	77	0.20027	1.770
80	4 — 79	1 — 76	76	0.10345	0.895
81	5 — 79	1 — 75	75	0.13326	1.149
82	6 — 79	1 — 74	74	0.19602	1.696
Mennonite (1)—Bellino (2)					
76	1 — 78	3 — 80	78	0.16776	1.483
77	1 — 79	2 — 80	79	0.25415	2.306*
78	1 — 79	1 — 79	79	0.13443	1.190
79	2 — 79	1 — 78	78	0.09960	0.873
80	3 — 79	1 — 77	77	0.06180	0.536
81	4 — 79	1 — 76	76	0.14195	1.234
Acceglio (1)—Bellino (2)					
77	1 — 79	2 — 80	79	0.14144	1.254
78	1 — 79	1 — 79	79	0.43461	4.235**
79	2 — 79	1 — 78	78	0.21673	1.935

constitute an initial step of the investigation of the relationship between mortality and ecological factors. We have attempted to delineate the observed mortality fluctuations with regression and sinusoidal models with the assumption that the mortality fluctuations are composed of measurable components. Component extraction consists of curve fitting by linear or polynomial regression methods so that the data are decomposed into (1) an underlying "trend," (2) local variations or cyclical components, and (3) a superimposed "noise," that is, error or random component.

The long-term secular trends, detected in all three series by curve fitting, show a general decline in mortality frequencies, which is primarily a function of corresponding decrease in population size. The original

series are detrended in order to minimize, if not eliminate, the effects of this exogenous variable (population size). In search of this periodicity, autocorrelation and spectral analysis are performed on the residuals, obtained from the detrending. The results indicate that each series is characterized by a number of cycles and subcycles. We have earlier proposed to call the major cyclical component the "mortality" cycle. The mortality cycle is approximately 10 years in the Mennonite series and slightly over 13 years in the Acceglio and Bellino series. The discrepancy of some three years is attributed largely to the difference in complexity of population structure and ecological conditions.

Despite the variation in periodicity, the patterns of the mortality cycles of the three populations are basically similar, as shown by the crosscorrelations. It may be suggested that there are underlying factors commonly linked to the cyclical mortality fluctuations normal to the three populations. We propose to call this as yet undelineated common condition across the Atlantic Ocean "panthanasia," broadly analogous to pandemics.

### CONCLUSIONS

From these data it is apparent that there are similar periodicities of mortality in populations living under different ecological conditions with vastly distinctive patterns of subsistence. Populations that are closer geographically, such as Acceglio and Bellino, exhibit remarkable similarities in mortality patterns with high crosscorrelations at the point of maximum fit between these two time series.

There is evidence to suggest that, with the developments in transportation and communication at the turn of this century, mortality patterns began cycling due to pandemic diseases. However, as the role of infectious diseases in mortality trends began to diminish due to improvements in medical technology, the periodicity of mortality once again became distinctive in human populations. Thus, we have gone full cycle in death periodicity from unique mortality peaks occurring in small isolated communities to similar mortality periodicities due to pandemic disease, back to the unique although less fluctuating patterns now resulting from degenerative diseases.

A number of factors contribute to the duration and size of the mortality periods. Turner (1976) in the study of mortality of the Tlaxcaltecs of Cuanalan noted the existence of a 15 year mortality cycle. The most common source of this mortality was attributed primarily to smallpox,



although influenza predominates in 1916 and 1918. With the exception of influenza, the periodicity was probably due to a build-up of a cohort of children who had not previously been exposed to the pathogen. Thus, the build-up of sufficient numbers of at-risk individuals contributes to the observed periodicity. The degree of geographical isolation of the population plays a role in the introduction of the pathogen into the community. The nature of the disease and whether the older members of the community have been exposed to it contribute to the mortality periodicity. Finally, various climatic fluctuations such as severe droughts and unusually hot summers or cold winters may be contributing factors to mortality fluctuations.

These results suggest that the time series methodology may prove to be a useful tool in comparing mortality patterns on populational and global levels. The examination of individual population life tables and age specific rates was a useful beginning in mortality studies. However, power spectral analysis, crosscorrelation and autocorrelation are powerful methods that permit the formulation of broader questions concerning mortality and evolution.

#### APPENDIX

In the methods section of the text, the regression lines have been fitted by solving the normal equations with the least-squares criterion for the values of  $b_0, b_1, b_2, \dots, b_n$ . The regression model programmed into POLYNM and the assumptions necessary for its appropriate application are more specifically described in the following.

The formal statement of the polynomial expansion regression equation, of which the simple linear regression is a special case, is:

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_t^2 + \dots + \beta_m X_t^m + e_t \quad (1a)$$

$$t = 1, 2, \dots, T$$

where  $Y_t$  = value of the observed dependent (mortality) variable;  $\beta_0$  and  $\beta_m$  = parameters in the model;  $X_t$  = value of the independent (year) variable; and  $e_t$  = random error term, in the  $t^{\text{th}}$  observation (or "trial"). The subscript  $t$  denotes that our three sets of data  $X_t$  ( $t = 1, 2, \dots, T$ ) are time series with the interval between  $X_t$  and  $X_{t+1}$  being fixed and constant at a one-year period. The order in which  $Y_t$ 's occur through time is as important as the particular values they assume. The terms  $\beta_0$  and  $\beta_m$ 's in the model are parameters whose values are unknown and are to be

estimated from the sample data. In the simple linear regression model, the properties of the least square point estimators,  $b_0$  and  $b_1$ , are established by the Gauss-Markov theorem: these estimators are unbiased and have minimum variance among all linear unbiased estimators of  $\beta_0$  and  $\beta_1$ . It should be mentioned that the Gauss-Markov theorem may not be applicable to a non-linear estimator because of a possibility that an estimator may be biased and have a lower mean square error than the unbiased estimator (Pindyck and Rubinfeld, 1976, p. 52). The  $X_t$ 's are known constants (years) in the model. The mean of the conditional probability distribution of  $Y$  given a value of  $X$ , denoted by  $\mu_{Y.X}$  is equal to

$$\beta_0 + \beta_1 X_t + \beta_2 X_t^2 + \dots + \beta_m X_t^m \quad (2a)$$

when the expected value of the error component  $e_t$  is zero [ $E(e_t) = 0$ ] for the  $t^{\text{th}}$  observation. Thus, the observed value of  $Y$  in the  $t^{\text{th}}$  observation is larger or smaller than  $\mu_{Y.X}$  by the amount  $e_t$ , that is, the value of the error component in the  $t^{\text{th}}$  observation

$$e_t = Y_t - Y'_t \quad (3a)$$

The polynomial expansion equations are based on the assumption that the observed value of  $Y$  in the  $t^{\text{th}}$  observation,  $Y_t$ , is the sum of two components—a random variable,  $e_t$ , and a constant which comprises all the remaining terms on the right side of the equation (1a) given above.

In passing, the "error model," described by Anderson (1971, pp. 31-32),  $y_t = f(t) + u_t$  (4a), specifies that the observable  $y_t$  is the sum of a trend in time,  $f(t)$ , and an (unobservable) error,  $u_t$ , and assumes that the trend is a polynomial of degree  $q$

$$f(t) = a_0 + a_1 t + \dots + a_q t^q \quad (4a)$$

In this error model, a trend is fitted "in order to obtain a base-line from which to measure the aspects of the series that change in small units of time," that is, short-term cycles, fluctuations, and irregularities. Similarly, in our model, the error component in each event is the deviation between the observed value of mortality frequency,  $Y$ , and  $\mu_{Y.X}$  (the mean of the conditional probability distribution of  $Y$  given a value of  $X$ ).

## APPENDIX II

A number of methods have been proposed for examining residuals (Anscombe 1961; Anscombe and Tukey 1963; Draper and Smith 1966; Larsen and McCleary 1972). We employ graphical techniques of analysis

of residuals. For each case,  $k$ , the standardized observed value of the dependent variable  $X_q$ , is

$$Z_k = \frac{(X_{qk} - \bar{X}_q)}{s_q} \quad (5a)$$

where  $\bar{X}_q$  is the mean, and  $s_q$  the standard deviation. The standardized predicted value is

$$\bar{Z}_k = \sum_{i=1}^p B_i \left( \frac{X_{pk} - \bar{X}_i}{s_i} \right) \quad (6a)$$

where  $X_p$  are independent variables,  $s_i$  their standard deviations, and  $B_i$  their regression coefficients. The standardized residual is

$$s_k = Z_k - \bar{Z}_k. \quad (7a)$$

That is, the standardized residual thus defined is the deviation of the standardized predicted value from the standardized observed value, not *standardized* residual *per se* (Norusis, 1978, p. 41).

### APPENDIX III

As mentioned earlier, our analytical procedures consist of constructing a regression model in which a mortality rate is related to the time variable, after which a time series model is constructed to explain the behavior of the residual terms for the regression. In the frequency interpretation (Anderson, 1971, p. 3), the decomposition means, in part, the repetitions of the entire situation, yielding a new set of observations, the random error terms, as a new realization of the stochastic process,  $u_t$ , in the equation (4a). Such random terms, obtained from the decomposition, are our time series, *stricto sensu*, upon which autocorrelation, Fourier series analysis, and cross-correlation are performed.

### ACKNOWLEDGMENTS

This research was supported in part by NIH grand AGO1646-02 and one of us (MHC) was on a PHS Research Career Development Award No. K04DE00028-05. The authors acknowledge the UCLA Health Sciences

Computing Facility for providing program BMDP2T with the support of the NIH Special Research Resources Grant RR-3.

#### LITERATURE CITED

- ANDERSON, T. W. 1971 *The statistical analysis of time series*. Wiley, New York.
- ANSCOMBE, F. J. 1961 Examination of residuals. *Proceedings of the Sixth Conference on the Design of Experiments in Army Research, Development and Testing* (U.S. Army Research Office, Durham, N.C.), 7-19.
- ANSCOMBE, F. J. AND J. W. TUKEY 1963 The examination and analysis of residuals. *Technometrics*, 5: 141-160.
- BLISS, G. I. 1967 *Statistics in biology: Statistical methods for research in the natural sciences*. Volume One. McGraw-Hill, New York.
- BOX, G. E. P. AND G. M. JENKINS 1970 *Time series analysis: Forecasting and control*. Holden-Day, San Francisco.
- DAVIS, J. C. 1973 *Statistics and data analysis in geology*. Wiley, New York.
- DIXON, W. J. 1960 Simplified estimation from censored normal samples. *Annals of Mathematical Statistics*, 31: 385-391.
- DIXON, W. J. AND F. J. MASSEY 1969 *Introduction to statistical analysis*, 3rd ed. McGraw-Hill, New York.
- DIXON, W. J. AND J. W. TUKEY 1968 Approximate behavior of the distribution of winsorized  $t$  (trimming winsorization 2). *Technometrics*, 10: 83-98.
- DIXON, W. J. ET AL. 1981 BMDP Statistical software. University of California Press.
- DRAPER, N. R. AND H. SMITH 1966 *Applied regression analysis*. Wiley, New York.
- FINNEY, D. J. 1971 *Probit analysis*. Cambridge University Press.
- JENKINS, G. M. AND D. G. WATTS 1968 *Spectral analysis and its applications*. Holden-Day, San Francisco.
- KENDALL, M. 1976 *Time-series*, 2nd ed. Charles Griffin, London.
- KIM, J. O. AND F. J. KOHOUT 1975 Special topics in general linear models. *In*: N. H. Nie et al., *SPSS Statistical package for the social sciences*. McGraw-Hill, New York, 368-397.
- LARSEN, W. A. AND S. J. MCCLEARY 1975 The use of partial residual plots in regression analysis. *Technometrics*, 14: 781-790.
- LIN, P. M. n.d. MSGP: Multivariate statistical and graphic computer programs. Unreleased.
- LIN, P. M. AND M. H. CRAWFORD n.d. Mortality as a stochastic process: A time series analysis with a probabilistic model. Unpublished manuscript.
- LUCCHETTI, E., A. MOTTURA, P. BIGATTI, L. SOLIANI, F. CONTERIO, B. CHIARELLI 1977 Aspects antropodemographiques de la population de Bellino, Val Varaita (CN, Italie). *Bulletin d'Etudes Prehistoriques Alpines*. VIII-IX, Aoste. pp. 159-167.
- NORUSIS, M. J. 1979 *SPSS Statistical algorithms*. SPSS, Inc., Chicago.
- PINDYCK, R. S. AND D. L. RUBINFELD 1981 *Econometric models and economic forecasts*. McGraw-Hill, New York.
- TURNER, K. R. 1976 Computer simulation of transplanted Tlaxcaltecan populations. *In*: M. H. Crawford (ed.) *The Tlaxcaltecs: prehistory, demography, morphology and genetics*. University of Kansas Publications in Anthropology, 7, 45-58.