Section IV

Applied Ecology of Invertebrate Pathogens

12

Modeling Insect Epizootics and their Population-Level Consequences

Bret D. Elderd

Department of Biological Sciences, Louisiana State University, Baton Rouge, LA, USA

12.1   Introduction

Mechanistic models of species interactions and disease transmission (e.g., Kermack and McKendrick, 1927; Rohani et al., 2002; Wootton, 2005) play an important role in helping us understand the patterns we see in nature and the processes responsible (Kendall, 2015). The importance of grounding ecological experiments in a testable theoretical framework is often understated (Scheiner, 2013), but can prove fruitful when done correctly. With regard to insect epizootics, interest in insect pathogens can be traced back to ancient China, where silkworm populations started to succumb to a then unknown pathogen. While the implications for infected individuals were obvious, the culprits were not discovered for centuries (Cory and Myers, 2003). In fact, it was only a few decades ago that insect pathogens were linked, via mechanistic models, to the boom-and-bust dynamics often seen in naturally occurring insect populations (Anderson and May, 1980). Early on, Brown (1987) described the importance of considering the theoretical perspective and using simulation models to gain a better understanding of epizootic dynamics; this field continues to grow (e.g., Onstad and Carruthers, 1990; Dwyer et al., 2000; Stuart et al., 2006; Hesketh et al., 2010; Shapiro-Ilan et al., 2012). The ideas and concepts associated with model formulation and testing have also continued to develop. These newly developed tools, along with standard analytical tools, will prove to be incredibly useful to any individual interested in insect epizootics, regardless of their training or research focus.

   In general, epizootic models tend to focus on either the short-term dynamics associated with a single epizootic (e.g., Dwyer et al., 1997) or the long-term population-level consequences of multiple epizootic events (e.g., Dwyer et al., 2004). By combining models with data regardless of the time scale considered, the study of insect epizootics has led to a better understanding of the processes responsible for driving the population dynamics of outbreaking insects. From a management/application perspective, the construction of mechanistic models helps to highlight the potential benefits or unforeseen consequences of using insect pathogens to control population outbreaks (Hochberg, 1989; Reilly and Elderd, 2014). From a broader perspective, insect epizootics have helped to develop both models and ideas that are central to a great deal of ecology and

to address fundamental ecological questions such as what causes populations to cycle (Barraquand et al., 2017). These insights continue to develop as new data are collected and models are refined.

Regardless of the approach one takes toward research, everyone uses models. There are three basic categories of models, which are not necessarily mutually exclusive: empirical, theoretical, and simulation-based (Hobbs and Hooten, 2015). The vast majority of science is conducted using empirical approaches. The empirical approach describes relationships and patterns between variables that have been measured and is exemplified by familiar statistical tools like regression and analysis of variance. Empirical approaches summarize relationships from a phenomenological perspective and not necessarily a mechanistic one. For example, in an empirical setting, a regression conducted on disease incidence and some independent variable (e.g., host density) may show that disease incidence increases as the independent variable increases. Here, the slope simply describes the rate of change or the pattern observed in the data, but, in this example, the model does not explicitly imply a mechanism. This measured relationship advances our understanding of the epizootic process but does not describe the mechanism driving the process.

On the other hand, theoretical models are mechanistic in nature. Historically, theoretical models have been considered overly simplistic, since they often focus on only a single interaction, or on a limited number of parameters. While theoretical models have provided a great deal of insight into ecological dynamics (e.g., the Lotka–Volterra equations), strictly theoretical approaches alone can easily become focused on the elegance of the method and lose sight of the ecology involved (Levin, 2012). However, when these models are confronted with data, the range of possible behaviors (e.g., population cycles or no cycles) quickly narrows and the models become quite powerful. Simulation-based models also take advantage of data that has been collected, but these models contain multiple interactions and can easily become relatively complex. Thus, simulation models are parameter-rich or high-dimensional. They often fit the data well, but it can be difficult to determine which of the many parameters drives the observed patterns (Elderd et al., 2006). Models constructed to understand epizootic dynamics or other ecological processes do not have to fall into a single category and may draw from multiple categories (Hobbs and Hooten, 2015). Thus, the three categories serve as useful heuristics for thinking about model development and linking models to data.

These three modeling perspectives are often seen as strictly distinct from other. However, this viewpoint draws on historical precedence and does not take into account how prevalent it has become to combine mechanistic models of increasing complexity with data. In the world of modeling insect epizootics, when Anderson and May (1980) compared theoretical model outcomes to empirical data, they opened up a new window from which to see the world. In their frequently cited works, Anderson and May (1979, 1980) combined ideas from classic predator–prey models and susceptible–infected–recovered (SIR) models, which essentially combined long-term dynamic models with short-term epizootic models, to understand natural observations of insect populations. Using this approach, Anderson and May (1980) demonstrated that pathogens could be responsible for the boom-and-bust cycles associated with the long-term dynamics of insect populations. They continued to expand upon the mechanisms responsible for short-term epizootic events and long-term population dynamics for species controlled by pathogen outbreaks (Anderson and May, 1981). Thus, they started a rich literature in

which the idea of combining models with data serves as a cornerstone for understanding insect epizootics.

In their original work, Anderson and May (1980) showed that insect population models that invoked host–pathogen interactions qualitatively displayed the same dynamics as observational data collected in the field. Following this, more rigorous methods of analyzing observational time series and field data began to take hold. These methods often advocated a likelihood-based approach that simply asked how likely were the data that had been collected to be correct, given the predicated dynamics from the model. If the model did not do well in terms of predicting the actual data, it was either set aside or refined, and the investigative process continued. The model, in this instance, can be considered equivalent to a hypothesis that is tested with data. The framework for this likelihood-based approach culminated in the influential book, *The Ecological Detective* (Hilborn and Mangel, 1997), which advocated not only confronting models with data but also testing multiple models at once. One method for doing so, which comes from the field of information theory, is the Akaike information criterion (AIC) (Burnham and Anderson, 2002). The AIC operates on the principle of parsimony to choose the best model. The most complicated model (i.e., the model with the most parameters) will always fit the data better (i.e., the model will be more likely) than less complicated ones. However, fitting the more complicated model comes at a cost, because a more complicated model will do a poor job of predicting the next set of data to be collected. Thus, the AIC balances between model fit and model complexity (Section 12.4.2). Just as the AIC draws upon likelihood-based methods, so Bayesian methods are based on likelihood approaches. Bayesian methods, particularly hierarchical Bayesian models, have become increasingly popular and provide the flexibility to analyze data and compare multiple models (Hobbs and Hooten, 2015). For a worked example from the Bayesian perspective, see Section 12.5.1. Overall, a lot of progress has been made since the initial explorations of purely theoretical models and the fitting of empirical models to data, and this progress will continue in the future. While likelihood-based frameworks work well for short-term epizootic events, different tactics are often required when examining long-term boom-and-bust cycles (Section 12.6.2) (e.g., Kendall et al., 1999; Turchin, 2003; Dwyer et al., 2004). Together, each of the methods mentioned here that focuses on combining models with data represents a suite of tools for understanding what drives epizootics over both the short and the long term.

12.2   The Pathogen and its Hosts

While there are a number of diseases described throughout this book that have a great deal of impact on invertebrate populations and epizootic dynamics, I will focus on diseases caused by baculoviruses (see also Chapter 7). Although the focus here is on baculoviruses, models have been developed to describe epizootic dynamics for a variety of insect pathogens. These include, but are certainly not limited to, fungal pathogens (e.g., Edelstein et al., 2005; Scholte et al., 2005; Hesketh et al., 2010) and nematodes (e.g., Stuart et al., 2006). For example, Scholte et al. (2005) used an enthomopathogenic fungal model to highlight the efficacy of a vector control method to decrease the population of a malaria vector, *Anopheles gambiae*. Thus, even though baculoviruses provide the biological motivation for model development here, the

methods outlined can be applied to many other biological systems. However, since all good mechanistic models need to be motivated by the biology of the system, baculoviruses represent a good place to start, given their importance in driving epizootic dynamics (Cory and Myers, 2003) and the use of mechanistic models in describing these dynamics (e.g., Dwyer et al., 1997; Elderd et al., 2013). To reiterate, while the biology and the associated models throughout draw on baculoviruses as examples, the methodologies discussed have quite a broad use in enhancing our understanding of epizootic dynamics as a whole.

Baculovirus infections begin when a susceptible individual consumes occlusion bodies (OBs), often containing multiple copies of the virus. If enough OBs are consumed, the individual becomes fatally infected. Sublethal or covert infections also occur (Roy et al., 2009), but at relatively low levels (Myers et al., 2000). Covert infections may contribute to the persistence of pathogens at low host densities (Roy et al., 2009) and function in a manner similar to vertical transmission between mothers and their offspring, which also allows pathogens to persist at relatively low host densities (Anderson and May, 1981). However, covert infections likely do not drive the boom-and-bust cycles associated with epizootic dynamics. If a lethal rather than a sublethal infection occurs, the infection process moves through a number of stages before the death of the host, which can release millions of OBs into the environment; transmission resulting from a sublethal infection would be minimal in comparison unless that infection became lethal.

In Lepidoptera, the infection process begins when a larva consumes foliage on which OBs reside. Upon entering the midgut, the outer protein coat of the OB dissolves, releasing the virions. The virions then infect the host's midgut cells and the infection eventually becomes systemic. After a period of time, the virus essentially liquefies the internal structure of the host as a result of producing more virus. Upon the host's death, the outer larval integument splits open and releases virus into the environment to begin the infection process again (Cory and Hoover, 2006; Elderd, 2013). Over time, the virus degrades due to ultraviolet (UV) light exposure.

An epizootic can begin when first-instar larvae become infected (Dwyer et al., 1997). However, there is a delay between infection of the first instars and release of the next round of virus. During this delay, which varies depending upon the virus and the species infected, the healthy larvae molt to second, third, or fourth instars. Due to the baculovirus infection, the infected individuals do not molt (Miller, 1997). After the infected larvae die and OBs are released on to the surrounding leaf tissue, the larger healthy larvae that have developed into later instars consume the virus, since it now resides on the leaf tissue on which they are feeding, and disease prevalence increases. The epizootic ends due to either larval pupation or "epizootic burnout," which occurs when there is a lack of infected individuals in the population to continue the spread of the disease (Dwyer et al., 2000; Fuller et al., 2012). The delay between infection and host death has important consequences for both the short- and the long-term dynamics of pathogen-driven host populations, such as lepidopteran species prone to epizootics.

While observational data in these systems can be readily collected, baculovirus systems also lend themselves to experimental approaches that test hypotheses surrounding transmission dynamics. To initiate an experiment that follows the natural progression of a baculovirus-driven epizootic, first instars can be lethally infected with a dose of OBs. Once infected, the first instars are confined on an experimental plant or branch using mesh bags. Mesh bags stop larvae from escaping and prevent the virus from

degrading due to UV light exposure. The infected individuals then die and, after death, release OBs on to the leaf tissue. Once the first instars have died, healthy third or fourth instars are placed into the mesh bag. These individuals are then allowed to feed for a period of time. Afterwards, the larvae are collected and reared in individual cups until death or pupation (Dwyer et al., 1997; Elderd et al., 2008; Elderd and Reilly, 2014). Infection can be easily diagnosed visually given the drastic manner in which the infection process slowly consumes the larva. Additionally, since the OBs are quite large and can be seen under a light microscope (Elderd, 2013), any potential infections can be readily confirmed. In the simplest approach, one can manipulate the amount of pathogen in the system (the independent variable) and record the fraction of insects surviving (the dependent variable). Thus, experiments that manipulate multiple factors such as temperature and the amount of pathogen in the system (Elderd and Reilly, 2014) can be readily performed. The data produced can then be combined with any suite of models to test the associated hypothesis.

12.3   Modeling Disease Transmission: A Single Epizootic

The models used to understand short-term epizootic dynamics associated with a single event can be traced back to Kermack and McKendrick (1927), who developed the SIR model to describe epidemic dynamics. Instead of SIR dynamics, baculovirus systems consist of susceptible individuals, infected individuals, and pathogen, since there is little evidence that infected individuals recover. If the simplifying assumption is made that all baculovirus infections are lethal, we need only consider the number of susceptibles and the amount of pathogen in the system, since all infected individuals eventually become pathogen (Dwyer et al., 2000). This assumption is met by the experimental methods described earlier. Mathematically, the equation for the susceptible larvae S takes the form of the following differential equation:

$$\frac{dS}{dt} = -\beta SV. \tag{12.1}$$

Here, the change in susceptible larvae over time is simply a product of the disease transmission coefficient β times the number of susceptibles S and the amount of virus in the system V. The transmission parameter β encompasses the whole of the infection process and can be thought of as the fraction of encounters between the virus and a susceptible larva that leads to an instantaneous infection. As with all models (empirical, mechanistic, or simulation-based), it is important to consider all the assumptions. The main ones for the model of susceptible populations given here are that per capita transmission $\left(\text{i.e.,} \ \frac{1}{S}\frac{dS}{dt}\right)$ is linear and that all individuals are equally susceptible to becoming infected. Relaxing this assumption, or changing the model structure to better fit the biology of the system and the data, leads to new insights into the transmission process. Equation 12.1, however, serves as a useful starting point.

By integrating equation 12.1 and using experimental data, estimates of the transmission rate β can be easily calculated. In an experiment, the amount of virus or the number of cadavers in the system at the beginning of the experiment $V(0)$ is known, as

is the initial number of susceptibles in the experimental treatment $S(0)$. Here, 0 refers to the start of the experiment. After conducting the experiment until time T, the number of susceptible individuals (i.e., the number of individuals that pupate rather than die from an infection) is also known, $S(T)$. These data can be easily plugged into the integral of equation 12.1, which is integrated from time 0 to T. The integral of equation 12.1 is simply:

$$-\ln\left[\frac{S(T)}{S(0)}\right] = \beta V(0)T. \tag{12.2}$$

Thus, by regressing the cadaver density against the negative natural log of the fraction uninfected ($-\ln\left[S(T)/S(0)\right]$) with no intercept term, an estimate of the transmission rate can be calculated directly from the data (Elderd et al., 2008); this is simply the slope of the line (Fig. 12.1, dashed line). But unlike standard regression models, which calculate the slope (and the intercept) as a phenomenological relationship, here, the slope is linked directly to the disease transmission rate β. Thus, the mechanism can be explicitly inferred from the mathematical model of the process.

12.3.1 Phenomenological and Mechanistic Models

While equation 12.1 represents a simple model of transmission dynamics, it is a reasonable representation of epizootic dynamics, and it and similar forms have been used extensively (e.g., Hochberg, 1989; Hochberg and Waage, 1991; Boots, 2004). However, the linear model does not always fit the data collected. Some baculovirus epizootic data show a decidedly nonlinear or curvilinear fit (Dwyer et al., 1997; Elderd and Reilly, 2014), such that infection rates at higher pathogen levels are less than expected if the
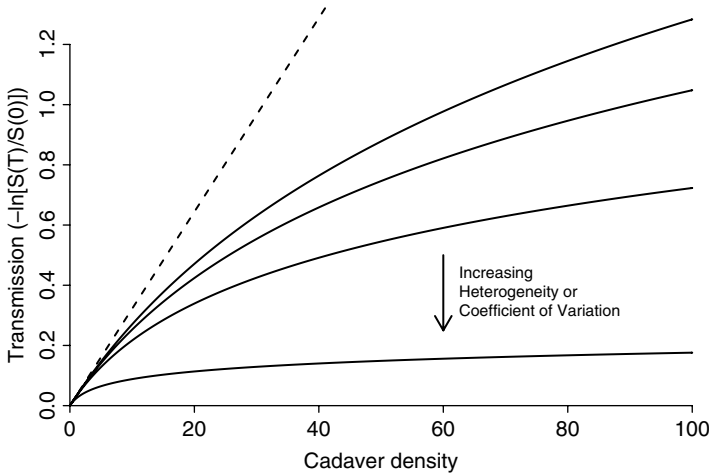

Fig. 12.1 Effect on transmission of increasing the transmission rate's coefficient of variation C in equation 12.5. The solid lines represent populations in which risk varies across individuals. The dashed line represents a population in which all individuals are equally at risk (equation 12.2). Each line uses the same value for the transmission rate of the virus.

linear model held true (equation 12.1). A simple solution to this problem would be to raise the number of susceptibles or the amount of virus by a power (Hochberg, 1991), which would result in a nonlinear model that could better fit the data. This phenomenological model then takes the form:

$$\frac{dS}{dt} = -\beta S^g V^h.$$
(12.3)

Here, g and h are the nonlinear effects on transmission of susceptible and infected population densities, respectively (Hochberg, 1991). However, while this power model will fit the nonlinear data better, the biological mechanism or mechanisms driving the nonlinear fit remain unknown. In this instance, what exactly does g or h mean from a biological standpoint?

A potential mechanism that may drive the nonlinearity in infection rates goes back to one of the main of assumptions of the linear model: that all individuals have the same transmission rate β. In Dwyer et al. (1997), the authors assumed that individuals differ in their susceptibility to virus. Essentially, some individuals are more susceptible than average and others are less susceptible than average. Thus, there was not a single transmission rate, but a mean transmission rate with some variability about the mean rate. Therefore, the transmission rate became a distribution rather than a single point estimate. The modified equation accounting for differences in susceptibility (i.e., heterogeneity in the transmission rate) thus becomes:

$$\frac{dS}{dt} = -\bar{\beta}\left[\frac{S(t)}{S(0)}\right]^{C^2} SV.$$
(12.4)

Here, $\bar{\beta}$ is the mean transmission rate. The transmission rate is scaled by the ratio of the number of susceptibles currently in the population $S(t)$ divided by the number of susceptibles at the start of the epizootic $S(0)$. The ratio is raised to the square of the coefficient of variation C associated with the transmission rate. Integrating equation 12.4 results in:

$$-\ln\left[\frac{S(T)}{S(0)}\right] = \frac{1}{C^2}\ln\left(1 + \bar{\beta}C^2 V(0)T\right).$$
(12.5)

Here, T is once again the time that the experiment ran. For equation 12.5, instead of estimating just β from the data, two parameters need to be estimated, $\bar{\beta}$ and C. For any single level of heterogeneity, at low pathogen levels, highly susceptible individuals become infected and transmission rises quickly (Fig. 12.1, solid lines). However, as pathogen levels increase, transmission tapers off, since only highly resistant individuals remain in the population. As the heterogeneity in the population increases, the coefficient of variation C increases, which results in fewer individuals becoming infected at the end of the epizootic as pathogen levels increase (Fig. 12.1). If, instead, C decreases and goes to zero (i.e., little variability in C), the dynamics become similar to the linear equation (equation 12.2). While equation 12.5 was developed with epizootics in mind, it borrows from work by Anderson and May (1991) on HIV spread and how varying contact rates influence HIV transmission. Thus, equation 12.4 represents

another example of the give and take between epizootiological and epidemiological research.

Once a model is developed, it is important to test it. Dwyer et al. (1997) exemplified this approach by showing the stepwise process of confronting models with data. In a series of experiments on the invasive gypsy moth (*Lymantria dispar*) and its species-specific baculovirus, Lymantria dispar multinucleopolyhedrovirus (LdMNPV), the authors tested whether the linear (equation 12.1) or the nonlinear (equation 12.4) model explained the data better, using a series of experimental epizootics. However, it should be noted that baculoviruses do not represent the only pathogen in the system. *Entomophaga maimaiga*, a fungal pathogen, also infects gypsy moth larva (Hajek, 1999), but infection rates can be either density-independent (Liebhold et al., 2013) or density-dependent (Hajek et al., 2015) according to the weather conditions (Hajek and van Nouhuys, 2016). LdMNPV, unlike *E. maimaiga*, is always strongly density-dependent (Liebhold et al., 2013). Thus, the linear and nonlinear models focused on baculovirus transmission along with host and pathogen densities (Dwyer et al., 1997) are appropriate given the biology of the baculovirus system.

For short-term epizootic events, the experimental data clearly support the nonlinear model for the gypsy moth (see Dwyer et al., 1997, fig. 3). Interestingly, in a comparison of lab-reared larvae with feral larvae, the degree of heterogeneity in transmission was much less in the former. Since the lab-reared larvae were not exposed to virus, the authors hypothesized that the level of heterogeneity should be less than for gypsy moth larvae reared from eggs collected from the wild. As new models have been developed and tested with empirical data, new mechanistic insights into what drives gypsy moth baculovirus epizootics continue to be gained. Other models have shown that rapid host evolution (Elderd et al., 2008), interactions between plant defensive chemicals and virus (Elderd et al., 2013), and heterogeneity in the pathogen (Fleming-Davies et al., 2015) are all important factors driving gypsy moth epizootic dynamics. All of the preceding models rest upon the mechanistic backbone of equation 12.4, which has continued to move the field forward. They also highlight that by combining mechanistic model development with experimental data, new insights can be continually gained.

12.4   Fitting Models to Data

To determine whether heterogeneity or other factors can be invoked as drivers of epizootic dynamics, a model must be compared to either experimental or observational data. Standard approaches fit the model using the well-trodden path of frequentist statistics. These methods determine fit either through the amount of variation explained or through the significance of a term in the model (e.g., the slope in a regression), as dictated by a null-hypothesis test and its subsequently generated p-value. However, the use of p-values continues to fall out of favor, as evidenced by a policy statement from the American Statistical Association (ASA) (Wasserstein and Lazar, 2016). This statement cautions against overreliance on the p-value as a valid statistic to indicate whether an effect drives the patterns seen in data. Additional problems arise when using this approach to compare multiple non-nested models. Often, when there are multiple mechanisms suspected of driving the dynamics in a variety of nonlinear or linear ways, non-nested models quickly accumulate. The question of how best to compare multiple

models to the data remained a problem until information theory began to gain a foothold in the wildlife literature (Anderson et al., 2000) and was highlighted in two influential books (Hilborn and Mangel, 1997; Burnham and Anderson, 2002).

12.4.1 Akaike Information Criterion

An information-theoretic approach to data analysis became widely used after the publication of Burnham and Anderson (2002). This approach allows a researcher to compare multiple models (i.e., alternative hypotheses) and determine which best fit the data. This is in direct contrast to classical statistics, which focuses on either accepting or rejecting a null hypothesis. The rejection of the null hypothesis simply means that the null model does not fit the data and is not an acceptance of the alternative model. Thus, all of our inference is based on the null model. In comparing multiple models, Burnham and Anderson (2002) focused on the use of the AIC, which operates according to the principle of parsimony or Occam's razor. That is, in deciding which is the best model, the researcher must shave away all that is unnecessary. Thus, in constructing a mechanistic model, one wants to construct a model with the smallest number of parameters that best explains the data. Essentially, there needs to be a balance between underfitting (i.e., too few parameters) and overfitting (i.e., too many parameters) models.

Hirotu Akaike, a Japanese statistician, developed a simple formula that corrects for constructing models that are too simple or too complex. The formula states:

$$\text{AIC} = -2L\left(\text{Data}\,|\,\Theta\right) + 2K. \tag{12.6}$$

Here, $L\left(\text{Data}\,|\,\Theta\right)$ is the log likelihood of the data given the model parameters Θ, and K is the number of parameters in the model (Burnham and Anderson, 2002). In a standard regression model, the log likelihood of the slope and the intercept is often calculated using the sum of squares of the difference or the error between the data and the model's predictions, assuming that the error is normally distributed. To calculate the log likelihood for data associated with infections, the error in the model often follows a binomial distribution, since the data are counts of infected and non-infected individuals (e.g., Elderd and Reilly, 2014). By using the equation 12.6 and, if necessary, the associated correction for small sample sizes AIC_c (see Burnham and Anderson, 2002), the raw AIC score can be calculated. The model with the lowest score is the best-fit model. Models with too few parameters are less likely and have a low log likelihood. Models with too many parameters (i.e., larger values of K), while fitting the data better, are penalized by adding to their AIC score. The best model thus represents a balance between model fit and complexity.

To fully gauge the degree of support for a model or for one of the alternative hypotheses, the differences between the best-fit model and the other models – the ΔAIC – should first be calculated. The formula for calculating the ΔAIC is:

$$\Delta\text{AIC}_i = \text{AIC}_i - \min\left(\text{AIC}\right), \tag{12.7}$$

where i is the model being considered and min(AIC) is the minimum of all AIC model scores. Thus, the best-fit model, which is the model with the lowest AIC score, has a ΔAIC of 0. Models that have $\Delta\text{AIC} > 10$ are considered poor fits to the data, those

with values between 4 and 7 have little support, and those with values greater than 0 but less than 2 have substantial support (Burnham and Anderson, 2002).

ΔAIC scores can, in turn, be used to calculate AIC weights, which are the weights of evidence for the relative likelihood of particular models given the models considered. AIC weights are calculated using:

$$w_i = \frac{\exp(-0.5\Delta\mathrm{AIC}_i)}{\Sigma_{r=1}^{R}\exp(-0.5\Delta\mathrm{AIC}_r)}, \tag{12.8}$$

where w_i is the weight of evidence for model i given all R models (i.e., better models are reflected by higher weights). These weights allow for a direct comparison between alternative hypotheses and can be used to gain further insight via multimodel inference (Burnham and Anderson, 2002).

12.4.2 An Example of the AIC in Action

To give a concrete example of the use of the AIC in analyzing epizootic data, I will draw on a series of experiments examining the effects of global climate change on baculovirus transmission in the fall armyworm (*Spodoptera frugiperda*) (Elderd and Reilly, 2014). The fall armyworm is a multivoltine crop pest that overwinters in Florida and Texas (Pitre and Hogg, 1983). As springtime temperatures increase, the fall armyworm reinvades the entire extent of its range by migrating northward until it reaches Ontario, Canada. As adults, female fall armyworms lay eggs in clusters. After the eggs hatch, there are six larval instars (Pitre and Hogg, 1983). They then pupate for 7–37 days depending upon the temperature (Sparks, 1979), emerge to mate, and continue their lifecycle. The species, like many lepidopterans, exhibits boom-and-bust dynamics. As the population increases during the boom phase, infestations occur, which can be widespread (Fuxa, 1982) and potentially devastating to farmers (Hinds and Dew, 1915).

Spodoptera frugiperda nucleopolyhedrovirus (SfNPV), a species-specific baculovirus, represents an important mortality source for the fall armyworm (Richter et al., 1987). Prior to an epizootic, a viral reservoir in the soil provides the initial inoculation of virus (Fuxa and Geaghan, 1983). After 4–6 days, initially infected larvae die (De Oliveira, 1999), while uninfected larvae grow to third or fourth instars (Pitre and Hogg, 1983). The older instars become infected by consuming the contaminated foliage on which the first instars died. Over time, virus particles degrade due to UV light exposure (Miller, 1997). The epizootic dynamics in the fall armyworm are very similar to those in other baculovirus-driven populations, like the gypsy moth. Thus, we can start with the same base model and modify it to answer whatever questions are posed. In this instance, how will rising temperatures affect transmission dynamics?

To determine how increasing temperatures affect epizootic dynamics, we established a series of control and experimental plots. In the experimental plots, we manipulated temperature using open-top chambers (OTCs) (Marion et al., 1997), which significantly raised temperatures in experimental as compared to control plots (Elderd and Reilly, 2014). In each of the $40.1\,\mathrm{m}^2$ plots, we placed a single soybean plant (*Glycine max*) of the same variety. Each plant was covered in a mesh bag, and a varying number of first-instar baculovirus-infected larvae (0, 15, 30, or 60) were placed on top. Once the first instars died, we placed 20 healthy fourth instars on the plant and allowed them

Table 12.1 AIC and WAIC results for the single-epizootic model. AIC_c scores, ΔAIC_c, AIC_c weights, and WAIC for each model. For the models considered, climate effect is due to OTCs raising temperatures in the experimental plots. The nonlinear model (equation 12.5) assumes heterogeneity in disease risk. For the linear model (equation 12.2), when $C = 0$, no difference in risk is assumed. $\bar{\beta}$ refers to estimates of mean transmission rate. Best-fit model is in bold.

Model	AIC_c	ΔAIC_c	AIC_c wt	WAIC
1. No climate effect, linear model	57.8	10.2	0.00	108.8
2. No climate effect, nonlinear model	52.7	5.1	0.03	106.4
3. Climate effect, linear model	52.5	4.9	0.04	107.4
4. Climate effect, nonlinear model	50.1	2.5	0.12	106.2
5. Climate effect, differences in C only	**47.6**	**0.0**	**0.42**	**103.3**
6. Climate effect, differences in $\bar{\beta}$ only	48.6	1.0	0.25	106.1

Note: AIC_c is based on an information theoretic approach to comparing multiple models. WAIC can be considered the Bayesian equivalent.

to feed. After 2 days, we collected the larvae and reared them until pupation or death. Baculovirus deaths were confirmed and recorded.

For heuristic purposes, in order to demonstrate the utility of the AIC approach, I will only consider a set of simple models (Table 12.1) based on equations 12.2 and 12.5. Additionally, although the experiment was conducted three separate times over the course of a number of years, I will only analyze the data from a single year. A complete analysis of the experiments and data is presented in Elderd and Reilly (2014).

Using the AIC approach for the climate-change experiment, the best-supported model, given the plausible models considered, is the one where $\bar{\beta}$ is the same for both treatments but the treatments have different values of C (Table 12.1). Note that the results are presented in terms of the small sample correction AIC_c, as recommended given the sample size (Burnham and Anderson, 2002). By examining ΔAIC_c, the support for the best-fit model (as compared to the null models) is strong, given that the null linear and nonlinear models have ΔAIC_c values between 5 and 10. Thus, increasing temperature has an effect on disease transmission. This does not hold true for all of the models considered. There is also support for the model where only $\bar{\beta}$ differs. The same pattern can be seen when examining the AIC_c weights (Table 12.1). For all of the data associated with these experiments, the general trend holds that C differs and that as temperatures rise, C exponentially decreases (Elderd and Reilly, 2014). Thus, as the climate warms, the nonlinear dynamics of the host–pathogen interaction become more and more similar to those of interactions governed by the linear model (Fig. 12.1).

12.5   A Bayesian Approach

Bayesian analysis has become another increasingly popular approach to fitting models to data. Like information-theoretic approaches and classic statistical approaches, Bayesian approaches are likelihood-based. That is, the results of the analysis hinge on how likely the data are given the model. However, there is an added component: prior

information about the system. The basis for the approach stems from Bayes' theorem, which states:

$$P(\Theta \,|\, \text{Data}) \propto \pi(\Theta)\mathcal{L}(\text{Data} \,|\, \Theta) \tag{12.9}$$

where the posterior probability of the model parameters Θ given the data is proportional to ($\propto$), the prior probability of the parameters $\pi(\Theta)$ times the likelihood of the data given the model parameters $\mathcal{L}(\text{Data} \,|\, \Theta)$. In the past, the implementation of a Bayesian approach was often limited due to the complexity of the computations associated with the analysis. Recently, a proliferation of Bayesian books with ecological perspectives (e.g., Clark, 2007; Kéry, 2010; Hobbs and Hooten, 2015) and the availability of freeware programs (e.g., WinBugs, JAGS, STAN) have made Bayesian approaches much more accessible.

A distinct advantage of Bayesian methods is that they provide a framework for incorporating prior information about a system (e.g., preliminary studies), which is especially valuable when data are sparse. Typically, prior information enters into the classical analysis framework in the discussion when the authors state whether their current findings are similar to or different from those of previous studies (Hille Ris Lambers et al., 2005). In a Bayesian approach, the prior contains quantitative information and becomes a parameter in the analysis ($\pi(\Theta)$ in equation 12.9). If no prior information is available, vague priors can be used, which contain relatively little information. Explicitly stating a prior can be controversial to some, but if individuals are uncomfortable selecting a prior, the easiest way to minimize prior influence is to overwhelm it with data (Hobbs and Hooten, 2015). However, the use of informed priors makes the most of previously hard-won data and represents a powerful approach to developing mechanistic models for understanding epizootic dynamics.

A fundamental difference between a Bayesian approach and more classical approaches stems from the difference in how the parameters are treated. Classic frequentist approaches assume that a parameter's value is fixed and that the exact estimate becomes better resolved as sample size increases (Hobbs and Hooten, 2015). In contrast, Bayesian approaches assume that a parameter is a random variable drawn from a distribution. This is the difference between a single value for quantifying disease transmission rates, which is estimated with increasing precision, and a distribution of uncertainty reflecting the inherent variability of the transmission rate (Ellison, 2004; Hobbs and Hilborn, 2006). A more in-depth examination of Bayesian analysis from a philosophical perspective, as touched upon earlier, can be found elsewhere in the literature (e.g., Dennis, 1996; Ellison, 1996, 2004).

12.5.1 Fitting a Bayesian Model

For the linear model (equation 12.1), and assuming there is no difference in the treatment effects, a simple Bayesian model can be constructed such that:

$$y_i \sim \text{binomial}(p_i, N_i), \tag{12.10}$$

$$\ln(p_i) = -\beta V(0)T, \tag{12.11}$$

$$\beta \sim \text{lognormal}(0, 1000). \tag{12.12}$$

The number of survivors or non-infected larvae is distributed ($\sim$) binomially with a probability p_i given an initial number of healthy larvae N_i. Here, p_i is simply the fraction of uninfected larvae ($S(T)/S(0)$). Thus, equation 12.11 is equivalent to equation 12.1. The disease transmission rate β has a prior probability that is log normally distributed with a mean of 0 and a variance of 1000. Thus, the prior is considered vague and contains little information. The resulting posterior for each replicate i becomes:

$$\overbrace{P(\beta \mid y_i)}^{\text{Posterior}} \propto \underbrace{\text{binomial}\Big(y_i \mid e^{-\beta V(0)T}, N_i\Big)}_{\text{Likelihood}} \overbrace{\text{lognormal}\big(\beta \mid 0, 1000\big)}^{\text{Prior}}, \qquad (12.13)$$

which, following equation 12.9, explicitly shows the relationship between the posterior and its Bayesian components, the likelihood and the prior.

To obtain posterior estimates of the disease transmission rate, one needs to fit the models (Table 12.1) to the data using Markov chain Monte Carlo (MCMC) methods. This can be done via a freeware package called JAGS (or "Just another Gibbs sampler") (Plummer et al., 2003), which can be run directly in R (Yu-Sung and Masanao, 2015), or else by writing the MCMC code directly in R (R Core Team, 2015). The standard approach to constructing the posterior consists of running multiple chains at various starting points, trimming their beginnings, and combining them. For each model (Table 12.1), five separate chains, each for 50 000 iterations, were run in JAGS, with the initial conditions for each chosen randomly. The first 10 000 iterations were discarded as "burn-in" to eliminate any transients associated with the initial conditions. All other iterations were retained to serve as estimates of the posterior distribution. The chains were not thinned, where thinning entails keeping only every *m*th iteration of the chain and discarding all others, as can be common practice. Link and Eaton (2012) showed that thinning is inefficient and reduces the precision of the parameter estimates. Using standard metrics, MCMC convergence was assessed by examining both within-chain and between-chain convergence. If the chains do not converge, the model is doing a poor job of estimating the parameters associated with the data. Two common metrics involve calculating the Brooks–Gelman–Rubin and the Heidelberger–Welch diagnostics. The Brooks–Gelman–Rubin statistic compares within- and between-chain variation (Brooks and Gelman, 1998), with values less than 1.1 indicating good between-chain convergence (Gelman and Hill, 2006). The Heidelberger–Welch diagnostic tests for stationarity (Heidelberger and Welch, 1983). Specifically, it tests whether or not a sample chain's mean changes over the entire MCMC sample. If the mean does not change, the chain is stationary. It is always a good idea to visually inspect the chains in addition to making sure that they have converged (i.e., that the draws from the various chains overlap and that the chains are stationary). When the chains converge, all are combined to produce the posterior distribution.

To assess overall model fit to the data, Gelman and Hill (2006) recommend carrying out posterior predictive checks. Posterior predictive checks use a standard discrepancy statistic, such as the sum of squared deviations of observed values from predictions, to examine how well the fitted model can generate new data. The simulated new data based on MCMC draws and the actual data are compared by measuring the lack of fit to model predictions. Large differences in fit between the two data sets indicate that the model misfits the actual data and should be modified. Lack of fit can be examined

visually or can be used to compute a Bayesian p-value (p_B), which quantifies the frequency with which the discrepancy for the simulated data is greater than the discrepancy for the actual data. Values between 0.15 and 0.85 indicate that the model fits the data well (Hobbs and Hooten, 2015).

Like AIC, there are equivalent Bayesian methods for comparing multiple non-nested models. The approach used will depend on the data collected and the analysis to be conducted (Hooten and Hobbs, 2015). For the epizootic field experiments previously described, the Watanabe–Akaike information criterion (WAIC) is perfectly suitable. The WAIC is similar to the AIC in that there are two components of the formula to parsimoniously balance model fit and model complexity. The WAIC takes a similar form to the AIC and is calculated by computing:

$$\text{WAIC} = -2\text{lppd} + 2p_w, \tag{12.14}$$

where lppd is the log posterior predictive density and p_w is the associated parameter penalty (Gelman et al., 2014) for overfitting the model. The posterior predictive density is based on how well the posterior estimates of the model (e.g., transmission rate) predict new data. Since there are no new data, we simply ask how well each estimate of the transmission rate from the posterior MCMC sample does in predicting the data at hand. The second term, which determines the effective number of parameters in the Bayesian model, is the sum of the variance associated with the log posterior predictive density (Hobbs and Hooten, 2015). Formally, the WAIC can be written as:

$$\text{WAIC} = -2\sum_{i=1}^{n}\log\left(\frac{\sum_{j=1}^{J}\left(y_i\mid\Theta^j\right)}{J}\right) + 2\sum_{i=1}^{n}\text{Var}\left(\log\left(y_i\mid\Theta\right)\right),$$

where n is the number of observations, J is the number of samples of the posterior, y_i represents each data point, and Θ^j are the parameter estimates from a single sample j of the posterior (Hobbs and Hooten, 2015). The WAIC represents just one metric that can be used to validate Bayesian models. While this formula can appear daunting, numerous resources exist that can help one in either understanding or calculating the WAIC (e.g., Gelman et al., 2014; Hobbs and Hooten, 2015; Hooten and Hobbs, 2015). Another popular method is cross-validation, whereby some data are used to fit the model and others are left out to test how well the model does in fitting them. In fact, the penalty term for the WAIC can be considered an approximation to cross-validation (Gelman et al., 2014). The most appropriate metric will depend upon the model being fit, the data, and the manner in which the data have been collected (Hooten and Hobbs, 2015).

12.5.2 An Example of the WAIC in Action

The same experiment and data from the AIC example can also be analyzed from a Bayesian perspective, and the associated WAIC scores calculated. For the models considered, the rankings are similar to the AIC results (Table 12.1). The model with the lowest WAIC scores is still the model where C differs but the transmission rate $\bar{\beta}$ stays
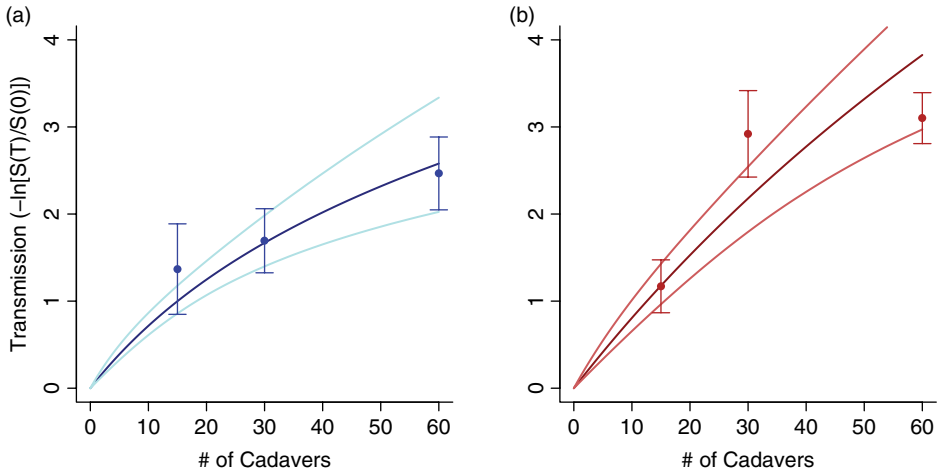
Fig. 12.2 Best-fit model (Table 12.1) for (a) control and (b) warmed plots. The solid dark line represents the predicted fraction infected (equation 12.5) given the median Bayesian estimates of $\overline{\beta}$ and C for each treatment. The lighter shaded lines correspond to the 90% credible intervals (CIs) associated with these estimates. The points represent the mean of the data. The bars are the standard errors.

the same across treatments. Note, for the WAIC, that there are no equivalent metrics associated with model comparisons, such as weights used in the AIC (see earlier). However, the use of the WAIC continues to be developed and refined. When applied to fall armyworm virus data, the model results show that the coefficient of variation C increases as temperatures increases, which results in an increase in overall transmission at higher cadaver densities (Fig. 12.2). Using the WAIC, the same conclusion can be drawn: that when temperatures rise the coefficient of variation associated with transmission declines and the dynamics become more and more similar to linear transmission dynamics (Elderd and Reilly, 2014). At the end of the day, both the AIC and the WAIC result in the same best model. The advantage of using a Bayesian framework becomes more readily apparent as the models considered become increasingly complicated.

12.6 Long-Term Dynamics

The focus, so far, has been on single occurrences of a high prevalence of disease in a population (i.e., a single epizootic). Considerable research also focuses on modeling the long-term dynamics of insect populations driven by semiregular epizootic events. As this research has shown via the use of mechanistic models, epizootics drive or help drive the boom-and-bust population cycles often associated with insects, particularly those of economic concern.

As previously mentioned, Anderson and May's (1980) seminal paper combined ideas from two often disparate fields of research: predator–prey dynamics and epidemiology. Most previous efforts in modeling disease outbreaks focused on single epizootic events. These models are best exemplified in the epidemiological literature as the SIR models (Kermack and McKendrick, 1927), in which a main assumption is that the overall population size does not change over the course of an epidemic or epizootic (Anderson and

May, 1979). This approach works well with questions focused on near-term consequences, such as, "How many individuals will become infected over the course of an epizootic?" On the other hand, predator–prey models focus on the long-term population dynamics of prey and their predators, which are based on the classic work of Lotka (1932) and Volterra (1926). Anderson and May used ideas from both fields to construct a model showing that larch bud moth (*Zeiraphera diniana*) outbreaks could be driven by host–pathogen interactions (Anderson and May, 1980). Surprisingly, prior to their work, ecologists generally ignored the ability of pathogens to control the population dynamics of an insect (Anderson and May, 1981). Interestingly, more recent work on the same larch bud moth system has shown that parasitoids, not pathogens, drive the boom-and-bust cycles (Kendall et al., 1999; Turchin, 2003). When expanding the model to include spatial dynamics, dispersal, along with plant quality, can play an important role (Bjørnstad et al., 2002). The change in the driver of the cycle from the pathogen to the parasitoid exemplifies the importance of continually confronting observational data with mechanistic models and modifying a model as new data and new hypotheses emerge.

12.6.1 Long-Term Dynamics: Confronting Models with Data

For the univoltine gypsy moth, the short-term dynamics associated with epizootics during the larval phase and the long-term dynamics associated with adult reproduction can be considered separately. First, the epizootic occurs (a within-generation process), and then reproduction occurs (a between-generation process). A number of mechanistic models have been developed to describe this within- and between-generation process (e.g., Dwyer et al., 2004; Bjørnstad et al., 2010; Elderd et al., 2013). The general gestalt of these models is summarized nicely by Fuller et al. (2012).

To start off, consider the short-term or within-generation dynamics, which are governed by a series of differential equations that track the entirety of the epizootic process. The equations are:

$$\frac{dS}{dt} = -\overline{\beta} \left[\frac{S(t)}{S(0)} \right]^{C^2} SV, \tag{12.15}$$

$$\frac{dE_1}{dt} = \overline{\beta} \left[\frac{S(t)}{S(0)} \right]^{C^2} SV - m\delta E_1, \tag{12.16}$$

$$\frac{dE_i}{dt} = m\delta E_{i-1} - m\delta E_i \left(i = 2, \ldots, m \right), \tag{12.17}$$

$$\frac{dV}{dt} = m\delta E_m - \mu V. \tag{12.18}$$

Here, the equivalent terms have the same meanings as before (see equation 12.4). A major change from the classic SIR model is reflected in the fact that there is now an exposed class E, within which there are a number of different stages E_i. The individuals in each i th stage, E_i, have consumed enough virus to become infected but have not yet succumbed to the virus and become pathogen, V. If there is only a single infected

class in the model, some larvae will instantly become pathogen, as exposed individuals continually move at an exponential rate out of the single exposed class (Keeling and Rohani, 2008). By allowing for m total stages, the infected stages becomes a sum of exponential distributions, which is a gamma distribution with a mean of $1/\delta$, where δ is the average speed of kill, and a variance of $1/m\delta$. The number of stages depends upon both the mean and the variance estimates of the speed of kill. For gypsy moth larvae, the best estimates are $1/\delta$ = 12 days and m = 20 (Fuller et al., 2012). To reiterate, equations 12.15–12.18 only describe the within-season dynamics of the insect host when it is susceptible and succumbs to the baculovirus.

Long-term or between-season dynamics of the host population track host reproduction after the epizootic ends. Recall, the epizootic ends either due to the uninfected individuals pupating or due to epizootic burnout (Dwyer et al., 2000; Fuller et al., 2012). At the end of the epizootic, the equations describing the long-term dynamics are:

$$N_{n+1} = \lambda N_n \left[1 - I\left(N_n, Z_n \right) \right] \left(1 - \frac{ab N_n}{b^2 + N_n^2} \right), \tag{12.19}$$

$$Z_{n+1} = f N_n I\left(N_n, Z_n \right) + \gamma Z_n. \tag{12.20}$$

Here, N_n and Z_n are the densities of the hosts and the cadavers before the epizootic in generation n and $I\left(N_n, Z_n \right)$ is the fraction of the larvae that become infected (equations 12.15–12.18). The net reproductive rate is λ. For outbreaking insects, population densities are kept at low levels during inter-outbreak periods by generalist predators or parasitoids (Dwyer et al., 2004). For gypsy moth populations, this can take the form of a Type III functional response. The fraction surviving predation is represented by the term $1 - ab N_n / \left(b^2 + N_n^2 \right)$, where a is the maximum predation rate and b is the saturation constant. Baculovirus densities depend upon the survival f of virus derived from the current generation and the survival of virus γ from previous generations. While it is likely that sublethal or covert infections play only a small role in the long-term dynamics, the preceding model also adequately describes covert infections. It assumes that some fraction of the virus survives from one generation to the next, which could be derived from covert infections. As long as this fraction is density-independent, the model provides an accurate accounting of covert infections (Elderd et al., 2013). Over the course of multiple generations, the modeling consists in stringing together the short-term (e.g., one season for univoltine gypsy moths) epizootic followed by adult reproduction, which sets the stage for the next epizootic.

12.6.2 Time-Series Diagnostics

While fitting models to data using results from short-term experiments draws directly from the standard statistical literature, long-term data sets represent a different problem from an analytical perspective. They are often observational and constitute a classic example of an "inverse problem" (Kendall et al., 1999), such that the data collected may arise due to many different mechanistic processes (e.g., intraspecific density-dependent regulation vs. host–pathogen interactions). How best to decide which mechanisms may

be responsible for the observed data is central to understanding what drives the boom-and-bust cycles associated with long-term epizootic dynamics.

For many of these observational data sets, the data are not directly fitted to the model. For instance, a number of papers exploring gypsy moth long-term dynamics use defoliation data as a proxy for gypsy moth population numbers (e.g., Dwyer et al., 2004; Elderd et al., 2008; Bjørnstad et al., 2010). To compare the model output with the observational data, authors often rely on matching various metrics associated with the time series of the data (e.g., average period between peak outbreaks or defoliation events) with the model output. Directly fitting the model to the data becomes increasingly problematic if the dynamics of the system are chaotic, since the model and the data are sensitive to initial conditions (Dwyer et al., 2004). Thus, instead of directly fitting the data to determine which model drives the observed dynamics, "time-series" probes are advocated (Kendall et al., 1999; Turchin, 2003).

Kendall et al. (1999) were among the first advocates in the ecological literature to push for the use of "time-series" probes by combining time-series statistics with mechanistic population models. Previous to this paper, most time-series analyses consisted of fitting nonmechanistic models that could be considered biologically naïve to observational data. On the other side of the coin were the theoretical population ecologists who constructed biologically explicit models that elicited general patterns seen in the data but often did not use standard goodness–of-fit metrics to see if their models stood up to the data (Kendall et al., 1999). The use of "time-series" probes blends the two historic approaches by combining biologically reasonable models with time-series analytical approaches.

There are three steps to this approach. The first consists in constructing the mathematical model describing the long-term dynamics (e.g., equations 12.19–12.20). In the second, once the model is constructed, it is parameterized using independent data and/or other time-series data. In the third step, the model predictions are compared to the time series using a suite of statistical probes, such as average period, amplitude, auto-correlation, and spectral density functions. This is done by simulating the model to generate a suite of synthetic time series and comparing the simulated dynamics to the actual time series. In the end, if the model fits the data using the time-series probes, the hypothesis driving the mechanistic models is worth pursuing. While Kendall et al. (1999) used data from Nicholson's classic blowfly populations to demonstrate the utility of this approach, the usefullness of time-series probes in understanding long-term epizootic dynamics has been demonstrated time and time again (e.g., Dwyer et al., 2004; Johnson et al., 2006; Abbott and Dwyer, 2008; Elderd et al., 2008; Bjørnstad et al., 2010).

While others have directly compared model output to data using classical goodness-of-fit measures for non-epizotically driven time series (Ives et al., 2008), the time-series probe approach remains popular. However, methods described by Ives et al. (2008) hold promise from both a standard-likelihood perspective and from a more Bayesian one (Barraquand et al., 2017) for examining cyclic dynamics. Thus, likelihood- and Bayesian-based approaches may also hold promise in the realm of modeling epizootic dynamics.

To understand how and whether a method works, as outlined in Ives et al. (2008), simulating data represents a useful first step. Exploratory analyses consist simply in using a model (e.g., equations 12.19–12.20) to create fake data by adding process or measurement error to the model's deterministic skeleton. The simulated data are analyzed using

a method of choice (e.g., a Bayesian approach), and the results are compared to the known simulated truth (Kéry, 2010).

To examine the methods used in Ives et al. (2008) from a Bayesian perspective, data were simulated using the nondimensionalized version of the burnout approximation model in Dwyer et al. (2000). The simulation used three equations to represent the dynamics, as follows:

$$N_{n+1} = \lambda \varepsilon_n N_n \left[1 - I\left(N_n, Z_n\right) \right],$$
(12.21)

$$Z_{n+1} = \phi N_n I\left(N_n, Z_n\right) + \gamma Z_n,$$
(12.22)

$$1 - I\left(N_n, Z_n\right) = \left[1 + C^2 N_n I\left(N_n, Z_n\right) + Z_n \right]^{-1/C^2}.$$
(12.23)

Here, ϕ is the product of pathogen survival and mean susceptibility of newly emerging larvae (Dwyer et al., 2000) and ε_n is a log normally distributed random variable with a median of 1 and a standard deviation of σ. For simplicity of presentation, γ is set to 0 and there are no generalist predators in the model. To understand the boom-and-bust dynamics of the insect host population given the preceding, there are only three parameters in the nondimensionalized model that matter: λ, C, and ϕ. All of the other parameters simply move the population mean up or down, and do not affect the period or amplitude of the population cycles. In terms of the simulated data, the analysis only uses the time series associated with the host population, N, as an input. Overall, the Bayesian approach
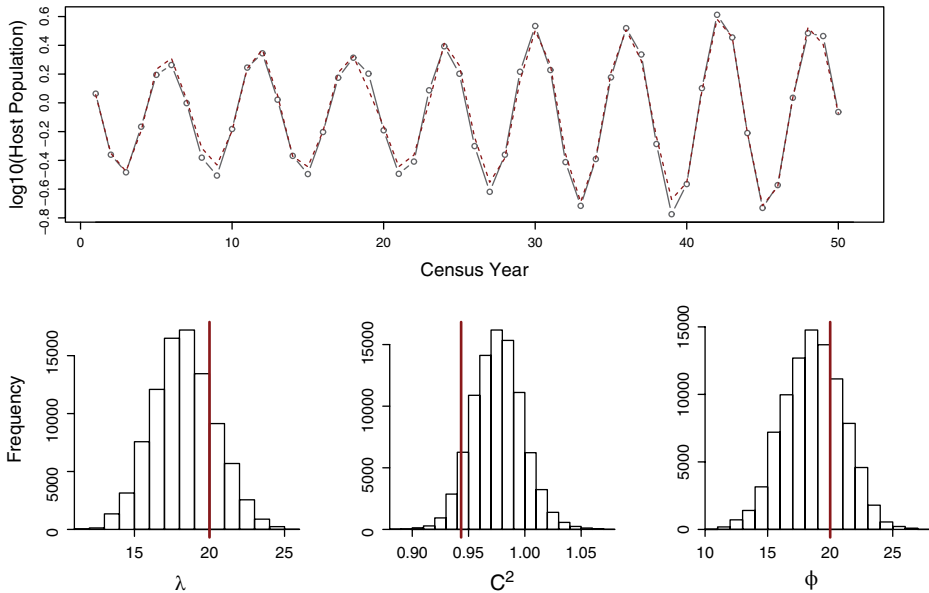

Fig. 12.3 Bayesian estimates of a simulated time series using equations 12.15–12.18, assuming epizootic burnout (Dwyer et al. 2000) without generalist predators. In the top panel, the simulated data are represented by the solid line and the best estimate of the dynamics by the dashed line. The bottom three panels show the associated histograms for each of the parameters estimated from the model. The vertical lines represent the true values of the parameters used to simulate the data.

does a reasonable job of predicting the dynamics and the model's associated true parameter values (Fig. 12.3). The next step is to confront the same modeling framework with data.

12.7   Modifying and Applying the Model

It is important to remember that models, like hypotheses, are not static. The development of epizootic models, whether regarding single epizootic events or the long-term dynamics of the host population, have continued and will continue to evolve as more data are collected and new hypotheses emerge. For the gypsy moth, the development of changes in the basic long-term model still continues. While earlier models focused on host–pathogen dynamics alone (Dwyer et al., 1997, 2000), later models included the importance of generalist predators (equations 12.19–12.20) (Dwyer et al., 2004), the effect of host–density on disease resistance (Reilly and Hajek, 2008), host evolutionary dynamics (Elderd et al., 2008), and host food resources (Bjørnstad et al., 2010; Elderd et al., 2013). The models that incorporated changes in host food source or food quality also incorporated spatial components to characterize outbreak dynamics in different forest types.

While the preceding highlights model development with regard to gypsy moth populations, model development and modification are also important to understanding epizootic dynamics in other insect species (e.g., Hochberg, 1989; Boots and Mealor, 2007; Elderd and Reilly, 2014).

Epizootic models can also be used to examine applied problems, which can be studied by modeling epizootic dynamics before enacting management strategies or policies. As a number of insect species are pests, pathogens and parasites represent potential biocontrol agents. While the easiest way to deal with a short-term infestation may be to release pathogens into the environment, the long-term effects often remain unknown. The main reasons for this stem from the need for long-term observational data collected over multiple epizootic periods (often decades) and the lack of data collected during a series of long-term experimental trials (i.e., controls and experimental plots). A prime example of such a successful combination of experimental trials and model development in a non-insect system is found in Hudson et al. (1998), who focused on controlling population declines in the red grouse (*Lagopus lagopus scoticus*) caused by a parasitic nematode. The field component of this research took almost a decade at multiple sites, which might not be possible in the midst of an insect infestation. However, the use of mechanistic models of insect epizootic dynamics for applied problems within the confines of one's computer allows the researcher to conduct experiments via simulation. This approach has led to insights into the effectiveness of pathogens as biocontrol agents for a number of insect taxa (Hochberg, 1989; Hochberg and Waage, 1991; Reilly and Elderd, 2014). Reilly and Elderd (2014), in their model of the gypsy moth–baculovirus system, showed that spraying focused only on the short-term effects of controlling an insect population may lead to unexpected and undesirable consequences. For certain spraying regimes, the gypsy moth population can be maintained at constant population levels instead of exhibiting boom-and-bust cyclic dynamics, but the average population size is relatively high. This, in turn, may result in constant defoliation of the forests that these biocontrol efforts are trying to protect. Overall, while the mechanistic models provide a deeper understanding of what drives

epizootics in insect populations, they also provide a useful tool for asking questions of an applied nature.

12.8   Conclusion

The use of models to understand epizootic dynamics has a long history in the ecological literature. Much of the past debate concerning which methodology is best suited for moving the field forward centered on the historic false dichotomy between empirical and theoretical approaches, while sometimes invoking simulation-based methods. However, the ability to confront models with data has led to new and exciting developments in the field, since models can now be used as hypotheses to drive research questions. While using the preceding techniques and ideas may seem easy to some and daunting to others, they do not necessarily need to be mastered by all. Instead, they represent a framework to begin a conversation about questions that can be answered, how to design empirical studies, and how best to use the data produced. The reason the false dichotomy of empiricism and theory continues to blur stems from more individuals being able to speak in multiple languages. Thus, mastering each technique is not essential, but being able to communicate across the false divide is. As the dialogue advances and individuals speak across their own expertise, the biology of the system becomes better connected to the mechanistic framework, which leads to a better understanding of what drives the epizootic process.

Acknowledgments

References

Abbott, K.C., Dwyer, G., 2008. Using mechanistic models to understand synchrony in forest insect populations: the North American Gypsy Moth as a case study. Am. Nat. 172, 613–624.

Anderson, D., Burnham, K., Thompson, W.L., 2000. Null hypothesis testing: problems, prevalence, and an alternative. J. Wildlife Manage. 64, 912–923.

Anderson, R.M., May, R.M., 1979. Population biology of infectious diseases: part I. Nature 280, 361–367.

Anderson, R.M., May, R.M., 1980. Infectious diseases and population cycles of forest insects. Science 210, 658–661.

Anderson, R.M., May, R.M., 1981. The population dynamics of microparasites and their invertebrate hosts. Philos. Trans. R. Soc. B Biol. Sci. 291, 451–524.

Anderson, R.M., May, R.M., 1991. Infectious Disease of Humans: Dynamics and Control. Oxford University Press, Oxford.

Barraquand, F., Louca, S., Abbott, K.C., Cobbold, C.A., Cordoleani, F., DeAngelis, D.L., et al., 2017. Moving forward in circles: challenges and opportunities in modeling population cycles. Ecol. Letters. In Press.

Bjørnstad, O.N., Peltonen, M., Liebhold, A.M., Baltensweiler, W., 2002. Waves of larch budmoth outbreaks in the European Alps. Science 298, 1020–1023.

Bjørnstad, O.N., Robinet, C., Liebhold, A.M., 2010. Geographic variation in North American gypsy moth cycles: subharmonics, generalist predators, and spatial coupling. Ecology 91, 106–118.

Boots, M. 2004. Modelling insect diseases as functional predators. Physiol. Entomol. 29, 237–239.

Boots, M., Mealor, M., 2007. Local interactions select for lower pathogen infectivity. Science 315, 1284–1286.

Brooks, S.P., Gelman A., 1998. General methods for monitoring convergence of iterative simulations. J. Comp. Graph. Stat. 7, 434–455.

Brown, G., 1987. Modeling, in: Fuxa, J., Tanada, Y. (eds.), Epizootiology of Insect Diseases. John Wiley & Sons, Chichester, pp. 43–68.

Burnham, K., Anderson, D., 2002. Model selection and multimodal inference: a practical information-theoretic approach. Springer, New York.

Clark, J.S., 2007. Models for Ecological Data: An Introduction. Princeton University Press, Princeton, NJ.

Cory, J.S., Hoover, K., 2006. Plant-mediated effects in insect-pathogen interactions. Trends Ecol. Evol. 21, 278–286.

Cory, J.S., Myers, J.H., 2003. The ecology and evolution of insect baculoviruses. Annu. Rev. Ecol. Evol. Syst. 34, 239–272.

De Oliveira, M. 1999. South America, in: Hunter-Fujita, F., Entwistle, P., Evans H., Crook, N., (eds.), Insect Viruses and Pest Management. John Wiley & Sons, Chichester, pp. 339–355.

Dennis, B., 1996. Discussion: should ecologists become Bayesians? Ecol. Appl. 6, 1095–1103.

Dwyer, G., Dushoff, J., Elkinton, J.S., Levin. S.A., 2000. Pathogen-driven outbreaks in forest defoliators revisited: building models from experimental data. Am. Nat. 156, 105–120.

Dwyer, G., Dushoff, J., Yee, S.H., 2004. The combined effects of pathogens and predators on insect outbreaks. Nature 430, 341–345.

Dwyer, G., Elkinton, J.S., Buonaccorsi, J.P., 1997. Host heterogeneity in susceptibility and disease dynamics: tests of a mathematical model. Am. Nat. 150, 685–707.

Edelstein, J.D., Trumper, E.V., Lecuona, R.E., 2005. Temperature-dependent development of the entomopathogenic fungus *Nomuraea rileyi* (Farlow) Samson in *Anticarsia gemmatalis* (Hubner) larvae (Lepidoptera : Noctuidae). Neotropical Entomol. 34, 593–599.

Elderd, B.D., 2013. Developing models of disease transmission: insights from the ecology of baculovirus-driven systems. PLoS Pathog. 9, e1003372.

Elderd, B.D., Dukic, V.M., Dwyer, G., 2006. Uncertainty in predictions of disease spread and public health responses to bioterrorism and emerging diseases. Proc. Natl. Acad. Sci. U.S.A. 103, 15 693–15 697.

Elderd, B.D., Dushoff, J., Dwyer, G., 2008. Host-pathogen dynamics, natural selection for disease resistance, and forest-defoliator outbreaks. Am. Nat. 172, 829–842.

Elderd, B.D., Haynes, K., Rehill, B., Dwyer, G., 2013. Interactions between an induced plant defense and a pathogen drive outbreaks of a forest insect. Proc. Natl. Acad. Sci. U.S.A. 110, 14 978–14 983.

Elderd, B.D., Reilly, J., 2014. Warmer temperatures increase disease transmission and outbreak intensity in a host-pathogen system. J. Anim. Ecol. 83, 838–849.

Ellison, A.M., 1996. An introduction to Bayesian inference for ecological research and environmental decision-making. Ecol. Appl. 6, 1036–1046.

Ellison, A.M. 2004. Bayesian inference in ecology. Ecol. Letters 7, 509–520.

Fleming-Davies, A.E., Dukic, V., Andreasen, V., Dwyer, G., 2015. Effects of host heterogeniety on pathogen diversity and evolution. Ecol. Letters 18, 1252–1261.

Fuller, E., Elderd, B., Dwyer, G., 2012. Pathogen persistence in the environment and insect-baculovirus interactions: disease-density thresholds, epidemic burnout, and insect outbreaks. Am. Nat. 179, E70–E96.

Fuxa, J.R., 1982. Prevalence of viral infections in populations of fall armyworm, *Spodoptera frugiperda*, in Southeastern Louisiana. Environ. Entomol. 11, 239–242.

Fuxa, J.R., Geaghan, J.P., 1983. Multiple-regression analysis of factors affecting prevalence of nuclear polyhedrosis virus in *Spodoptera frugiperda* (Lepidoptera, Noctuidae) populations. Environ. Entomol. 12, 311–316.

Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 2014. Bayesian data analysis. Taylor & Francis, Philadelphia, PA.

Gelman, A., Hill, J., 2006. Data analysis using regression and multilevel/hierarchical models. Cambridge University Press, Cambridge.

Hajek, A.E., 1999. Pathology and epizootiology of *Entomophaga maimaiga* infections in forest Lepidoptera. Microbiol. Mol. Biol. Rev. 63, 814–835.

Hajek, A.E., Tobin, P.C., Haynes, K.J., 2015. Replacement of a dominant viral pathogen by a fungal pathogen does not alter the collapse of a regional forest insect outbreak. Oecologia 177, 785–797.

Hajek, A.E., van Nouhuys, S., 2016. Fatal diseases and parasitoids: from competition to facilitation in a shared host. Proc. R. Soc. B Biol. Sci. 283, 20160154.

Heidelberger, P., Welch, P.D., 1983. Simulation run length control in the presence of an initial transient. Oper. Res. 31, 1109–1144.

Hesketh, H., Roy, H.E., Eilenberg, J., Pell, J.K., Hails, R.S., 2010. Challenges in modelling complexity of fungal entomopathogens in semi-natural populations of insects. Biocontrol 55, 55–73.

Hilborn, R., Mangel, M., 1997. The Ecological Detective: Confronting Models with Data. Princeton University Press, Princeton, NJ.

Hille Ris Lambers, J., Clark, J.S., Lavine, M., 2005. Implications of seed banking for recruitment of southern Appalachian woody species. Ecology, 86 85–95.

Hinds, W., Dew, J., 1915. The Grass Worm or Fall Army Worm. Technical Report Bulletin no. 186, Alabama Agricultural Experiment Station.

Hobbs, N.T., Hilborn, R., 2006. Alternatives to statistical hypothesis testing in ecology: a guide to self teaching. Ecol. Appl. 16, 5–19.

Hobbs, N.T., Hooten, M.B., 2015. Bayesian models: a statistical primer for ecologists. Princeton University Press, Princeton, NJ.

Hochberg, M.E., 1989. The potential role of pathogens in biological control. Nature 337, 262–265.

Hochberg, M.E., 1991. Nonlinear transmission rates and the dynamics of infectious disease. J. Theor. Biol. 153, 301–321.

Hochberg, M.E., Waage, J.K., 1991. A model for the biological control of *Oryctes rhinoceros* (Coleoptera, Scarabaeidae) by means of pathogens. J. Appl. Ecol. 28, 514–531.

Hooten, M.B., Hobbs, N.T., 2015. A guide to Bayesian model selection for ecologists. Ecol. Monogr. 85, 3–28.

Hudson, P.J., Dobson, A.P., Newborn, D., 1998. Prevention of population cycles by parasite removal. Science 282, 2256–2258.

Ives, A.R., Einarsson, A., Jansen, V.A.A., Gardarsson, A., 2008. High-amplitude fluctuations and alternative dynamical states of midges in Lake Myvatn. Nature 452, 84–87.

Johnson, D.M., Liebhold, A.M., Tobin, P.C., Bjørnstad, O.N., 2006. Allee effects and pulsed invasion by the gypsy moth. Nature 444, 361–363.

Keeling, M.J., Rohani, P., 2008. Modeling infectious diseases in humans and animals. Princeton University Press, Princeton, NJ.

Kendall, B.E., 2015. Some directions in ecological theory. Ecology 96, 3117–3125.

Kendall, B.E., Briggs, C.J., Murdoch, W.W., Turchin, P., Ellner, S.P., McCauley, E., et al., 1999. Why do populations cycle? A synthesis of statistical and mechanistic modeling approaches. Ecology 80, 1789–1805.

Kermack, W.O., McKendrick, A.G., 1927. A contribution to the mathematical theory of epidemics. Proc. R. Soc. A Math. Phys. Eng. Sci. 115, 700–721.

Kéry, M., 2010. Introduction to WinBUGS for Ecologists. Academic Press, Boston, MA.

Levin, S.A., 2012. Towards the marriage of theory and data. Interface Focus 2, 141–143.

Liebhold, A.M., Plymale, R., Elkinton, J.S., Hajek, A.E., 2013. Emergent fungal entomopathogen does not alter density dependence in a viral competitor. Ecology 94, 1217–1222.

Link, W.A., Eaton, M.J., 2012. On thinning of chains in MCMC. Methods Ecol. Evol. 3, 112–115.

Lotka, A.J., 1932. Contribution to the mathematical theory of capture I. Conditions for capture. Proc. Natl. Acad. Sci. U.S.A. 18, 172–178.

Marion, G.M., Henry, G.H.R., Freckman, D.W., Johnstone, J., Jones, G., Jones, M.H., et al. 1997. Open-top designs for manipulating field temperature in high-latitude ecosystems. Global Change Biol. 3, 20–32.

Miller, L.K. (ed.). 1997. Baculoviruses. Kluwer Academic, Dordrecht.

Myers, J.H., Malakar, R., Cory, J.S., 2000. Sublethal nucleopolyhedrovirus infection effects on female pupal weight, egg mass size, and vertical transmission in gypsy moth (Lepidoptera : Lymantriidae). Environ. Entomol. 29, 1268–1272.

Onstad, D.W., Carruthers, R.I., 1990. Epizootiological models of insect diseases. Annu. Rev. Entomol. 35, 399–419.

Pitre, H.N., Hogg, D.B., 1983. Development of the fall armyworm (Lepidoptera, Noctuidae) on cotton, soybean and corn. J. Georgia Entomol. Soc. 18, 182–187.

Plummer, M. et al., 2003. JAGS: a program for analysis of Bayesian graphical models using Gibbs sampling, in: Proceedings of the 3rd International Workshop on Distributed Statistical Computing, Vol. 124, p. 125. Technische Universit at Wien, Vienna.

R Core Team., 2015. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. Available from: https://www.R-project. org/ (accessed May 8, 2017).

Reilly, J.R., Elderd, B.D., 2014. Effects of biological control on long-term population dynamics: identifying unexpected outcomes. J. Appl. Ecol. 51, 90–101.

Reilly, J.R., Hajek, A.E., 2008. Density-dependent resistance of the gypsy moth *Lymantria dispar* to its nucleopolyhedrovirus, and the consequences for population dynamics. Oecologia 154, 691–701.

Richter, A.R., Fuxa, J.R., Abdelfattah, M., 1987. Effect of host plant on the susceptibility of *Spodoptera frugiperda* (Lepidoptera, Noctuidae) to a nuclear polyhedrosis virus. Environ. Entomol. 16, 1004–1006.

Rohani, P., Keeling, M.J., Grenfell, B.T., 2002. The interplay between determinism and stochasticity in childhood diseases. Am. Nat., 159 469–481.

Roy, H.E., Hails, R.S., Hesketh, H., Roy, D.B., Pell, J.K., 2009. Beyond biological control: non-pest insects and their pathogens in a changing world. Insect Conserv. Div. 2, 65–72.

Scheiner, S.M., 2013. The ecological literature, an idea-free distribution. Ecol. Letters 16, 1421–1423.

Scholte, E.J., Ng'habi, K., Kihonda, J., Takken, W., Paaijmans, K., Abdulla, S., et al., 2005. An entomopathogenic fungus for control of adult African malaria mosquitoes. Science 308, 1641–1642.

Shapiro-Ilan, D., Bruck, D., Lacey, L., 2012. Principles of epizootiology and microbial control, in: Tanada, Y., Kaya, H.K. (eds.), Insect Pathology. Academic Press, Amsterdam, pp. 29–72.

Sparks, A.N., 1979. Review of the biology of the fall armyworm (Lepidoptera, Noctuidae). Florida Entomol. 62, 82–87.

Stuart, R.J., Barbercheck, M.E., Grewal, P.S., Taylor, R.A.J., Hoy, C.W., 2006. Population biology of entomopathogenic nematodes: concepts, issues, and models. Biol. Control 38, 80–102.

Turchin, P., 2003. Complex Population Dynamics: A Theoretical/Empirical Synthesis. Princeton University Press, Princeton, NJ.

Volterra, V., 1926. Fluctuations in the abundance of a species considered mathematically. Nature 118, 558–560.

Wasserstein, R.L., Lazar, N.A., 2016. The ASA's statement on p-values: context, process, and purpose. Am. Stat. 70, 129–133.

Wootton, J.T., 2005. Field parameterization and experimental test of the neutral theory of biodiversity. Nature 433, 309–312.

Yu-Sung, S., Masanao, Y., 2015. R2jags: using R to Run "JAGS." Available from: https://CRAN.R-project.org/package=R2jags (accessed May 8, 2017).