Population Modeling of Greenbug Aphids in Spring Wheat

Marie Ann Coffin

Follow this and additional works at: https://openprairie.sdstate.edu/etd

Recommended Citation
https://openprairie.sdstate.edu/etd/4500

This Thesis - Open Access is brought to you for free and open access by Open PRAIRIE: Open Public Research Access Institutional Repository and Information Exchange. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of Open PRAIRIE: Open Public Research Access Institutional Repository and Information Exchange. For more information, please contact michael.biondo@sdstate.edu.
POPULATION MODELING OF GREENBUG APHIDS
IN SPRING WHEAT

BY

MARIE ANN COFFIN

A thesis submitted in partial fulfillment
of the requirements for the degree
Master of Science
Major in Mathematics
South Dakota State University
1988
POPULATION MODELING OF GREENBUG APHIDS
IN SPRING WHEAT

This thesis is approved as a creditable and independent investigation by a candidate for the degree, Master of Science, and is acceptable for meeting the thesis requirements for this degree. Acceptance of this thesis does not imply that the conclusions reached by the candidate are necessarily the conclusions of the major department.

Robert J. Lacher
Major Advisor
Date

Norman Elliot
Thesis Advisor
Date

Kenneth Yocom
Head, Mathematics Department
Date
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Tables and Figures</td>
<td>i</td>
</tr>
<tr>
<td>Abstract</td>
<td>1</td>
</tr>
<tr>
<td>Chapter One: Literature Review</td>
<td>2</td>
</tr>
<tr>
<td>Chapter Two: The Model</td>
<td>4</td>
</tr>
<tr>
<td>Chapter Three: Testing the Model</td>
<td>23</td>
</tr>
<tr>
<td>Chapter Four: Sensitivity Analysis</td>
<td>29</td>
</tr>
<tr>
<td>Chapter Five: Suggestions for Further Research</td>
<td>32</td>
</tr>
<tr>
<td>Appendix A: Degree Day Approximations</td>
<td>34</td>
</tr>
<tr>
<td>Appendix B: Model Program</td>
<td>40</td>
</tr>
<tr>
<td>Appendix C: SAS Program to Fit Gamma Curves</td>
<td>49</td>
</tr>
<tr>
<td>Bibliography</td>
<td>55</td>
</tr>
</tbody>
</table>
List of Tables and Figures

Table 1: Parameters of the Gamma Distribution . . . . 12
Figure 1a: Gamma Curve for the First Life-Stage . . . 13
Figure 1b: Gamma Curve for the Second Life-Stage . . 14
Figure 1c: Gamma Curve for the Third Life-Stage . . 15
Figure 1d: Gamma Curve for the Fourth Life-Stage . . 16
Figure 1e: Gamma Curve for the Fifth Life-Stage . . 17
Figure 2a: Performance for Site I, 1963 . . . . . . . 25
Figure 2b: Performance for Site II, 1963 . . . . . . . 26
Figure 2c: Performance for Site I, 1966 . . . . . . . 27
Figure 2d: Performance for Site I, 1967 . . . . . . . 28
Abstract

A computer model was developed which used weather data and initial field counts of green-bug aphids to predict the behavior of the aphid population throughout the season. The model considered the effects of heat, food supply, predation and rainfall on the population. Data were obtained on the life-cycle of aphids grown in a controlled laboratory setting. These data were used to postulate a stochastic model of aphid growth which could be applied to a population in the field. A modified Leslie matrix model was used to predict the population's future behavior. Factors were then added to this basic model to account for mortality due to old age and predation, and the fecundity of the adult population. The completed model was compared with the actual field data obtained for several years. This comparison showed that the model generally performed well. As a last step, sensitivity analysis was done to determine which factors in the model would most profit by closer scrutiny. It was found that predation, in particular, needed to be analyzed more carefully.
Chapter One

Literature Review

Population modeling proceeds in the following general manner. A cohort of aphids is reared in the laboratory under controlled conditions and constant temperature. Regular observation of the aphids' development yields data on the length of time each individual spends in each life-stage. This is used to postulate the behavior of the aphids in a natural environment.

The crudest estimation methods utilize only the observed mean development time to estimate the development time in nature. It is possible, however, to get much better estimates by using more of the information that the laboratory study yields.

Two different methods exist for doing this. Manetsch [3] proposed setting up a system of differential equations which describe the changes in the population. Solving the system at time $t$ will yield the size of the population at that time.

Leslie [1] devised a different approach. His is based on the assumption that population growth is controlled by a small number of factors acting on the present population. The cumulative effect of these factors can be expressed as survival and reproductive probabilities on the present...
population. Thus, the survival and reproductive probabilities, when multiplied by the current size of the population, will yield the future population size.

Manetsch proved that solutions to the set of differential equations he proposed would necessarily be distributed as a gamma random variable. For this reason, it was decided that incorporating the gamma distribution into the Leslie model would be more efficient and require less computer time than repeatedly solving differential equations.
Chapter Two

The Model

We are attempting to model a greenbug aphid population. Aphids do not over-winter in South Dakota, because the cold is too intense. Instead, winged aphids migrate in on the southerly winds of spring. They land on small grain plants and start feeding and reproducing.

The aphid life-cycle can be summarized as follows. Newborn aphids begin to feed on plant juices and grow. As they grow, they must occasionally shed their exoskeletons for larger ones. The period between two such molts is called an instar. Greenbugs have four immature instars before reaching adulthood. Shortly after adulthood, the aphids begin reproducing, and continue to do so until death.

When the aphid population begins to experience crowding, some of the aphids produced will be winged (alatae). Alate aphids are winged when they reach adulthood, and immediately ride the winds to a new area. Upon landing at a new site, their wing muscles degenerate, and the aphids begin reproducing.

When the cereal plants mature in summer, the aphids migrate to later-growing crops and non-crop grasses.

It should be noted that cereal crops will be harmed only if aphid populations become sufficiently large in the early season, while the plants are still small. In larger
plants, it is not possible for the aphids to remove enough of the plant's nutrients to affect crop growth and yield.

The Model

This model attempts to consider all aspects of the aphid life-cycle. Since aphids do not over-winter in South Dakota, the life-cycle can be considered to begin when aphids are first sighted in the field. At this time, an initial count or estimate is made, and is used to start the model.

Leslie's matrix model was chosen. Leslie [1] assumed that the age-structure of the population at some time t was known. He then used this information to predict the age-structure of the population at a future time.

In practice, however, the age structure is usually not known; the stage-structure is much more easily obtained. Woodward [2] suggested that Leslie's method could be modified to be used on a population where only the stage-structure was known. Woodward began by assuming that development times in each stage were normally distributed. This was done so that the age-structure of the population could be "recovered" (actually estimated) from the stage-structure. Woodward proceeded by dividing each life-stage into a number of sub-stages. This is done so that one can approximate the amount of time that an individual has spent in its current life-stage. Then the Leslie matrix is modified to include all of these
sub-stages.

For our purposes, it did not seem necessary to estimate the age-structure of the population, and so normality assumptions were unnecessary. In fact, they are probably unwarranted, since visual examination of developmental data show that development times are markedly skewed to the left. (Manetsch [3], using a distributed delay method, derives the result that development times have a gamma distribution.)

When constructing the model it is necessary first to estimate how many sub-stages will be needed. Clearly, too few sub-stages will lead to a crude approximation, while too many will make the model cumbersome to operate. Matetsch suggested that good results could be obtained by choosing the number of sub-stages to be equal to the shape parameter of the gamma distribution. This proved to be feasible, since the gamma distributions we modeled all had shape parameters between one and ten.

The model is based on the Leslie matrix model. Briefly, the Leslie matrix is a modified Markov Chain in which the diagonal elements represent the probabilities of an individual's surviving to the next time-step. In addition, the first row of the matrix contains the average numbers of female births to individuals in each age class in the time-step that survive until the next time-step. The remaining elements of the matrix are zero. This matrix is
called the A matrix. We also define a vector whose elements are the numbers of individuals in each life-stage. This is called the F vector. If we take the matrix product \( F^* = A \times F \), the \( F^* \) will contain the number of individuals in each life-stage after one time-step.

**Example:** Suppose the numbers of aphids in each of five life-stages are 5, 3, 2, 1 and 6, respectively. Also suppose the average number of females born to a female in age-class 5 is 0.78, and the average number of females born to younger females is 0. Finally, suppose that the probability of surviving to the next age-class is 0.95 for all age classes. Then

\[
A = \begin{bmatrix}
0 & 0 & 0 & 0 & 0.78 \\
0.95 & 0 & 0 & 0 & 0 \\
0 & 0.95 & 0 & 0 & 0 \\
0 & 0 & 0.95 & 0 & 0 \\
0 & 0 & 0 & 0.95 & 0
\end{bmatrix}, \quad F = \begin{bmatrix}
5 \\
3 \\
2 \\
1 \\
6
\end{bmatrix}
\]

and

\[
F^* = \begin{bmatrix}
4.68 \\
4.75 \\
2.85 \\
1.90 \\
0.95
\end{bmatrix}
\]

For our purposes, it was considered more realistic to
modify the model in several ways. The original Leslie model assumes that survival and fecundity probabilities are constants which are known or estimated. In fact, it appears that the probabilities (e.g. 0.78 and 0.95 in the example above) are functions which may depend upon the present size and structure of the population, as well as the quantity and quality of food available. Although this is only a small change in theory, it necessitates a major change in the computations, since the elements of the A matrix must be re-calculated at each time-step, to reflect changes in the population and environment. Some of the necessary considerations are discussed below.

**Time Measurements**

Aphid developmental rate is known to be linearly related to temperature [11]. This linear relationship holds true over a wide range of temperatures. An aphid has a base temperature, below which no development will occur, and an upper limit temperature beyond which no additional increase in developmental rate will result from higher temperatures. Between these two, developmental rate is essentially linear. However, since temperatures in nature fluctuate greatly in time, aphid developmental rate is not in any sense linear, or indeed predictable, with respect to time. Because of this, a "time scale" in degree-days (DD) was used. One degree-day is defined to be a temperature of one degree
celcius above base temperature, experienced for 24 hours.
The total degree-days accumulated in a day can be estimated
from the maximum and minimum temperatures for the day.
Degree-day estimation was done by the method of Allen [4].
For details, see Appendix A.

**Laboratory Study**

A cohort of aphids was reared in the laboratory under
constant temperature and infinite food supply. The constant
temperature made the calculation of accumulated degree-days
particularly simple. It is known that a cohort of aphids
that enter a life-stage simultaneously will emerge
distributed in time, where "time" is measured in
degree-days. Thus, the time at which an individual matures
to the next life-stage can be considered a random variable.

To construct the model, we need to know the
distribution of that random variable. Previously, this has
often been assumed to have a gamma-type distribution. On
the other hand, Wagner, et. al. [5] postulated a Weibull
distribution and obtained good results from it. There is
some theoretical justification for this, since the Weibull
distribution was developed specifically to model
failure-time data. The gamma distribution has density
function

\[ f(x, \alpha, \beta) = \frac{1}{\alpha \Gamma(\alpha)} x^{\alpha - 1} e^{-x/\beta} \]

while the Weibull distribution has density function

\[ f(x, \alpha, \beta) = \frac{1}{\alpha \beta^\alpha} x^{-(\alpha/\beta)} e^{-(x/\beta)^\alpha} \]
Preliminary examination of the laboratory data suggested that either a Gamma or Weibull distribution might be appropriate (the two are similar in shape), but no method was found to evaluate which distribution would provide a better description. In fact, no adequate method was found to estimate the parameters of such a Weibull distribution. This estimation is decidedly non-trivial, and it may be that the Weibull is not commonly used for just such computational reasons. In contrast, the parameters of a Gamma distribution can be easily estimated from the sample mean and standard deviation of the population in question.

Such estimates were made from the laboratory data, and the resulting Gamma distribution was tested against the laboratory data, in an attempt to measure how well the proposed distribution fit the facts. A value of $R^2$ was calculated for each life-stage. This is usually interpreted as representing the fraction of the total variation in the random variable which can be accounted for by the distribution. In five of the six life-stages, the value of $R^2$ was in excess of 0.90. This was felt to be an excellent fit; therefore, the Gamma distribution was used to construct the model.

The poor fit of the fifth set of parameters can be explained by the fact that we are attempting to predict when an individual moves from the fifth life-stage.
(pre-reproductive adult) to the sixth (reproducing adult).
The time that an individual spends in the fifth life-stage
is typically quite short -- often only a few hours. Since
the aphids were monitored at eight-hour intervals, it is
likely that the data collected were not very accurate on
this point. However, it was also felt that even a poor
estimate would be adequate, since the estimate would not be
wrong by more than a few hours. Table 1 contains a list of
the parameters and $R^2$ values for all six life-stages.

**Maturation Probabilities**

Aphids have four immature life-stages, and a mature
life-stage. For ease in computation, the mature life-stage
was divided into two: pre-reproductive matures, and
reproducing matures. This gives us a total of six
life-stages. Since individuals who enter a life-stage at
the same time will leave it at different times, each
life-stage is divided into a number of sub-stages, so that a
better estimate could be made of the length of time an
individual had spent in its present life-stage. Knowing the
parameters of the underlying distribution enables us to
estimate how many of the individuals in any sub-stage will
mature to the next life-stage in a time-step. These
individuals are transferred by the model to the first
sub-stage of the next life-stage. Figure 1 shows
graphs of the empirical distribution functions, and the
gamma curves that were fitted to them.
# Table One

Parameters of the Gamma Distribution

<table>
<thead>
<tr>
<th>Life-stage</th>
<th>Scale Parameter</th>
<th>Shape Parameter</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.16987</td>
<td>9.19488</td>
<td>0.901</td>
</tr>
<tr>
<td>2</td>
<td>2.47643</td>
<td>8.96503</td>
<td>0.924</td>
</tr>
<tr>
<td>3</td>
<td>3.78789</td>
<td>4.14873</td>
<td>0.915</td>
</tr>
<tr>
<td>4</td>
<td>3.08195</td>
<td>7.29780</td>
<td>0.976</td>
</tr>
<tr>
<td>5</td>
<td>6.73449</td>
<td>1.97160</td>
<td>0.604</td>
</tr>
<tr>
<td>6</td>
<td>59.7507</td>
<td>9.85034</td>
<td>0.956</td>
</tr>
</tbody>
</table>
Figure 1a
Gamma Curve for the First Life-Stage

Degree-Days
Figure 1B

Gamma Curve for the Second Life-Stage

Degree-Days

0.0 10 20 30 40 50 60 70 80 90 100

0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
Figure 1c
Gamma Curve for the Third Life-Stage

0.015

0.010

0.005

Degree-Days

0 10 20 30 40 50 60 70 80 90 100
Figure 1e
Gamma Curve for the Fifth Life-Stage

![Graph showing a gamma curve for the fifth life-stage. The x-axis represents Degree-Days ranging from 0 to 100, and the y-axis represents a range from 0.005 to 0.015. The curve peaks around Degree-Days 20-30.]
Mortality

Aphid mortality is caused by several factors: predation, inadequate food supply, old age and excessive rainfall. In the model, each of these factors needs to be considered separately.

Food Supply. Studies by Kiekiefer [12] indicate that within limits, the quality of food supplied has no significant effect on aphid development. That is, the growth stage of the plant and the nutritive quality can vary greatly and have only a minor effect on aphid development. For this reason, the plant was considered to be an infinite food source for the aphids. Of course, the plant's maturing and drying in summer does affect the aphid population, but this effect was not considered, since we are not concerned with aphid development at such a late date.

Excessive Rainfall. It is known that a hard rain-storm can be damaging to the aphid population. A very hard rain can wash the aphids off the plant entirely. However, no quantitative data on the effect were available. Therefore, the model contains a "dummy" subprogram to calculate the probability of an individual's being killed by excessive rainfall. Currently, the subprogram computes a probability of zero. If data become available on this subject, the subprogram can be re-written

Age-Specific Mortality. A cohort of aphids was raised
in the laboratory, and careful counts were made of how many
died in any given time-span. Non-linear regression was used
to fit this data to a curve, using an equation described by
Graf, et al. [6]. The equation is:

\[ \text{MORT} = 1 - a \times \exp[b \times \exp(c \times DD)] \]

This method appeared to yield a very good fit. The
estimated parameters are \( a = 0.908589 \), \( b = -0.006861 \), \( c = 0.006561 \)

**Predation.** Aphids have two major predators: **Hippodamia convergens** and **H. tridecimpunctata tibialis** (abbreviated HC and HT in the program). It was not
considered practical or necessary to model the entire
predator life-cycle, since in the early part of the season,
predator numbers remain essentially constant. Since there
was very little data on which to base any assumptions, the
simplest possible functional model was used. A Type I
functional response [7] was chosen, in which the amount (in
weight) of aphids eaten is assumed to be a linear function
of the size of the aphid population, unless the population
is effectively infinite, in which case the amount eaten is
constant.

Data were obtained in the following manner. Counts
were taken of how many aphids of a particular life-stage
were eaten in one day by a starved predator. This was
considered to be the maximum amount that a predator could
eat in one day. Aphids of different life-stages were weighed to determine their relative average weights.

Since aphids of different life-stages have greatly different average weights, a new measurement of "aphid units" was introduced into the model. An aphid unit (AU) is defined to be the average weight of a first instar aphid. The average weights of all the instars, expressed as aphid units, appear in Table 3. Thus for predator purposes, the population can be considered to consist of a certain number of aphid units. In addition, the predators' "appetites" can also be expressed in aphid units. Then only two cases can occur.

Case 1: The aphid population is so large that the predators can eat as much as they wish. In this case, the predators will eat a constant amount of aphid units. Then for any individual aphid, the probability of being preyed upon is simply the fraction of the total population that is eaten. Unfortunately, we do not know how large the aphid population must be for these conditions to exist. An estimate of 25 adult aphids per predator per day was used. This is called the critical number of aphids, and is abbreviated in the program as NCA.

Case 2: The predators cannot eat as much as they wish. The number of aphids eaten when the aphid population is zero will, of course, be zero. The number of aphids eaten when
the aphid population reaches the critical number is also known. Between these two values, consumption is assumed to be a linear function of the size of the population. Hence, the probability of an individual aphid's falling prey to a type HT predator during one degree-day is proportional to \( \frac{T\text{UNITS}}{\text{NCA}_1} \). Similarly, the probability of an individual aphid's falling prey to a type HC predator is proportional to \( \frac{T\text{UNITS}}{\text{NCA}_2} \). Here \( T\text{UNITS} \) is the total weight of the aphid population, in aphid units; \( \text{NCA}_1 \) and \( \text{NCA}_2 \) are the critical numbers of aphids for type HT and HC predators respectively.

Now, if the number of HT predators is \( \text{NHT} \), and the number of HC predators is \( \text{NHC} \), the probability of an individual aphid's not being preyed upon (i.e., surviving) is

\[
P(\text{survival}) = (1 - P_1)^{\text{NHT}} (1 - P_2)^{\text{NHC}}
\]

Then, the probability of being preyed upon is

\[
P(\text{pred}) = 1 - (1 - P_1)^{\text{NHT}} (1 - P_2)^{\text{NHC}}
\]

where \( P_1 \) and \( P_2 \) represent the probability of being eaten by an HT and HC predator, respectively. We assume that all of the predators are acting independently. Now for any individual in the population, if we let \( M \) represent age-specific mortality, \( P \) represent predation, and \( R \) represent rainfall mortality (all expressed as probabilities), and we assume that these factors operate independently on the
population, then the probability of the individual's surviving to the next time-step is

\[ P(\text{survival}) = 1 - M - P - R + M*P + M*R + P*R - M*P*R. \]

Fecundity.

The same cohort used to model mortality above was also used to model the fecundity function. This was also modeled according to Graf's equation:

\[ \text{Fertility} = \frac{[a \times DD]}{[b^{DD}].} \]

The estimated parameters were \( a = 0.008 \), \( b = 1.001 \).
Chapter Three
Testing the Model

Field data were gathered over a period of years at a number of experimental sites in South Dakota. Information gathered included the number of nymphs, adult alate and adult apterate aphids of four different species present, as well as the number of predators sighted. These were recorded as number present in a quadrat. At each location, 40 quadrats were sampled at weekly intervals, beginning in early spring and continuing through the summer.

Four of these sampling periods were chosen to test the model. These four periods were felt to represent particularly complete and accurate sampling. Those chosen were Site I, 1963; Site II, 1963; Site I, 1966 and Site I, 1967. For each sampling period, the counts were averaged over the 40 quadrats sampled, and the averages were then converted from number per quadrat to number per 500 tillers.

The first aphid counts and corresponding predator counts for the year were then used as initial conditions to start the model running. The model would then predict the future behavior of the population, which could be compared to the observed behavior. This comparison is useful in assessing the utility of the model.

The model runs made in this way showed several
discrepencies in the program. The model consistently underestimated the observed population. This suggests that some factor, such as predation or age-specific mortality, may be given too much weight in the model. Also, adult aphids in the model die off much more quickly than is consistent with observed behavior. Clearly, these are areas deserving of more attention. Graphs of the actual and predicted population growth appear in Figure 2.
Figure 2a
Site I, 1963

Diagram showing the total number of aphids over days in the Gregorian calendar from 1963.
Figure 2b
Site II, 1963

Total # of Aphids

Days, Gregorian Calendar

actual
predicted
Figure 2c
Site I, 1966

Total # of Aphids

Days, Gregorian Calendar

actual
predicted
Figure 2d
Site I, 1967

![Graph showing actual and predicted aphid populations over 67120 to 67170 days on the Gregorian Calendar.](image)
Chapter Four
Sensitivity Analysis

The next step in analyzing the model is to assess which factors may be the most important, and therefore suggestive of further attention. Typically, this is done by introducing small changes into various factors. Those factors in which a relatively small change produces a markedly different result are considered to be factors which should be examined more closely. In this model, the factors chosen for sensitivity analysis were predation, rainfall, age-specific mortality, and fecundity.

**Rainfall.** It proved almost impossible to analyze a dummy subroutine. As might be expected, a small constant probability of rainfall-induced mortality produced a small change in the population, but did not affect its over-all behavior. Introducing small rainfall effects at random seemed to have little effect. In reality, however, rainfall effects, although occurring at irregular intervals, might be quite large and have a significant effect. This topic clearly needs further research.

**Age-Specific Mortality.** The proposed model for age-specific mortality is

\[
\text{MORT} = 1 - a \times \text{EXP}[b \times \text{EXP}(c \times DD)]
\]

Even slight changes in the parameters affect the population
size. Also, since this function has three parameters, the interaction of changes in the three parameters must be investigated. Furthermore, if age-specific mortality is over-estimated for adults, this affects the reproductive capability of the population, and hence the age structure as well as the size of the population.

**Fecundity.** The proposed fecundity function is

\[
\text{FERT} = \frac{[a \times \text{DD}]}{\text{b}^\text{DD}}.
\]

Again, it appears that slight changes in the parameters have a significant effect. It would be worthwhile to estimate these parameters more precisely, if only because they have extremely small values, so that a small change produces a large effect.

**Predation.** In examining the predation function, a number of problems come to light. The number of predators present in the early season appears to be quite high in relation to the number of aphids available for consumption. This suggests one of two situations: 1. We have estimated how many aphids must be present for the population to be effectively infinite. Perhaps our estimate is much too high. 2. There may be other insects at the site which also serve as food for the predators. In this case, we must be able to make some sort of estimate of how many of these "other" insects the predators are consuming, or incorporate some sort of correction factor into the model.
Clearly, both of the above possibilities need to be investigated further. It must also be considered that the aphid model in use is very crude. It does not take into consideration the searching time, eating time, resting time, etc. which are generally considered important in estimating predation.

In summary, it appears that these problems are causing predation to be over-estimated, which produces a damping effect on population growth. The whole predation function undoubtedly merits further attention.
Chapter Five

Suggestions for Further Research

Predation. It is generally felt that either a Type II or Type III predator/prey model is a better predictor than a Type I model. [7] These models take into account such factors as searching time, resting time and feeding time. They also allow for the fact that at very low prey densities, predation is generally lower than a linear model would predict.

It is also necessary to explore the predator/prey interaction more deeply. For instance, does the number of prey eaten per day depend only on the biomass of the prey? We have assumed this to be true.

Finally, some way must be found to quantify the presence of other prey in the field. Perhaps there is some way to estimate this from available data.

Hypothesis Testing. Frequently, it is assumed that development times for the first three instars come from the same underlying distribution. This hypothesis should be tested. If we are assuming a gamma distribution, this is equivalent to testing the hypothesis that three means are equal and that at the same time, three standard deviations are also equal. One possibility for this would be a Kolmogorov-Smirnov Test [8].

It should also be noted that a separate base
temperature was found for each life-stage. Since these temperatures are fairly close together, the hypothesis that all or several of the life-stages have the same base temperature should be tested.

**Choosing the Underlying Distribution.** In this model the underlying distribution of development times was assumed to be gamma. A more rigorous approach to choosing a distribution could be used.

The data should be fitted to a generalized three-parameter gamma distribution [9]. This distribution may be: Gamma, Lognormal or Weibull depending on the choice of parameters.

Fitting the data to this model would provide a way quantitatively assessing the goodness-of-fit. Unfortunately, we found no way to fit the parameters of such a model, but feel sure that some numerical method could be found.

**Rainfall.** Some way must be found to work deaths due to excessive rainfall into the model. There is also some suggestion in the literature that prolonged periods of high humidity may make the aphids more susceptible to fatal fungal infections.[10]
Appendix A
Degree-Day Approximations

If we let \( y(t) = \) temperature (in °C) at time \( t \), \( 0 \leq t \leq 24 \), and if \( y \) is a continuous function, then the accumulated degree-hours for the day will be the area between the curve \( y = y(t) \) and \( y = t_b \), where \( t_b \) is the base temperature (i.e. the temperature below which no development will take place). Then we can obtain degree-days by dividing degree-hours by 24. That is,

\[
DD = \frac{1}{24} \int_0^{24} (y - t_b) \, dt
\]

In actual practice, the function \( y = y(t) \) is not known. Instead, we must find an approximation for \( y \), and several methods exist for doing this. The methods discussed here have in common that they require knowing only the maximum and minimum temperatures for the day.

**Method 1**: We assume that \( y = (\text{MAX} + \text{MIN})/2 \). That is, we use a rough approximation of the average temperature for the day as an estimate for \( y \). Note that although this is a very crude approximation for \( y \), it may yield a good approximation for accumulated degree-days. This method has the advantage of quick and easy calculation, and has long been used with good
results. The formula is:

\[ DD = \frac{(\text{MAX} + \text{MIN})}{2} - t_b. \]

**Method 2:** Temperature appears to vary in a generally sinusoidal fashion. We can assume that \( y \) is a sine wave with minimum at \( t = 0 \) (midnight) and maximum at \( t = 12 \) (noon). We obtain

\[ y(t) = \frac{\text{MAX} - \text{MIN}}{2} \sin((12/\pi)t + 6) + \frac{\text{MAX} + \text{MIN}}{2}. \]

If we assume that \( y(t) \geq t_b \) for all values of \( t \), this will exactly the same approximation as Method 1. However, if \( y(t) \leq t_b \) for any value of \( t \), the graphs look like this:

The desired degree-days will be only the shaded areas. It appears that the two methods may not yield the same result. To obtain degree-days, we must first find \( t_1 \) and \( t_2 \). We can find \( t_1 \) by observing that \( y(t_1) = t_b \), or
\[ t_b = \frac{(\text{MAX} - \text{MIN})}{2} * \left[ \sin \left( \frac{(12/\pi)t_1 + 6}{\pi} \right) \right] + \frac{(\text{MAX} + \text{MIN})}{2}. \]

Solving for \( t_1 \) yields:

\[ t_1 = \frac{(\pi/12)}{\sin^{-1} \left[ \frac{t_b - \frac{(\text{MAX} + \text{MIN})}{2}}{\frac{(\text{MAX} - \text{MIN})}{2}} \right]} - 6 \]

This equation has infinitely many solutions, and we must choose one such that \( 0 \leq t_1 \leq 12 \). Then, by symmetry of the curve, \( t_2 = 24 - t_1 \). Thus, we can calculate

\[ \text{DD} = \frac{1}{24} \int_{t_1}^{t_2} (y(t) - t_b) \, dt. \]

Method 2 is also fairly easy to compute, and has been in general use until recent times.

**Method 3:** Allen [4] noted that an assumption of Method 2, that the minimum temperature occurs at midnight and that the maximum temperature occurs at noon, is generally untrue. As a rule of thumb, the minimum temperature will be at dawn (about 6 a.m.) and the maximum at about 2:00 p.m. (or \( t = 14 \)). Then the sinusoidal curve will look like this:
Note that we have constructed the wave by assuming that the maximum and minimum temperatures do not vary from one day to the next. If we allow these to vary from day to day, the curve becomes:

Then

\[
Y = \begin{cases} 
  m_1 + m_2 \sin\left(\frac{t + 48}{16}\right) & 0 \leq t \leq 6 \\
  m_3 + m_4 \sin\left(\frac{t + 6}{8}\right) & 6 \leq t \leq 14 \\
  m_5 + m_6 \sin\left(\frac{t - 6}{16}\right) & 14 \leq t \leq 24 
\end{cases}
\]
where

\[ m_1 = \frac{(\text{MAX}_1 + \text{MIN}_1)}{2} \quad m_4 = \frac{(\text{MAX}_2 - \text{MIN}_2)}{2} \]

\[ m_2 = \frac{(\text{MAX}_1 - \text{MIN}_1)}{2} \quad m_5 = \frac{(\text{MAX}_2 + \text{MIN}_2)}{2} \]

\[ m_3 = \frac{(\text{MAX}_2 + \text{MIN}_1)}{2} \quad m_6 = \frac{(\text{MAX}_2 - \text{MIN}_2)}{2}. \]

Now suppose that for some \( t, y(t) \leq t_b \). For example, suppose \( \text{MIN}_1 \leq t_b \). Then our curve looks like this:

![Graph showing curve and annotated areas with \( t_1, t_2, A_1, A_2 \).]

and we have

\[ DD = \frac{(A_1 + A_2)}{24}. \]
To find this area, we must first find $t_1$ and $t_2$. To find $t_1$, note that

$$t_b = m_1 + m_2 \sin\left[\pi\left(t_1 + 48\right)/16\right]$$

Solving for $t_1$:

$$t_1 = \left(16/\pi\right) \sin^{-1}[t_b - m_1/m_2] - 48.$$ 

Again, we must choose the appropriate value so that $0 \leq t_1 \leq 6$. If no such $t_1$ exists, then $A_1 = 0$. Otherwise,

$$A_1 = \int_0^{t_1} \left[m_1 + m_2 \sin\left(\pi(t + 48)/16\right) - t_b\right]dt.$$ 

This process must be repeated to find $t_2$ and $A_2$.

The situation becomes further complicated if either $\text{MAX}_1$ or $\text{MAX}_2$ falls below $t_b$.

At this point, the algorithm is not practical to use "by hand". As Allen points out, however, the method can be easily adapted to a computer. A subroutine to utilize this method is cumbersome to write, since so many different possibilities must be considered, but it will be quite efficient to run.
Appendix B

Model Program
This is a Fortran program to model aphid development times.

Name: HODEL

Written by: Marie Coffin

First attempt: 2/11/87

Latest revision: 3/20/88

INTEGER TAU, DD, N DAYS, STAGE, SUBSTG, GDD, DAY, MN, DY, YR, CONJUL, JDATE
INTEGER YTEMP, MTEMP, DTEMP, HTEMP, LTEMP
REAL A, F, FTEMP, FIN, TOTAL, FERT, MORT, EXTRA, M, P, R, HIGH, LOW
DIMENSION A(60,60), F(60), FTEMP(60), FIN(60), HIGH(60), LOW(60)
DIMENSION BTEMP(7), RELDD(6), TOTALS(7), ATEMP(60), INIT(6)
DIMENSION DEGDAY(7)

DATA A, ATEMP, DEGDAY/3600*0.0,60*0.0,7*0.0/
DATA BTEMP/6.29,6.54,10.36,7.59,6.56,8.17,0.0/

This is the initial cohort.

READ 5, MN, DY, YR
5 FORMAT (140,3(12,1X))
JDATE = CONJUL(MN, DY, YR)
PRINT 7
7 FORMAT ('1',P40,JDATE)
READ 10, INIT
10 FORMAT (27X,615).
READ 15, NHT, NHC
15 FORMAT (54X,215)
READ 20, NDAYS
20 FORMAT (29X,13/)
READ 25, YTEMP, MTEMP, DTEMP, HTEMP, LTEMP
25 FORMAT (5X,3(12,1X),2(13,1X))
WHILE (MTEMP, LT, MN, OR, DTEMP, LT, DY) DO
READ 25, YTEMP, MTEMP, DTEMP, HTEMP, LTEMP
ENDWHILE

HIGH(1) = CENTIG(HTEMP)
LOW(1) = CENTIG(LTEMP)
DO 110 I = 2, NDAYS
110 CONTINUE

WHILE (HIGH(I), LT, YTEMP, OR, LOW(I), GT, LOW(I), OR, LTEMP, LT, MTEMP) DO
110 CONTINUE

HIGH(I) = CENTIG(HTEMP)
LOW(I) = CENTIG(LTEMP)
DO 120 I = 2, NDAYS
120 CONTINUE

NEWDAY = NDAYS + 1
DO 120 I = 2, NEWDAY
120 CONTINUE
ADD = ACCDD([HIGH(I),LOW(I),HIGH(I-1),LOW(I+1),BASTMP(J)])
DEGDAY(J) = DEGDAY(J) + ADD
115 CONTINUE
ATEMP(I-1) = DEGDAY(7)
120 CONTINUE
C NSTEPS = INT(DEGDAY(7)/12.0)
DO 130 I = 1,6
REDD(I) = DEGDAY(I)/FLOAT(NSTEPS)
130 CONTINUE
REDD(6) = REDD(6) * 24.0
ABSDD = 12.0
C CALL DIST(F,INIT,REDD,TOTALS)
C PRINT 50,(TOTALS(I),I = 1,6)
PRINT 60,TOTALS(F)
C DAY = 1
DO 210 TAU = 1,NSTEPS
********** -***********************************
C THIS SECTION CALCULATES THE NUMBER
C THAT HAVE MATURATED TO THE NEXT LIFESTAGE
********** -***********************************
C DO 140 I = 1,60
FIN(I) = 0.0
FTEMP(I) = 0.0
140 CONTINUE
C DO 150 I = 1,50
CALL CALSTG(I,STAGE,STAGE)
GDD = INT(REDD(STAGE) * SUBSTG)
PROMAT = PHAT(STAGE,GDD)
NMAT = PROMAT * F(I)
LSTAGE = NXTSTG(I)
F(I) = F(I) - NMAT
FIN(LSTAGE) = FIN(LSTAGE) + NMAT
150 CONTINUE
C + THIS SECTION CALCULATES THE FERTILITY
C AND MORTALITY ELEMENTS OF THE A MATRIX
C +**********************************************************
C DO 160 J = 51,60
CALL CALSTG(J,STAGE,STAGE)
A(1,J) = FERT(SUBSTG,REDD(6))
BABIES = A(1,J) * F(J)
FTEMP(1) = FTEMP(1) + BABIES
160 CONTINUE
C DO 170 I = 2,60
J = I - 1
CALL CALSTG(J,STAGE,STAGE)
M = MORT(STAGE,STAGE,REDD(STAGE))
P = PRED(J,STAGE,TOTALS,F,ABSDD,HTH,HTC)
R = RAIN(J)
A(I,J) = 1 - M - P - R + M*P + M*R + P*R - M*P*R
SURVIV = A(I,J) * F(J)
IF (I.LT.51) THEN
  FTMP(I) = FTMP(I) + SURVIV
ELSE
  FRACT = SURVIV/24.0
  FTMP(I) = FTMP(I) + FRACT
  FTMP(J) = FTMP(J) + 23.0 * FRACT
ENDIF
CONTINUE

DO 180 I = 1,60
  F(I) = FTMP(I) + FIN(I)
CONTINUE

TOTALS(7) = 0.0
DO 200 I = 1,6
  TOTALS(I) = 0.0
DO 190 J = 1,10
  K = (I - 1) * 10 + J
  IF (F(K).LE.0) F(K) = 0
  TOTALS(I) = TOTALS(I) + F(K)
CONTINUE

TEMP = TAU * ABSDD
IF (TEMP.GE.ATEMP(DAY)) THEN
  DAY = DAY + 1
  JDATE = JDATE + 1
  PRINT 40,JDAT
  FORMAT (99,D15.6)
  PRINT 50, (TOTALS(I),I = 1,6)
  FORMAT ('11,6(F8.1,2X)')
  PRINT 60,TOTALS(7)
  FORMAT ('11,F12.1')
ENDIF
CONTINUE

STOP
END

FUNCTION PMAT(STAGE,TIME)

* THIS FUNCTION CALCULATES THE PROBABILITY *
* OF MATURING TO THE NEXT LIFE-STAGE, AS *
* A FUNCTION OF ACCUMULATED DEGREE-DAYS *

INTEGER STAGE,TIME

IF (STAGE.EQ.1) THEN
  B = 3.16987
  C = 9.19488
ELSEIF (STAGE.EQ.2) THEN
  B = 2.47683
  C = 8.96503
ELSEIF (STAGE.EQ.3) THEN
  B = 3.78789
  C = 4.14813
ELSEIF (STAGE.EQ.4) THEN
  B = 3.08195
  C = 7.2978
ELSEIF (STAGE.EQ.5) THEN
B = 6.73449
C = 1.9716
ELSE
B = 59.7507
C = 9.85034
ENDIF
PMAT = 0.0
DO 100 I = 1, TIME
PMAT = PMAT + F(I, B, C)
100 CONTINUE
RETURN
END
FUNCTION F(N, B, C)
*** THIS FUNCTION CALCULATES THE PROBABILITY *
*** DENSITY FUNCTION (GAMMA DISTRIBUTION) *
F = (N/B) ** (B-1) * EXP(-N/B)/(B * GAMMA(B))
RETURN
END
FUNCTION FERT(SUBSTG, RELDD)
*** THIS FUNCTION CALCULATES THE FERTILITY *
*** ASSOCIATED WITH THE NUMBER OF DD *
INTEGER SUBSTG
DO = RELDD * SUBSTG
FERT = (0.008 * DD)/(1.001 ** DD)
RETURN
END
REAL FUNCTION MORT(STAGE, SUBSTG, RELDD)
*** THIS FUNCTION CALCULATES THE MORTALITY *
*** ASSOCIATED WITH THE NUMBER OF DD *
INTEGER STAGE, SUBSTG
REAL DD
DIMENSION AVGDD(5)
DATA AVGDD/29.1466, 22.2013, 15.7149, 22.4915, 13.2777/
LAST = STAGE - 1
DO = 0.0
I = 1
WHILE (I.LE.LAST) DO
DO = DD + AVGDD(I)
I = I + 1
ENDWHILE
DO = DO + RELDD * SUBSTG
IF (DD.GE.1500) THEN
MORT = 1.0
ELSE
MORT = 1.0 - 0.908589 * EXP(-0.006861*EXP(0.006561*DD))
ENDIF
RETURN
END
FUNCTION ACCDD(MAX2,MIN1,MAX1,MIN2,TB)

C *********************************************************************************************************************
C * THIS FUNCTION CALCULATES ACCUMULATED DEGREE-DAYS FOR THE DAY. *
C *********************************************************************************************************************
C
REAL MIN1,MIN2,MAX1,MAX2,TB,M1,M2,M3,M4,M5,M6
DATA PI/3.14159/
C IF (MIN1.GE.TB) THEN
   Y1 = 3*(MAX1 + MIN1) - 1.5719925*(MAX1 - MIN1) - 6*TB
   Y2 = 4*(MAX2 + MIN1) - 8*TB
ELSE
   IF (MAX1.LT.TB) THEN
      Y1 = 0.0
   ELSE
      M1 = (MAX1 + MIN1)/2.0
      M2 = (MAX1 - MIN1)/2.0
      X1 = 16.0/PI * ARSIN((TB - M1)/M2) - 48.0
      CALL QUAD(X1)
      FACTOR = 1 + COS(PI/16.0 * (X1 + 48.0))
      Y1 = X1 * (M1 - TB) - 16.0/PI * M2 * FACTOR
   ENDIF
ENDIF
C IF (MAX2.LT.TB) THEN
   Y2 = 0
ELSE
   M3 = (MAX2 + MIN1)/2.0
   M4 = (MAX2 - MIN1)/2.0
   X2 = 8.0/PI * ARSIN((TB - M3)/M4) - 6.0
   CALL QUAD(X2)
   IF (X2.LT.6.0) X2 = 12.0 - X2
   FACTOR = COS(PI/8.0 * (X2 + 6.0))
   Y2 = (14 - X2) * (M3 - TB) + 8.0/PI * M4 * FACTOR
ENDIF
C ELSE
   IF (MIN2.GE.TB) THEN
      Y3 = 5*(MAX2 + MIN2) + 2.35264*(MAX2 - MIN2) - 10*TB
   ELSE
      IF (MAX2.LT.TB) THEN
         Y3 = 0.0
      ELSE
         M5 = (MAX2 + MIN2)/2.0
         M6 = (MAX2 - MIN2)/2.0
         X3 = 16.0/PI * ARSIN((TB - M5)/M6) - 6.0
         CALL QUAD(X3)
         IF (X3.LT.18.0) X3 = 36.0 - X3
         X3 = AMIN1(24.0,X3)
         FACTOR = -COS(PI/16.0 * (X3 - 6.0))
         Y3 = (24 - X3) * (M5 - TB) + 16.0/PI * M6 * FACTOR
      ENDIF
   ENDIF
ENDIF
C ACCDD = (Y1 + Y2 + Y3)/24.0
C RETURN
END

FUNCTION PRED(J,STAGE,TOTALS,APHIDS,DD,NHT,NHC)

C *********************************************************************************************************************
C * THIS FUNCTION COMPUTES THE PROBABILITY THAT AN INDIVIDUAL WILL BE KILLED *
C *********************************************************************************************************************
C
`REAL NAU, NCON
INTEGER STAGE
DIMENSION APHIDS(60), UNITS(6), TOTALS(7), AU(6)
DATA AU/1.0, 2.26, 5.08, 7.63, 12.2, 12.2/

NCA = 305.0 • (NHC + NHT)

IF (APHIDS(J).LT.0.01 .OR. TOTALS(STAGE).LT.0.01) THEN
  PRED = 0.0
ELSE
  TUNITS = 0.0
  DO 100 I = 1, 6
    UNITS(J) = TOTALS(I) * AU(I)
    TUNITS = TUNITS + UNITS(J)
  CONTINUE
  IF (TUNITS.GE.NCA) THEN
    NAU1 = 0.85 • DD
    NAU2 = 1.13 • DD
  ELSE
    NAU1 = 0.85 • DD • TUNITS / NCA
    NAU2 = 1.13 • DD • TUNITS / NCA
  ENDIF
  PRED = 1.0 • (1.0 • NAU1 / TUNITS) • NHT • (1.0 • NAU2 / TUNITS) • NHC
ENDIF
RETURN
END

FUNCTION RAII(I)
  RAII = 0.0
RETURN
END

SUBROUTINE DIST(F, INIT, LENGTH, TOTALS)
  REAL LENGTH(6)
  INTEGER TIME
  DIMENSION F(60), CDF(10), INIT(6), TOTALS(7)

  TOTALS(7) = 0.0
  DO 200 J = 1, 6
    TOTALS(J) = FLOAT(INIT(J))
  DO 100 I = 1, 10
    K = (J - 1) • 10 + I
    TIME = INt(I • LENGTH(J))
    CDF(I) = PMAT(J, TIME)
  IF (1.EQ.1) THEN
    ...
\[
\begin{align*}
F(K) &= CDF(1) \times \text{INIT}(J), \\
\text{ELSE} \quad F(K) &= (CDF(1) - CDF(I-1)) \times \text{INIT}(J) \\
\text{ENDIF}
\end{align*}
\]

100 CONTINUE

\[
\text{TOTALS}(7) = \text{TOTALS}(7) + \text{TOTALS}(7)
\]

C RETURN

END INTEGER FUNCTION CONJUL(MONTH, DAY, YEAR)

C *******************************************************
C THIS FUNCTION CONVERTS AN ORDINARY
C (GREGORIAN) DATE TO A JULIAN DATE.
C ************************************************************************

C INTEGER MONTH, DAY, YEAR, DAYSUM
DIMENSION NDAYS(12)
DATA NDAYS/31, 28, 31, 30, 31, 30, 31, 31, 30, 31, 30, 31 /

C IF (MOD(YEAR, 4) .EQ. 0) NDAYS(2) = NDAYS(2) + 1

C DAYSUM = 0
I = 1
WHILE (I LT MONTH) DO
DAYSUM = DAYSUM + NDAYS(I)
I = I + 1
ENDWHILE

DAYSUM = DAYSUM + DAY

CONJUL = 1000 * YEAR + DAYSUM
C RETURN
END

SUBROUTINE CALSTG(I, STAGE, SUBSTG)

C *******************************************************
C THIS SUBROUTINE TAKES THE SUBSCRIPT FROM
C THE I ARRAY AND CALCULATES THE STAGE
C AND SUBSTAGE CORRESPONDING TO IT.
C ************************************************************************

C INTEGER STAGE, SUBSTG, REM
REM = MOD(I, 10)
C IF (REM .EQ. 0) THEN
STAGE = I/10
SUBSTG = 10
ELSE
STAGE = I/10 + 1
SUBSTG = REM
ENDIF
C RETURN
END

FUNCTION NXTSTG(I)

C *******************************************************
C THIS FUNCTION CALCULATES WHAT STAGE THE
C INDIVIDUALS FROM ANY STAGE WILL MATURE
C TO (J-VALUES ARE CALCULATED, NOT STAGES).
C ************************************************************************
J = 0
WHILE (J*10.LT.1) DO
   J = J + 1
ENDDO
NXTSTG = J * 10 + 1
C
RETURN
END

SUBROUTINE QUAD(X)
C
RETURN
END

FUNCTION CENTIG(FTEMP)
C
RETURN
END
Appendix C

SAS Program to Fit Gamma Curves
PROC SORT DATA=A;
BY DEV1;
DATA B;
SET A;
IF DEV1 = . THEN DELETE;
ELSE NMAT = 1;
IF LAST.DEV1 THEN OUTPUT;
PROC MEANS MEAN STD SUM N NOPRINT;
VAR DEV1 NMAT;
OUTPUT OUT = STATS MEAN = MEANA MEANB
STD = STDA STDB
SUM = NA NB
N = NOBSA NOBSB;
DATA VARBLS; SET STATS;
MEAN = MEANA;STD = STDA;N = NB;NOBS = NOBSA;
B = STD**2/MEAN;
C = (MEAN/STD)**2;
KEEP MEAN STD B C N NOBS;
PROC PRINT DATA = VARBLS;
TITLE2 'PARAMETERS FOR FIRST LIFE-STAGE: GAMMA MODEL';
DATA C;
IF N = 1 THEN SET VARBLS;
SET B;
PROC PRINT DATA = C;
DATA COMPLETE;
SET C;
CF + NMAT/N;
PCF = 0;
DO X = 1 TO DEV1;
PCF + (X/B)**(C-1) * EXP(-X/B)/(B*GAMMA(C));
END;
OSQR = (CF - PCF)** 2;
SSE + OSQR;
YSUM + CF;
YSQR + CF ** 2;
IF N = NOBS THEN DO;
SST = YSQR - (YSUM ** 2)/NOBS;
RSQR = 1 - SSE/SST;
OUTPUT;
END;
PROC PRINT DATA = COMPLETE;
PROC SORT DATA=A;
BY DEV2;
DATA B;
SET A;BY DEV2;
IF DEV2 = . THEN DELETE;
IF FIRST.DEV2 THEN NMAT = 1;
ELSE NMAT = 1;
IF LAST.DEV2 THEN OUTPUT;
PROC MEANS MEAN STD SUM N NOPRINT;
VAR DEV2 NMAT;
OUTPUT OUT = STATS MEAN = MEANA MEANB
STD = STD0 STD00
SUM = NA NB
N = NOBSA NOBSB;
DATA VARBLS; SET STATS;
MEAN = MEANA;STD = STD0;N = NB;NOBS = NOBSA;
B = STD**2/MEAN;
C = (MEAN/STD)**2;
KEEP MEAN STD B C N NOBS;
PROC PRINT DATA = VARBLS;
TITLE2 'PARAMETERS FOR SECOND LIFE-STAGE: GAMMA MODEL';
DATA C;
IF _N_ = 1 THEN SET VARBLS;
SET B;
PROC PRINT DATA = C;
DATA COMPLETE;
SET C;
CF = NMAT/N;
PCF = 0;
DO X = 1 TO DEV2;
PCF + (X/B)**(C-1) * EXP(-X/B)/(B*GAMMA(C));
END;
DSQR = (CF - PCF) ** 2;
SSE + DSQR;
YSUM + CF;
YSQR + CF ** 2;
IF N = NOBS THEN DO;
SST = YSQR - (YSUM ** 2)/NOBS;
RSQR = 1 - SSE/SST;
OUTPUT;
END;
PROC PRINT DATA = COMPLETE;
PROC SORT DATA=A;
BY DEV3;
DATA B;
SET A;BY DEV3;
IF DEV3 = . THEN DELETE;
IF FIRST.DEV3 THEN NMAT = 1;
ELSE NMAT = 1;
IF LAST.DEV3 THEN OUTPUT;
PROC MEANS MEAN STD SUM N NOPRINT;
VAR DEV3 NMAT;
OUTPUT OUT = STATS MEAN = MEANA MEANB
STD = STD0 STD00
SUM = NA NB
N = NOBSA NOBSB;
DATA VARBLS; SET STATS;
MEAN = MEANA;STD = STD0;N = NB;NOBS = NOBSA;
B = STD**2/MEAN;
C = (MEAN/STD)**2;
KEEP MEAN STD B C N NOBS;
PROC PRINT DATA = VARBLS;
TITLE2 'PARAMETERS FOR THIRD LIFE-STAGE: GAMMA MODEL';
DATA C;
  IF _N_ = 1 THEN SET VARBLS;
  SET B;
  PROC PRINT DATA = C;
  DATA COMPLETE;
  SET C;
  CF + NMAT/N;
  PCF = 0;
  DO X = 1 TO DEV4;
    PCF + (X/B)**(C-1) * EXP(-X/B)/(B*GAMMA(C));
  END;
  DSQR = (CF - PCF)**2;
  SSE + DSQR;
  YSUM + CF;
  YSQ = CF**2;
  IF _N_ = NOBS THEN DO;
    SST = YSQ - (YSUM**2)/NOBS;
    RSQR = 1 - SSE/SST;
    OUTPUT;
  END;
  PROC PRINT DATA = COMPLETE;
  PROC SORT DATA=A;
  BY DEV4;
  DATA B;
  SET A; BY DEV4;
  IF FIRST.DEV4 THEN NMAT = 1;
  ELSE NMAT + 1;
  IF LAST.DEV4 THEN OUTPUT;
  PROC MEANS MEAN STD SUM N NOPRINT;
  VAR DEV4 NMAT;
  OUTPUT OUT = STATS MEAN = MEANA MEANB
      STD = STDA STDB
      SUM = NA NB
      N = NOBSA NOBSB;
  DATA VARBLS; SET STATS;
  MEAN = MEANA;STD = STDA;N = NB;NOBS = NOBSA;
  B = STD**2/MEAN;
  C = (MEAN/STD)**2;
  KEEP MEAN STD B C N NOBS;
  PROC PRINT DATA = VARBLS;
  TITLE2 'PARAMETERS FOR FOURTH LIFE-STAGE: GAMMA MODEL';
  DATA C;
  IF _N_ = 1 THEN SET VARBLS;
  SET B;
  PROC PRINT DATA = C;
  DATA COMPLETE;
  SET C;
  CF + NMAT/N;
  PCF = 0;
  DO X = 1 TO DEV4;
    PCF + (X/B)**(C-1) * EXP(-X/B)/(B*GAMMA(C));
  END;
  DSQR = (CF - PCF)**2;
  SSE + DSQR;
  YSUM + CF;
  YSQ = CF**2;
  IF _N_ = NOBS THEN DO;
    SST = YSQ - (YSUM**2)/NOBS;
    RSQR = 1 - SSE/SST;
    OUTPUT;
  END;
  PROC PRINT DATA = COMPLETE;
  PROC SORT DATA=A;
BY DEV5;
DATA B;
SET A; BY DEV5;
IF DEV5 = 1 THEN DELETE;
IF FIRST.DEV5 THEN NMAT = 1;
ELSE NMAT + 1;
IF LAST.DEV5 THEN OUTPUT;
PROC MEANS MEAN STD SUM N NOPRINT;
VAR DEV5 NMAT;
OUTPUT OUT = STATS MEAN = MEANA MEANB
STD = STDA STDB
SUM = NA NB
N = NOBSA NOBSB;
DATA VARBLS; SET STATS;
MEAN = MEANA; STD = STDA; N = NB; NOBS = NOBSA;
B = STD**2/MEAN;
C = (MEAN/STD)**2;
KEEP MEAN STD B C N NOBS;
PROC PRINT DATA = VARBLS;
TITLE2 'PARAMETERS FOR FIFTH LIFE-STAGE: GAMMA MODEL';
DATA C;
IF _N_ = 1 THEN SET VARBLS;
SET B;
PROC PRINT DATA = C;
DATA COMPLETE;
SET C;
CF + NMAT/N;
PCF = 0;
DO X = 1 TO DEV5;
PCF + (X/B)**(C-1) * EXP(-X/B)/(B*GAMMA(C));
END;
DSQR = (CF + PCF)**2;
SSE + DSQR;
YSQ = CF + PCF;
YSQR + CF**2;
IF N = NOBS THEN DO;
SST = YSQR - (YSQ**2)/N0BS;
RSQR = 1 - SSE/SST;
OUTPUT;
END;
PROC PRINT DATA = COMPLETE;
DATA DEATH;
INPUT T6 @@;
DEV6 = T6 * 13.84;
CARDS;
PROC SORT DATA = DEATH;
BY DEV6;
DATA B; SET DEATH;
BY DEV6;
IF FIRST.DEV6 THEN NDEAD = 1;
ELSE NDEAD + 1;
IF LAST.DEV6 THEN OUTPUT;
PROC MEANS MEAN STD SUM N NOPRINT;
VAR DEV6 NDEAD;
OUTPUT OUT = STATS MEAN = MEANA MEANB
STD = STDA STDB
SUM = NA NB
N = NOBSA NOBSB;
DATA VARBLS; SET STATS;
MEAN = MEANA; STD = STDA; N = NB; NOBS = NOBSA;
B = STD**2/MEAN;
C = (MEAN/STD)**2;
KEEP MEAN STD B C N NOBS;
PROC PRINT DATA = VARBLS;
TITLE2 'PARAMETERS FOR SIXTH LIFE-STAGE: GAMMA MODEL';
DATA C;
IF _N_ = 1 THEN SET VARBLS;
SET B;
PROC PRINT DATA = C;
DATA COMPLETE;
SET C;
CF * NDEAD/N;
PCF = 0;
DO X = 1 TO DEV6;
   PCF * (X/B)**(C-1) * EXP(-X/B)/(B*GAMMA(C));
END;
DSQR = (CF - PCF) ** 2;
SSE + DSQR;
YSUM + CF;
YSQR + CF ** 2;
IF _N_ = NOBS THEN DO;
   SST = YSQR - (YSUM ** 2)/NOBS;
   RSQR = 1 - SSE/SST;
   OUTPUT;
END;
PROC PRINT DATA = COMPLETE;
Bibliography


