Time Series Analysis of Mosquito Population Data

Article *in* Journal of Medical Entomology · January 1974

DOI: 10.1093/jmedent/10.6.533 · Source: PubMed

J. Med. Ent. Vol. 10, no. 6: 533–543 30 December 1973

TIME SERIES ANALYSIS OF MOSQUITO POPULATION DATA

By **Carl S. Hacker**[1,2], **David W. Scott**[2] **and James R. Thompson**[2]

Abstract: A statistical technique, time series analysis, has been introduced, allowing the development of objective and quantitative insights into the behavior of relative mosquito densities as a function of time. This method is particularly useful for detecting periodicities in mosquito densities as well as the relationship between meteorological phenomena and mosquito densities.

Following the discovery during the late 19th and early 20th centuries that mosquitoes are an essential link in the life cycle of a number of human parasites, it was appreciated that these parasites might be controlled by either eliminating or reducing the density of their mosquito vectors (Ross 1910). Much effort was subsequently exerted to control mosquito vectors particularly through the reduction or alteration of their larval habitats. The introduction of residual pesticides in the 1940's for the control of a number of insect pests appeared to assure the eventual elimination of human misery resulting from mosquito-borne diseases. However, mosquitoes proved to be quite plastic, and during the 1950's strains resistant to these pesticides arose in a number of field populations (Brown 1958). However, by developing new classes of pesticides, it has still been possible to keep mosquitoes under control.

[1]Population Studies Module, School of Public Health, University of Texas, Health Science Center at Houston, P. O. Box 20186, Houston, Texas 77025, U.S.A.

[2]Department of Mathematical Sciences, William Marsh Rice University, Houston, Texas 77001, U.S.A.

Two difficulties have arisen that complicate the problem of mosquito control: (1) new pesticides are not being developed rapidly enough to counter the rate at which resistant genotypes are appearing in some populations, (2) there is a growing concern over the long-range effects of persistent pesticides on the environment. To counter these problems considerable attention is being directed to other techniques, collectively termed *biological control* or *integrated control.*

Very little knowledge is required about the dynamics of the target population for the successful use of pesticides. Unfortunately, this is not considered to be the case when biological and integrated control techniques are used. Instead, a thorough understanding of the interactions of several factors which affect mosquito densities may be required. To facilitate this understanding it seems reasonable to build a hierarchy of mathematical models of increasing complexity. As each level in the hi-

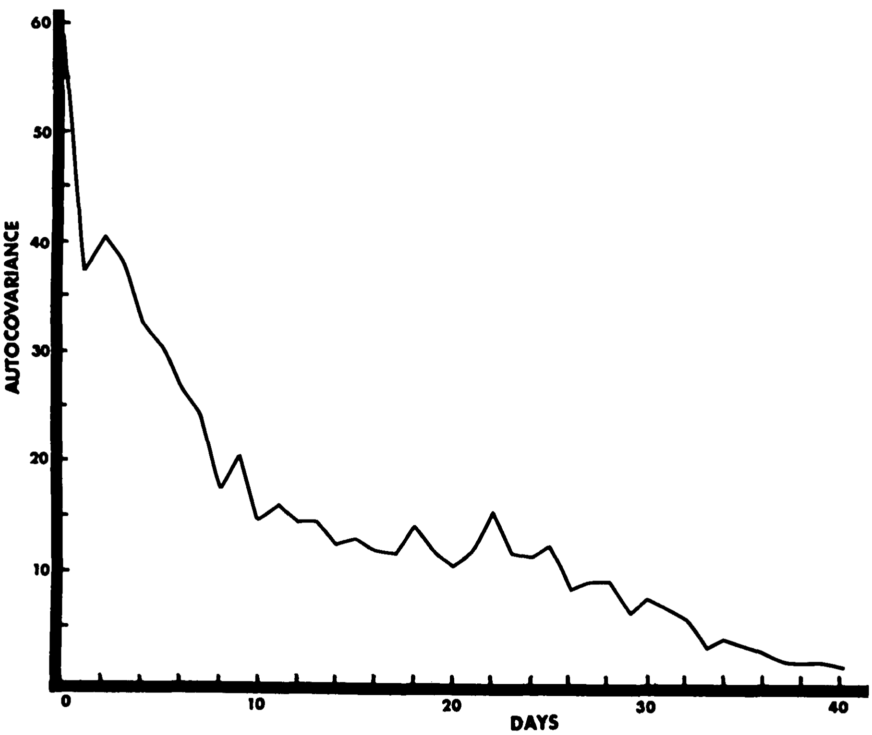

FIG. 1. Autocovariance function for 3 years of measurements of *Aedes vexans* densities in Malvern, Iowa.

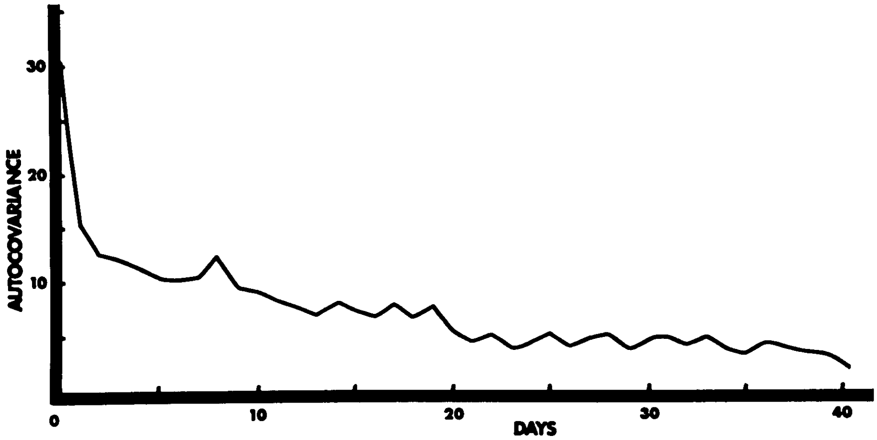

FIG. 2. Autocovariance function for 3 years of measurements of *Culex tarsalis* densities in Malvern, Iowa.

erarchy is reached, one expects the model to more nearly reflect some true state of a natural population of mosquitoes (de Figueiredo et al. 1973).

As a first step in developing models of mosquito population fluctuations, we propose to apply time series analysis to regularly spaced estimates of mosquito densities. Time series analysis has been applied to a number of biological problems, but as far as we can determine our study is the first concerned with mosquito biology.

MATERIALS AND METHODS

Analytical techniques

We propose to use time series analysis to interpret the data we will describe below. This technique is not new, but its application to mosquito studies is, and we shall demonstrate its use with these data. We have included a technical appendix giving the formulae which we used. More detailed description of the technique of time series analysis can be

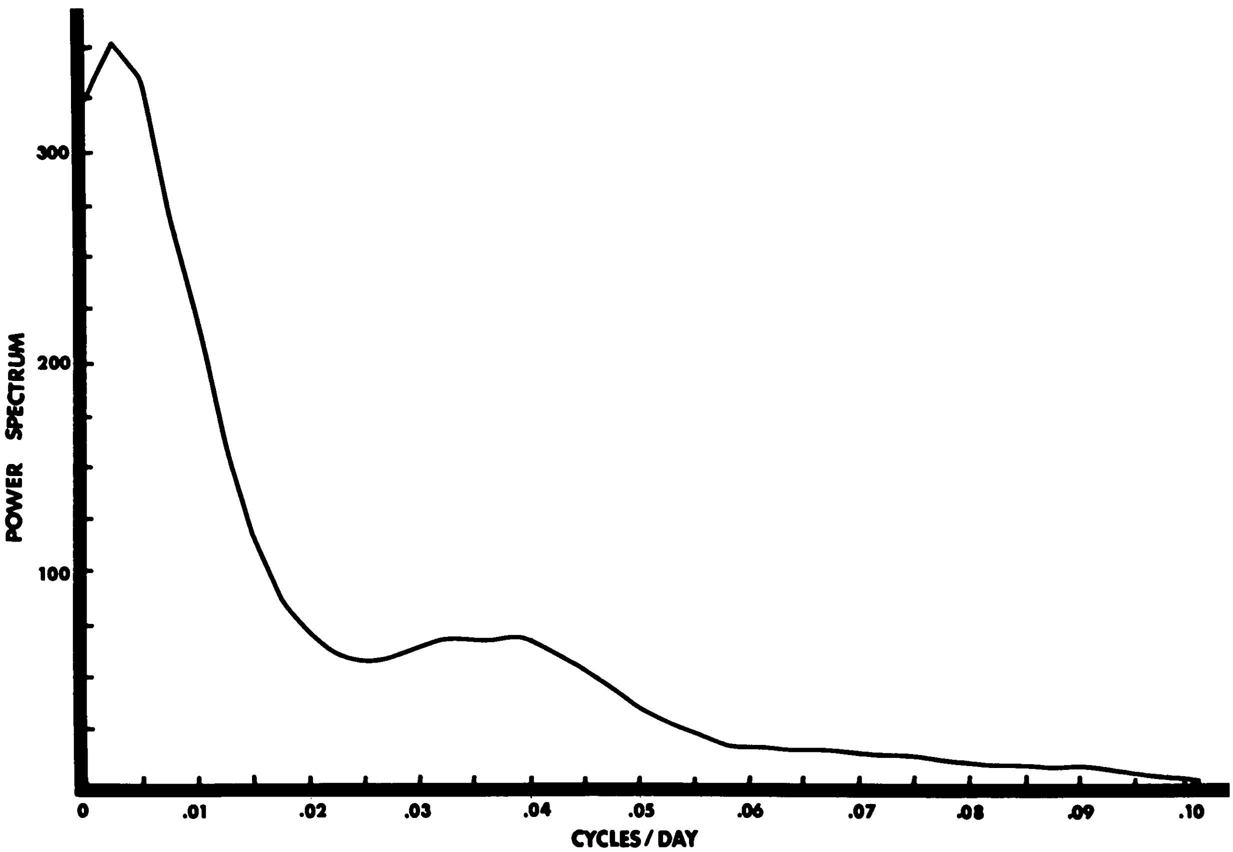

FIG. 3. Power spectrum for 3 years of measurements of *Aedes vexans* densities in Malvern, Iowa.

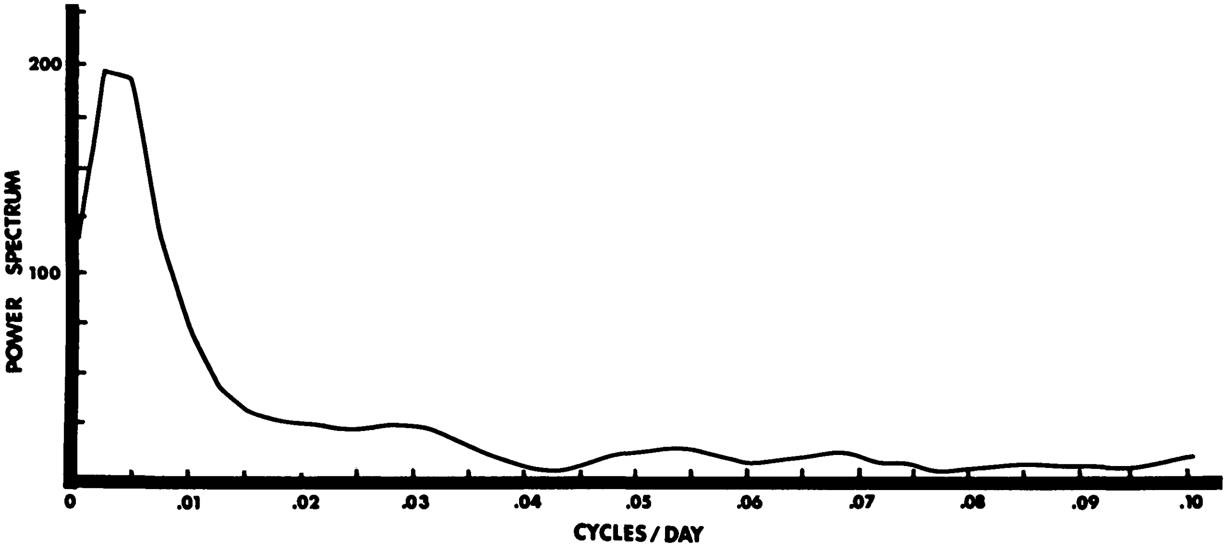

FIG. 4. Power spectrum for 3 years of measurements of *Culex tarsalis* densities in Malvern, Iowa.

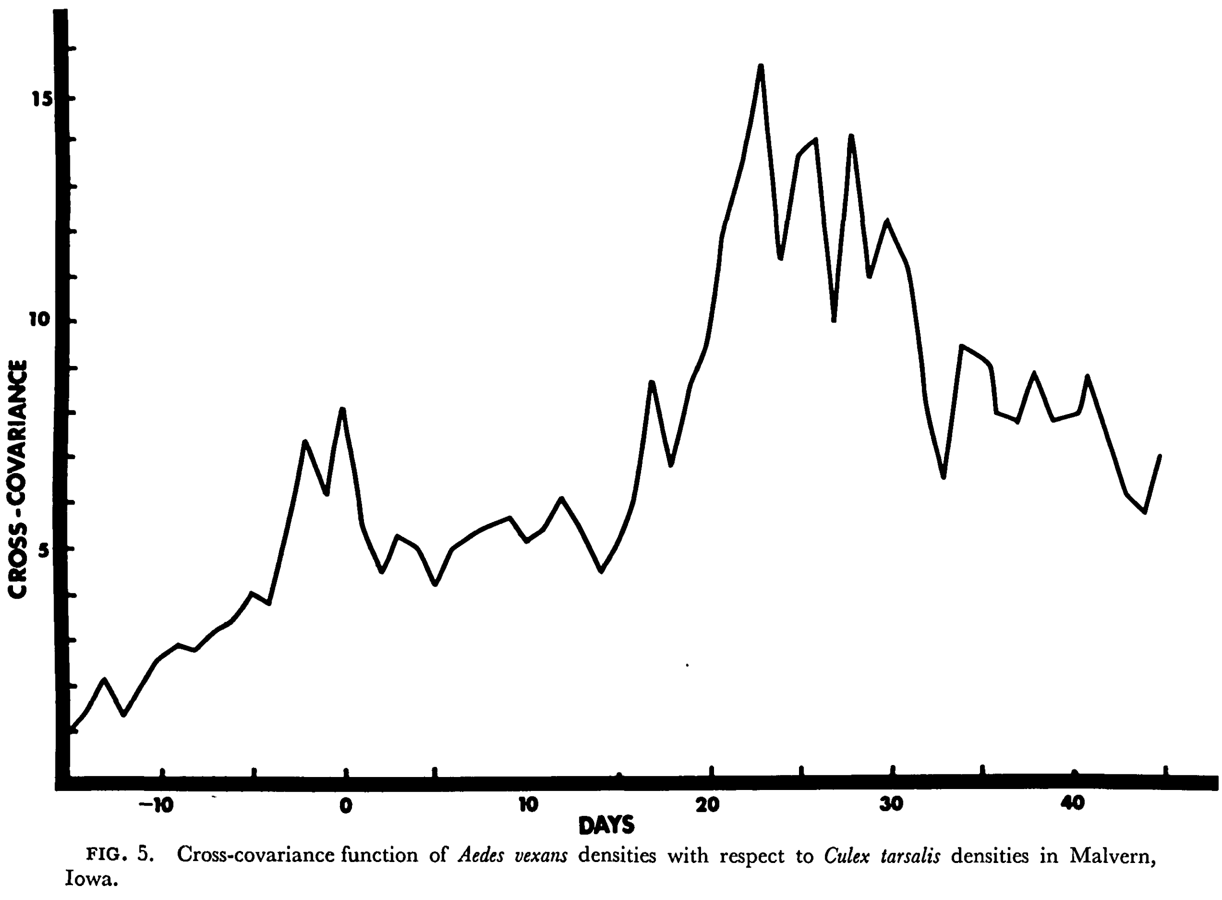

FIG. 5. Cross-covariance function of *Aedes vexans* densities with respect to *Culex tarsalis* densities in Malvern, Iowa.

found in such texts as Box & Jenkins (1970). We provide the following discussion for the benefit of our less statistically oriented readers.

A *time series* is a set of observations generated sequentially in time. If the observations are made at measured intervals of time (usually equal intervals) the series is said to be discrete. Such data have been collected for a number of biological systems. For example, mosquito populations are frequently sampled with light traps at regular (daily, bidaily, etc.) intervals for long periods of time (3 or more years). The set of the numbers of females of a given species collected in each interval would form a time series. Using the formulae explained in the Appendix, several functions can be derived from this series which should be of interest to mosquito biologists. The functions include the autocovariance function, the power spectrum, and the cross-covariance function.

The autocovariance function measures the correspondence in magnitude between observations separated by increasing intervals of time which are termed *lag times*. Generally, one finds the observations in the time series set separated by one time interval will correspond ("correlate") closer than those observations separated by 10 intervals of time. That is, as the lag time increases, the autocovariance function tends to decrease (FIG. 1).

Technically, the power spectrum is the Fourier transform of the autocovariance function. Practically, this spectrum allows one to detect periodicities in the time series set. The power spectrum is plotted against cycles/time interval. Where one detects a peak in the spectrum the corresponding cycle length can be determined from the inverse of the intercept of a perpendicular from this peak with the axis (FIG. 3). If several peaks occur, then the relative importance is related to their relative magnitudes.

The cross-covariance function is similar to the autocovariance function and measures the correspondence between 2 time series sets, for example the daily minimum temperature and mosquito densities. From this function one determines quantitatively the number of time intervals by which one series leads or lags a 2nd series (FIG. 5).

Sources of data

Since their development in the 1930's, light traps have been used to estimate relative mosquito densities. These traps work on the principle that

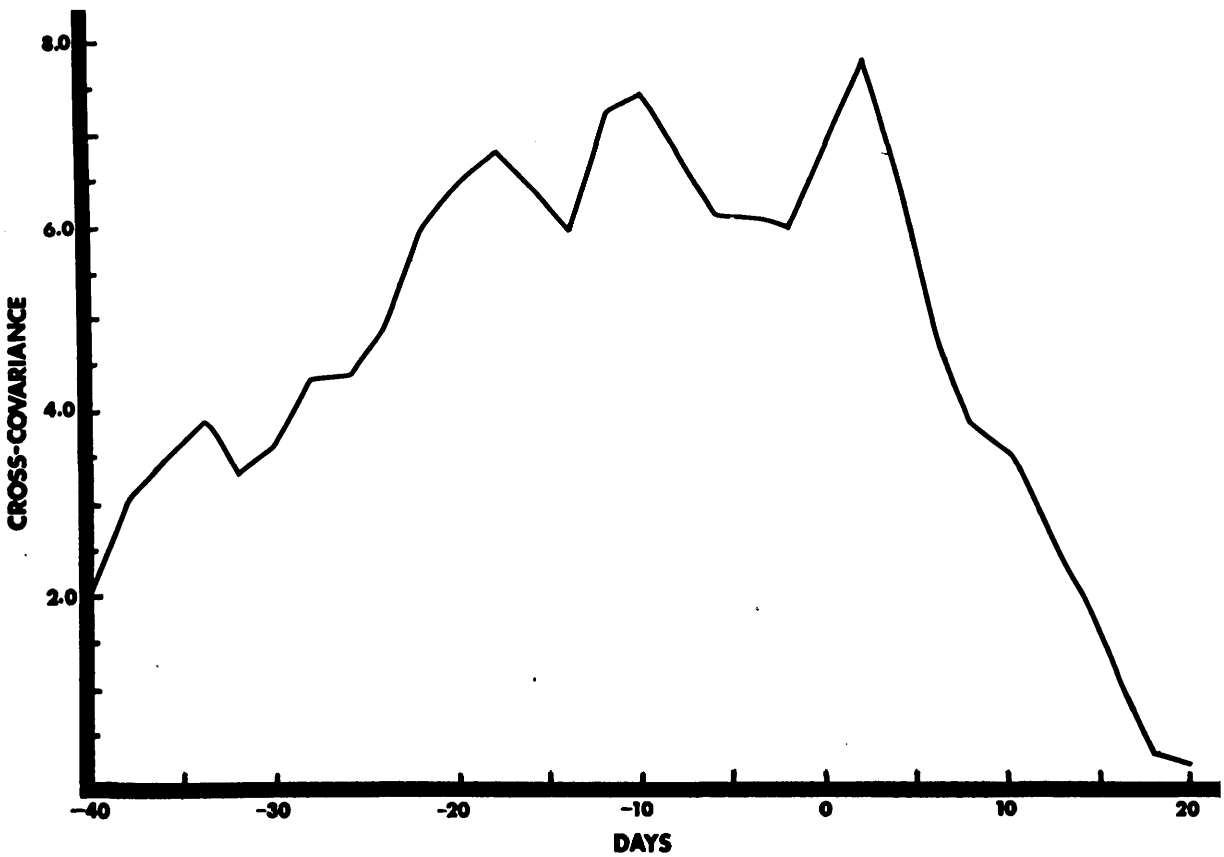

FIG. 6. Cross-covariance function for *Culex tarsalis* densities with respect to minimum temperature in Malvern, Iowa.

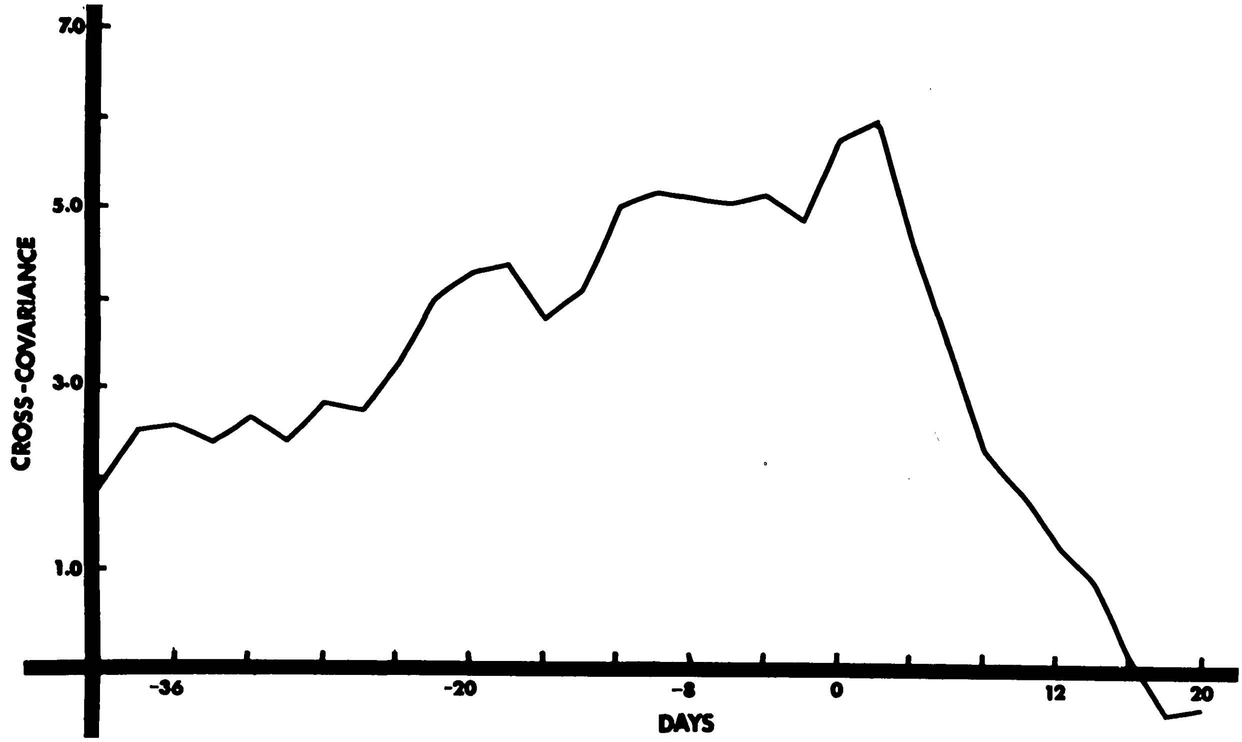

FIG. 7. Cross-covariance function for *Culex tarsalis* densities with respect to maximum temperature in Malvern, Iowa.

some species of mosquitoes are attracted to light. Once in the vicinity of the trap, the mosquitoes are blown into a collecting container by a downrush of air generated by a fan in the trap. The mosquitoes collected from a trap operated over an evening are then sorted into species and counted. Generally, more females than males are captured and analysis is often restricted to the former. The regularity with which a trap is operated depends in part upon the nature of the project as well as on the resources of the investigator. We are interested in a number of studies where traps have been operated at regular intervals in the same area for periods of time greater than a year. The data from these studies are usually presented in the form of a graph or table showing number of mosquitoes trapped as a function of time (Pinger & Rowley 1972). Frequently, meteorological data are recorded along with these data and correlations are sought by visual examination of the 2 sets of observations (Darsie et al. 1953).

We have obtained data such as these from several sources. One set from Dr W. Rowley, Iowa State University, includes collections made with New Jersey light traps near Malvern, Iowa, from May to October for 3 consecutive years (Pinger & Rowley 1972). A 2nd set of data was obtained from Mr W. Bartnett, director of the Harris County Mosquito Control District, Houston, Texas. These data include collections made twice weekly with New Jersey light traps and extend over a 7-year period. A 3rd set of data was collected under the direction of Dr Donald Roberts, and consisted of daily collections of egg rafts of the *Culex pipiens* complex from several artificial pools maintained near Edgewood, Maryland. This set of data was only collected for one 4-month period, May to September, an interval not optimally suited for time series analysis. However, we have included these data to illustrate the application of this method. Meteorological data for the Iowa and Texas series were obtained from the weather station nearest the trap as reported in *Climatological Data* published by the National Oceanographic and Atmospheric Administration. For the Maryland study the meteorological data recorded at the study site were used.

RESULTS

Iowa data

The data from Malvern, Iowa include observations on several species of mosquitoes. We have analyzed the data on female mosquito captures for the 2 mosquitoes which were trapped in the greatest numbers, *Aedes vexans* and *Culex tarsalis*. The

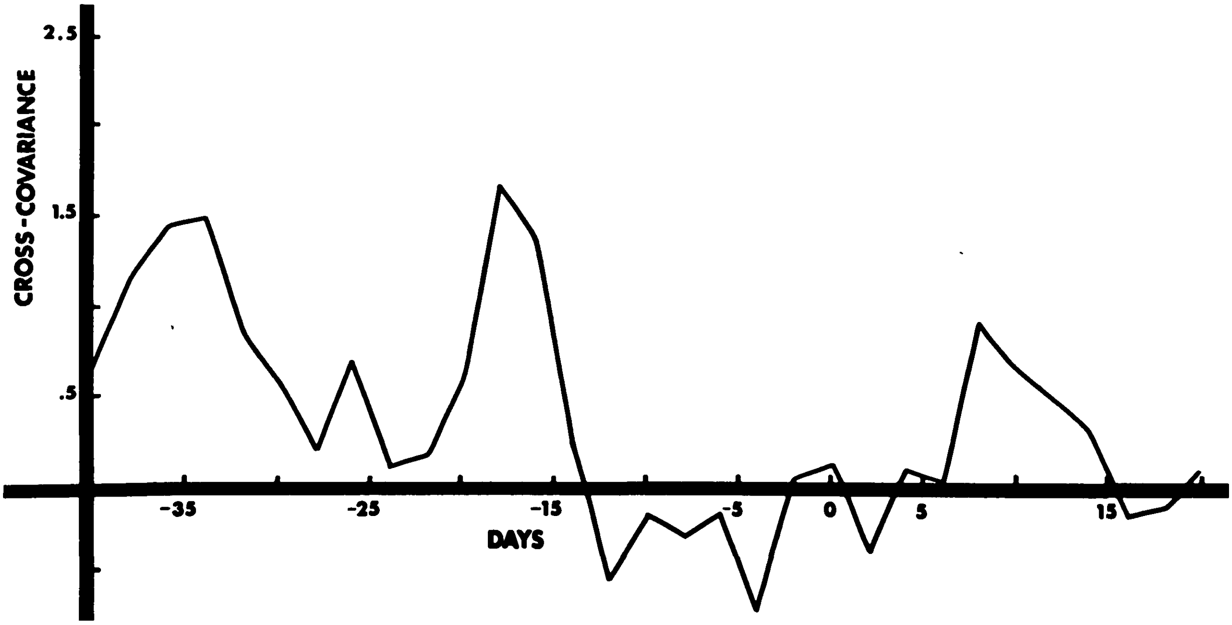

FIG. 8. Cross-covariance function for *Culex tarsalis* densities with respect to rainfall in Malvern, Iowa.

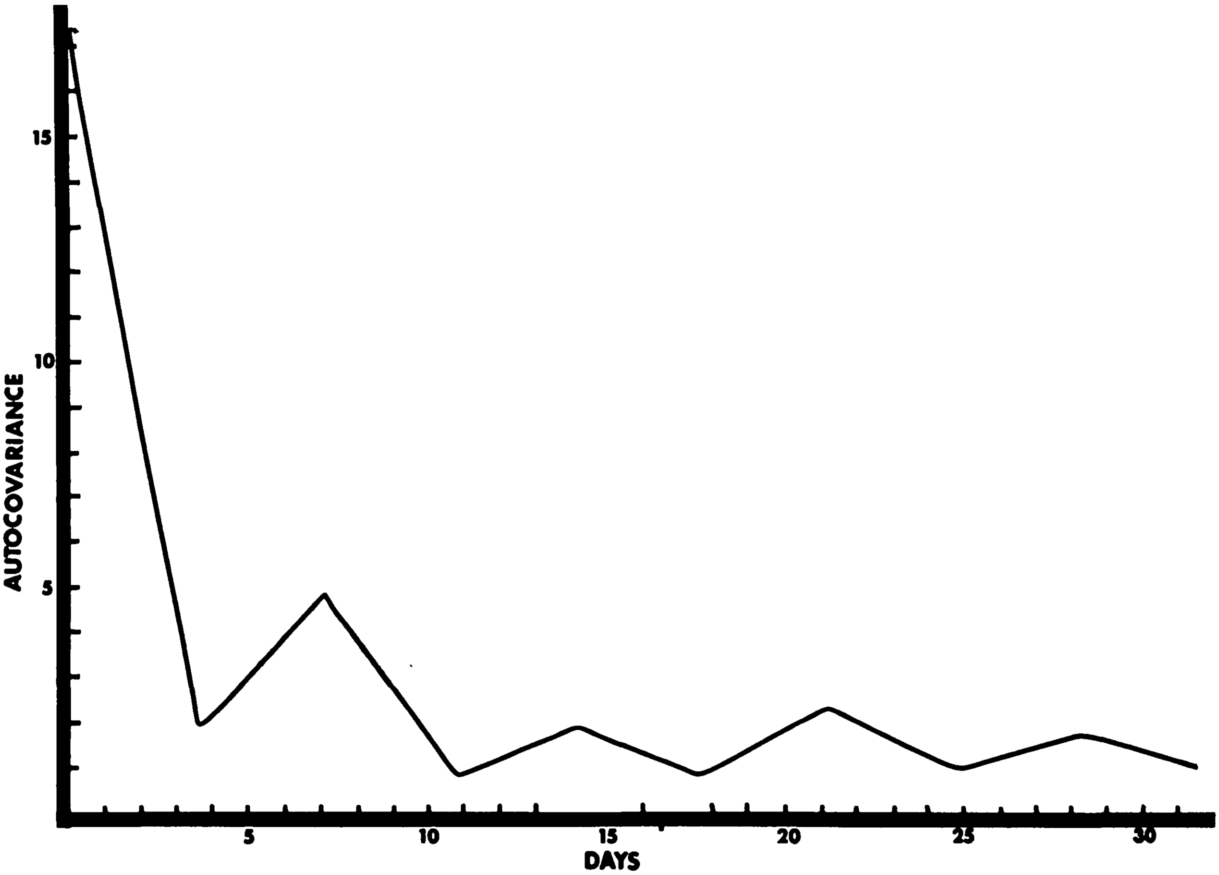

FIG. 9. Autocovariance function for 5 years of measurements of *Culex salinarius* densities in Houston, Texas.

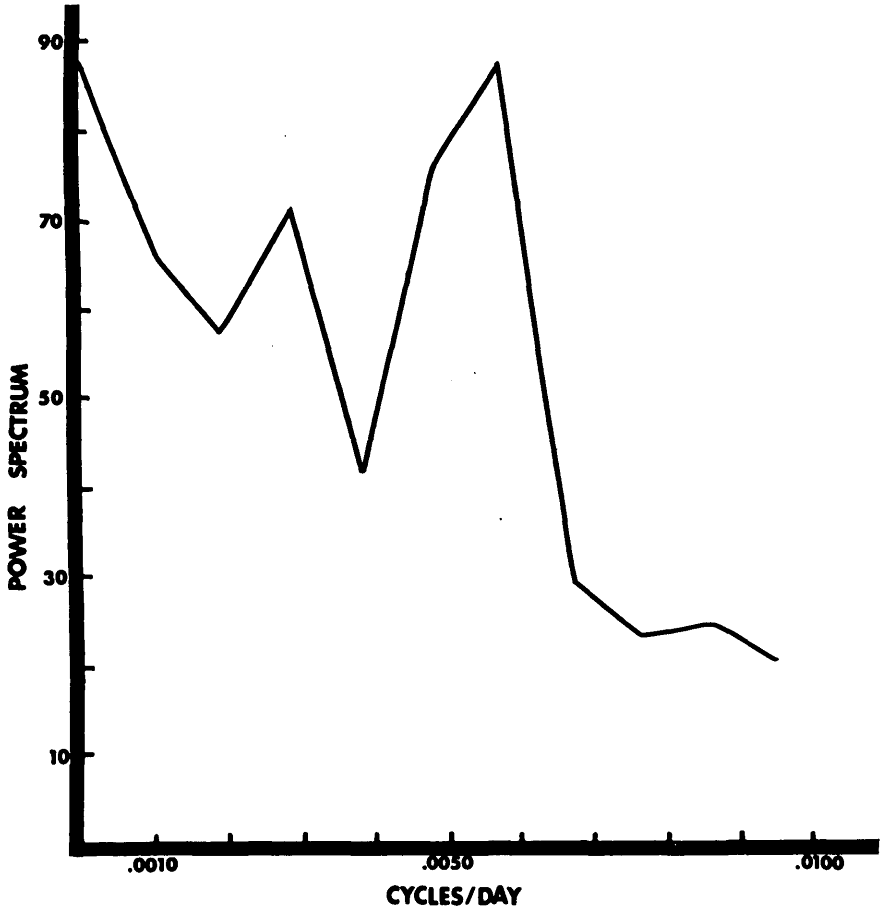

FIG. 10. Power spectrum for 5 years of measurements of *Culex salinarius* densities in Houston, Texas.

graphs of the autocovariance functions from these 2 species are presented in FIG. 1, 2.

The important feature to recognize in these figures is the slowly decaying nature of the autocovariance function. This means that the size of these populations (as measured by light traps) is heavily dependent upon the size of the population in the near past, and also as far back as several weeks. This is what would be expected intuitively. More importantly, from our viewpoint, it is encouraging for future modeling because we know that the time series is not an unpredictable series of random fluctuations.

The next tool to consider is the power spectrum for detecting cycles. The power spectrum for each species is plotted in FIG. 3, 4. There is only one high energy frequency for both species occurring between 0.0025 and 0.0050 cycles/day or between 200 and 400 days. This indicates a yearly cycle for these mosquitoes in Iowa which is to be expected. Also of interest is a secondary peak for *A. vexans* corresponding to about one month's time. This peak is small by comparison, r = (69/326 = 0.2), and may result from the influence of the bright full moon on the effectiveness of the mosquito trapping devices.

Although we have observed that *A. vexans* and *C. tarsalis* are both active during the same 5-month season, it is reasonable to seek a relationship between the level of one mosquito population and that of the other. To determine this, we examine the cross-covariance function which is plotted in FIG. 5. The cross-covariance function is non-zero, and in fact, displays 2 peaks leading to the conclusion that there is a secondary peak at a lag time of zero which implies that if *A. vexans* is high today, then probably so is the *C. tarsalis* population. There is also a primary peak at a lag time of 23 days which is probably of greater interest to biologists. This peak indicates that when the population density of *A. vexans* is high, then the density of the *C. tarsalis* population will be high in 23 days.

Cross-covariance analyses were also conducted between the level of the *C. tarsalis* population and each of 3 weather variables, minimum daily temperature, maximum daily temperature, daily rainfall. To lessen outlier peak effects, the square root of the mosquito number was taken. The results

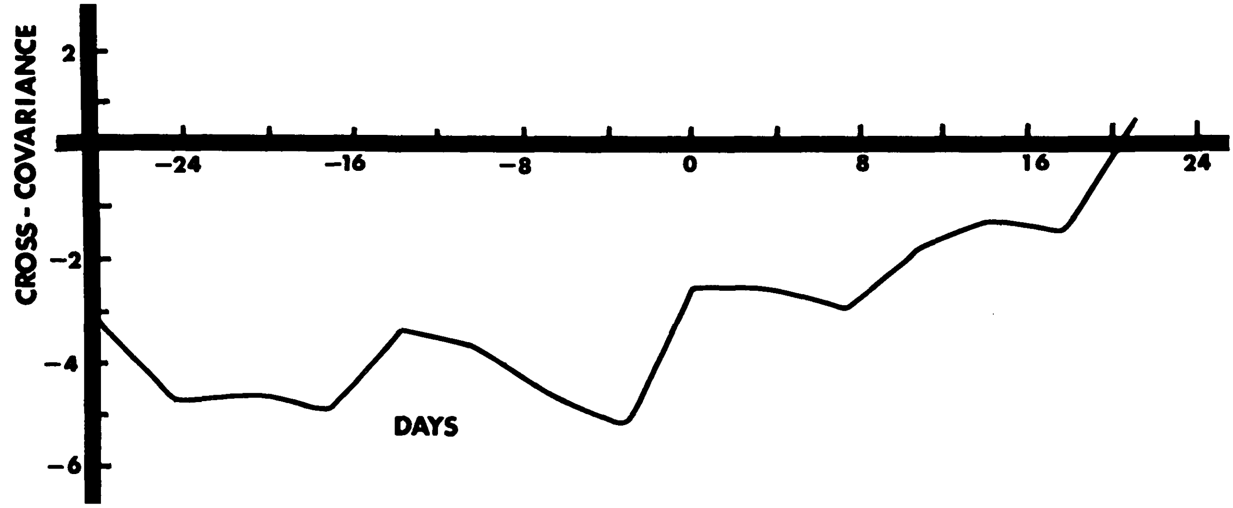

FIG. 11. Cross-covariance functions for *Culex salinarius* densities with respect to minimum temperature in Houston, Texas.

are given in FIG. 6, 7, 8.

The peak at −10 days in FIG. 6 indicates that the minimum temperature must be higher than a threshold level for about 2 weeks before a high level of *C. tarsalis* occurs. The secondary peak at +2 days is difficult to interpret biologically. The maximum temperature in summer is sufficiently high so that the cross-covariance function is highest at a lag time of 0 days (FIG. 7). From FIG. 8 the 2 peaks at −17 and −34 days indicate that high rainfall at 17 and/or 34 days prior to a mosquito density measurement will encourage a high reading. Perhaps this is due to the induction of a favorable environment at the time of an egg laying; 17 days would appear to be the average time of development into an adult.

Texas data

The data from one of several traps operated by the Harris County (Texas) Mosquito Control District have been treated to a similar analysis. In Texas more traps were operated, and since mosquitoes can be found in every month of the year, traps were operated over the entire year. However, the number of mosquitoes trapped per unit of effort was lower in Texas. We have chosen for analysis one trap which collected the largest number of mosquitoes over the longest period of time. From this trap we used the numbers of *Culex salinarius* captured. The autocovariance function for this species is plotted in FIG. 9. It should be noted that this function decays rapidly. The power spectrum for this species is plotted in FIG.

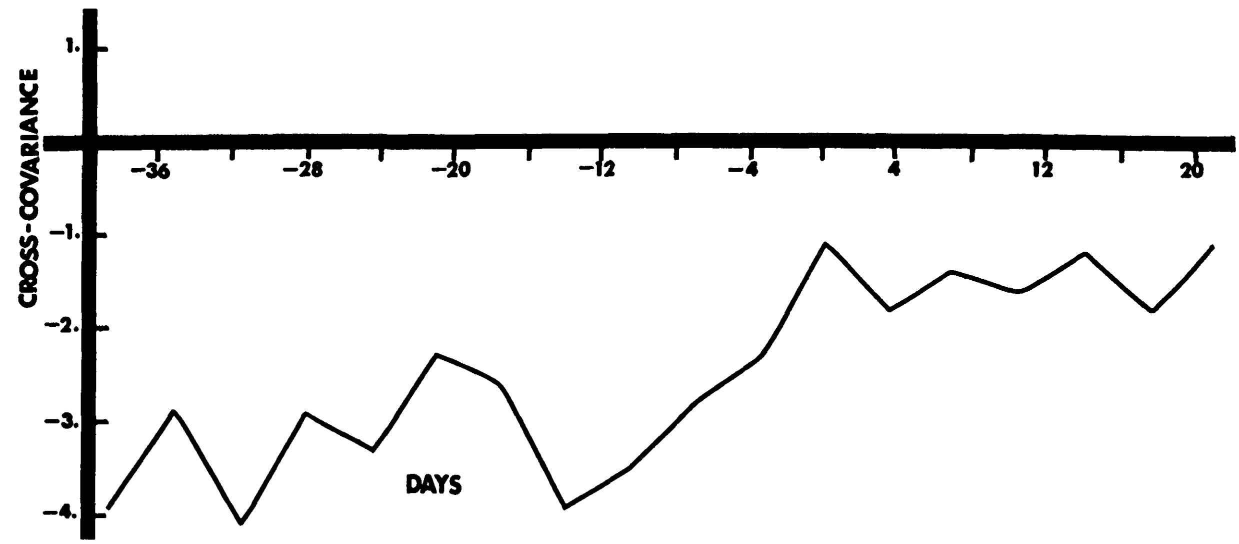

FIG. 12. Cross-covariance function for *Culex salinarius* densities with respect to maximum temperature in Houston, Texas.

10. An annual cycle is evident from this figure as well as a cycle of 175 days.

Cross-covariance data were calculated between the density of this mosquito and 3 meteorological variables, minimum temperature, maximum temperature, and precipitation. These functions are plotted in FIG. 11, 12, 13. The negative correlations between mosquito densities and the 2 temperature variables indicate that this species is prevalent in the cooler portion of the year. There are also no well-defined peaks. From FIG. 13, we observe that rainfall appears to lead the adult mosquito population by 10 to 12 days, approximately the minimum time from egg to adult.

Maryland data

The 3rd set of data differs from the above sets. In this set the number of egg rafts laid by female *C. pipiens* complex on artificial pools was determined daily. Therefore, the conclusions reached from the analysis of this series will relate to the density and behavior of ovipositing females. Because of the shortness of the series we can expect little from the power spectrum. When the cross-covariance functions were calculated we could detect no peaks when comparing maximum temperature, average temperature, or rainfall with egg raft densities. However, there was a peak at a lag time of 14 days in the cross-covariance spectrum comparing minimum temperature and egg raft densities (FIG. 14).

DISCUSSION

A number of techniques have been developed for estimating the relative numbers of mosquitoes in an area. In the past, the data accumulated using these techniques have been analyzed by subjective examination. We have introduced the use of a statistical technique, time series analysis, which allows us to develop objective and quantitative insights into the behavior of relative mosquito densities as a function of time.

In Iowa there was a clearly defined annual peak detected for both *A. vexans* and *C. tarsalis*. However, along the Gulf Coast of Texas where mosquitoes can be found in every month of the year, a 2nd cycle could be detected with a period of 175 days, though the biological significance of this cycle is obscure.

By using cross correlations between 2 time series, we have been able to determine objectively the number of days by which one series leads a 2nd. This is particularly useful for examining the relationship between environmental parameters and mosquito densities. For example, in Iowa the minimum temperature must be over a threshold

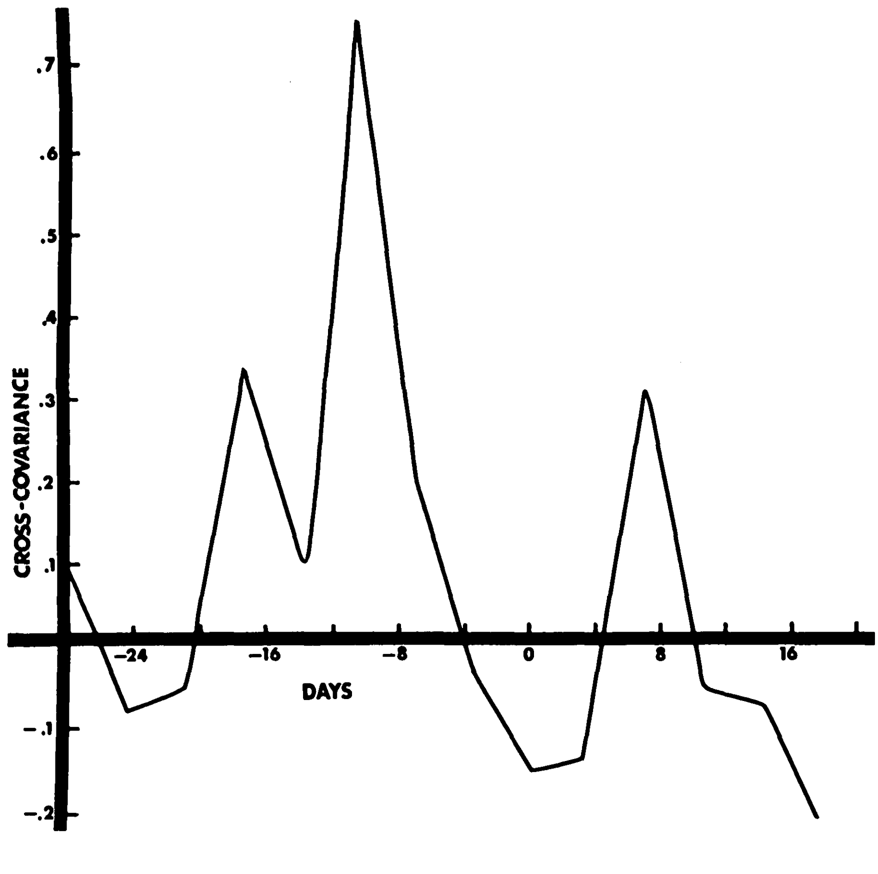

FIG. 13. Cross-covariance function for *Culex salinarius* densities with respect to rainfall in Houston, Texas.

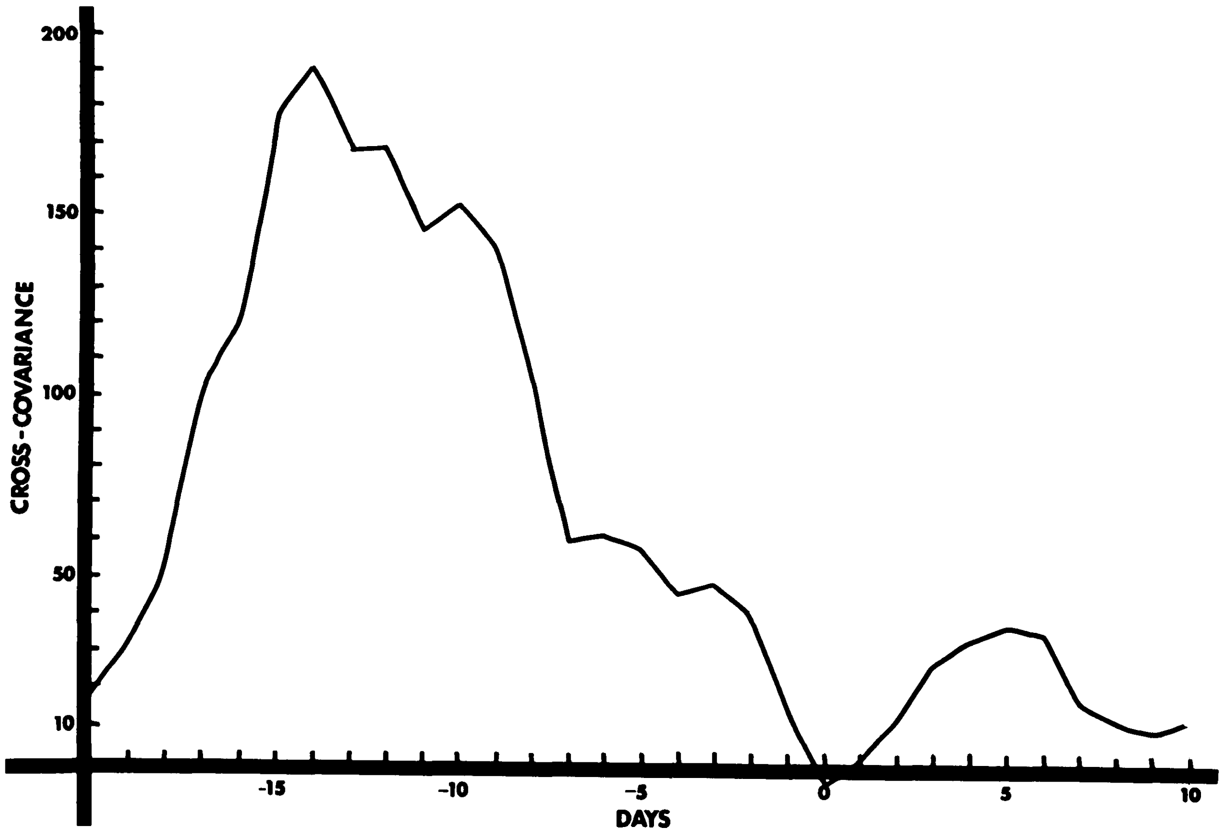

FIG. 14. Cross-covariance functions for *Culex pipiens* complex egg densities with respect to minimum temperature in Edgewood, Maryland.

or at least 2 weeks before a peak in *C. tarsalis* numbers. Also, rainfall is required 18 and/or 34 days prior to a high level in the density of this species in Iowa. On the other hand, *C. salinarius* occurs in high densities during the cooler part of the year in South Texas and the peak in the cross-covariance function is negative. With our Maryland data we examined the relationship between egg raft densities and several meteorological variables and found that only minimum temperature was a significant factor in predicting mosquito densities. Generally, we found that average measures of environmental variables were of little value in predicting densities. We suspect that we may be only able to observe the response of mosquitoes to threshold values of environmental variables.

Cross-covariance analysis is also of value in comparing the changes in density of 2 mosquito populations. It would be useful in some cases to be able to forecast a species of mosquito by monitoring the level of a 2nd species. This would be particularly true when the peak in density of a vector species followed the peak of a non-vector or simple pest species, and if the pest species were easier to monitor. From our Iowa data we were able to demonstrate such a capability. In this case the peak in density of *A. vexans* preceded the peak of *C. tarsalis*. These 2 species have similar biological habitats and it would be of interest to determine if any biological bases, such as competition, can be found for this phenomenon.

We have provided here a few sets of data to illustrate the application of time series analysis to mosquito population data. The results so far are of local interest only. However, as more data are analyzed, especially from different geographic regions, and using different collection methods, we expect to reveal more general patterns in the behavior of mosquito populations. Time series analysis also provides a method for developing predictive models of mosquito densities. This is discussed in a companion paper (Hacker et al. 1973).

Acknowledgments: We wish to express our gratitude to the several researchers who made our research possible by generously permitting us to use portions of the data from their long-term studies of mosquito populations in their respective areas. These include Dr Wayne Rowley, Iowa State University; Mr Robert Bartnett, Harris County Mosquito Control District, Texas; and Dr Donald Roberts, University of Texas, School of Public Health. We also wish to thank Mrs Dana Crisp for preparation of the graphs. This study was supported by grants to Drs Hacker and Thompson from the National Science Foundation (GB-33688 and GB-33689) and by the Office of Naval Research (NR042-283).

TECHNICAL APPENDIX

Formation of the time series: A graph of the volume of a lake plotted against time would give rise to a continuous line, which could be represented by x(t). This would be a continuous time series where the lake volume is known for all past time. However, such data are not available for a mosquito population, but rather discrete population counts are made at regular intervals of time starting at a given time in the past. Such a discrete time series can be represented by z(t), t = 0,1,2,...T, where z(T) is the last bit of data available.

This series is assumed to represent a stochastic process which may be partly described by its mean, μ, which is best estimated by

$$\bar{z} = \frac{1}{T+1} \sum_{t=0}^{T} z(t)$$

Such a stochastic process is said to be first-order stationary if the process essentially fluctuates about its mean. If the process tends to grow or diminish or any combination of the 2, then the process is said to exhibit non-stationary trends. These concepts are important for modelling considerations using techniques presently available in time series analysis and will be discussed in a companion paper (Hacker et al. 1973).

A stochastic process $\{X_t\}$, is a collection of random variables indexed by the subscript *t*. It is frequently natural to use *t* to indicate time. A common problem is to take observations x(t) from $\{X_t\}$ over an interval of time and use them to make inferences about the structure of $\{X_t\}$. If no assumptions are made about the relationship between X_{t_1} and X_{t_2} where $t_1 \neq t_2$, the inferential problem is more or less hopeless. However, we shall in the following make the usual stationarity and ergodicity assumptions which will render the inferential problem possible.

$\{X_t\}$ is said to be *first order stationary* if

$$\text{Average } X_t = E(X_t) = \mu \text{ for all } t \in T.$$

If, in addition

$$E[(X_t - \mu)\ (X_{t+\tau} - \mu)] = C_{xx}\ (\tau) \text{ for all } t \in T,$$

$\{X_t\}$ is said to be *second order stationary*.

The function C_{xx} (τ) is referred to as the autocovariance function of the stochastic process $\{X_t\}$. If we normalize C_{xx} (τ) by dividing by C_{xx} (0), we obtain the *autocorrelation function*.

$$\rho xx\ (\tau) = C_{xx}(\tau)/C_{xx}\ (0)$$

The Fourier transform P (f) of C_{xx} (τ) is referred to as the *power spectrum* or *power spectral density*.

$$P(f) = 2\int_0^\infty C_{xx}\ (\tau)\ \cos\ (2\pi f\tau) d\tau$$

When the vector $(X_{t_1}, X_{t_2}, \ldots, X_{t_n})$ has a Gaussian distribution for all *n* and all $(t_1, t_2, \ldots, t_n)$, then $\{X_t\}$ is called a *Gaussian stochastic process*. A stationary Gaussian stochastic process is completely characterized if we know μ_x and C_{xx} (τ). In practice, most stochastic processes are not strictly Gaussian. However, μ_x and C_{xx} (τ) will tell us a great deal about the process.

The assumption of *ergodicity* will enable us to estimate ensemble averages by time averages. Thus, for example, if X(t), $t \in [0,T]$ is a realization from a stationary ergodic process $\{X_t\}$:

$$\mu_x = \lim_{T\to\infty} \frac{1}{T} \int_0^T x(t)dt$$

$$C_{xx}(\tau) = \lim_{T\to\infty} \frac{1}{T-\tau} \int_0^{T-\tau} [x(t+\tau) - \mu_x][X(t) - \mu_x]dt.$$

For 2 jointly stationary and ergodic stochastic processes, $\{X_t\}$ and $\{Y_t\}$, we may measure interrelation via the *cross-covariance* function

$$C_{xy}(\tau) = E[(X_t - \mu_x)(Y_{t+\tau} - \mu_y)]$$
$$= \lim_{T\to\infty} \frac{1}{T-\tau} \int_0^{T-\tau} (x(t) - \mu_x)(y(t+\tau) - \mu_y)dt$$

The Fourier transform of C_{xy} (τ) is a useful tool, but we shall not employ it here.

The calculations of estimates of the above function were accomplished using the BMD02T program from the Biomedical Program Package developed at the University of California at Los Angeles.

LITERATURE CITED

Box, G. E. P. & G. M. Jenkins. 1970. *Time series analysis forecasting and control.* Holden-Day, San Francisco.

Brown, A. W. A. 1958. The spread of insecticide resistance in pest species. *Advances in Pest Control Research* Vol. 2, p. 351–418. Interscience, New York.

Darsie, R. F., Jr., D. MacCreary & L. S. Stearn. 1953. Analysis of mosquito-trap collections at Delaware City and Lewes, Delaware, for the twenty-year period, 1932–1951. *N. J. Mosq. Exterm. Assoc. Proc.* 40th Ann. Meet. p. 169–90.

de Figueiredo, R., C. S. Hacker & J. R. Thompson. 1973. *A systems approach to mosquito control.* Rice Univ. Tech. Rept (in press).

Hacker, C. S., D. W. Scott & J. R. Thompson. 1973 A forecasting model for mosquito population densities. *J. Med. Ent.* **10:** 544–51.

Pinger, R. R. & W. A. Rowley. 1972. Occurrence and seasonal distribution of Iowa mosquitoes. *Mosquito News* **32:** 234–41.

Ross, Ronald. 1910. *The prevention of malaria.* John Murray, London.