Methods in Ecology and Evolution

British Ecological Society

Methods in Ecology and Evolution 2012, **3**, 466–470 doi: 10.1111/j.2041-210X.2012.00186.x

APPLICATION

BaSTA: an R package for Bayesian estimation of age-specific survival from incomplete mark–recapture/recovery data with covariates

Fernando Colchero*, Owen R. Jones and Maren Rebke

Max Planck Institute for Demographic Research, Konrad Zuse strasse 1, Rostock, 18057, Germany

Summary

1. Understanding age-specific survival in wild animal populations is crucial to the study of population dynamics and is therefore an essential component of several fields including evolution, management and conservation.

2. We present Bayesian survival trajectory analysis (BaSTA), a free open-source software package for estimating age-specific survival from capture–recapture/recovery data under a Bayesian framework.

3. The method copes with low recapture probabilities, unknown ages (e.g. because of left-truncation) and unknown ages at death (e.g. because of right-censoring). It estimates survival and detection parameters as well as the unknown birth and death times (i.e. latent states) while allowing users to test a range of survival models. In addition, the effect of continuous or categorical covariates can be evaluated.

4. This tool facilitates the analysis of age patterns of survival in long-term animal studies and will enable researchers to robustly infer the effect of covariates, even with large amounts of missing data.

Key-words: Bayesian inference, capture–recapture, capture–recovery, free software, long-term individual-based data sets, R project, survival analysis

Introduction

Understanding how survival trajectories change with age is crucial to population dynamics and is therefore an essential component of ecology, evolutionary biology and conservation science (Martin 1995; Clutton-Brock & Sheldon 2010). Commonly, survival in wild animal populations is inferred from capture–recapture and capture–recovery data (Catchpole *et al.* 1998). However, these data often include numerous records with missing birth and death times, which makes inference on age patterns of survival challenging (Ricklefs & Scheuerlein 2001; Metcalf *et al.* 2009). Recently, ecologists and evolutionary biologists have acknowledged the importance of developing analytical methods that account for these limitations. As a consequence, several alternative models have been developed (Schofield & Barker 2008; Pledger *et al.* 2009; Zajitschek *et al.* 2009; Colchero & Clark 2012). Unfortunately, these methods are not easily implemented and require considerable conceptual and computational investments. This highlights the need to provide user-friendly software.

Here we present Bayesian Survival Trajectory Analysis (BaSTA), a free open-source package that runs on the R platform (R Development Core Team 2011) and implements the hierarchical Bayesian model described by Colchero & Clark (2012). This package facilitates drawing inference on age-specific survival from capture-recapture/recovery (CRR) data when large proportions (or even all) of the records have missing information about times of birth and death. In addition, BaSTA allows users to evaluate the effect of both continuous and categorical covariates on survival as well as time differences in recapture probabilities.

Missing data in CRR data sets

In CRR studies, a population is sampled repeatedly for a number of periods (e.g. years), where each sampling occasion is regarded as a discrete snapshot in time of the state of the population. Typically at each sampling occasion, the sampling scheme consists of marking newborns and individuals that have not previously been tagged and then, in the following occasions, these individuals are either recaptured, not detected, or recovered if found dead (Catchpole *et al.* 1998). As a

*Correspondence author. E-mail: colchero@demogr.mpg.de result, CRR data sets commonly include not only records of individuals with known birth and death times (i.e. uncensored), but also numerous records from individuals for which times of birth are unknown, for example because they were born before the study (i.e. left-truncated), and individuals for which times of death are unknown, for instance those dying after the study has ended (i.e. right-censored; Fig. 1). In addition, low recapture probabilities can reduce the number of years an individual is observed, while the probability of finding dead individuals is commonly low, extending censoring within the study. For some very long-term studies (e.g. Jones *et al.* 2008) these issues are alleviated somewhat, but generally, such problems are commonplace and are particularly troublesome to shorter studies (Nisbet 2001; Colchero & Clark 2012). Our package, BaSTA, has been designed to address these unavoidable problems, allowing users to retain the entire data set without needing to discard records from individuals with unknown age.

Components of BaSTA

BaSTA consists of routines initialized by the user through data input and the definition of basic model settings (Fig. 2). Although BaSTA provides default values, users can define additional variables such as initial parameters, jump standard errors, priors and general setup variables for the main algorithm. The package then performs a range of error checks on the input data and the user-defined settings. If no errors are found, the model runs one or multiple Markov chain Monte Carlo (MCMC) algorithms (see Colchero & Clark 2012). When the MCMC runs finish, BaSTA calculates diagnostics that include measures of serial autocorrelation on parameter traces, parameter update rates, convergence, parameter overlap and preliminary model selection. See the Supporting Information for an extended description of the models and a step-by-step tutorial where every section below is described in detail.

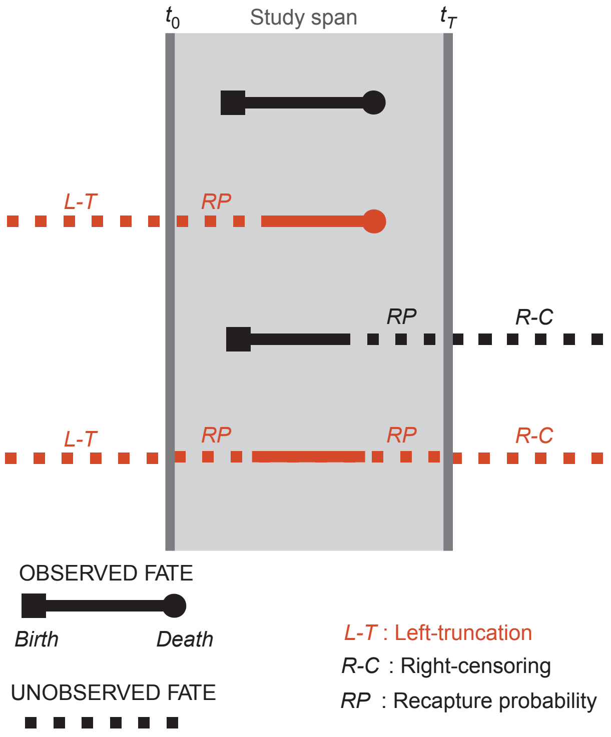

Fig. 1. Types of capture histories commonly found in ecological studies. Histories outlined in red are usually discarded from studies of age-specific survival. Bayesian Survival Trajectory Analysis includes these records into the inference process.

DATA FORMAT

BaSTA's input data format is compatible with other CRR programs such as MARK (White & Burnham 1999). The data takes the form of a table (i.e. an R data frame) where each row represents one individual. This table includes individual times of birth and death (if known) and the recapture history matrix that assigns 1 for every year an individual is detected and 0 otherwise (Table 1). If users wish to evaluate covariates, an optional covariate matrix can be included (see pages 3–4 in Supporting Information). BaSTA provides functions to coerce conventional individual survey tables and covariate matrices into the proper format.

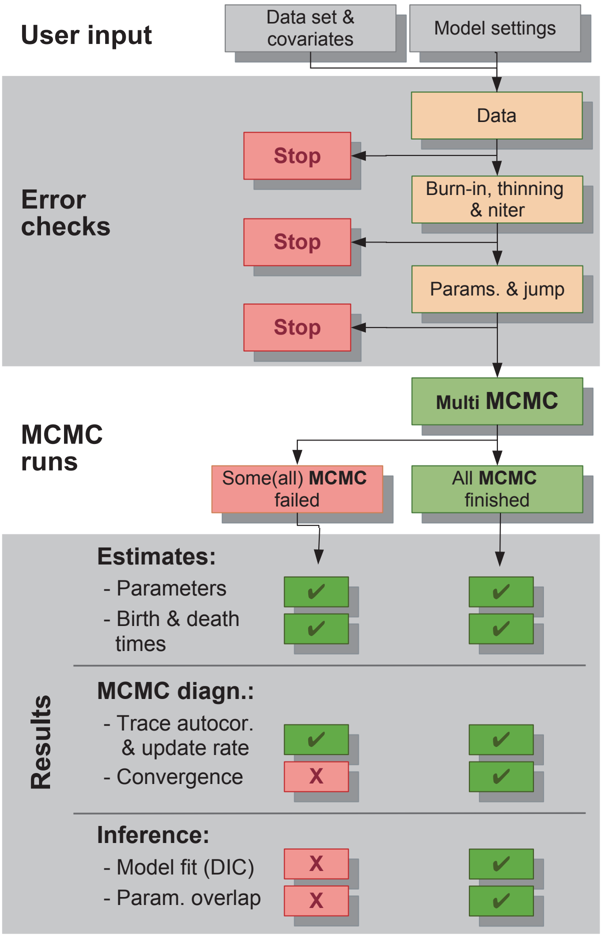

Fig. 2. Work flow of Bayesian Survival Trajectory Analysis main function after data input and argument definition from the user. During the initial 'error check' sequence, if an error is found, the function is stopped and an error message is printed.

Table 1. Data format required by Bayesian Survival Trajectory Analysis. Column 1 refers to each individual's ID, columns 2 and 3 are the times of birth and death, columns 4–7 are the capture-history matrix and the last two columns correspond to covariates for location

ID	Birth	Death	Year 1	Year 2	Year 3	Year 4	Location 1	Location 2
1	b_1	d_1	1	0	1	1	1	0
2	b_2	0	0	1	1	1	1	0
3	0	d_3	1	1	1	1	0	1
4	0	0	0	1	1	0	0	1

MORTALITY MODELS AND RECAPTURE PROBABILITIES

BaSTA includes four basic mortality functions : (i) exponential (Cox & Oakes 1984); (ii) Gompertz (Gompertz 1825; Pletcher 1999); (iii) Weibull (Pinder *et al.* 1978); and (iv) logistic (Vaupel *et al.* 1979; Pletcher 1999) (see Table S3 in Supporting Information). Each function can describe different trends in age-specific mortality, giving BaSTA considerable flexibility when estimating these vital rates (Fig. 3).

In addition, BaSTA allows users to extend these basic functions to include more complex shapes (Fig. 3). Specifically, three general forms can be defined: (i) *simple*, which uses only the basic functions described above; (ii) *Makeham* (Pletcher 1999); and (iii) *bathtub* (e.g. Siler 1979).

Also, BaSTA allows the user to test whether recapture probabilities should be constant or whether they should change as a function of time. This could be useful when the recapture effort is not equal every year.

CONDITIONING ON A MINIMUM AGE

In some species, fates of individuals younger than a certain age are typically unknown. For example, after fledging, juvenile seabirds disperse for several years, during which they can experience high mortality. After this time, they may settle in a different colony from the colony where they were born and are thus never detected again. Consequently, uncertainty in the fate of juveniles can inflate early mortality estimates. Accordingly, BaSTA allows users to condition survival estimation on having reached a minimum age while keeping the integrity of the data set (i.e. no records need to be discarded).

COVARIATES

BaSTA also allows users to define three optional structures to evaluate the effect of covariates on age patterns of survival: (i) *fused*, where categorical covariates are included as linear functions of the survival parameters (analogous to generalized linear models) and continuous covariates are evaluated under a proportional hazards framework (Klein & Moeschberger 2003); (ii) all covariates as *proportional hazards*; and (iii) *all-in-mortality*, where all covariates are evaluated as linear functions of the survival parameters (see Fig. S1 in Supporting Information).

MULTIPLE RUNS AND DIAGNOSTICS

Multiple runs

An important feature of BaSTA is the ability to run multiple simulations, in parallel or in series. Running multiple simulations allows users to test whether the final MCMC parameter chains (i.e. traces) all converged to the same final values, irrespective of initial parameters (Gelman *et al.* 2004). BaSTA performs these multiple runs in parallel with the SNOWFALL package (Knaus 2010), which reduces computing time proportionally to the number of processors or cores used.

Diagnostics

If multiple simulations are implemented and they all run to completion, BaSTA calculates potential scale reduction for each parameter to estimate convergence (Gelman *et al.* 2004). If all parameters have converged, BaSTA calculates the deviance information criterion (DIC; Spiegelhalter *et al.* 2002), which has been described as a measure of predictive power and criterion for model fit. However, we emphasize that the use of DICs is still controversial and the results should be taken with caution (see responses in Spiegelhalter *et al.* 2002). BaSTA's DIC is calculated as an approximation to the group-marginalized DIC presented by Millar (2009). Additionally, BaSTA includes a diagnostic based on Kullback–Leibler discrepancies

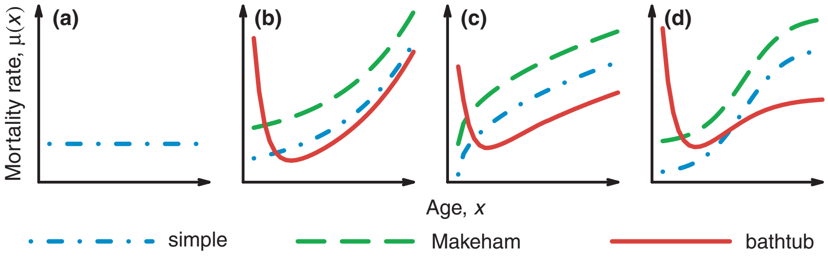

Fig. 3. Mortality rates, $\mu(x|\theta)$, resulting from the four basic models included in Bayesian Survival Trajectory Analysis (BaSTA): (a) exponential; (b) Gompertz; (c) Weibull; and (d) logistic. The three different lines in each plot (except in a) show examples of the shapes that can be tested with BaSTA, namely: 'simple'; 'Makeham'; and 'bathtub'.

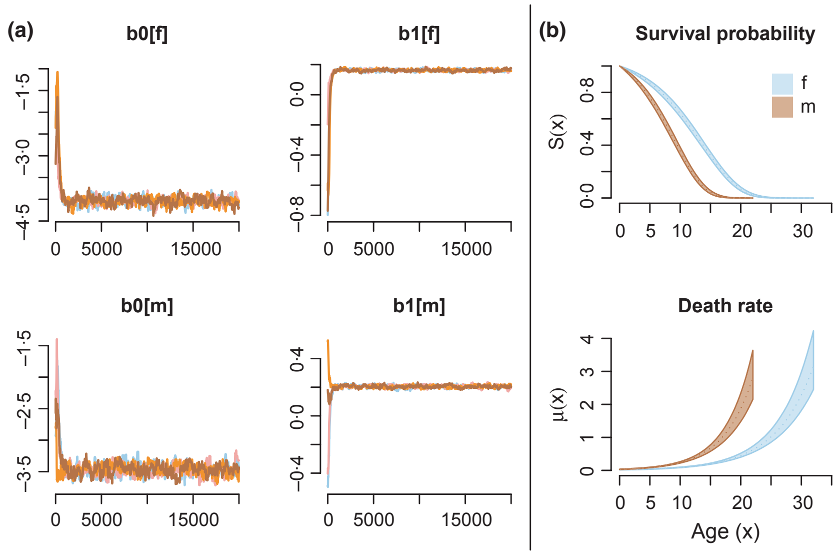

Fig. 4. Plots produced from a Bayesian Survival Trajectory Analysis output with the generic R function `plot()`: (a) trace plots for mortality parameters; and (b) predicted mortality (death rates) and survival probabilities. The model ran for 20 000 iterations and four parallel simulations. Each colour in (a) corresponds to a parallel run.

(KLD; Kullback & Leibler 1951; McCulloch 1989), which provides users with a measure of how differently (or similarly) each categorical covariate affects survival (e.g. effects of males and females on mortality parameters). This measure is based on a calibration proposed by McCulloch (1989), which simplifies the interpretation of KLD values.

Retrieving and visualizing results

The outputs of BaSTA can be explored using R's generic functions (`print()`, `summary()` and `plot()`) to obtain summary statistics, parameter trace plots and figures of the predicted mortality and survival probabilities (Fig. 4). Furthermore, BaSTA can optionally produce life tables from the predicted ages at death for each of the categorical covariates.

Conclusions

Data from wild populations are often characteristically 'messy'. For numerous records, the variables of interest are either masked (e.g. measurement error) or hidden (e.g. nondetection vs. death). Consequently, researchers are forced to choose between reducing the data set to complete records, or using analytical methods that, although powerful, require a great deal of investment and can be challenging to implement. To bridge the gap between developers and potential users, we have produced BaSTA as a 'user-friendly' option for the analysis of CRR data when the aim is to understand age patterns of survival in wild populations.

BaSTA is available at CRAN (cran.r-project.org/) and can be installed in R with the following command: `install.packages(‘‘BaSTA’’)`.

Users are encouraged to join and submit enquiries or comments to the BaSTA users mailing list at: `http://lists.r-forge.r-project.org/mailman/listinfo/basta-users`.

Acknowledgements

This package was conceived at a modelling workshop held at the Mediterranean Institute For Life Sciences, Split, Croatia, sponsored by the Max Planck Independent Research Group on Modeling the Evolution of Aging (MEA). We would like to thank Jim Clark, Annette Baudisch, Alex Scheurlein, Dalia A. Conde, Adam Lenart, James W. Vaupel and the MPIDR IT department.

References

Catchpole, E., Freeman, S., Morgan, B. & Harris, M. (1998) Integrated recovery/recapture data analysis. *Biometrics*, **54**, 33–46.

Clutton-Brock, T.H. & Sheldon, B.C. (2010) Individuals and populations: the role of long-term, individual-based studies of animals in ecology and evolutionary biology. *Trends in Ecology & Evolution*, **25**, 562–573.

Colchero, F. & Clark, J.S. (2012) Bayesian inference on age-specific survival for censored and truncated data. *The Journal of Animal Ecology*, **81**, 139–149.

Cox, D.R. & Oakes, D. (1984) *Analysis of Survival Data*. Chapman and Hall, London, UK.

Gelman, A., Carlin, J., Stern, H. & Rubin, D. (2004) *Bayesian Data Analysis*, chap. 11, 2nd edn, pp. 283–310. Chapman & Hall/CRC, Washington, DC.

Gompertz, B. (1825) On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philosophical Transactions of the Royal Society of London*, **115**, 513–583.

Jones, O.R., Gaillard, J.M., Tuljapurkar, S., Alho, J.S., Armitage, K.B., Becker, P.H., Bize, P., Brommer, J., Charmantier, A., Charpentier, M., Clutton-Brock, T., Dobson, F.S., Festa-Bianchet, M., Gustafsson, L., Jensen, H., Jones, C.G., Lillandt, B.G., McCleery, R., Merila, J., Neuhaus, P., Nicoll, M.A.C., Norris, K., Oli, M.K., Pemberton, J., Pietiainen, H., Ringsby, T.H., Roulin, A., Saether, B.E., Setchell, J.M., Sheldon, B.C., Thompson, P.M., Weimerskirch, H., Wickings, E.J. & Coulson, T. (2008) Senescence rates are determined by ranking on the fast-slow life-history continuum. *Ecology Letters*, **11**, 664–673.

Klein, J. & Moeschberger, M. (2003) *Survival analysis. Techniques for censored and truncated data*, 2nd edn. Springer, New York, New York, USA.

Knaus, J. (2010) snowfall: Easier cluster computing (based on snow). R package version 1.84. URL http://CRAN.R-project.org/package=snowfall

Kullback, S. & Leibler, R.A. (1951) On information and sufficiency. *Annals of Mathematical Statistics*, **22**, 79–86.

Martin, K. (1995) Patterns and mechanisms for age-dependent reproduction and survival in birds. *American Zoologist*, **35**, 340–348.

McCulloch, R.E. (1989) Local model influence. *Journal of the American Statistical Association*, **84**, 473–478.

Metcalf, C.J.E., Stephens, D.A., Rees, M., Louda, S.M. & Keeler, K.H. (2009) Using Bayesian inference to understand the allocation of resources between sexual and asexual reproduction. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **58**, 143–170.

Millar, R.B. (2009) Comparison of hierarchical bayesian models for overdispersed count data using DIC and Bayes' factors. *Biometrics*, **65**, 962–969.

Nisbet, I.C. (2001) Detecting and measuring senescence in wild birds: experience with long-lived seabirds. *Experimental Gerontology*, **36**, 833–843.

Pinder III, J.E., Wiener, J.G. & Smith, M.H. (1978) The Weibull distribution: a method of summarizing survivorship data. *Ecology*, **59**, 175–179.

Pledger, S., Efford, M., Pollock, K., Collazo, J. & Lyons, J. (2009) Stopover duration analysis with departure probability dependent on unknown time since arrival. *Ecological and Environmental Statistics*, **3**, 349–363, Springer.

Pletcher, S. (1999) Model fitting and hypothesis testing for age-specific mortality data. *Journal of Evolutionary Biology*, **12**, 430–439.

R Development Core Team (2011) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. URL http://www.R-project.org/

Ricklefs, R. & Scheuerlein, A. (2001) Comparison of aging-related mortality among birds and mammals. *Experimental Gerontology*, **36**, 845–857.

Schofield, M.R. & Barker, R.J. (2008) A unified capture-recapture frame work. *Journal of Agricultural, Biological, and Environmental Statistics*, **13**, 458–477.

Siler, W. (1979) A competing-risk model for animal mortality. *Ecology*, **60**, 750–757.

Spiegelhalter, D., Best, N., Carlin, B. & Linde, A.V.D. (2002) Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B, Statistical Methodology*, **64**, 583–639.

Vaupel, J., Manton, K. & Stallard, E. (1979) The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, **16**, 439–454. URL http://www.jstor.org/stable/2061224

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

Zajitschek, F., Brassil, C., Bonduriansky, R. & Brooks, R. (2009) Sex effects on life span and senescence in the wild when dates of birth and death are unknown. *Ecology*, **90**, 1698–1707.

Received 26 October 2011; accepted 12 January 2012
Handling editor: Satu Ramula

Supporting Information

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

Data S1. Tutorial and technical details of the models used in BaSTA.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials may be reorganized 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.