Methods in Ecology and Evolution

British Ecological Society

Methods in Ecology & Evolution 2010, **1**, 192–198 doi: 10.1111/j.2041-210X.2010.00018.x

A primer on the application of Markov chains to the study of wildlife disease dynamics

Elise F. Zipkin[1]*, Christopher S. Jennelle[2] and Evan G. Cooch[1]

[1]*Department of Natural Resources, Cornell University, Ithaca, NY 14853, USA and* [2]*Department of Forestry and Wildlife Ecology, University of Wisconsin, Madison, WI 53706, USA*

Summary

1. For wildlife researchers, disease specialists and policy analysts unfamiliar with the mathematical/statistical language of disease models, translation of probability statements into meaningful terms for disease research and control may be challenging. Markov chain models are powerful tools, applicable to the study of disease dynamics that allow straightforward calculations of easily interpretable metrics of interest including probabilities of infection/recovery, expected times to initial infection, duration of illness and life expectancies for susceptible and infected individuals.

2. We present the basic principles and assumptions behind Markov chain modelling with an intuitive interpretation of parameter estimates and a step-by-step guide (including software code) for implementing this approach in the study of wildlife diseases. We also include an explanation of the estimation process necessary to implement Markov chain modelling (i.e. estimating the probability of state transitions between consecutive time steps) from typical survey data.

3. We demonstrate the usefulness and ease of calculation of Markov chains through an example using a house finch *Carpodacus mexicanus–Mycoplasma gallisepticum* (MG) system. Our results show how semi-weekly transition estimates of susceptible and infected individuals can be used to estimate a wide array of seasonal disease-associated metrics.

4. Markov chain modelling can provide a basic understanding of parameters estimated from wildlife disease studies, and can aid in understanding the implications of disease on wildlife populations and in evaluation of control measures. We envision this paper serving as an entry point into the extensive literature and potential applications of Markov chains in epidemiological modelling.

Key-words: Epidemiological model, house finch *Carpodacus mexicanus,* Markov process, *Mycoplasma gallisepticum*, wildlife disease

Introduction

Exploring the dynamics of wildlife diseases at a population level usually entails the collection of demographic data including the capture (and possible marking) of numerous individuals over a temporal and spatial gradient. Using such data, it is often desirable to quantify parameters such as disease prevalence, disease incidence, force of infection, disease amplification and reproductive rate, or infection and recovery probabilities. While mathematical and statistical methods are available to generate these parameters, many non-modelling specialists in the area of wildlife disease ecology (i.e. conservation practitioners, wildlife managers, veterinarians and policy analysts) are unfamiliar with both the available tools and the interpretation of relevant disease metrics. Given the complexity of the mathematical and statistical language required to estimate epidemiological quantities of interest, interpreting model outputs in the context of disease control and management policy can be a difficult task. Therefore, it is critically important to provide alternative and interpretable translations of such metrics for understanding theoretical and applied wildlife disease studies.

Markov chain models are one method that is used to describe disease dynamics at the individual level, which can be scaled up to make inferences regarding populations. The tractability and relative ease of calculating quantities of interest make Markov modelling attractive in the study of wildlife diseases, which are often characterized by disease states (e.g. uninfected and infected) that are sampled at discrete time intervals. While the theory and practice of Markov chain modelling

*Correspondence author.
Present address: USGS Patuxent Wildlife Research Center, 12100 Beech Forest Rd, Laurel, MD 20708, USA. E-mail: ezipkin@usgs.gov
Correspondence site: http://www.respond2articles.com/MEE/

© 2010 The Authors. Journal compilation © 2010 British Ecological Society

to epidemiology is not new (Bishop, Fienberg, & Holland 1971; Kay 1986; McElhany, Real, & Powers 1995; LeStrat & Carrat 1999), basic applications of this powerful quantitative tool can be valuable to non-modelling specialists. In this paper, we present the underling theory, assumptions and terminology for using Markov chain models in the study of wildlife disease dynamics. The calculations we outline are useful for deriving easily interpretable metrics in the study of wildlife diseases, which may facilitate hypothesis generation, understanding disease dynamics in wildlife populations, and future study design. We focus on discrete, first-order Markov models as they are most amenable to parameterization using typical encounter–re-encounter data from wild populations; for example, multi-state mark–recapture models (*sensu* Arnason 1973; Brownie *et al*. 1993; Schwarz, Schweigert, & Arnason 1993; Nichols & Kendall 1995) are in fact first-order discrete Markov models (for a general review of multi-state Markov models in both discrete and continuous time, see Schwarz 2009; for application of multi-state Markov models to situations where individuals are not marked, see Anderson *et al*. 1993; Commenges 1999). We provide a step-by-step approach to calculate: (1) the probability of initial infection and recovery for an individual, (2) the expected duration of susceptible and infected states and (3) the life expectancies for susceptible and infected individuals. These quantities are useful for characterizing basic disease dynamics and can help researchers explore key aspects of disease transmission within a population. Management agencies can use these metrics to explore the virulence of wildlife diseases and the potential for disease containment by examining how probabilities of infection and recovery would change under various control scenarios. We demonstrate the applicability of Markov modelling by presenting an application of this approach to data from a house finch *Carpodacus mexicanus* population exposed to the bacterial pathogen *Mycoplasma gallisepticum* (MG). We provide details of each analysis and have included software code (written in the freely available language R, http://www.r-project.org, see Appendix S1) to run the models for those interested in applying these methods to their own study systems.

Materials and methods

FUNDAMENTAL ASSUMPTIONS

The fundamental idea behind Markov chain modelling is that the current state of an individual (e.g. susceptible, infected or dead) is dependent only upon the state of the individual at the previous time step, measured in discrete intervals (e.g. days, weeks and months). It is possible to extend the approach to include information about the state of the individual in many previous time steps, but for the purposes of this paper (and for ease of calculations), we restrict the analysis to include only the last time step, referred to as a first-order Markov process (for an explanation of higher order Markov processes, see Taylor & Karlin 1998 and Ross 2006). A Markov chain is essentially a conditional probability, where the probability that the process under study, denoted as X, at time $n + 1$ is in the j state ($X_{n+1} = j$) given that it was in state i during the previous time step n. This is represented mathematically as:

$$p_{ij} = \Pr\{X_{n+1} = j | X_n = i\}. \qquad \text{eqn 1}$$

The expression in eqn 1 is labelled p_{ij} and is the probability of moving from state i to state j in one time step, called a one-step 'transition probability'. We make the simplifying assumption that the transition probabilities are stationary, or independent of the time period n so that the probability of transitioning from state i to j does not depend on when (over the course of the study period) the transition occurs. This assumption implies that we do not account for seasonal variations in the rates of disease transitions. While seasonality can be a major component in disease dynamics (Altizer *et al*. 2006), if we appropriately restrict our temporal scale then this assumption is tenable.

MARKOV CHAINS AS RELATED TO DISEASE DYNAMICS

In the context of disease, we often define three discrete states: susceptible (state 0), infected (state 1) and dead (or removed; state 2). This simplified scenario does not account for health history (an individual naïve to infection is considered the same as one that has recovered from infection), although additional states can be added to account for disparities between naïve and recovered individuals or to incorporate other details of interest. For example, an immune state can be included to encompass individuals that have recovered from infection and are incapable of contracting the disease in the future. Susceptible individuals are not infected with the disease and can remain susceptible, become infected or die in a given time step. Similarly, infected individuals can recover, remain infected or die. Once an individual dies, it of course remains dead. The susceptible and infected states are referred to as 'transient' states because an individual can enter and leave them many times while death is referred to as an 'absorbing' state because once an individual dies, it must remain dead. Through data collection that observes the state of individuals at regular time intervals, we can estimate each of the one-step state transition probabilities (see section *Estimating one-step transition probabilities* below) and arrange them into a 'transition probability matrix', denoted as P:

$$P = \begin{pmatrix} p_{00} & p_{01} & p_{02} \\ p_{10} & p_{11} & p_{12} \\ 0 & 0 & 1 \end{pmatrix}. \qquad \text{eqn 2}$$

The rows (indexed as $i = 0$–2) represent the state of a process for a given individual (0 = susceptible, 1 = infected and 2 = dead) at time n and the columns (indexed as $j = 0$–2) indicate the state of the process at the following time step $n + 1$. For example, p_{01} (the element in row 0 and column 1) is the probability of transitioning from the susceptible to infected state in one time step (conditional on survival), commonly referred to in disease literature as the discrete time force of infection. The elements p_{02} and p_{12} express mortality for uninfected and infected individuals, respectively, while p_{10} is the recovery or defection probability (Cohen 1973). The last row of the matrix represents the transition probabilities for a dead individual. As death is an absorbing state, the probability of becoming susceptible or infected is zero. When determining the time step unit (i.e. hours, days, weeks, etc.), it is important to consider the underlying time course of the disease to ensure that transition processes occur on a biologically meaningful time scale. For example, the time course of avian influenza (World Health Organization. 2005) may be in the order of days, whereas chronic wasting disease (Williams *et al*. 2002b) can be on the order of years. The time step for the one-step transition probabilities should be defined so that it is only possible for one transition to occur during each time interval. If the time step is unreasonable for a given disease, then the metrics calculated using Markov chain models may

be incorrect (e.g. if the time step is too large and multiple transitions can occur, then expected duration in each state may be overestimated).

At any given time, an individual must be in exactly one of these three states, which is reflected by the fact that each of the rows always sum to one. We can examine the n-step transition probabilities of the Markov process by raising the matrix to nth power, P^n (see Lay 2000 for details on matrix multiplication and the Appendix S1 for software code). The probability that a process initially in state i will be in state j after exactly n time steps is simply the elements of matrix P^n, denoted as p_{ij}^n.The elements of the matrix P^n only provide information on the Markov chain at the nth time step; nothing can be inferred about the state of the process during any of the $n - 1$ time steps.

It is now straightforward to use the elements (p_{ij}) of the transition matrix P to calculate several disease metrics. First, we examine the probability that a susceptible individual initially becomes infected during the interval between the $m - 1$ and m time steps. As the time steps have been appropriately defined for the disease of interest, it is only possible for one transition to occur within sequential time steps. We define the initial time step as n and examine the intervals between n, $n + 1$, $n + 2$,..., $n + m$. The probability that a susceptible individual becomes infected after one time step is simply p_{01} and the probability that it remains susceptible is p_{00}. Thus, the probability that a susceptible individual first becomes infected after two time steps is simply the probability that it remained susceptible for exactly one time step and then became infected during the next time step:

$$\Pr\{X_{n+2} = 1, X_{n+1} = 0 | X_n = 0\} = p_{00}p_{01}. \quad \text{eqn 3}$$

Following this path of logic, the probability that a susceptible individual becomes infected for the first time between the $m - 1$ and m time steps would be:

$$f_{01}^{(m)} = \Pr\{X_{n+m} = 1, X_{n+m-1}, \ldots, X_{n+1} = 0 | X_n = 0\} = p_{00}^{m-1}p_{01}. \quad \text{eqn 4}$$

for $1 \leq m < \infty$, and is formally referred to as a sub-distribution of infection probabilities (as the probabilities sum to less than one, $f_{01}^{(m)}$does not characterize the complete probability space). More specifically, $f_{01}^{(m)}$ is defined as the probability that a susceptible individual (in state 0) first becomes infected (moves to state 1) in exactly m time steps for all possible values of m (i.e. the distribution of infection probabilities for all possible values of m). The same logic follows for calculating the probability that an infected individual first recovers between the $m - 1$ and m time steps:

$$f_{10}^{(m)} = \Pr\{X_{n+m} = 0, X_{n+m-1} = 1, \ldots, X_{n+1} = 1 | X_n = 1\} = p_{11}^{m-1}p_{10} \quad \text{eqn 5}$$

and is similarly referred to as the sub-distribution probability of an infected individual recovering in exactly m time steps. As m gets large ($m \to \infty$), the probability of initial infection (or recovery) approaches zero, which implies that the total probability of becoming infected (or recovering) approaches a limit (a value between 0% and 100%). This allows us to calculate the overall probability that a susceptible individual will become infected (or an infected individual will recover) during the time period of the study and the speed at which the process occurs. In our three-state model, the overall probability that an individual transitions from state i to j (as $m \to \infty$) has a simple closed form solution:

$$\Pr\{i \to j\} = \frac{p_{ij}}{1 - p_{ii}}. \quad \text{eqn 6}$$

We can now use the probability distributions to determine the expected time to first infection (i.e. the average time to initial infection for a susceptible individual given that the individual does become infected) and the expected time to recovery (i.e. the average duration of infection). The expected time for an individual in a given state at time n to first enter another state, denoted $E[\tau_{ij}^{(1)}]$and read as the first time ($\tau^{(1)}$) that an individual moves from state i to state j is:

$$E[\tau_{ij}^{(1)}] = \frac{\sum m f_{ij}^{(m)}}{\Pr\{i \to j\}} \quad \text{eqn 7}$$

which is the sum of $mf_{ij}^{(m)}$for all possible values of m (time steps) divided by the lifetime probability of transitioning from i to j. In our three-state model, the expected time to state j (infection or recovery) has a closed form solution:

$$E[\tau_{ij}^{(1)}] = \frac{1}{1 - p_{ii}}. \quad \text{eqn 8}$$

Markov chain models also provide a useful approach for calculating the life expectancies for susceptible and infected individuals. In the context of Markov chains, we define the life expectancy as the expected time to death for an individual from the current time period n. Thus, life expectancy, in this context, is not necessarily equivalent to life expectancy from birth (unless the transition probabilities are specified specifically for newborns). In technical terms we are calculating the expected time to absorption for each state (i.e. absorption into the dead state), denoted as W_i for $i = 0, 1$. To examine life expectancies, we create a matrix Q that contains only the transition probabilities for the recurrent transitions for the transient states (i.e. processes of infection and recovery). In this model, we have defined two transient states, 0 and 1 (susceptible and infected) and one absorbing state, 2 (dead), therefore:

$$Q = \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix}. \quad \text{eqn 9}$$

The life expectancy for individuals is:

$$W = (I - Q)^{-1} \times \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{eqn 10}$$

where I is the identity matrix with the same number of rows and columns as Q, in this case:

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{eqn 11}$$

and the superscripted -1 indicates that we take the inverse of the matrix in parentheses (i.e. the matrix resulting from subtraction of the matrix Q from matrix I) and multiply ($\times$) the two quantities together using standard matrix multiplication (Lay 2000). The life expectancy term, W, is a vector containing the expected times to death for individuals starting as susceptible (the first element of W) and infected (the second element) respectively. The life expectancies incorporate the probabilities of multiple transitions from susceptible to infected states and vice versa (i.e. W represents the sum of the expected number of time periods in each transient state conditioned on the initial state).

ESTIMATING ONE-STEP TRANSITION PROBABILITIES

Markov models allow for the estimation of relevant disease metrics given one-step transition probabilities, but a critically important step is the estimation of transition probabilities from data. A common

data collection approach for studying disease dynamics involves observing individuals over time and recording the disease status during a given sampling occasion, such as encounter–reencounter methods (e.g. Faustino *et al.* 2004). The resulting time-series data can be represented as a vector (y) of 0, 1 and 2 s indicating whether a given individual was uninfected (0), infected (1) or dead (2) (i.e. a detection history of $y = 0, 0, 1, 0, 0, 2$ would indicate that an individual was infected only during the third sampling occasion (y_3) and was dead in the sixth sampling occasion (y_6) – as death is absorbing, no additional data for that individual are necessary). In this simple estimation procedure, we assume that state observations can be made at every sampling occasion and that states are always correctly classified (e.g. perfect detection). In such cases, the probability of observing a given series of 0, 1 and 2 s can be specified using the transition probabilities as defined in eqn 1. For example, if an individual starts in the susceptible state, the probability that it becomes infected and $y = 0, 1$ is observed, is simply p_{01}. If we encountered that same individual in the third sampling occasion and it was still infected ($y = 0, 1, 1$), the probability of such an occurrence can be specified by $p_{01}p_{11}$ and so on. As transition events are independent of one another (as defined by the Markov principle), the likelihood of the transition probability, p_{ij}, is that of a binomial model:

$$L(p_{ij}|N, y) = \binom{N_i}{x_{ij}} p_{ij}^{x_{ij}} (1 - p_{ij})^{N_i - x_{ij}} \qquad \text{eqn 12}$$

given the data (N and x), where N_i is the number of observed transitions that start in state i and x_{ij} is the number of transitions from i specifically to j. As all rows in a transition matrix must sum to one (i.e. $\sum_j p_{ij} = 1$),we can use the likelihood to estimate the transition probabilities for all $i \neq j$ and, in turn, calculate the probability of remaining within the same state by subtracting all probabilities of changing states from one.

Using the likelihood function and available data, there are several methods for estimating the one-step transition probabilities, including maximum likelihood and Bayesian approaches. The programs MARK (White & Burnham 1999) and M-SURGE (Choquet *et al.* 2005) are specifically designed for estimating parameters from encounter–reencounter data using maximum likelihood and are especially useful for estimating transition probabilities in more complicated multi-state models, such as when there is state uncertainty (e.g. difficulty in differentiating susceptible from infected) or interest in estimating transition probabilities for subsets of the population (e.g. grouping by age). Alternatively, parameters can be estimated using a Bayesian approach. The basic model, presented above, is easily implemented in freely available software such as WinBUGS (Spiegelhalter *et al.* 2003). Assessment of the effect of parameter uncertainty on the value of disease metrics (e.g. probability of infection and life expectancy) is straightforward in WinBUGS and similar programs by using the posterior distributions of the one-step transition probabilities to repeatedly calculate each metric. We provide the R code (which calls the program WinBUGS) for estimating the one-step transition parameters of the three-state model (susceptible, infected and dead) using a Bayesian approach, including code for calculating the full posterior distributions of relevant disease metrics (e.g. eqns 6, 8 and 10) in Appendix S2.

In most field studies, animals are detected imperfectly during sampling. In the context of wildlife diseases, encounter probabilities may vary as a function of demographic, environmental, spatial, temporal and disease state factors. Capture–mark–recapture (CMR) methods are a theoretically and empirically robust suite of analytical tools that can be used to estimate demographic parameters while accounting for imperfect detection of individuals (Lebreton *et al.* 1992; Williams *et al.* 2002a). If data are collected using a CMR framework, the above simple estimation model can be modified to incorporate imperfect detection and similarly relevant covariates to refine estimates of the one-step transition parameters. Multi-state CMR models that account for detection variability have been successfully developed for estimating disease parameters in previous studies of *M. gallisepticum* dynamics in house finches (Faustino *et al.* 2004; Jennelle *et al.* 2007; Conn & Cooch 2009) and are used to estimate the elements of our P matrix in the following application of Markov chain models.

Application to house finch – *Mycoplasma gallisepticum* system

We now present an example of Markov chain models using data from a house finch – *M. gallisepticum* (MG) system. The *M. gallisepticum* pathogen (common in the poultry industry) was the causal agent of a major epidemic of conjunctivitis in house finches in 1994 and is now widespread among eastern populations (Ley, Berkhoff, & Levisohn 1997; Dhondt, Tessaglia, & Slothower 1998). Besides inducing severe conjunctivitis, MG can cause upper respiratory difficulties, blindness and even death in infected individuals (Ley, Berkhoff, & McLaren 1996; Fisher *et al.* 1997). The disease appears to have stabilized in the eastern USA (Hochachka & Dhondt 2000), suggesting that a Markov chain approach with an assumption of stationary transition probabilities is reasonable. As part of a larger study of the dynamics of MG in finches (Dhondt *et al.* 2005), a capture–recapture field study was conducted in Ithaca, NY, to estimate disease state-specific survival, infection and recovery probabilities. A thorough description of multi-state capture–recapture modelling within this context is presented in Faustino *et al.* (2004). We used Markov chains to compare the virulence of the disease in two separate years: 2002 and 2003. To populate our transition matrices, we obtained weekly state-specific survival and transition probabilities estimated in program MARK (White & Burnham 1999) from Faustino *et al.* (2004) for the autumn–winter seasons of 2002 and 2003. Based on a horizontal transmission study of MG in captive finches, individuals contract the disease on a time frame of a minimum of 3–5 days post-exposure (Sydenstricker *et al.* 2006); as such we used a half-weekly time step for our transition probability matrix. These data produced the following transition probability matrices:

$$P_{2002} = \begin{pmatrix} 0 \cdot 92 & 0 \cdot 02 & 0 \cdot 06 \\ 0 \cdot 09 & 0 \cdot 71 & 0 \cdot 20 \\ 0 & 0 & 1 \end{pmatrix},$$

$$P_{2003} = \begin{pmatrix} 0 \cdot 88 & 0 \cdot 03 & 0 \cdot 09 \\ 0 \cdot 23 & 0 \cdot 63 & 0 \cdot 14 \\ 0 & 0 & 1 \end{pmatrix}. \qquad \text{eqn 13}$$

We calculated the probability distributions for infection of susceptible individuals and recovery of infected individuals according to eqns 4 and 5 respectively (Fig. 1a; 2002 in black, 2003 in grey). We then used the probability distributions of first transitions to determine the cumulative probabilities that a susceptible individual had become infected and an infected

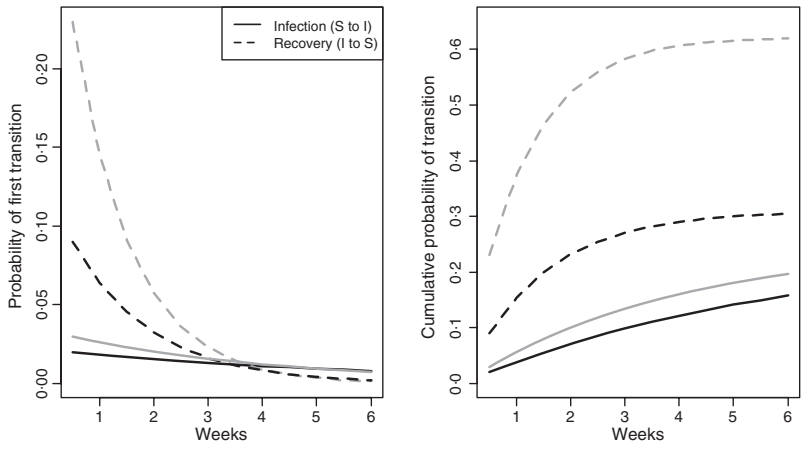

Fig. 1. The left panel shows the probability that a susceptible house finch first becomes infected with MG (2002: solid black line; 2003: solid grey line) and the probability that an infected individual first recovers (2002: dash black line; 2003: dash grey line) in exactly *m* weeks (calculated in half-week time steps). The right panel shows the cumulative probability that a susceptible individual has been infected (2002: solid black line; 2003: solid grey line) and an infected individual has recovered (2002: dash black line; 2003: dash grey line) over the duration of the study.

individual had recovered by a given time step (Fig. 1b; 2002 in black, 2003 in grey). Interestingly, the overall probability of infection for a susceptible individual (eqn 6) was 0·25 in both the 2002 and 2003 seasons, although the time frame of infection and the duration of illness varied between the years (Table 1). The long-term transition probabilities for both infection and recovery stabilized in less than the 3-month duration of the study period (Fig. 1). Given that a susceptible individual becomes infected, the expected time to first infection eqn 8 was longer in 2002 (approximately 44 days) compared with 2003 (approximately 29 days; Table 1). We estimated the life expectancies (eqn 10; expected time to death from the onset of the season) for susceptible and infected individuals in each of the seasons (Table 1) and found that susceptible individuals had longer life expectancy in 2002 compared with that in 2003 but the opposite was true for individuals starting in the infected state (for R code, see Appendix S1).

Discussion

Markov chain models are valuable tools for studying disease dynamics that can improve understanding of infectious wildlife diseases and optimal control strategies. It is useful for researchers and managers to determine the overall likelihood of infection and recovery (and the expected duration of infection) when assessing the virulence of a disease. Such analyses on the level of individuals can be used to develop and parameterize robust population models. Markov chain models can help determine how probabilities of infection vary among individuals of different ages/stages (or geographic regions) which can, in turn, be used to explore disease dynamics at a population level.

Table 1. Estimated disease metrics for house finches exposed to *Mycoplasmal conjunctivitis* during the 2002 and 2003 autumn–winter seasons in Ithaca, NY

Year	2002	2003
Probability of infection	0·25	0·25
Probability of recovery	0·31	0·62
Expected time to infection (days)	43·8	29·2
Expected duration of illness (days)	12·1	9·5
Life expectancy for healthy individuals (weeks)	7·2	5·3
Life expectancy for infected individuals (weeks)	4·0	4·7

In the context of the house finch – *M. gallisepticum* (MG) example, our use of Markov models permitted the estimation of semi-weekly probabilities of infection and recovery in terms of seasonal probabilities of infection and recovery, expected time to and duration of infection and state-dependent life expectancy from the onset of the survey period. The Markov model provided a convenient and accessible approach to calculate disease metrics that summarize relevant aspects of the host–pathogen interaction. The shorter expected time to infection in 2003 compared with that in 2002 is consistent with the notion that there was an increase in MG establishment in the NY study area between seasons. Despite this trend, the overall recovery probability increased and the duration of infection decreased over the same time period. These results are consistent with experimental work that has shown that while MG reinfection is possible in finches, recovery time increases and finches appear to mount a successful immune response (Sydenstricker *et al.* 2005).

Our estimates of life expectancy are unreasonably low (for both disease states), probably due to bias introduced into survival estimates (p_{02} and p_{12}) as a consequence of not accounting for emigration from the study area (Faustino *et al.* 2004). In addition, our Markov model did not account for differences among classes of individuals (e.g. young vs. old), but separate models for each life stage could be created to compare disease dynamics across an individual's lifetime. As the size of the study area did not change, it is reasonable to expect that emigration, while unknown, was consistent between years. Thus, while our estimates of state-specific life expectancy are probably negatively biased, the differences in life expectancy estimates between years may reflect true shifts and a potentially narrowing gap among susceptible and infected individuals. Given that individuals survive initial infection and can be reinfected (Sydenstricker *et al.* 2005), over time we would expect more individuals in the population to have been exposed to MG. Although we define MG-susceptible animals as not expressing conjunctivitis (a not wholly

reliable but convenient and efficient way to assess disease state), this state probably comprises a mixture of MG-naïve and MG-recovered birds. It is reasonable to expect that there may be physiological impacts that influence survival after clinical signs have subsided. The observed decrease in the difference in life expectancy between susceptible and infected birds between years may reflect similar survivorship among finches with and without conjunctivitis due to saturation of a local finch population with repeated exposure to the disease. An additional transient recovery state in the model could help determine how continued reinfection affects disease dynamics.

Although house finches are not a managed species, we believe that the metrics in this paper can be useful in evaluating wildlife disease control efforts. Such metrics can help management agencies evaluate the virulence of a given disease and its occurrence level within a population. Ultimately, determining the mode(s) of disease transmission (e.g. density-dependent and frequency-dependent) will provide a firm basis for evaluating the efficacy of management efforts. Yet, determining modes of disease transmission is not a trivial task and can take many years to unravel. Management agencies cannot afford such a delay in crucial information and must base control actions upon measurable and meaningful disease metrics in the interim time frame. Quantities such as expected time to infection and seasonal probabilities of infection and recovery have clearly interpretable meanings with respect to management efforts and can help set thresholds of acceptable disease persistence. The usefulness of control efforts (at least in the short term), whether host culling or vaccination, can be measured against temporal or treatment-specific comparison of the metrics we present, and can additionally be moulded into an adaptive management framework (Williams *et al*. 2002a).

The purpose of this paper was to demonstrate the usefulness and relative ease of Markov chain modelling in the study of wildlife diseases. We envision this paper can serve as an entry point into the extensive literature and potential applications of Markov chains in epidemiological modelling. We presented a structure for analysis of infectious disease data using first-order (one-step) discrete Markov processes, but it is also possible to accommodate more complex analyses using Markov chains such as higher order or continuous time processes (where a process is dependent upon disease state in time $n - 1$, $n - 2$, etc.; Anderson *et al*. 1993; Resnick 2005). Additionally, more states can be added to account for heterogeneity between naïve and recovered individuals, and state uncertainty can be accommodated during estimation of the one-step transition probabilities. While there are many other, more complicated methods to describe disease dynamics, the simplicity of the discrete time first-order Markov approach is ideal to determine preliminary quantities of interest in the study of disease dynamics.

LIMITATIONS TO THE APPROACH

The first-order Markov approach requires several simplifying assumptions. Most important for the study of wildlife disease are that probability of infection is only dependent on the current state of an individual (i.e. there is no conditioning on prior health history) and that transition probabilities are stationary (i.e. time invariant over the course of the study). While the former assumption may pose less difficulties in the interpretation of transition probabilities made from susceptible to infected states, it could limit the scope of understanding of the recovery process (as recovery is likely conditional on the amount of time spent in the infected state). In first-order Markov processes, the probability of first transition monotonically decreases over time (e.g. Fig. 1). Yet, it is likely that the sojourn time in the infected state is unimodal (i.e. peaks at an intermediate time step) for at least some diseases. More complicated analyses, including continuous time Markov modelling can be used to account for this process (Anderson *et al*. 1993; see also Schaub *et al*. 2001 for a discussion of 'stopover duration', or modelling extended time within a state, in the context of bird migrations). The assumption of stationary transition probabilities may be amenable over the course of short time periods, but care should be taken to choose transition probabilities that are relevant and adequately relate to the temporal scale over which the disease process occurs. In general, it is important to consider the ecology of a disease when applying this (or any other) model. While there are limitations to the application of the methodology, the tractability of Markov chain analyses as well as the ease of calculation and interpretation makes this approach a powerful tool to obtain a basic understanding of wildlife disease dynamics.

Acknowledgements

The authors thank J.A. Royle, E. Osnas, S. Robinson, M. Samuel and D. Storm and the anonymous reviewers for helpful comments that vastly improved the manuscript.

References

Altizer, S., Dobson, A., Hosseini, P., Hudson, P., Pascual, M. & Rohani, P. (2006) Seasonality and the dynamics of infectious diseases. *Ecology Letters*, **9**, 467–484.

Anderson, P.K., Borgan, O., Gill, R.D. & Keiding, N. (1993) *Statistical Models Based on Counting Processes*. Springer, New York, USA.

Arnason, N. (1973) The estimation of population size, migration rates and survival in a stratified population. *Research in Population Biology*, **15**, 1–8.

Bishop, Y.M., Fienberg, S. & Holland, P.W. (1971) Discrete multivariate analysis: theory in practice. *Models for Measuring Change: Contingency Tables and Markov Models*, pp. 257–259. MIT Press, Boston, Massachusetts, USA.

Brownie, C., Hines, J.E., Nichols, J.D., Pollock, K.H. & Hestbeck, J.B. (1993) Capture–recapture studies for multiple state including non-Markovian transitions. *Biometrics*, **49**, 1173–1187.

Choquet, R., Reboulet, A.M., Pradel, R., Gimenez, O. & Lebreton, J.-D. (2005) M-SURGE: new software specially designed for multistate capture–recapture models. *Animal Biodiversity and Conservation*, **27**, 207–215.

Cohen, J.E. (1973) Host mortality in a catalytic model applied to schistosomiasis. *American Naturalist*, **107**, 199–212.

Commenges, D. (1999) Multi-state models in epidemiology. *Lifetime Data Analysis*, **5**, 315–327.

Conn, P. & Cooch, E.G. (2009) Multistate capture–recapture analysis under imperfect state observation: an application to disease models. *Journal of Applied Ecology*, **46**, 486–492.

Dhondt, A.A., Tessaglia, D.L. & Slothower, R.L. (1998) Epidemic *Mycoplasmal conjunctivitis* in house finches from eastern North America. *Journal of Wildlife Diseases*, **3**, 265–280.

Dhondt, A.A., Altizer, S., Cooch, E.G., Davis, A.K., Dobson, A.P., Driscoll, M.J.L., Hartup, B.K., Hawley, D.M., Hochachka, W.M., Hosseini, P.R., Jennelle, C.S., Kollias, G.V., Ley, D.H., Swarthout, E.C.H. & Sydenstricker, K.V. (2005) Dynamics of a novel pathogen in an avian host: *Mycoplasmal conjunctivitis* in House Finches. *Acta Tropica*, **94**, 77–93.

Faustino, C.R., Jennelle, C.S., Connolly, V., Davis, A.K., Swarthout, E.C., Dhondt, A.A. & Cooch, E.G. (2004) *Mycoplasma gallisepticum* infection dynamics in a house finch population: seasonal variation in survival, encounter and transmission rate. *Journal of Animal Ecology*, **73**, 651–669.

Fisher, J.J., Stallknecht, D.E., Lutrell, M.P., Dhondt, A.A. & Converse, K.A. (1997) *Mycoplasmal conjunctivitis* in wild songbirds: the spread of a new contagious disease in a mobile host population. *Emerging Infectious Diseases*, **31**, 69–72.

Hochachka, W.M. & Dhondt, A.A. (2000) Density-dependent decline of host abundance resulting from a new infectious disease. *Proceedings of the National Academy of Sciences*, **97**, 5303–5306.

Jennelle, C.S., Cooch, E.G., Conroy, M.J. & Senar, J.C. (2007) State-specific detection probabilities and disease prevalence. *Ecological Applications*, **17**, 154–167.

Kay, R. (1986) A Markov model for analysing cancer markers and disease states in survival studies. *Biometrics*, **42**, 855–865.

Lay, D.C. (2000) *Linear Algebra and its Applications*. Addison-Wesley, New York.

Lebreton, J.D., Burnham, K.P., Clobert, J. & Anderson, D.R. (1992) Modeling survival and testing biological hypotheses using marked animals: a unified approach with case studies. *Ecological Monographs*, **62**, 67–118.

LeStrat, Y. & Carrat, F. (1999) Monitoring epidemiologic surveillance data using hidden Markov models. *Statistics in Medicine*, **18**, 3463–3478.

Ley, D.H., Berkhoff, J.E. & McLaren, J.M. (1996) *Mycoplasma gallisepticum* isolated from house finches (*Carpodacus mexicanus*) with conjunctivitis. *Avian Diseases*, **40**, 480–483.

Ley, D.H., Berkhoff, J.E. & Levisohn, S. (1997) Molecular epidemiological investigations of *Mycoplasma gallisepticum* (MG) conjunctivitis in songbirds by random amplified polymorphic DNA (RAPD) analyses. *Emerging Infectious Diseases*, **3**, 375–380.

McElhany, P., Real, L.A. & Powers, A.G. (1995) Vector preference and disease dynamics: a study of barley yellow dwarf virus. *Ecology*, **76**, 444–457.

Nichols, J.D. & Kendall, W.L. (1995) The use of multistate capture–recapture models to address questions of evolutionary ecology. *Journal of Applied Statistics*, **22**, 835–846.

Resnick, S.I. (2005) *Adventures in Stochastic Processes*. Birkhauser, Boston, Massachusetts, USA.

Ross, S.M. (2006) *Introduction to Probability Models*, 9th edn. Academic Press, San Diego, California, USA.

Schaub, M., Pradel, R., Jenni, L. & Lebreton, J.D. (2001) Migrating birds stop over longer than usually thought: an improved capture–recapture analysis. *Ecology*, **82**, 852–859.

Schwarz, C.J. (2009) Migration and movement – the next stage. *Modeling Demographic Processes in Marked Populations* (eds D.L. Thomson, E.G. Cooch & M.J. Conroy), Environmental and Ecological Statistics Series, pp. 323–348. Springer, Berlin.

Schwarz, C.J., Schweigert, J.F. & Arnason, A.N. (1993) Estimating migration rates using tag-recovery data. *Biometrics*, **49**, 177–193.

Spiegelhalter, D.J., Thomas, A., Best, N.G. & Lunn, D. (2003) *WinBUGS Version 1.4 User Manual*. MRC Biostatistics Unit, Cambridge, UK.

Sydenstricker, K.V., Dhondt, A.A., Ley, D.H. & Kollias, G.V. (2005) Re-exposure of captive house finches that recovered from *Mycoplasma gallisepticum* infection. *Journal of Wildlife Diseases*, **41**, 326–333.

Sydenstricker, K.V., Dhondt, A.A., Hawley, D.M., Jennelle, C.J., Kollias, H.W. & Kollias, G.V. (2006) Characterization of experimental *Mycoplasma gallisepticum* infection in captive house finch flocks. *Avian Disease*, **50**, 39–44.

Taylor, H.M. & Karlin, S. (1998) *An Introduction to Stochastic Modeling*, 3rd edn. Academic Press, San Diego, California, USA.

White, G.C. & Burnham, K.P. (1999) Program MARK: survival estimation from populations of marked animals. *Bird Study*, **46**, 120–139.

Williams, B.K., Nichols, J.D. & Conroy, M.J. (2002a) *Analysis and Management of Animal Populations*. Academic Press, San Diego, California, USA.

Williams, E.S., Miller, M.W., Kreeger, T.J., Kahn, R.H. & Thorne, E.T. (2002b) Chronic wasting disease of deer and elk: a review with recommendations for management. *Journal of Wildlife Management*, **66**, 551–563.

World Health Organization. (2005) Avian influenza A (H5N1) infection in humans. *New England Journal of Medicine*, **353**, 1374–1385.

Received 12 October 2009; accepted 9 February 2010
Handling Editor: Robert P. Freckleton

Supporting information

Additional Supporting Information may be found in the online version of this article:

Appendix S1. R code to calculate disease metrics in the house finch–MG example.

Appendix S2. R and WinBUGS code to estimate one-step transition probabilities from time-series data.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.