Modeling nest–survival data: a comparison of recently developed methods that can be implemented in MARK and SAS

J. J. Rotella, S. J. Dinsmore & T. L. Shaffer

Abstract

Modeling nest–survival data: a comparison of recently developed methods that can be implemented in MARK and SAS.— Estimating nest success and evaluating factors potentially related to the survival rates of nests are key aspects of many studies of avian populations. A strong interest in nest success has led to a rich literature detailing a variety of estimation methods for this vital rate. In recent years, modeling approaches have undergone especially rapid development. Despite these advances, most researchers still employ Mayfield's *ad–hoc* method (Mayfield, 1961) or, in some cases, the maximum–likelihood estimator of Johnson (1979) and Bart & Robson (1982). Such methods permit analyses of stratified data but do not allow for more complex and realistic models of nest survival rate that include covariates that vary by individual, nest age, time, etc. and that may be continuous or categorical. Methods that allow researchers to rigorously assess the importance of a variety of biological factors that might affect nest survival rates can now be readily implemented in Program MARK and in SAS's Proc GENMOD and Proc NLMIXED. Accordingly, use of Mayfield's estimator without first evaluating the need for more complex models of nest survival rate cannot be justified. With the goal of increasing the use of more flexible methods, we first describe the likelihood used for these models and then consider the question of what the effective sample size is for computation of AICc. Next, we consider the advantages and disadvantages of these different programs in terms of ease of data input and model construction; utility/flexibility of generated estimates and predictions; ease of model selection; and ability to estimate variance components. An example data set is then analyzed using both MARK and SAS to demonstrate implementation of the methods with various models that contain nest–, group– (or block–), and time–specific covariates. Finally, we discuss improvements that would, if they became available, promote a better general understanding of nest survival rates.

Key words: Nest success, Nest survival, Generalized linear models, Avian demography, Mayfield.

Resumen

Modelización de datos de supervivencia en nidos: estudio comparativo de varios métodos desarrollados recientemente que pueden implementarse en MARK y SAS.— La estimación del éxito de nidificación y la evaluación de los factores potencialmente relacionados con las tasas de supervivencia de los mismos son aspectos clave de numerosos estudios sobre poblaciones de aves. El gran interés por el éxito de nidificación se ha traducido en una rica literatura que detalla varios métodos de estimación de esta tasa vital. En los últimos años, los enfoques de modelización han experimentado un rápido desarrollo. No obstante, pese a estos avances, la mayoría de los investigadores siguen empleando el método *ad–hoc* de Mayfield (Mayfield, 1961) o, en algunos casos, el estimador de probabilidad máxima de Johnson (1979) y Bart & Robson (1982). Tales métodos permiten el análisis de datos estratificados, pero, en cambio, no permiten modelos más complejos y realistas de la tasa de supervivencia en nidos cuando se incluyen covariantes que cambian según el individuo, la edad del nido, el tiempo, etc., y que pueden ser continuas o categóricas. Actualmente, con la ayuda de Program MARK, así como de Proc GENMOD y Proc NLMIXED de SAS, es posible implementar métodos que permiten a los investigadores evaluar rigurosamente la importancia de varios factores biológicos susceptibles de incidir en las tasas de

supervivencia en nidos. Por consiguiente, no está justificada la utilización del estimador de Mayfield sin antes evaluar la necesidad de emplear modelos más complejos para determinar la tasa de supervivencia en nidos. Con objeto de incrementar el empleo de métodos más flexibles, primero describimos la probabilidad empleada para estos modelos, para posteriormente tomar en consideración cuál es el tamaño de muestra eficaz para el cálculo de AICc. Seguidamente, tomamos en consideración las ventajas y desventajas de estos programas por lo que respecta a la facilidad de introducción de datos y de construcción de modelos, la utilidad/flexibilidad de las estimaciones y predicciones generadas, la facilidad de la selección de modelos y la capacidad para estimar los componentes de la varianza. A continuación, analizamos un conjunto de datos de ejemplo utilizando los programas MARK y SAS con objeto de demostrar la implementación de los métodos con varios modelos que contienen nido–, grupo– (o bloque–), y covariantes específicas al tiempo. Por último, comentamos varias mejoras que, si estuvieran disponibles, fomentarían una mejor comprensión general de las tasas de supervivencia en nidos.

Jay J. Rotella, Ecology Dept., Montana State Univ., Bozeman, Montana 59717, U.S.A.– Stephen J. Dinsmore, Dept of Wildlife and Fisheries, Box 9690, Mississippi State, Mississippi 39762, U.S.A.– Terry L. Shaffer, Northern Prairie Wildlife Research Center, U S Geological Survey, 8711 37th Street Southeast, Jamestown, North Dakota 58401, U.S.A.

Introduction

Nest success is a key component of reproductive rate for many species of birds, which can be defined as the probability that a nest survives from initiation to completion and has at least one offspring leave the nest. Accordingly, estimates of nest success are crucial to many studies of avian populations, and methods for estimating this vital rate have received considerable attention (e.g., Mayfield, 1961; Johnson, 1979; Hensler & Nichols, 1981; Bart & Robson, 1982; Pollock & Cornelius, 1988; Bromaghin & McDonald, 1993; Aebischer, 1999; Natarajan & McCulloch, 1999; Rotella et al., 2000; Dinsmore et al., 2002). Williams et al. (2002) provide a useful review of historical development, available approaches, and estimation programs. In most studies, nests are found at various (and perhaps unknown) ages, i.e., days since the first egg was deposited in the nest, and estimation is done on a nest's daily survival rate during the period of time it is under observation. The estimated probabilities of surviving each day in the entire nesting cycle are then multiplied together to estimate nest success.

The Mayfield method, either in its original form or as expanded by Johnson (1979) and Bart & Robson (1982), requires the assumption of a constant *daily* nest survival rate for all nests in a sample over the time period being considered. Thus, nest success data are frequently divided into groups for analysis with the Mayfield method, e.g., stratification by stage of the nesting cycle, season, and habitat conditions. But, stratification can lead to small samples for many strata if multiple covariates are used to classify data, which is common because most nesting studies investigate how daily survival rates of nests vary in relation to multiple explanatory variables, many of which are measured on continuous scales.

Despite these limitations, the Mayfield method is still frequently used to analyze nest success data, e.g., Chase (2002), Liebezeit & George (2002), Moorman et al. (2002), Tarvin & Garvin (2002). Accordingly, many studies fail to fully explore their interesting biological questions regarding spatial and temporal variation in daily nest survival. Yet, researchers clearly appear to be interested in such questions. For example, Chase (2002) and Liebezeit & George (2002) both used multiple–logistic regression to compare features of successful and unsuccessful nests. Given that these authors employed the Mayfield method to estimate nest success, it seems that they were aware of the bias inherent in conducting logistic regression analyses on what is essentially apparent nest success, i.e., the proportion of nests in the sample that are successful. Such an analysis is only valid if inactive nests can be found with the same probability as active ones, a condition rarely met in studies of real nests (see Shaffer [2004] for a detailed explanation).

Such problematic analyses are no longer necessary, however, given recent advances in analysis methods. Concurrently, Dinsmore et al. (2002), Stephens (2003), and Shaffer (2004) developed methods for modeling daily survival rates of nests as functions of hypothesized nest–, group– and time–specific covariates. The likelihood–based methods allow visitation intervals to vary among observations and make no assumptions about when nest failure occurs within an interval. Values for time–specific explanatory variables, such as nest age, date, disturbance, and weather, are allowed to vary daily. Models can contain variables that are measured on categorical or continuous scales. Thus, these methods provide a highly–flexible and powerful alternative to traditional constant–nest–survival methods and allow a wide variety of competing models to be assessed via likelihood–based information–theoretic methods (see Burnham & Anderson 2002).

Although the modeling methods presented by Dinsmore et al. (2002), Stephens (2003), and Shaffer (2004) all use the likelihood presented by Dinsmore et al. (2002: pp 3478), there are important differences among the approaches and how they are implemented. For example, methods of Dinsmore et al. (2002) are implemented in program MARK (White & Burnham, 1999), whereas those of Stephens (2003) and Shaffer (2004) are executed in SAS (SAS Institute, 2000). Key differences also exist in terms of input data, how covariates are handled, ease of model specification, the need to provide starting values for parameters, the potential to model random effects, simulation capabilities, and how desired model output is obtained. To facilitate use of these methods, this paper describes the general approach, points out the advantages of implementing the method in different programs, and, where possible, provides methods for re–coding data such that future projects can easily switch from one method to another. In this way, researchers can take advantage of the best features of each method based on project goals and investigator experience with the relevant software packages.

A generalized linear models approach for nest survival

The approaches to modeling nest–survival described by Dinsmore et al. (2002), Stephens (2003), and Shaffer (2004) extend the model described by Bart & Robson (1982). Each employs a generalized linear modeling approach (McCullagh & Nelder, 1989) based on the same binomial likelihood, where daily survival rates are modeled as a function of nest–, group, and/or time–specific covariates. Daily survival rates can then be estimated from the resulting model and multiplied together, as appropriate, to estimate nest success.

To illustrate the model likelihood, let S_i denote the probability that a nest survives from day i to day $i + 1$ (i.e., S_i is a daily survival rate). Consider a nest that was found on day k, was active when revisited on day l, and was last checked on day m ($k < l < m$). Because the nest is known to have survived the

first interval, its contribution to the likelihood for that interval is $S_k S_{k+1}...S_{l-1}$. During the second interval, the nest either survives with probability $S_l S_{l+1}...S_{m-1}$ or fails with probability $(1 - S_l S_{l+1}...S_{m-1})$. The likelihood is thus proportional to the product of probabilities of observed events for all nests in the sample (see Dinsmore et al., 2002).

In the nest survival models discussed here, a link function is used to characterize the relationship between daily survival rate and the covariates of interest. A variety of link functions can be used (White & Burnham, 1999; Williams et al., 2002). Here, focus will be on use of the logit link (and its inverse) as it is the natural link for the binomial distribution (McCullagh & Nelder, 1989). The logit link is frequently used in mark–resight modeling, provides a flexible form, bounds estimates of survival in the (0,1) interval, and is available in MARK and SAS. Dinsmore et al. (2002), Stephens (2003), and Shaffer (2004) all used the logit link in their work, and Lebreton et al. (1992) presented methods for estimating confidence intervals and back–transforming to model parameters and estimates of their variances and covariances when the logit link is used.

Until very recently, the log link was most commonly used in likelihood–based extensions of the Mayfield method, e.g., Johnson (1979), Bart & Robson (1982), and Rotella et al. (2000). The log link is convenient because, unlike for the logit link, covariate values can be summed across a visitation interval to obtain the appropriate covariate value for the interval being modeled (Dinsmore et al., 2002). However, convergence problems can arise when the log link is used, and the log link does not constrain survival to the interval (0,1). Further, the methods described herein provide alternative methods of obtaining the appropriate covariate value for the interval being modeled such that there is no loss of utility when using the logit link instead of the log link. Thus, use of the log link will not be considered here.

With the logit link, daily survival rate of a nest on day i is modeled as

$$\frac{\exp(\beta_0 + \sum_j \beta_j x_{ji})}{1 + \exp(\beta_0 + \sum_j \beta_j x_{ji})}$$

where the x_{ji} ($j = 1,2,...,J$) are values for J covariates on day i and the $\{\beta_j\}$ are coefficients to be estimated from the data. Logit transformation of the above expression yields $\beta_0 + \sum_j \beta_j x_{ji}$. Thus, it can be seen that the relationship between the logit of S_i, i.e., $\ln(S_i / (1 - S_i))$, and the covariates is linear, whereas the relationship between S_i and the covariates is logistic or S–shaped. Once the $\{\beta_j\}$ have been estimated, the $\{S_i\}$ for given values of $\{x_{ij}\}$ can be estimated from the inverse–link function. Note that the above formulation allows daily survival rates to vary among groups of nests (i.e., group–specific covariates), among individual nests (i.e., nest–specific covariates), and among days (i.e., time–specific covariates).

The model likelihood provides insight into the effective sample size for nest–survival data collected from periodic nest visits. A nest that survives an interval of t days is modeled as t Bernoulli trials, whereas a nest that fails the interval is modeled as a single Bernoulli trial. The effective sample size is thus the sum of (1) the total number of days that all nests were known to have survived (*each day* that a nest was known to survive contributes 1 to effective sample size, i.e., the outcome of each Bernoulli trial is known) and (2) the number of intervals that ended in failure (*each interval* in which a nest was known to fail contributes 1 to effective sample size, i.e., the exact day of failure was not known but it is certain that the nest failed during the interval). For example, a nest that survived 45–day intervals and subsequently failed would contribute 21 to the study's effective sample size, regardless of the length of the interval during which the failure occurred.

The parameters $\{\beta_j\}$ of competing models are estimated iteratively by the method of maximum likelihood in software designed for generalized linear models. Accordingly, a variety of likelihood–based methods are available for obtaining parameter estimates and evaluating competing models. In the rare case of control–treatment experiments where nests are randomly allocated to treatment groups, likelihood ratio tests can be used to formally test hypotheses about whether specific covariates are associated with variation in nest survival. If an *a priori* set of candidate models is posed, then information–theoretic measures such as Akaike's information criterion (AIC) can be used to select which model or models to use for inference (Burnham & Anderson, 2002).

Assumptions of the daily nest–survival models described here are: (1) homogeneity of daily survival rates; (2) nest fates are correctly determined; (3) nest discovery and subsequent nest checks do not influence survival; (4) nest fates are independent; (5) all visits to nests are recorded; and (6) nest checks are conducted independently of nest fate. If data are available for > 1 interval length, an extension of the model presented by Rotella et al. (2000) can be used to evaluate and possibly relax assumption 3 as shown below. Assumption 1, by virtue of the fact that daily survival rates can be modeled as a function of group–, nest–, and time–specific covariates, is far less restrictive than is necessary for Mayfield's method. If nest age is to be considered in models of daily survival rate then it is also assumed that nests can be correctly aged when they are first found (Dinsmore et al., 2002).

Three approaches to modeling nest survival

Nest survival model in program MARK (approach 1)

The nest survival model of Dinsmore et al. (2002) is implemented in Program MARK (White & Burnham, 1999). Minimally, five pieces of information are re-

quired for each nest: (1) the day the nest was found; (2) the last day the nest was checked when alive; (3) the last day the nest was checked; (4) the fate of the nest (0 = successful, 1 = depredated), and (5) the number (frequency) of nests that had this history. In MARK, these pieces of information are used to generate an encounter history for each nest in live/dead (LDLD…) format. Encounter histories, which are assigned to groups, can also include individual and group– and time–specific covariates. Group covariates can also be incorporated through the design matrix (White & Burnham, 1999).

For example, to model daily nest–survival rates for Mountain Plovers (*Charadrius montanus*), Dinsmore et al. (2002) standardized 19 May as day 1 of the nesting season and numbered all nest–check dates sequentially thereafter. They assigned each encounter history to one of 12 groups (two sex groups in each of six years). For each nest, they included 78 individual covariates. Seventy six of the covariates accounted for the daily age of the nest on each of the 76 days of the nesting season as a continuous covariate. Beginning on the day the nest was found, nest age (in days) was entered sequentially until hatching age (29 d); all other values were zero. An example encounter history (see tables 1 and 2) for a single nest was: 53 59 63 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27; where the first five numbers represent the critical information for the nest, and the remaining values indicate daily values of nest age (i.e., the x_{ij} values for the ith nest). [Information about the last two covariates, i.e., sex and year, was provided by group assignment (see above).] The use of 76 individual covariates for nest age is a straightforward way of providing information to Program MARK about the age of each nest on each day of the nesting season; however, other options exist for handling nest age if one is proficient with the special functions (i.e., programming statements) that are allowed as entries in the design matrix of Program MARK. For example, one could simply enter the nest's age on the first day of the nesting season (a single value instead of 76 values) and subsequently use this value and special functions in the design matrix to create the ages on other days (see the program's help files for details). In Program MARK, the design matrix allows additional constraints to be placed on parameter estimates. In the Mountain Plover example, daily values for maximum daily temperature and daily precipitation were entered in the design matrix. The design matrix can also be used to specify individual covariates to be included in the model.

If one were interested in estimating the effects of observer visits to nests on nest survival (Rotella et al., 2000), an additional 76 covariates could be added to the encounter history to indicate whether the nest was visited (1 = visited) or not (0 = not visited) on each day. As was done for nest age by

Dinsmore et al. (2002), observer effects could be modeled as a single parameter. More complex models of observer effects could be developed if additional covariates contained information on the nature of the visit, e.g., how closely the nest was approached, how long the visit lasted, and whether or not nest contents were handled.

Clearly, as shown by Dinsmore et al. (2002), a diverse collection of models can be considered with this method. Simpler models that have been commonly employed in past studies of nest survival can also be evaluated. A model that includes a single value of S_i (i.e., constant for all groups, nest ages, dates, and weather conditions) is similar to that of Johnson (1979) and Bart & Robson (1982). Evaluating a model that allows $\{S_i\}$ to vary among groups is analogous to (but more analytically efficient than) conducting a stratified analysis with methods of Johnson (1979) and Bart & Robson (1982) and testing for homogeneity among group–specific survival rates with methods of Sauer & Williams (1989).

Nest survival models in SAS (approaches 2 and 3)

Stephens (2003) and Shaffer (2004) each presented methods for analyzing nest survival data containing multiple interval lengths in SAS (SAS Institute, 2000). These two different approaches have many similarities and several key differences. As was the case for Dinsmore et al.'s (2002) approach, these methods require no assumptions about when nest losses occur during an observation interval for which a nest failure is recorded and can handle diverse types of covariates.

The data input for SAS is different from that used in MARK but ultimately provides the same information. For SAS, each row of input typically contains information for one observation interval for an individual nest. An observation interval is the length of time (t; an integer, typically measured in days) between any two successive nest visits. Note that for a given nest, different observation intervals do not need to be of the same length. The minimum data that must be provided are the length of the interval (t) and the nest's fate for the interval (*IFate*; 1 = successful, 0 = unsuccessful). In addition, individual and group– and time–specific covariates can be included. For example, the date (*StartDate*) and age of the nest (*StartAge*) at the start of the interval might be recorded. Other individual covariates such as habitat measures associated with the nest site could be included. Group covariates such as habitat type or year could also be included.

The example Mountain Plover nest encounter history presented earlier consists of data from two observation intervals and thus, would be entered as two SAS observations (observation one: 6 1 53 4; and observation two: 4 0 59 10; where the variables are *t*, *IFate*, *StartDate*, and *StartAge*). Here, the nest (1) survived the first interval, which was 6 days long and began on day 53 of the

nesting season when the nest was 4 days old and (2) failed during the second interval, which was 4 days long and began on day 59 of the nesting season when the nest was 10 days old. The group assignments used by Dinsmore et al. (2002) can be achieved by adding two additional covariates for sex and year.

Because interval lengths typically are > 1 d, it is necessary to use SAS procedures that allow logistic regression to be done iteratively for each day in an interval. Stephens (2003) and Shaffer (2004) achieved this in slightly different ways with similar end results.

Proc GENMOD – approach 2

Shaffer (2004) used Proc GENMOD (SAS Institute, 2000) to achieve this by (1) specifying use of a binomial probability distribution, (2) defining the inverse link function between an interval's fate and the covariates of interest as

$$\left(\frac{\exp(\beta_0 + \sum_j \beta_j x_{ji})}{1 + \exp(\beta_0 + \sum_j \beta_j x_{ji})} \right)^t$$

and (3) specifying the response as the ratio of the outcome ($IFate$ = 1 or 0) to the number of trials (n = 1) for each interval. For covariates that varied across an interval, e.g., nest age or weather conditions, Shaffer (2004) averaged daily values within each observation interval and used the averages as explanatory variables. For nest age and date, the average value across the interval can easily be obtained using the values for these variables at the start of the interval, the interval length, and SAS's DATA Step. Thus, a variety of models can then be readily evaluated using various combinations of covariates and, if desired, transformed values of covariates, e.g., squared values for evaluating quadratic terms in models.

Proc NLMIXED – approach 3

Stephens (2003) implemented nest–survival models in Proc NLMIXED of SAS (SAS Institute, 2000) using the same data input format as Shaffer (2004) in approach 2. However, this method uses programming statements from within NLMIXED to iteratively do logistic regression for each of the days in an interval (see below). Through programming statements, covariates that vary across an interval in a predictable fashion, e.g., date and age, can be calculated for each day of the interval thus, avoiding the need to work with values that are averaged across an interval.

Consider a model that includes (1) a covariate x_1 that does not vary by time, (2) nest age, and (3) date. This method models the probability that a nest survives a given interval (i.e., probability that $IFate_i$ = 1) as:

$$\prod_{k=0}^{t-1} \left(\frac{\exp(\beta_0 + \beta_1 x_{1i} + \beta_2 (StartAge_i + k) + \beta_2 (StartDate_i + k))}{1 + \exp(\beta_0 + \beta_1 x_{1i} + \beta_2 (StartAge_i + k) + \beta_2 (StartDate_i + k))} \right)^t$$

Applying this model to a 2–d observation interval that started on the 20th day of the nesting season for a nest that was 15 days old at the start of the interval and whose value for covariate x_1 was 8 would yield:

$$\left(\frac{\exp(\beta_0 + \beta_1 8 + \beta_2 15 + \beta_2 20)}{1 + \exp(\beta_0 + \beta_1 8 + \beta_2 15 + \beta_2 20)} \right) \left(\frac{\exp(\beta_0 + \beta_1 8 + \beta_2 16 + \beta_2 21)}{1 + \exp(\beta_0 + \beta_1 8 + \beta_2 16 + \beta_2 21)} \right)$$

Because Stephens'(2003) method allows covariates to vary for different days within an interval, observer effects on nest survival can be modeled in a straightforward manner. Specifically, an index variable (*visit*) is created with programming statements such that it takes on a value of 1 for the first day of an interval (day the nest was visited) and 0 otherwise. This variable can then be used to evaluate whether variation in daily survival rates was associated with observer visits. If additional covariates contain information on the nature of a nest visit, these covariates can be allowed to interact with the *visit* variable to test for their potential influence on survival rate. To illustrate, consider a 2–d interval and a model that includes the effect of (1) an observer visit and (2) a single covariate (x_1) on daily survival rate. With Stephens' (2003) method,

$IFate_i =$

$$\left(\frac{\exp(\beta_0 + \beta_0 visit + \beta_1 x_{1i})}{1 + \exp(\beta_0 + \beta_0 visit + \beta_1 x_{1i})} \right) \left(\frac{\exp(\beta_0 + \beta_1 x_{1i})}{1 + \exp(\beta_0 + \beta_1 x_{1i})} \right)$$

Procedures in SAS allow for examination of a rich collection of models for nest survival and provide an alternative to Program MARK. The NLMIXED procedure also allows models to include random effects as well as fixed effects, i.e., mixed models (SAS Institute, 2000). Mixed models are appropriate if levels of some covariates (i.e., fixed effects) represent all possible levels, or at least the levels for which inferences are desired, i.e., fixed factors, whereas for others covariates (i.e., random effects), the levels observed are only a random sample of a larger set of potential levels of interest (Breslow & Clayton, 1993; Littell et al., 1996; Pinheiro & Bates, 2000). Examples of covariates that might be treated as random effects are study site, year, or individual. This is true because it will often be the case that the sites, years, or individuals studied, i.e., the particular experimental units, are selected at random from the population of interest, i.e., the population of sites, years, or individuals. As stated by Pinheiro & Bates (2000: 8), "they are *effects* because they represent a deviation from an overall mean". Thus, the effect of choosing a particular site, year, or individual may be a shift in the expected response value for observations made on that experimental unit relative to those made on other experimental units experiencing the same levels for the fixed effects. In other words, multiple observations made on the same site, year, or individual may be correlated, and, if so, this should be accounted for in the

analysis. In a broad discussion of data analysis, Littell et al. (1996: vii) stated that, "we firmly believe that valid statistical analysis of most data sets requires mixed model methodology."

Proc NLMIXED allows inclusion of a single random factor, which can then be modeled as an influence on daily survival rate (SAS Institute, 2000). The random effects are assumed to follow normal distributions, typically with zero mean and unknown variances. Examples of the utility of mixed models are provided by Stephens (2003), who reported strong support for models that considered study site as a random factor in a multi–site data set, and by Shaffer (2004) who illustrated use of a random site effect to analyze data from a randomized block study design. The NLMIXED procedure can be used regardless of whether Stephens (2003) or Shaffer's (2004) approach to handling values of time–varying covariates is used.

Strengths and weaknesses of each approach

It is clear that each of the methods described above can consider a wide variety of models and allow for substantial improvements over analyses that have been typically employed in nesting studies. All methods will yield the same results for models that do not consider covariates that vary during an observation interval (Shaffer, 2004). For models that do consider time–varying covariates such as nest age, differences in results obtained from Shaffer's (2004) method and those of Dinsmore et al. (2002) or Stephens (2003) will be slight, especially if intervals are short and effects of time–varying covariates are modest (Shaffer, 2004). Regardless of these similarities, there are advantages and disadvantages to each approach that may influence which method is most appropriate for use in a specific study.

Program MARK is readily available at no cost at the following URL: http://www.cnr.colostate.edu/~gwhite/mark/mark.htm. Also, its use is well documented (e.g., White & Burnham, 1999; Dinsmore et al., 2002), and formal coursework and internet teaching materials are readily available to those interested in using the software. Software support is available through an electronic analysis forum at: http://www.phidot.org/forum/index.php. Further, Program MARK provides a consistent approach to implementing a broad variety of mark–resight analyses. Once the software is learned, MARK allows users to build a variety of models and to easily employ different link functions. AIC model–selection and model–averaging capabilities are built into the software. In addition, because Program MARK has a module specifically designed for analyzing nest–survival data, effective sample size is automatically calculated and used in AIC_c calculations.

One drawback to use of Program MARK is the method in which nest–specific, time–varying

covariates are entered. If nesting seasons comprise many days and a number of individual and time–varying covariates are to be considered, input files can be cumbersome. Further complications with the handling of covariates can be encountered during the modeling process if one is interested in evaluating various transformations of the covariates, including interactions or polynomial terms. The difficulty can be mitigated if all covariates of interest, including transformed values and products of multiple variables (for interaction terms) are included in the input file. Also, the tools available for transforming covariates and creating interaction terms within Program MARK are continually being improved.

Monte Carlo simulation studies are useful for evaluating various properties of estimation and model–selection methods under different modeling scenarios. Program MARK does not currently have simulation capabilities for nest–survival data. Thus, Monte Carlo methods cannot yet be employed in this software.

Because SAS is commonly used for a many types of statistical analyses, many researchers have ready access to the software and are familiar with its use. In SAS, the formatting of input data is flexible, and powerful tools are available for subsequent data manipulation through the DATA Step. In addition, programming capabilities are available within the relevant procedures. Thus, link functions other than the logit can be implemented (though not as easily as in Program MARK), and model specification is straightforward. Specification of models involving categorical covariates can be somewhat tedious in Proc NLMIXED, as categorical covariates must be recoded as indicator variables in a DATA Step. Proc GENMOD allows the use of a CLASS statement for specifying categorical covariates. This can save time and frustration if there are multiple categorical covariates, especially if those covariates take on numerous values, and models with and without interactions are being considered.

As noted above, Proc NLMIXED allows random effects to be incorporated, which may be of interest in some studies. A variety of output data sets can be created in SAS that can be useful for examining predicted values across ranges of covariate values. AIC model–selection and model–averaging capabilities are not built into the software, but macros for performing the computations are available for Proc GENMOD at http://www.npwrc.usgs.gov/resource/tools/nestsurv/nestsurv.htm. A macro that generates a table of AIC model–selection results from PROC NLMIXED is located at http://www.montana.edu/rotella/nestsurv/. Because SAS procedures are not specifically designed for analyzing nest–survival data, the effective sample size is not automatically calculated and used in AIC_c calculations. Rather, the number of observations (intervals) present in the input file is considered to be the sample size. Fortunately, the effective sample size is easily calculated in a DATA Step, and the value can then be used to correctly

calculate AIC$_c$ (see code at http://www.npwrc.usgs.gov/resource/tools/nestsurv/nestsurv.htm or http://www.montana.edu/rotella/nestsurv/).

Monte Carlo simulations are possible in SAS using various programming statements and macros (Fan et al., 2003). Simulations can be useful during the planning stages of a study by allowing the researcher to understand how results may vary as a function of the generating model, estimating models, sampling designs, and sampling effort. SAS programs for conducting simulations and summarizing results of analyses with Proc NLMIXED can be obtained at http://www.montana.edu/rotella/nestsurv/

Program MARK and Proc NLMIXED both allow one to input starting values for each parameter in the model. By default, when the logit link is used and one does not specify starting values for the parameters, Program MARK starts the optimization of all parameter estimates at 0.01; and NLMIXED uses a default value of 1. Model convergence can be affected by the choice of starting values, and researchers may have to experiment with different starting values to achieve convergence. This is usually not an issue unless sample sizes are small, or models are fairly complex or involve a random effect. Non–convergence can also occur when strong collinearity exists among covariates or with erroneously coded data (e.g., interval lengths < 1). Useful strategies for researchers who are experiencing convergence problems with MARK or NLMIXED is to obtain starting values by first fitting a simpler model that is nested within the model you are currently building. In our experience, this approach is critical when working in NLMIXED with models containing random effects. If convergence problems still exist, then we suggest obtaining starting values for the model's parameters using Proc GENMOD. In most cases, starting values obtained in this manner will be very similar to or identical to final estimates produced by MARK or Proc NLMIXED (see Shaffer, 2004). A limitation of this approach is that models involving observer effects or random effects cannot be fitted in Proc GENMOD. However, Proc GENMOD can still be used to estimate starting values for other covariates in the model.

Given that each of the methods described above has slight advantages and disadvantages, it is worth noting that it is not particularly difficult to convert a dataset prepared for analysis in SAS into an input file appropriate for use in Program MARK. This can be accomplished using SAS' DATA Step and programming statements to re–configure a SAS input file that consists of multiple data records per nest into a MARK input file that contains a single record for each nest. Such a conversion, which is implemented in the example below, allows users to readily analyze data in either program MARK or SAS and thus, make the best use of each program's advantages. Converting a dataset prepared for analysis in Program MARK into an input file appropriate for use in SAS can also be accomplished (see below). However, the input file for Program MARK will not necessarily include the information about the dates of all nest visits, and thus, in such cases, estimation of observer effects on daily nest survival rate will not be possible unless the input data are modified to include such information.

An example

To demonstrate the implementation of the methods described above in both SAS and MARK, an analysis of data for Mallard (*Anas platyrhynchos*) nests that were monitored during 2000 in the Coteau region of North Dakota as part of a larger study (Stephens, 2003) is presented for various models that contain nest–, group– and time–specific covariates. The data set contains information from a total of 1,585 observation intervals made on 565 nests that were monitored on 18 sites during a 90–d nesting season. Interval lengths were typically 4, 5, or 6 d (average = 4.66 d, SD = 1.41 d). Here, the following subset of the covariates measured by Stephens (2003) was considered: (1) nest age (Age; 1 to 35 d), (2) date (*Date*; 1 to 90), (3) vegetative visual obstruction at the nest site (*Robel*; Robel et al., 1970), (4) the proportion of grassland cover (*PpnGr*) on the 10.4–km^2 study site that contained the nest, (5–7) the habitat type in which the nest was located (3 indicator variables, each coded as 0 or 1, that were used to distinguish among nests found in native grassland [*NatGr*], planted nesting cover [*PlCov*], wetland vegetation [*Wetl*], and roadside right–of–ways [*Road*]), (9) study site (*Site*), and (10) nest–visitation status (Ob–an indicator variable coded as 1 on the day a nest was visited and 0 otherwise).

Data were originally recorded in interval–specific form, i.e., each row of data contained information for one observation interval for an individual nest (table 1). Data in this format were appropriate for analysis by approaches 2 and 3 in SAS but needed to be re–formatted before being analyzed in Program MARK. Accordingly, SAS DATA Steps and programming statements (appendix 1) were used to re–configure the original data file into a MARK input file with a single record per nest (table 2). Likewise, the MARK input file can be re–configured appropriately for analysis in SAS (appendix 2).

Competing models of nest survival were first evaluated using Program MARK. Because these analyses were done for illustrative purposes, the model list was kept brief, and justification of the models is not provided. No model in the list included study site because this was considered to be a random effect and so was beyond the current capabilities of nest–survival analyses in MARK. Although effects of observer visits on nest survival could have been considered in MARK, they were not because we considered it best to estimate observer effects in models that considered as many other sources of heterogeneity as possible, i.e., in models that contained the full suite of both fixed and random effects.

Table 1. Input format for interval–specific nest–survival data: ID. Nest number; Species. Species code; Site. Study site; Hab. Habitat code; Int. Observation interval; t. Interval length (d); *IFate*. Nest fate for the interval; *SDate*. Date at the start of the interval; *SAge*. Nest age at the start of the interval; *Robel*. Vegetative visual obstruction at nest site; *PpnGr*. Proportion of grassland cover on the 10.4 km² study site. (Also, see table 2.)

Tabla 1. Formato para la introducción de datos específicos a un intervalo, relacionados con la supervivencia en: ID. Número del nido; Species. Código de la especie; Site. Lugar del estudio; Hab. Código del hábitat; Int. Intervalo de observación; t. Duración del intervalo (d); IFate. Destino del nido para el intervalo; SDate. Fecha al inicio del intervalo; SAge. Edad del nido al inicio del intervalo; Robel. Obstrucción vegetativa visual en el lugar del nido; PpnGr. Proporción de cobertura de pasto en los 10,4 km² que ocupa el lugar del estudio. (Véase también la tabla 2.)

ID	Species	Site	Hab	Int	t	IFate	SDate	SAge	Robel	PpnGr
1	Mall	14	PlCov	1	5	1	1	1	4.50	0.96
1	Mall	14	PlCov	2	5	1	6	6	4.50	0.96
1	Mall	14	PlCov	3	4	1	11	11	4.50	0.96
1	Mall	14	PlCov	4	6	1	15	15	4.50	0.96
1	Mall	14	PlCov	5	5	1	21	21	4.50	0.96
1	Mall	14	PlCov	6	5	1	26	26	4.50	0.96
1	Mall	14	PlCov	7	4	1	31	31	4.50	0.96
2	Mall	14	PlCov	1	5	1	1	3	0.88	0.96
2	Mall	14	PlCov	2	5	1	6	8	0.88	0.96
2	Mall	14	PlCov	3	4	1	11	13	0.88	0.96
2	Mall	14	PlCov	4	6	0	15	17	0.88	0.96
…										
2,206	Mall	16	Road	1	4	1	73	13	6.00	0.80
2,206	Mall	16	Road	2	5	1	77	17	6.00	0.80
2,206	Mall	16	Road	3	4	1	82	22	6.00	0.80
2,206	Mall	16	Road	4	3	1	86	26	6.00	0.80

The most parsimonious fixed–effects model of nest survival included Age and *PpnGr* (table 3). This model was 1.06 AIC_c units better than the second–best model, which included Age but not *PpnGr,* and was $\geq$ 2.90 AIC_c units better than all other models evaluated. The best model indicated that daily survival rate increased with nest age ($\hat{\beta}$ = 0.0188, SE = 0.007) and grassland extent ($\hat{\beta}$ = 0.369, SE = 0.211) (fig. 1). Models that held daily survival rate constant or simply allowed it to vary by habitat type, i.e., the only model types that have been used in many recent publications on nest survival (see above), received little support ($\Delta AIC_c \geq$ 6.11).

Shaffer (2004) showed that analyses using his methods in Proc GENMOD yield (1) the same results as those obtained in MARK for fixed–effects models with time–invariant covariates and (2) similar answers to those provided by MARK for models with time–varying covariates, e.g., nest age or date. Results from the GENMOD approach are not presented here (see appendix 3 for example code), but they were indeed nearly identical to those obtained with MARK.

Proc NLMIXED was used to evaluate the same list of fixed–effects models; results were virtually identical to those obtained from MARK. For example, t to at least 4 decimal places ($\hat{\beta}$ = 0.3689, SE = 0.2105).

To evaluate the importance of considering more complex models, three additional models were evaluated using Proc NLMIXED (appendix 4). These models were created by adding observer effects, a random effect of site, or both effects to the most parsimonious fixed–effects model. Of the 12 models considered, the two most parsimonious models both included a random effect of site (table 4). Spatial process variance was estimated as 0.089 (SE = 0.052) by the better of these two models. The second–most parsimonious model (ΔAIC_c = 0.33) provided some evidence of a negative effect of observer visits on daily survival rate for the day immediately following a nest visit ($\hat{\beta}$ = –0.844, SE = 0.629). The point estimate indicates that the effect was potentially of a size that is of interest, but the lack of precision makes inference difficult. For example, on a site with 50% grassland cover, daily survival

Table 2. Input format for nest–survival data to be analyzed in Program MARK (White & Burnham, 1999; Dinsmore et al., 2002) where the numeric variables are: 1. Nest identification number; 2. The date the nest was found; 3. The last date the nest was known to be alive; 4. The date that the nest's final fate was determined; 5. The nest's fate (0 = successful, 1 = depredated); 6. The number of nests with this history; 7–95. The age of the nest on the first 89 days of the 90–day nesting season (a 0 is used for all dates preceding the date the nest was found and following age 35, i.e., the maximum age); 96. Vegetative visual obstruction at the nest site; 97. Proportion of grassland cover at nest site (10.4 km^2); 98–101. Indicator variables indicating the habitat type a nest was in. (See table 1 for alternative format.)

Tabla 2. Formato para la introducción de datos relativos a la supervivencia en nidos que se analizarán con Program MARK (White & Burnham, 1999; Dinsmore et al., 2002), donde las variables numéricas son: 1. El número de identificación del nido; 2. La fecha en que se encontró el nido; 3. La última fecha en que se tenía conocimiento de que el nido estaba vivo; 4. La fecha en que se determinó el destino final del nido; 5. El destino del nido (0 = satisfactorio, 1 = depredado); 6. El número de nidos con este historial; 7–95. La edad del nido durante los primeros 89 días de la estación de anidamiento de 90 días (para todas las fechas anteriores a la fecha en que se encontró el nido, se emplea un 0, seguido de la edad 35; es decir, la edad máxima); 96. Obstrucción vegetativa visual en el lugar del nido; 97. Proporción de cobertura de hierba en el lugar del nido (10,4 km^2); 98–101. Variables indicadoras del tipo de hábitat en que se encontraba el nido. (Para un formato alternativo, ver tabla 1.)

Nest Survival Group = 1;

/* 1 */ 1 35 35 0 1 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 4.5 0.9616 0 1 0 1;

/* 2 */ 1 15 21 1 1 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0.875 0.9616 0 1 0 1;

...

/* 2206 */ 73 89 89 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 6 0.8002 0 0 0 1;

for a 15–d old nest would be predicted as 0.911 (SE = 0.033, 95% CI = 0.842 to 0.981) if it were visited and 0.960 (SE = 0.010, 95% CI = 0.939 to 0.981) otherwise, where the estimates were obtained using the ESTIMATE statement (1 statement for each of the 2 scenarios) of Proc NLMIXED (see appendix 4).

When the random effect of site was added to various fixed–effects models, estimates of coefficients for fixed effects were quite stable with one notable exception. The coefficient for grassland extent was reduced from 0.369 to 0.086, while the estimated standard error increased slightly (0.211 to 0.233), which alters the inferences that can be drawn from this data set in important ways, i.e., the importance of grassland extent is called into question. Because our emphasis here is on illustrating the various analysis methods, these changes are not discussed further, and other models that might be suggested by the data are not explored. However, Stephens (2003) conducted a detailed analysis of both *a priori* and exploratory models for the larger, multi–year dataset that included the data analyzed here.

Recommendations

The methods presented by Dinsmore et al. (2002), Stephens (2003) and Shaffer (2004), which are elaborated on here, allow a variety of competing models to be assessed via likelihood–based information–theoretic methods. Thus, they provide excellent alternatives to traditional constant–survival methods, and these three approaches can be used interchangeably as best suits a particular problem. The methods presented here: (1) can be used to conduct analyses of stratified data (appropriate if the simplifying assumptions of constant survival apply) and provide estimates that are almost identical to Mayfield estimates (or various refinements), (2) perform comparisons of survival rates among groups, (3) allow a much broader variety of covariates and competing models to be evaluated, and (4) should be employed in most nest–survival studies. It is worth noting that these methods can also be used for analyzing survival data collected from radiomarked individuals using ragged (uneven) intervals among animals and over time.
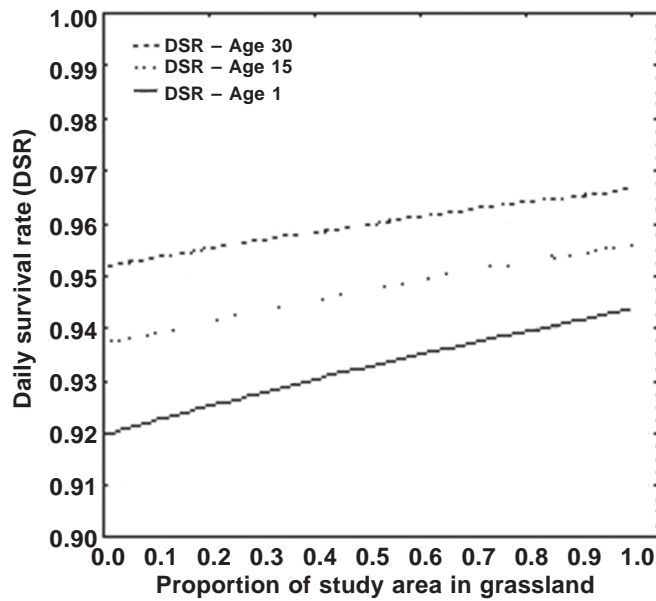
Fig. 1. Estimated relationship between daily survival rate (S) and the proportion of a study site consisting of grassland (*PpnGr*) for Mallard nests of different ages. Estimates from the best fixed-effects model where $\log (S / (1 - S)) = 2.43 + 0.019 \cdot Age + 0.369 \cdot PpnGr$.

*Fig. 1. Relación estimada entre la tasa de supervivencia diaria (*S*) y la proporción de un lugar de estudio formado por pastos (*PpnGr*) en nidos de distintas edades de Mallard. Las estimaciones de los mejores modelos de efectos fijos fueron $\log (S / (1 - S)) = 2.43 + 0.019 \cdot Age + 0.369 \cdot PpnGr$.*

Table 3. Summary of model–selection results obtained in Program MARK (White & Burnham, 1999; Dinsmore et al., 2002) for fixed–effects models of daily survival rate for Mallard nests studied by Stephens (2003) in North Dakota. K is the number of parameters in the model, and w_i is the model weight.

Tabla 3. Resumen de los resultados sobre la selección de modelos obtenidos con Program MARK (White & Burnham, 1999; Dinsmore et al.*, 2002) para los modelos de efectos fijos de la tasa de supervivencia diaria en nidos de Mallard estudiados por Stephens (2003) en Dakota del Norte. K es el número de parámetros que contiene el modelo, y* w_i *es el peso del modelo.*

Model	K	AICc	ΔAIC_c	w_i
$\beta_0 + \beta_1{}^*Age + \beta_2{}^*PpnGr$	3	1563.010	0.000	0.465
$\beta_0 + \beta_1{}^*Age$	2	1564.066	1.056	0.274
$\beta_0 + \beta_1{}^*Age + \beta_2{}^*Robel$	3	1565.906	2.896	0.109
$\beta_0 + \beta_1{}^*Age + \beta_2 NatGr + \beta_3{}^*CRP + \beta_4{}^*Wetl$	5	1567.344	4.334	0.053
$\beta_0 + \beta_1{}^*PpnGr$	2	1567.368	4.358	0.053
β_0	1	1569.117	6.107	0.022
$\beta_0 + \beta_1{}^*Robel$	2	1570.775	7.765	0.010
$\beta_0 + \beta_1{}^*Date$	2	1570.826	7.817	0.009
$\beta_0 + \beta_1{}^*NatGr + \beta_2{}^*CRP + \beta_3{}^*Wetl$	4	1571.957	8.948	0.005

Table 4. Summary of model–selection results obtained in Proc NLMIXED (SAS Institute, 2000) for fixed–effects and mixed models of daily survival rate for Mallard nests studied by Stephens (2003) in North Dakota.

Tabla 4. Resumen de los resultados sobre la selección de modelos obtenidos con Proc NLMIXED (SAS Institute, 2000) para modelos mixtos y de efectos fijos de la tasa de supervivencia diaria correspondiente a los nidos de Mallard estudiados por Stephens (2003) en Dakota del Norte.

Model	K	AICc	ΔAIC$_c$	w_i
$\beta_0 + \beta_1{*}Age + \beta_2{*}PpnGr + b_1{*}$site	4	1554.013	0.000	0.529
$\beta_0 + \beta_1{*}Age + \beta_2{*}PpnGr + \beta_3{*}Ob + b_1{*}$site	5	1554.340	0.327	0.449
$\beta_0 + \beta_1{*}Age + \beta_2{*}PpnGr + \beta_3{*}Ob$	4	1562.265	8.252	0.009
$\beta_0 + \beta_1{*}Age + \beta_2{*}PpnGr$	3	1563.010	8.996	0.006
$\beta_0 + \beta_1{*}Age$	2	1564.066	10.053	0.003
$\beta_0 + \beta_1{*}Age + \beta_2{*}Robel$	3	1565.906	11.892	0.001
$\beta_0 + \beta_1{*}Age + \beta_2 NatGr + \beta_3{*}CRP + \beta_4{*}Wetl$	5	1567.344	13.330	0.001
$\beta_0 + \beta_1{*}PpnGr$	2	1567.368	13.355	0.001
β_0	1	1569.117	15.103	0.000
$\beta_0 + \beta_1{*}Robel$	2	1570.775	16.762	0.000
$\beta_0 + \beta_1{*}Date$	2	1570.826	16.813	0.000
$\beta_0 + \beta_1{*}NatGr + \beta_2{*}CRP + \beta_3{*}Wetl$	4	1571.957	17.944	0.000

Despite these advances, further analysis improvements would be useful. The methods presented here do not consider age–specific encounter probabilities, where "age" refers to the age of a newly encountered nest, as do some methods for survival analysis (Pollock & Cornelius, 1988; Williams et al., 2002). Information on age–specific nest encounter probabilities can provide information about survival probabilities prior to encounter. The utility of such information is presented by Williams et al. (2002).

Improved methods of estimating goodness–of–fit and for detecting and estimating overdispersion, or extra–binomial variation, would be useful given that a variety of factors may cause overdispersion. Nest–success data are commonly collected according to multilevel designs that result in grouped data, e.g., multiple observations on at least some nests, multiple nests per site, and multiple sites within each year. Thus, undetermined random effects of individuals, sites, and years could cause overdispersion or within–group correlations in daily survival rates, e.g., nest fates from multiple nests from within a colony or from a given study plot may not be independent. In addition, the spatial clustering of covariate levels could generate spatial correlation in nest survival rates and thus cause overdispersion. The random–effects model implemented in Proc NLMIXED offers an improvement as it can estimate random effects due to one source, e.g., site. However, current methods in NLMIXED do not accommodate multi–level nonlinear mixed models (e.g., some random effects associated with site, some associated with year, and others associated with individuals), although, as mentioned above, they will be of interest in some studies. Recently, Bayesian techniques (Cam et al., 2002; Link et al., 2002; Williams et al., 2002) were used to address individual heterogeneity in mark–resight analysis, and to estimate age–specific nest–survival rates (He et al., 2001; He, 2003). Thus, Bayesian techniques may hold potential for improving future modeling of nest survival data. A composite likelihood approach (Lele & Taper, 2002) has been used successfully for nest–success data (M. Taper, S. Lele, & J. J. Rotella) and should allow more thorough treatment of multiple random effects in nest–survival data in the future, e.g., through simultaneous consideration of factors such as individuals, sites, and years. In some studies, uncertainty will exist about nest ages and when transitions among nest stages occur (Williams et al., 2002). This problem has been addressed for stratified data (Stanley, 2000) but not yet for data sets containing more complex sets of covariates.

References

Aebischer, N. J., 1999. Multi–way comparisons and generalized linear models of nest success: extensions of the Mayfield method. *Bird Study,* 46

(supplement): S22–S31.

Bart, J. & Robson, D. S., 1982. Estimating survivorship when the subjects are visited periodically. *Ecology,* 63: 1078–1090.

Breslow, N. E. & Clayton, D. G., 1993. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association,* 88: 9–25.

Bromaghin, J. F. & McDonald, L. L., 1993. Weighted nest survival models. *Biometrics,* 49: 1164–1172.

Burnham, K. P. & Anderson, D. R., 2002. *Model Selection and Multi–model Inference: a Practical Information–Theoretic Approach,* 2nd ed. Springer–Verlag, New York.

Cam, E., Link, W. A., Cooch, E. G., Monnat, Y.–A., & Danchin, E., 2002. Individual covariation in life–istory traits: seeing the trees despite the forest. *American Naturalist,* 159.

Chase, M. K., 2002. Nest site selection and nest success in a song sparrow population: the significance of spatial variation. *Condor,* 104: 103–116.

Dinsmore, S. J., White, G. C. & Knopf, F. L., 2002. Advanced techniques for modeling avian nest survival. *Ecology,* 83: 3476–3488.

Fan, X., Felsovalyi, A., Sivo, S. A. & Keenan, S. C., 2003. *SAS for Monte Carlo studies: a guide for quantitative researchers.* SAS Institute, Inc., Cary, North Carolina.

He, C. Z., 2003. Bayesian modeling of age–specific survival in bird nesting studies under irregular visits. *Biometrics,* 59: 962–973.

He, C. Z., Sun, D. & Tra, Y., 2001. Bayesian modeling age–specific survival in nesting studies under Dirichlet priors. *Biometrics,* 57: 281–288.

Hensler, G. L. & Nichols, J. D., 1981. The Mayfield method of estimating nest success: A model, estimators and simulation results. *Wilson Bulletin,* 93: 42–53.

Johnson, D. H., 1979. Estimating nest success: the Mayfield method and an alternative. *Auk,* 96: 651–661.

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.

Lele, S. & Taper, M. L., 2002. A composite likelihood approach to estimation of (co)variance components. *Journal of Statistical Planning and Inference,* 103: 117–135.

Liebezeit, J. R., & George, T. L., 2002. Nest predators, nest–site selections, and nesting success of the dusky flycatcher in a managed ponderosa pine forest. *Condor,* 104: 507–517.

Link, W. A., Cam, E., Nichols, J. D. & Cooch, E., 2002. Of BUGS and birds: an introduction to Markov chain Monte Carlo. *Journal of Wildlife Management,* 66: 277–291.

Littell, R. C., Milliken, G. A., Stroup, W. W. & Wolfinger, R. D., 1996. *SAS system for mixed models.* SAS Institute Inc., Cary, North Carolina.

Mayfield, H. F., 1961. Nesting success calculated from exposure. *Wilson Bulletin,* 73: 255–261.

McCullagh, P. & Nelder, J. A., 1989. *Generalized linear models,* 2nd ed. Chapman and Hall, New York.

Moorman, C. E., Guynn Jr., D. C. & Kilgo, J. C., 2002. Hooded warbler nesting success adjacent to group–selection and clearcut edges in a southeastern bottomland forest. *Condor,* 104: 366–377.

Natarajan, R. & McCulloch, P. C. E., 1999. Modeling heterogeneity in nest survival data. *Biometrics,* 55: 553–559.

Pinheiro, J. C. & Bates, D. M., 2000. *Mixed–effects models in S and S–plus.* Springer–Verlag, New York.

Pollock, K. H. & Cornelius, W. L., 1988. A distribution–free nest survival model. *Biometrics,* 44: 397–404.

Robel, R. J., Briggs, J. N., Dayton, A. D. & Hulbert, L. C., 1970. Relationships between visual obstruction measurements and weights of grassland vegetation. *Journal of Range Management,* 23: 295–298.

Rotella, J. J., Taper, M. L. & Hansen, A. J., 2000. Correcting nesting success estimates for possible observer effects: maximum–likelihood estimates of daily survival rates with reduced bias. *Auk,* 117: 92–109.

SAS Institute, 2000. *SAS/STAT user's guide, Version 8.* SAS Institute, Inc., Cary, North Carolina.

Sauer, J. R. & Williams, B. K., 1989. Generalized procedures for testing hypotheses about survival or recovery rates. *Journal of Wildlife Management,* 53: 137–142.

Shaffer, T. L., 2004. A unified approach to analyzing nest success. *Auk,* 121: 526–540.

Stanley, T. R., 2000. Modeling and estimation of stage–specific daily survival probabilities of nests. *Ecology,* 81: 2048–2053.

Stephens, S. E., 2003. The influence of landscape characteristics on duck nesting success in the Missouri Coteau Region of North Dakota. Ph. D. Dissertation, Montana State Univ.

Tarvin, K. A. & Garvin, M. C., 2002. Habitat and nesting success of blue jays (*Cyanocitta cristata*): importance of scale. *Auk,* 119: 971–983.

White, G. C. & Burnham, K. P., 1999. Program MARK: survival estimation from populations of marked animals. *Bird Study,* 46 Supplement: 120–138.

Williams, B. K., Nichols, J. D. & Conroy, M. J., 2002. *Analysis and management of animal populations: modeling, estimation, and decision making.* Academic Press, New York.

Appendix 1. Code for converting an example dataset prepared for analysis in SAS (SAS Institute, 2000) into an input file for Program MARK (White & Burnham, 1999).

Apéndice 1. Código para convertir un conjunto de datos de un ejemplo preparado para ser analizado en SAS (SAS Institute, 2000) a un archivo de datos para Program MARK (White & Burnham, 1999).

```
proc sort data = sasuser.mall2000nd;
by id int;
run;

data mark1; set Sasuser.mall2000nd;
        retain firstday; retain firstage;
    by id;
        if first. id then do;
            firstday= sdate;  /*date at start of 1st interval = date found*/
            firstage=sage;    /*age at start of 1st interval = age when found*/
        end;
        if last. id then do;
            lastdaylive= sdate; /*date at start of last interval for nest*/
            lastday=sdate + t;  /*date at end of last interval for nest*/
        end;
        if firstday=. then delete;
        if lastday=. then delete;
        if lastdaylive=. then delete;
/* create indicator variables for different nesting habitats*/
    if hab=1 then NatGr=1; else NatGr=0;   /*Native Grassland*/
    if hab=2 or hab=3 or hab=9 then PlCov=1; else PlCov=0;    /*Planted Cover*/
    if hab=7 or hab=22 then Wetl=1; else Wetl=0;  /*Wetland sites*/
    if hab=20 or hab=8 then Road=1; else Road=0;       /*Roadside sites*/
drop sdate sage;
run;

data mark2; set mark1;
array age{90};/*where 90 is the number of days in the study's nesting season*/
    do i=1 to 90;
        if i<firstday then age{i}=0;
        else  age{i}=firstage+i-firstday;
    end;
    do i=1 to 90;
        if age{i}>35 then age{i}=0; /*where 35 is the species' maximum nest*/
    end;
run;
data markinp; set mark2;
/*Create a text file with the necessary output for MARK*/
/*The directory used in the statement below must exist on the computer being used*/
file 'C:\My Documents\nest success\MallMARK.inp' ;
/*Use ID number for first nest in the data set to put header line on file for MARK*/
/*the next line can be deleted if ID numbers aren't in the dataset. But the header line*/
/*for MARK must then be put in by hand before using it in MARK.*/
if id=1 then put «Nest Survival Group=1 ;» ;
/*Code below assumes that ifate is 1 for a nest that survives an interval and*/
/*that ifate is 0 for nests that fail during an interval between 2 visits.*/
if ifate = 1 then put «/* « id « */ « firstday lastday lastday « 0 1 « age1-age89 robel
PpnGr NatGr PlCov Wetl Road «;»;
    else put «/* « id « */ « firstday lastdaylive lastday « 1 1 « age1-age89 robel
PpnGr NatGr PlCov Wetl Road «;»;
run;
```

Appendix 2. Code for converting a particular dataset prepared for analysis in Program MARK (White & Burnham, 1999) into an input file for SAS (SAS Institute, 2000).

Apéndice 2. Código para convertir un conjunto de datos concreto preparado para ser analizado con Programa MARK (White & Burnham, 1999) a un archivo de datos para SAS (SAS Institute, 2000).

```
* This program reads the MARK input file shown in table 3 and creates a SAS
* data set that can be analyzed with Proc Genmod or Proc Nlmixed.
* Because the MARK format does not contain information on all intermediate visits
* to a nest, a maximum of 2 records is generated for each nest.;

data mark2sas;
  array age(89) age1-age89;
  Infile 'C:\My Documents\nest success\MallMARK.inp' firstobs=2 lrecl=750;
  Input junk $ id junk $ firstday lastdaylive lastday markfate freq age1-age89 Robel PpnGr NatGr PlCov Wetl Road;
  If markfate=0 then do;  /* successful nest - generate 1 interval with ifate=1 */
    ifate=1;
    sdate=firstday;
    t = lastday - firstday;
    sage = age{sdate};
    output;
  end;
  If markfate=1 then do; /* unsuccessful nest - generate 1 interval with ifate=1 and 1 interval with ifate=0 */
    ifate=1;
    sdate=firstday;
    t = lastdaylive - firstday;
    sage = age{sdate};
    output;
    ifate=0;
    sdate=lastdaylive;
    t = lastday -lastdaylive;
    sage = age{sdate};
    output;
  end;
  keep id Robel PpnGr NatGr PlCov Wetl Road ifate sdate t sage;
run;

proc print; run;
```

Appendix 3. Example code for analyzing fixed–effects models of nest–survival data from periodic nest visits with Proc GENMOD in SAS (SAS Institute, 2000). Note: code will have to be modified accordingly for other datasets, e.g., variable names will need to be adjusted.

Apéndice 3. Código de ejemplo para analizar con Proc GENMOD en SAS (SAS Institute, 2000) modelos de efectos fijos correspondientes a los datos de supervivencia en nidos a partir de visitas periódicas a los mismos. Nota: el código deberá modificarse en función de los otros conjuntos de datos; es decir, se deberán ajustar los nombres de las variables.

```
* Sample code for fitting logistic–exposure models (Shaffer, 2004) to
* interval nest–visit data.
*
* Variables are as follows:
*
* Ifate = 0 if nest fails the interval, and 1 if it survives.
* Trials = 1 in all cases.
* Avgage = age (d) of the nest at interval midpoint.
* PctGr = proportion of grassland cover.
* t = interval length (d)
*
* Macros for generating AIC analyses and model-averaged parameter
* estimates are available from
*  http://www.npwrc.usgs.gov/resource/tools/nestsurv/nestsurv.htm
*;

Proc Genmod Data=Mall;
    Fwdlink link = log((_mean_**(1/t))/(1-_mean_**(1/t)));
    Invlink ilink = (exp(_xbeta_)/(1+exp(_xbeta_)))**t;
    Model Ifate/Trials = Avgage PpnGr / Dist=bin;
    Ods output modelfit=modelfit;
    Ods output modelinfo=modelinfo;
    Ods output ParameterEstimates=ParameterEstimates;
    Title 'b0 + b1*avgage + b2*pctgr4';
Run;
```

Appendix 4. Code for analyzing fixed– and random–effects models of nest–survival data from periodic nest visits with Proc NLMIXED in SAS (SAS Institute, 2000). Note: code will have to be modified accordingly for other datasets, e.g., variable names will need to be adjusted.

Apéndice 4. Código para analizar con Proc NLMIXED en SAS (SAS Institute, 2000) modelos de efectos fijos y aleatorios correspondientes a los datos de supervivencia en nidos a partir de visitas periódicas. Nota: el código deberá modificarse en función de los otros conjuntos de datos; es decir, se deberán ajustar los nombres de las variables.

```
*    This file:
*           1. inputs a dataset containing nest survival information,
*           2. calculates the effective sample size for the dataset,
*           3. runs a variety of models of NEST SURVIVAL in NLMIXED,
*           4. creates an AICc table for model selection, &
*           5. outputs the AICc table to HTML, RTF, and pdf files.
*    Step 1: calculate the effective sample size according to the methods
*           of Dinsmore et al. (2002). Here, n–ess is incremented by 1 for
*           each day a nest was under observation and survived and by 1 for
*           each interval for which a nest was under observation and failed.;
*    This step:
*           1. calculates the contribution to n–ess for each observation
*           interval and adds that contribution to the sum of ness, i.e.,
*           ness column is a running total
*           2. creates dummy/indicator variables for each of 4 habitat types;
data Mall; set Sasuser.mall2000nd;
 if ifate=0 then ness+1;
 else if ifate=1 then ness+t;
/* create indicator variables for different nesting habitats */
          /* Native Grassland */
    if hab=1 then NatGr=1; else NatGr=0;
          /* CRP & similar */
    if hab=2 or hab=3 or hab=9 then CRP=1; else CRP=0;
          /* Wetland sites */
    if hab=7 or hab=22 then Wetl=1; else Wetl=0;
          /* Roadside sites */
    if hab=20 or hab=8 then Road=1; else Road=0;
run;

* This step finds the actual n-ess for the dataset,
* which is the maximum value in the ness column.;
Proc Univariate data=Mall;
 var ness;
 output out=ness max=ness;
run;

* This step sorts the data by site, which is used as a random factor in * some models below. PROC
    NLMIXED assumes that a new realization
* occurs whenever the SUBJECT= variable changes from the previous
* observation, so your input data set should be clustered according
* to this variable. You can accomplish this by running PROC SORT
* prior to calling PROC NLMIXED using the SUBJECT=variable as the
* BY variable. ;
Proc Sort data=Mall;
by site; run;

* This step reformats the Fit Statistics table of NLMIXED
* so it displays more decimal places in the created tables;
proc template;
    define table Stat.Nlm.FitStatistics;
          notes «Fit statistics»;
```

Appendix 4. (Cont.)

```
        column Descr Value;
        header H1;
        define H1;
                text «Fit Statistics»;
                space = 1;
        end;
        define Descr;
                header = «Description»;
                width = 30;
                print_headers = OFF;
                flow;
        end;
        define Value;
                header = «Value»;
                format = 12.4;
                print_headers = OFF;
        end;
    end;
run;

* Run the most parsimonious fixed-effects model for this datset;
Proc Nlmixed data=Mall tech=quanew method=gauss maxiter=1000;
parms B0=0, B1=0, B2=0;
    p=1;
        do i=0 TO t-1;
    logit=B0+B1*(SAge+i)+B2*PpnGr;
            p=p*(exp(logit)/(1+exp(logit)));
        end;
model ifate~binomial(1,p);
Run;

* Run the most parsimonious fixed-effects model for this datset
*   with the addition of an observer effect on dsr for day 1
*   of each interval (done with a dummy variable called 'Ob');
Proc Nlmixed data=Mall tech=quanew method=gauss maxiter=1000;
parms B0=0, B2=0, B4=0, B8=0;
    p=1;
        do i=0 TO t-1;
            if i=0 then Ob=1;
            else Ob=0;
            logit=(B0+(B8*Ob))+B2*(SAge+i)+B4*PpnGr;
            p=p*(exp(logit)/(1+exp(logit)));
        end;
model ifate~binomial(1,p);
Run;

* Run the most parsimonious fixed-effects model for this datset
*   with the addition of a random effect of site, i.e., run a
*   mixed model where the random effect influences the intercept.;
Proc Nlmixed data=Mall tech=quanew method=gauss maxiter=1000;
parms B0=2.42, B2=0.019, B4=0.38, vsite=0.5;
    p=1;
        do i=0 TO t-1;
        if i=0 then Ob=1;
        else Ob=0;
                logit=(B0+u)+B2*(sage+i)+B4*PctGr4;
            p=p*(exp(logit)/(1+exp(logit)));
        end;
```

Appendix 4. (Cont.)

```
model  ifate~binomial(1,p);
random  u~normal(0,vsite)  subject=site;
Run;

* Run the most parsimonious fixed-effects model for this datset
*   with the addition of an observer effect on dsr for day 1
*   and the addition of a random effect of site.;
Proc Nlmixed data=Mall tech=quanew method=gauss maxiter=1000;
parms B0=2.42, B2=0.019, B4=0.38, B8=-1, vsite=0.5;
    p=1;
      do i=0 TO t-1;
      if i=0 then Ob=1;
      else Ob=0;
             logit=(B0+u+B8*Ob)+B2*(sage+i)+B4*PctGr4;
        p=p*(exp(logit)/(1+exp(logit)));
      end;
model  ifate~binomial(1,p);
random  u~normal(0,vsite)  subject=site;

* The following lines of code estimate the daily survival rate
*  (DSR) for a 15-day old nest on a site with a grassland
*   proportion of 0.5 on (1) the day of a nest visit and
*   (2) a day without a nest visit;

estimate 'dsr-visited'
   exp(B0+B2*(15)+B4*0.5+B8*1)/(1+exp(B0+B2*(15)+B4*0.5+B8*1));
estimate 'dsr-not visited'
   exp(B0+B2*(15)+B4*0.5+B8*0)/(1+exp(B0+B2*(15)+B4*0.5+B8*0));

Run;
```