Pro

Introduction Documentation Downloading and Installation Recent Changes
Reporting Problems Workshops Known Problems

Program MARK

Announcements: Upcoming intermediate level workshop at Colorado State University, Fort Collins, Colorado, USA, 2-6 June, 2003.

Upcoming workshop on advanced techniques for monitoring animal populations at the 3rd International Wildlife Management Congress in Christchurch, New Zealand, 1-5 December, 2003.

The Program MARK hypertext-based online discussion forum, Analysis of Data from Marked Individuals, is found at: http://canuck.dnr.cornell.edu/HyperNews/get/marked/marked.html .

Introduction

Program MARK, a Windows 95, 98, NT, 2000, or XP program, provides parameter estimates from marked animals when they are re-encountered at a later time. Re-encounters can be from dead recoveries (e.g., the animal is harvested), live recaptures (e.g. the animal is re-trapped or re-sighted), radio tracking, or from some combination of these sources of re-encounters. The time intervals between re-encounters do not have to be equal, but are assumed to be 1 time unit if not specified. More than one attribute group of animals can be modeled, e.g., treatment and control animals, and covariates specific to the group or the individual animal can be used. The basic input to program MARK is the encounter history for each animal. MARK can also provide estimates of population size for closed populations. Capture (p) and re-capture (c) probabilities for closed models can be modeled by attribute groups, and as a function of time, but not as a function of individual-specific covariates.

Parameters can be constrained to be the same across re-encounter occasions, or by age, or by group, using the parameter index matrix (PIM). A set of common models for screening data initially are provided, with time effects, group effects, time*group effects, and a null model of none of the above provided for each parameter. Besides the logit function to link the design matrix to the parameters of the model, other link functions include the log-log, complimentary log-log, sine, log, and identity.

Program MARK computes the estimates of model parameters via numerical maximum likelihood techniques. The FORTRAN program that does this computation also determines numerically the number of parameters that are estimable in the model, and reports its guess of one parameter that is not estimable if one or more parameters are not estimable. The number of estimable parameters is used to compute the quasi-likelihood AIC value (QAICc) for the model.

Outputs for various models that the user has built (fit) are stored in a database, known as the Results Database. The input data are also stored in this database, making it a complete description of the model building process. The database is viewed and manipulated in a Results Browser window.

Summaries available from the Results Browser window include viewing and printing model output (estimates, standard errors, and goodness-of-fit tests), deviance residuals from the model (including graphics and point and click capability to view the encounter history responsible for a particular residual), likelihood ratio and analysis of deviance (ANODEV) between models, and adjustments for over dispersion. Models can also be retrieved and modified to create additional models.

These capabilities are implemented in a Microsoft Windows 95 interface. Context-sensitive help screens are available with Help click buttons and the F1 key. The Shift-F1 key can also be used to investigate the function of a particular control or menu item. Help screens include hypertext links to other help screens, with the intent to provide all the necessary program documentation on-line with the Help System.

The theory and methods used in Program MARK are described in more detail in an "electronic book".

Sixteen different parameterizations of encounter data are provided in Program MARK.

Cormack-Jolly-Seber Model

Live recaptures are the basis of the standard Cormack-Jolly-Seber. Marked animals are released into the population, often by trapping them from the populations. Then, marked animals are encountered by catching them alive and re-releasing them. If marked animals are released into the population on occasion 1, then each succeeding capture occasion is one encounter occasion. Consider the following scenario:

Release ----S(1)-----> Encounter 1 -------S(2)------> Encounter 2

Animals survive from initial release to the first re-encounter with probability S(1), and from the first encounter occasion to the second encounter occasion with probability S(2). The recapture probability at encounter occasion 1 is p(2), and p(3) is the recapture probability at encounter occasion 2. At least 2 encounter occasions are required to estimate the survival rate between the first release occasion and the first encounter occasion, i.e., S(1). The survival rate between the last two encounter occasions is not estimable because only the product of survival and recapture probability for this occasion is identifiable.

Generally, the survival rates of the CJS model are labeled as phi(1), phi(2), etc., because the quantity estimated is the probability of remaining available for recapture. Thus, animals that emigrate from the study area are not available for recapture, so appear to have died in this model. Thus, phi(i) = S(i)(1 - E(i)), where E(i) is the probability of emigrating from the study area.

Lebreton et al. (1992) develop this model, and use SURGE (Pradel and Lebreton 1993) to provide parameter estimates. MARK provides the same capabilities as SURGE, plus additional types of models. Another program applicable to live recaptures is POPAN, which provides for estimation of population size and recruitment with the Jolly-Seber model. A third program is SURPH, which is similar in its capability to MARK for live recapture and known fate data. None of the above 3 programs will handle the band recovery models, the joint live recapture and dead recovery models, robust design model, or the multi-strata model.

Band Recovery Model

With dead recoveries, marked animals are released into the population, and re-encountered as dead animals, typically harvested. This theory has been developed by Brownie et al. (1985). Parameters estimated are survival rate, S(i), and band reporting rate, r(i), following Seber (1970). The primary model used by MARK differs somewhat from the parameterization of Brownie et al. (1985) because the f(i) of Brownie et al. are reparameterized as (1 - S(i))r(i). The primary parameterization of MARK results in better numerical estimation properties, plus, makes the band recovery models consistent with the parameterization of the CJS models. In particular, the use of covariates with the S(i) and r(i) is reasonable, because each parameter represents a particular process in the the overall band recovery process (unlike the f(i) parameter of the Brownie et al. model). However, the last S(i) and r(i) are confounded. In addition, with the S(i) and r(i) parameterization, S(i) is always estimated between zero and one. However, when the estimate of S(i) is at the boundary, i.e., close to or equal to one, the standard error is not estimated correctly. An equivalent situation occurs with the binomial distribution when either no successes occur in the data, or all successes occur in the data, and the standard error is estimated as zero. Both the S(i), r(i) and S(i), f(i) parmeterizations of the band recovery model are included in MARK.

Joint Live and Dead Encounters

The joint live and dead model is based on theory developed by Burnham (1993). The parameter space consists of survival rates [S(i)], recapture rates [p(i)], reporting rates [r(i)], and fidelity [F(i)]. An extension developed by Barker (1997) that allows live resightings during the interval between live recaptures is also available. Barker's model extends the capability of Burnham's model, plus allows for the option of no dead recoveries and live recaptures and live resightings.

Known Fate Model

Known fate data assumes that there are no nuisance parameters involved with animal captures or resightings. The data derive from radio-tracking studies, although some radio-tracking studies fail to follow all the marked animals and so would not meet the assumptions of this model. A diagram illustrating this scenario is

Release -----S(1)----> Encounter 2 -----S(2)----> Encounter 3
-----S(3)----> Encounter 4 ...

where the probability of encounter on each occasion is 1 if the animal is alive or dead.

Closed Captures Models

The closed captures models allow the modeling of the initial capture probability (*p*) and the recapture probability (*c*) to estimate population size (*N*). This data type is the same as is analyzed with Program CAPTURE (White et al. 1982). All the likelihood models in CAPTURE can be duplicated in MARK. However, MARK allows additional models not available in CAPTURE, plus comparisons between groups and the incorporation of time-specific and/or group-specific covariates into the model.

The main limitation of Program MARK for closed capture-recapture models is the lack of models incorporating individual heterogeneity. Individual Covariates cannot be used with the closed captures data type because animals that were never captured (and hence, whose individual covariates could never be measured) are incorporated into the likelihood as part of the estimate of population size (N). Models that can incorporate individual covariates existing in the literature (Huggins 1989, 1991) have been implemented in MARK. Estimates of population size are given for the Huggins' models, but these estimates are not quite as efficient as the closed captures data type where the statistical models are equivalent to those in Program CAPTURE. However, the ability to incorporate individual covariates makes the Huggins' models more appropriate if individual heterogeneity exists in the data.

Robust Design Models

Robust Design Models are a combination of the CJS live recapture model and the closed capture models, and are described in detail by Kendall et al. (1997, 1995) and Kendall and Nichols (1995). Instead of just 1 capture occasion between survival intervals, multiple (>1) capture occasions are used that are close together in time. These closely-spaced encounter occasions are termed "sessions".

For each trapping session (j), the probability of first capture (p(ji)) and the probability of recapture (c(ji)) are estimated (where i indexes the number of trapping occasions within the session), along with the number of animals in the population (N(j)). For the intervals between sessions, the probability of survival (S(j)), the probability of emigration from the study area or more precisely, the probability of the animal not being available for capture on the jth occasion given that it was available on the j-1st occasion (gamma' ' (j)), and the probability of staying away from the study area or the probability of an animal not being available for capture on the jth occasion given that it was not available for capture on the j-1st occasion (gamma' (j)) are estimated. Indexing of these parameters follows the notation of Kendall et al. (1997). Thus, gamma' '(2) applies to the second trapping session, and gamma' (2) is not estimated because there are no marked animals outside the study area at that time. To provide identifiability of the parameters for the Markovian emigration model, Kendall et al. (1997) suggest setting gamma' ' (k-1) = gamma' '(k) and gamma'(k-1) = gamma'(k), where k is the number of trapping sessions. To obtain the "No Emigration" model, set all the gamma parameters to zero. To obtain the "Random Emigration" model, set gamma'(i) = gamma' '(i).

The main limitation of MARK for robust design models is the lack of models incorporating individual heterogeneity in the estimation of population size. Individual Covariates can be used to model the parameters S, gamma' ', and gamma' in the Robust Design data type. Individual Covariates cannot be used with the Robust Design data type for the p's, c's, and N's because animals that were never captured (and hence, whose individual covariates could never be measured) are incorporated into the likelihood as part of the estimate of population size (N). Models that can incorporate individual covariates existing in the literature (Huggins 1989, 1991) have been implemented in MARK. Estimates of population size are given for the Huggins' models, but these estimates are not quite as efficient as the closed captures data type where the statistical models are equivalent to those in Program CAPTURE. However, the ability to incorporate individual covariates makes the Huggins' models more appropriate if individual heterogeneity exists in the data.

Multi-strata Models

The multi-strata model of Brownie et al. (1993) and Hestbeck et al. (1991) allows animals to move between strata with transition probabilities. At this time, only the movement model without memory is implemented. An extension to the multi-strata model to include dead recoveries is also implemented.

Jolly-Seber Models

Jolly-Seber Models (Jolly 1965; Seber 1965, 1982, 1986, 1992; Pollock et al. 1990, Schwarz and Arnason 1996) extend the CJS live recaptures models to include recruitment into the populations. In addition to the apparent survival and recapture probabilities of the Cormack-Jolly-Seber model (recaptures only model), the Jolly-Seber model allows estimation of the population size (*N*) at the start of the study, plus the rate of population change (lambda) for each interval. Also included in MARK are the 3 models developed by Pradel (1996) where only recruitment is estimated, both recruitment and apparent survival are estimated, and apparent survival and rate of population change are estimated.

Nest Survival Model

Estimation of nest survival has been a problem of interest since the Mayfield estimator. The nest survival model implemented into MARK allows estimation of daily nest survival rates as a function of both time of season and age of nest. The nest survival model is also useful for "sloppy" radio-tracking datasets, where all animals in the radioed population are not checked simultaneously, as required for the known fate model.

Occupancy Models

Estimation of the proportion of sites occupied is a common problem in ecology. MacKenzie et al. (2002) have formalized the model to incorporate the probability of detection of a species at a site. MacKenzie et al.'s model, plus a robust-design extension, have both been implemented into MARK.

Encounter Histories

The Encounter Histories File is the file that contains the encounter histories, i.e., the raw data needed by Program MARK. Format of the file depends on the data type and examples are given in the help file. The convention of Program Mark is that this file name must end in the INP suffix. The root part of the file name dictates the name of the dBASE file used to hold model results. For example, the input file MULEDEER.INP would produce a Results File with the name MULEDEER.DBF and 2 additional files (MULEDEER.FPT and MULEDEER.CDX) that would contain the memo fields and index orderings, respectively. MULEDEER.CDX will be erased upon exit from MARK.

Encounter Histories Files do not contain any PROC statements, but only encounter histories or recovery matrices. You can have group label statements and comment statements in the input file, just to help you remember what the file contains. The interactive interface adds the necessary program statements to produce parameter estimates with the numerical algorithm based on the model specified.

Once the encounter histories file is created with an ASCII text editor, the next step is to execute the program and select File, New. You then enter the number of Encounter Occasions, number of Groups, and the Data Type. After this input is provided, the Parameter Matrices are created, one for each parameter and group. These matrices default to Time matrices, which you can then modify to other possibilities using menu options. If you don't need any additional constraints, which can be specified via the Design Matrix, then choose the Run menu option to produce the numerical estimates. The Run Window has additional requests for input, including the Run Title, Model Name, Time Intervals, and Encounter Histories File Name. When you click the OK button to run compute the numerical estimates, you must wait for this process to complete before proceeding. At that time, a Results data base will be created (if you request it), and the output stored in the data base for comparison with other models you may provide.

The input file for the example data from American Fisheries Monograph No. 5 (Burnham et al. 1987) is provided as AFSMONGR.INP. This data set has 5 re-encounter occasions, 2 groups, and is live recapture data. Specify these values when you start the program from the File | New menu choices. In the Run Dialog Window, select the AFSMONGR.INP file as the Encounter Histories Input File. Alternatively, the results database for this example is also included with the program. Use the File | Open menu choices to open this file, and review the model results provided.

Documentation

No paper documentation is available for MARK. Electronic documentation is provided in the Windows 95 help file that accompanies the program and available here as HTML files. Open up the Help document with the program, and read some of the documentation, or check out the HTML version. You can print any of this material if you really want hard copy.

A reasonably complete description of Program MARK was developed for the Euring 97 conference, available as a PDF file. I consider this paper as the primary citation for Program MARK:

> White, G.C. and K. P. Burnham. 1999. Program MARK:
> Survival estimation from populations of marked animals.
> Bird Study 46 Supplement, 120-138.

An "Introductory User's Guide to MARK" is being developed by Evan Cooch at Cornell University. For the complete novice, this is the place to start to learn how to run MARK. This guide is a work in progress, so is not complete just yet.

Notes concerning the theory and use of MARK from the graduate course that David Anderson and Gary White teach at Colorado State University: FW663, Analysis of Vertebrate Populations, are available. This is the same material provided as "Technical Background" from Evan's site referenced in the preceding paragraph.

A set of slides that illustrate the concepts of MARK is available for viewing. These slides give a general overview, and portions of them are used in the slide talks listed below.

A one day workshop on Program MARK was given at the Second International Wildlife Management Congress in Gödölló, Hungary, July 2, 1999. The following are the slide talks given:

- Introduction to Program MARK -- Gary C. White
- Exploring Ecological Relationships in Survival and Estimating

Rates of Population Change Using Program MARK -- Alan B. Franklin
- The Robust Design for Capture-Recapture Studies: Analysis using Program MARK -- William L. Kendall
- Jointly Analyzing Live and Dead Encounters using MARK -- Richard J. Barker
- Advanced Features of Program MARK -- Gary C. White

In addition, the following papers are to be published from this workshop.

- First Steps with Program MARK: Linear Models -- Evan Cooch
- Exploring Ecological Relationships in Survival and Estimating Rates of Population Change Using Program MARK -- Alan B. Franklin
- The Robust Design for Capture-Recapture Studies: Analysis using Program MARK -- William L. Kendall
- Jointly Analyzing Live and Dead Encounters using MARK -- Richard J. Barker and Gary C. White
- Advanced Features of Program MARK -- Gary C. White, Kenneth P. Burnham, and David R. Anderson

One of the problems with obtaining software from the Web is that hard copy documentation is not available, such as is the case for Program MARK. The following sites provide information on how to cite electronic documents: MLA-Style Citations of Electronic Sources and Electronic Sources: APA Style of Citation .

Downloading MARK

New Version 1 September 2002

I developed a new version of the MARK interface with a new Visual Objects compiler on 1 September, 2002. Thus, to use this new version, you must download the 9 disk files as described below (or the one CDROM disk file as described). The update file is only compatible with the version of MARK after 1 September, 2002. A common error that occurs when you have mixed versions of the code (i.e., the DLL files are not completely compatible with the EXE file) is an error from Kernal32.DLL, particularly on NT machines.

New Version 22 November 2002 -- replace the version of 1 September 2002

Turns out that the new version of the Visual Objects compiler was causing more problems than it solved. In particular, users were not able to open previously created MARK files that had large outputs associated with them. Thus, I've had to go back to the previous version of the Visual Objects compiler, which will aggravate the problem of mixed versions of the code on a single machine. I suggest you uninstall MARK and then reinstall from scratch the setup.exe file described below, to insure that you don't end up with incompatibilities between the DLL files and the MARK_INT.EXE file.

Copy the single setup.exe file to your hard disk, and execute it to install MARK. This setup file should place a MARK icon on your desktop, register the graphics package, and put the examples distributed with the program in an Examples subdirectory under the Mark directory. If you have difficulty copying a single large file, the setup has been broken down into 7 pieces: setup.exe, setup.2, setup.3, setup.4, setup.5, setup.6, setup.7, setup.8, and setup.9. Copy down these pieces and install MARK by executing the setup.exe file.

Note that any of the setup files, are available via anonymous ftp from ftp.cnr.colostate.edu in the pub/mark subdirectory, in case you are having problems with your browser copying down the files. The single large file is in the cdrom/disk1 subdirectory, the 9-part setup files are in 144mb/disk1 through 144mb/disk9.

If you have a recent download installed, you can update just the critical files by installing them from update.zip.

Windows ME and XP Setup Problems

Some folks are having difficulties downloading MARK onto Windows ME and XP operating systems. The problem concerns the setup.exe program wanting to create a file entitled TGETUP9 when ME or XP already has one. Here's the work around from Jon Runge:

1. Through Window Explorer go to Tools: Folder Options: View. Check the "show hidden files and folder" box and uncheck the "hide protected operating system files" box.

2. Go to the folder C:\WINDOWS:\TEMP. Rename TGSETUP9.TMP to something like TGSETUP~9.TMP.

3. Run Setup.exe for MARK.

4. When done, go back and restore TGSETP9's original name.

Mac and Linux Machines

To run MARK on a Mac (from Evan Cooch):

Equipment Tested: Macintosh PowerBook G3 (Lombard) 333 MHz with 192 MB of ram (note that Mac clock speed numbers are NOT the same as Windows/Intel clock speeds, i.e., a 333 MHz Mac is faster than a comparable WinTel machine).

Software: Virtual PC version 3.0.3 with Windows 98. Able to use MARK under VirtualPC with Windows 98. Also able to use Microsoft Access under VirtualPC.

Recommendations: The more ram you have the better. Set your VirtualPC program's memory to as much ram as you can afford. The emulator program (VirtualPC) actually runs Windows using the amount of ram that you set aside for the emulator. I set the Virtual PC to use 69MB of memory and find that this allows Windows/Dos software to run as fast as a real contemporary WinTel machine. Also, I've had best results running the Mac OS with an abbreviated set of Extensions. You can easily do this by creating a reduced Extension set with the Extension Manager (this is a Control Panel).

To run MARK on a Linux machine (from Len Thomas):

Software: VMWare -- a BIOS emulator for both Linux and WinNT that effectively lets you run one or more "virtual computers" inside your current operating system. So, for example, you can open a Win95 window from your linux box, and everything within that window thinks its in Windows 95. Of course you do need a Win95 license for this, but at least it gets around the problem of wanting to run linux for most things, but having some legacy software in windows.

Many people use VMWare because they do most things in linux (SPlus, C++, F90), but then some people want or have to use MS Office for their word processing, for example. I use it the other way around: I do most things in WinNT (Visual Basic, etc), but need to be able to test my programs in "vanilla" Windows NT, 98, 95, 2000 systems, so I can run these inside my main machine.

Communication between virtual computers is via virtual networking.

At this time MARK has never been tested under VMWare in linux, but

MS Office works, so MARK is expected to work.

Recent Changes

Recent changes include the following:

April, 1998

1. Making the PIM Chart interactive, so that you can edit the PIM's by dragging blocks of parameters, or by right-clicking on a block and changing the PIM. This feature makes for rapid editing, and also allows editing of very large PIMs without actually opening up the PIM Window and all the edit boxes.

2. The design matrix now allows you to paste in a block of values contained in the clipboard. Either use the Ctrl+V command, or else select the Paste Clipboard command.

3. The robust design model now has a saturated model value computed, although I'm not quite sure its correct. However, without a saturated model value, the deviance was computed as zero, so that likelihood ratio tests were not possible. Turns out that some data sets still have a zero deviance, meaning that the saturated model deviance is not correct. The AIC values are okay, just the deviance is wrong.

4. Barker's model has been thoroughly checked, and seems to work correctly. Note that the model has been re-parameterized compared to earlier versions. See the Help file for details.

May, 1998

5. Because the identity design matrix is the default, I made the program "smarter" so that an actual matrix does not have to be passed between the interface and the numerical analysis program, and not stored in the output. Instead, the key work "identity" is added to the design matrix statement. No design matrix is printed in the output. More importantly, no design matrix pops up on the screen when you retrieve a model that had the "identity" key word on the design matrix statement.

6. I have modified the calculation of the threshold value for declaring a parameter not estimable. The old threshold value was 0.3E-6. The new value varies depending on the dimension of the variance-covariance matrix. Thus, some models may give a different number of parameters estimated than previously. If you find an error, let me know so that I can continue to refine this process.

June, 1998

7. Seven new models were added to the list of data types:

Huggins (1991) method where the closed captures likelihood is conditioned on only the animals captured has been added. This model allows individual covariates to be used for estimating the initial (*p*) and recapture (*c*) probabilities of the closed capture model. Estimates of population size (*N*) are generated as derived parameters, because *N* does not actually occur in the likelihood with Huggins' model as it does in the closed captures models of Program CAPTURE. Estimates of *N* are asymptotically equivalent to estimates of *N* from the closed captures models of Program CAPTURE, but are not quite as efficient because some information is lost by conditioning the likelihood.

Huggins (1991) method was also added to the robust design model, so that individual covariates can also be used with all the paramters.

The Jolly-Seber model has been added. The first parameterization is from Ken Burnham, and uses the first population size *N*1 and the rate of population change, lambda1 through lambda(*k*-1) for *k* encounter occasions. This model does not converge readily, so I suggest you use one of the Jolly-Seber models described below.

Pradel's (1996) model that estimates seniority from the encounter histories has been added.

Pradel's (1996) model that estimates both seniority and apparent survival from the encounter histories has been added. This model is another parameterization of the Jolly-Seber model.

Pradel's (1996) model that estimates apparent survival and rate of population change (lambda) has been added. Again, this model is another parameterization of the Jolly-Seber model.

Pradel's (1996) model was reparameterized to include apparent survival and fecundity (number of new individuals in the population at time *i*+1 per individual at time *i*). Again, this model is another parameterization of the Jolly-Seber model.

8. The Graphics Server graphics package was added as the default graphics package. The reason for this second graphics package was to provide better graphics when multiple sets of parameters were displayed. Both packages are available via from the Results Browser Window with the Output | Specific Model Output | Interactive Graphics and the Output | Specific Model Output | Interactive Graphics GS menu choices.

May, 1999

9. The variance components code was changed to use the full precision of estimates for the variance-covariance matrix. You will have to recompute parameter estimates of existing models to use the new code.

10. The older graphics package was eliminated. This change cut the size of the program distribution files.

11. An option to re-compute all the models in a Results Database was appended to the Run menu of the Results Browser.

June, 1999

12. PIM matrices can now be copied to the clipboard, and vice versa the clipboard can be pasted into a PIM. This feature also means that you can create PIMs in a spreadsheet, and then paste the values into a PIM Window (or also on the PIM chart).

13. An "All Different Matrix" option has been added to the PIM menus to allow you to create PIMs with every value different.

14. The Full menu choice has been changed when you request a Design Matrix. Now, the Full menu choice produces a design matrix with an intercept, group effects (if needed), time effects, and group*time effects (if needed). The limitation of the Full option is that the PIMs must have the number of parameters equal to a full g*t model. However, you can create age structure across PIMs, and still use the Full menu choice as long as the number of parameters is equal to a full g*t model. The columns of the design matrix are initially labeled, although these labels disappear if you manipulate the matrix, e.g., add or delete columns. I'm thinking about how to keep the labels around.

The old version of the Full menu choice has now been renamed to be a the Identity menu choice.

15. The list of Pre-Defined Models has now been expanded to

include both PIM and Design Matrix coding of the standard models. In addition, the Design Matrix coding list includes g+t models. All the Design Matrix models are constructed based on a full g*t set of PIMs. Hence, you still can't use the Pre-Defined Models option to construct age-structured models.

16. A menu choice has been added to the Output menu to list the *m(ij)* arrays of live recaptures and dead recoveries of the encounter history matrix in a NotePad window. This option is useful for identifying anomalies in the data. In the past, you could only run RELEASE for the CJS model to look for anomalies.

17. Variance components outputs are further developed, with additional options specified by check boxes for output other than the main numerical output and the graph of the estimates.

18. The problem with a semicolon in the encounter histories file between comment delimiters, e.g.,

/* stuff; stuff */,

has been fixed.

19. Problem behavior with the individual covariate standardization method has been uncovered. When you specify a model with a common a common intercept but 2 or more slopes for the individual covariate, e.g.,

1 weight 0

1 0 weight

and specify to standardize the individual covariate, you will get a different value of the deviance than from the model run with unstandardized individual covariates. This behavior is because the centering effect of the standardization method affects the intercept differently depending on the value of the slope parameter. The effect is caused by the nonlinearity of the logit link function. You get the same effect if you standardize variables in a logistic regression, and run them with a common intercept. The result is that the estimates are not scale independent, but depend on how much centering is performed by subtracting the mean value.

20. I think I have gotten rid of the problem that caused the message "Unique ID > 8000. Tell Gary". If you still get this message with the new version, let me know.

July, 1999

21. The ability to open one of the last 4 files previously used has been added to the File menu.

22. Real parameter estimates are now labeled as "Estimates" instead of "S(I)" in the output.

23. An option was added to make the default to tell you when a model is retrieved. With the File | Preferences menu choice, you can change this option back to not making you respond to this message.

24. The option to list the Akaike Weights in the Results Browser has now been made the default. You can turn this option off with the File | Preferences menu choice.

December, 1999

25. A problem with losses on capture in estimating *p* and *c* of the robust design and Huggins robust design models has been fixed. Previously, *S*, gamma", and gamma' were estimated correctly with

losses on capture, but not *p* and *c*, and hence *N*.

26. Options for export of parameter estimates (both beta and real), and SE's and confidence intervals to Excel spreadsheets or the clipboard were added.

27. Options for export of variance-covariance matrices to dBase files or the clipboard were added. These changes forced a re-organization of the menu choices under Output | Specific Model Output to keep the menu from becoming too clumsy. Also, the model averaging variance-covariance matrix can now be exported to a dBase file.

28. Extensive testing of the Jolly-Seber model demonstrates that it works correctly, but is just difficult to achieve convergence with for sparse data. With simulations and large sample sizes, it works fine.

29. More work was done on the simulation procedure. However, I still have not removed the double-click trap to keep the faint of heart out of the simulation procedures. I now believe that all models are simulated correctly except the robust design, Huggins robust design, and Pradel model with recruitment but no survival. Currently, you are not allowed to simulate these models. All the rest should work correctly, including the Jolly-Seber model. The problem with simulating the robust design models is how to specify the poulation size so that it is not less than what is consistent with the *S*, gamma", and gamma' parameters for the previous interval and preceding *N*. This problem involves fixing the input interface, not the actual numerical simulation procedure.

March, 2000 (updated to Version 1.7)

30. A "robust design" extension of the Barker model is now included in MARK.

31. A VPA - Virtual population model - is now included.

32. The ability to submit simulations to Program RELEASE is included in the Simulation menu.

33. The ability to submit simulations to Program CAPTURE is included in the Simulation menu.

34. The design matrix now displays different colors for cells with zero, one, and other values. This feature is handy for spotting the pattern in complex matrices. You can change the colors displayed with the File | Preferences menu choice.

35. The ability to run more than one data type within the same Results Browser window has been implemented. For example, you can now run both the dead recoveries {S, r} models and the Brownie et al. {S, f} models, and have the results appear in the same Results Browser window. To change the data type for models, select the PIM | Change Data Type menu choice from the Results Browser window. This convenience also introduces some problems, such as model averaging (not all the models have the same PIMs), so that the model averaging procedure only averages across the models with the same data type. Another example of where problems occur is when the likelihoods are very different between the data types, e.g., the closed captures models and Huggins' models. The AIC values are not on the same scale for these different data types. However, this feature is very useful to compare the 2 sets of estimates from the same Results Brower window. With this change, I updated the version number of MARK to 1.7.

36. An option to use multi-threading was added to the File | Preferences menu choice. This option allows you to be building a new model while numerical estimation is running for a previous model in background.

37. An option was added to the Run Window that specifies to use the mean values of individual covariates to compute the real parameter estimates reported in the output. Note that when individual covariates are standardized, the mean value for a standardized covariate is zero.

June, 2000 (updated to Version 1.8)

38. The "robust design" extension of the Barker model was modified to include the permanent emigration (F) parameter, and the notation for the temporary emigration parameters was changed to be gamma" (old F) and gamma' (old F') to be consistent with Lindberg et al. (In Press).

39. Calls to Programs ESTIMATE and BROWNIE were added to the Test menu choice in the Results Browser window. The programs are useful in assessing goodness-of-fit for dead recovery data.

40. The capability to "undelete" a model that was deleted from the Results Browser window was added. This menu choice is available under the Delete menu of the Results Browser window. You are presented with a list of all the models you have deleted during this execution of MARK, and allowed to select the models you want to undelete. However, once you end MARK, the deleted models permanently disappear.

41. A "Partial Zero" menu choice was added to the Design Matrix window to allow the user to "clear" or zero out a portion of a column. This option is useful when you copy a column, and only want to retain a portion of the values in the column.

42. The variance components output was modified to include a naive estimator of the process variance. Additionally, the estimate of the process variance and associated confidence intervals are now reported with negative values, if necessary. However, the estimate of the process standard deviation is reported as zero for negative process variances (and similarly for the lower confidence bound).

43. A new column has been added to the Results Browser -- Model Likelihood. The values in this column are the Akaike weight of the model in the row divided by the Akaike weight of the minimum AICc model. This value is the model likelihood, the strength of evidence for the model. The best model has a value of 1, with all other models <1.

44. The Variance Components module now has a check box to run the Random Effects model that corresponds to the variance components analysis you have requested. Specifically, a Random Effects model can be viewed as intermediate in its number of parameters between the time-specific model (t) and the dot (.) model where all values are the same. The approach developed by Ken Burnham is to fix the parameter values that were used to estimate the process variance to their S-tilde values from the variance components analysis. The number of parameters that this set now represents is estimated from the trace of the G matrix. Adding trace(G) to the number of parameters estimated from the numerical analysis run with fixed values from the S-tilde gives the number of parameters estimated for the random effects model. Because trace(G) is not an integer, the number of parameters for the random effects model appears as a non-integer. To accommodate these non-integers, existing Results Files will be updated automatically the first time they are accessed. You should see a message stating so. To be safe, you may want to back up your MARK files before opening them with this new version.

45. Although the background colors used in the design matrix to represent the pattern of the matrix are very helpful, especially at first, the cost is a lot of time watching design matrices being created

on the screen. Therefore, I put an option in the File | Preferences dialog to turn off the color in design matrices if you desire, which definitely increases speed for large matrices.

46. I fixed a bug with the probability of surviving the study (Kaplan-Meier estimate) for the Known Fate data type when time intervals were unequal. Previous versions assumed all time intervals were equal to 1. Now, the length of the time interval is used to compute the product of survivals and its standard error and confidence intervals.

47. Labels for the real parameter estimates are now provided in the output. Labels for real parameter estimates are taken from the "Parm" column that is added to the design matrix -- giving you some idea about what the parameter type is for each estimate. Note however, that some parameter types can be set to others, e.g., initial capture probability (*p*) and recapture probability (*c*) in the closed capture models. Hence, be careful to not be mislead. You do not need a design matrix to obtain these labels -- the process is automatic. Labels are generated from the parameter type for the first occurence of the parameter index in the PIMs. If you want to obtain these labels for existing model output, you need to re-run the model, i.e., from the Results Browser, Run | ReRun Models menu choices. The real parameter estimate labels are limited to 20 characters.

48. Labels for the beta parameter estimates and columns of the design matrix are now available. When in the design matrix, if you right-click the matrix, the menu choice "Label" allows you to enter 20 characters of information to include in the column heading of the column with the highlighted cell. If you start with the Design | Full menu choice, the headings in the design matrix will be automatically labeled. Likewise, the pre-defined design matrix models also have the columns automatically labeled. When the model is run, this information is passed to the numerical estimation process, and these labels will appear beside the beta parameter estimates. When models are retrieved, the design matrix column headings also are retrieved from the output.

49. Columns in the design matrix can be moved by dragging the column heading to a new position. The new ordering can be preserved by right-clicking the design matrix and selecting "Reorder Columns" from the pop-up menu. Basically reordering columns is equivalent to renumbering the beta parameters.

August, 2000

50. A bug introduced in June concerning parsing of real parameter estimates from the output file when derived parameter estimates were present was fixed. This bug was detected because of incorrect output in the model averaging procedure when derived parameter estimates were present.

September, 2000

51. The computer time required to process individual covariates has been dramatically reduced, at least 10 times.

52. A bug with the naive variance component estimator has been fixed -- previously, 2 consecutive runs of the variance components module would give incorrect estimates for the naive estimator for the second and additional runs.

January, 2001

53. The variance components estimator has been modified to show estimates of sigma^2 < 0, then to set the estimate of sigma to zero, and re-compute the beta-hats and their standard errors with sigma = 0.

54. Likelihood ratio tests when *c* > 0 have been modified to incorporate the quasi-likelihood parameter. Likewise, the deviance is now displayed in the results browser is corrected for *c*.

55. Design matrix window now has an option to set the font size, so that large matrices can be miniaturized and the colors used to determine if the pattern is correct.

56. Derived parameters are now handled separately from the real parameter estimates. Eventually, derived parameters will be a fully-supported third parameter type, and you will be able to graph them and perform variance components.

May, 2001 (updated to Version 2.0 -- Greater Prairie Chicken photo)

57. Two additional data types for closed captures have been added. Two heterogeneity models for closed captures have been created based on mixture distributions. The simple version of this model only has a constant capture probability for all occasions and for recaptures. The full heterogeneity model has a set of *p* and *c* parameters for each mixture. Encounter histories for these 2 models are identical to the closed captures data type. These models required a fairly major change to the PIM structure to allow for large rectangular PIMs.

58. A model for nest survival estimation was added. This model is different from the known fate data type because the exact time of failure of the nest is not known. Input for nest survival data can be in the form of a nest survival data type, where the time of first finding the nest, next to last time the nest was observed, and the last time the nest was observed are entered, followed by the fate of the nest (0=successful, 1=unsuccessful). Individual covariates may also be used. More details on the structure of this model and the encounter histories file can be found in the help file.

59. A bug that caused incorrect standard errors for the derived parameter estimates of *N* was fixed in the Huggins robust design model.

60. Some really large problems (e.g., very large PIMs or a large number of PIMs) will not work correctly with the existing model averaging parameter selection interface. Therefore, I added another interface that will work. To use this new interface, click the "Use non-interactive model averaging parameter specification window" choice from the File | Preferences menu choice. With this interface, you have to pick the PIM and then specify the parameter within the PIM to model average. Not as visual, but more effective for large problems.

June, 2001

61. A bug in computing the variance of the derived parameter estimate N-hat for the Huggins data type was fixed. Standard errors of N-hat from the Huggins models now correspond closely with standard errors from the closed captures data types.

62. The scaling of the gamma" and gamma' for the length of the interval was removed. These parameters should be viewed as the probability of being available for capture conditional their previous status, and hence are not a function of the length of the time interval.

63. When setting up a new analysis, you can now paste in time interval lengths from the clipboard. Thus, if you have a complicated set of time intervals, enter them into a spreadsheet, and then paste the values into MARK.

64. A "Save As" capability was added to the File menu choice to save

a copy of the current Results File.

65. The dialog to select a set of "Pre-defined Models" has been re-written to allow the user to pick models for each parameter, rather than select from a list of all possible models. This change allows pre-defined models for data types that previously would generate too long of potential models, allows the user to more easily select just models desired, and makes it harder to "Select All" models. In other words, rather than shamelessly data dredging with a few mouse clicks, you will now have to work at it!

66. The ability to summarize derived parameters from simulations was added.

67. Several bugs that caused incorrect estimates for the nest survival data type were fixed.

July, 2001

68. The nest success data type was further modified to produce warnings when invalid data are entered for a nest record. The help file under the 'Nest Survival' topic was updated to present examples of these invalid cases.

September, 2001

69. The multi-strata model with live and dead encounters is now working correctly after another bug was fixed on 3 September, 2001. The help file has also been updated to describe this model.

November, 2001

70. The occupancy estimation model of MacKenzie et al. (2002) has been implemented. The help file has also been updated to describe this model.

January, 2002

71. Minor fixes to several bugs that have surfaced, including aborts when no pre-selected models are selected, incorrect label on beta labeling window, and specification of default preferences when the program is installed.

March, 2002

72. The Run Window has been modified to allow a user to specify values of individual covariates to be used in computing the real parameter estimates. With these changes, 3 options are available for specifying values of the individual covariates: values from the first encounter history in the encounter histories file, mean values, and values the user specifies.

73. Paste buttons were added to the simulation input screens for beta parameter estimates and numbers of releases, and initial input screens for group labels, individual covariate names, and strata names and labels. Now, the user can generate a set of input values for any of these screens in a spreadsheet, and then paste the values into MARK.

74. The Run Window has been modified to allow a user to specify different link functions for parameters in a model. However, this procedure is tricky, and can lead to numerical problems if not done intelligently. Default initial parameter values are more difficult to set, so numerical convergence becomes more of an issue. The user will have to take more responsibility in specifying initial parameter values if multiple link functions are used in one model.

September, 2002 -- Version 3.0

75. The robust-design occupancy estimation model of MacKenzie et al. (submitted) has been implemented.

76. An updated Graphics Server graphics package has replaced the older version. The interface to the graphics package now includes legends for each set of parameters plotted.

77. Some minor bugs with labeling columns in the design matrix, PIM Charts off the top of the screen, and some other little annoyances that have been pointed out have been fixed.

78. The capability to construct interaction terms between individual covariates has been implemented with the product function in the design matrix. In addition, an add function has also been implemented. See the revised help file for details of how these functions operate within the design matrix.

79. The Pearson chi-square statistic is now computed routinely in the full output, and can be saved in the simulation and bootstrap files. The Pearson chi-square statistic is sometimes better behaved than the deviance statistic, and may provide an improvement in estimating the over-dispersion parameter, *c*.

November, 2002 -- Version 3.1

80. An entry in the design matrix menu now allows you to see the list of individual covariates available for the current data set, and to insert 1 or more of these individual covariate names in the current highlighted cell in the design matrix.

81. Added the capability to retrieve one or more columns from a previous model's design matrix and insert them into the current design matrix. Note that the number of rows in the design matrix has to be the same for both design matrices.

December, 2002

82. An option was added to the Run Window to allow estimation of profile likelihood confidence intervals for real parameters. However, numerical problems can cause the estimates to be wrong, so use this option carefully. The help file explains the procedure, and shows an example of incorrect profile likelihood confidence intervals.

March, 2003

83. The multinomial logit (MLogit) and cumulative logit (CLogit) link functions were added as options under the Parm.-Specific link function option of the Run Window. The MLogit link is particularly useful for the multi-strata model to constrain the transitions from a strata to sum to <=1. Another application is the probability of entry parameter of the POPAN model (see below). See the revised link functions help file write-up for details.

84. The POPAN model was revised, and is now working correctly for the real parameter values. This is the Jolly-Seber model from Schwarz and Arnason (1996), where recruitment to the population is modeled as the probability of entry (pent in MARK, beta in the original paper) from a super population.

85. Problems with convergence of the robust design occupancy model were corrected, plus 2 additional parameterizations of this model were added. See the help file for details.

86. An additional parameterization of the Jolly-Seber model developed by Link and Barker (submitted to Biometrika) was added. To use this parameterization, start with one of the other Jolly-Seber data types (e.g., Burnham's Jolly-Seber data type, one of the 3 Pradel data types, or the POPAN data type), and use the PIM | Change Data Type menu choice to obtain the Link-Barker

parameterization. Details are in the help file.

June, 2003

87. Fixed a bug in the Huggins closed-capture heterogeneity models so that *N* is correctly estimated.

August, 2003 -- Version 3.2

88. There are 6 different data types for estimation under the closed capture models: *Mt* (*p* and *c*), *Mh* (pi and *p*) and full Mh (pi, *p*, and *c*); with each of these data types available with population size (*N*) in the likelihood, or a Huggins version where the likelihood does not contain *N*. Now, when the Closed Captures data type is selected, you are given a choice of which of these models of capture probabilities to use for your initial model. However, you can select different data types with the Change Data Type menu choice that is available under the PIM menu. Note that the likelihood for the Huggins models is not comparable to the likelihood for the data types that include *N* in the likelihood, so that AIC model selection should only be done within the 2 categories, not between them.

89. These same 6 closed models are now available for the robust design model, as well as the robust design version of Barker's model. When you select either the robust design, or the Barker robust design, you are asked to select one of the 6 closed models. However, you can change the closed model with the Change Data Type menu choice.

90. The robust design multi-strata model has been added to MARK. The open model of Kendall and Bjorkland (2001 Biometrics 57:1113-1122) has been added, along with a closed version that can operate with any of the 6 closed models described above. When you select the closed robust design multi-strata data type, you are asked to select one of the 6 models, but you can change between them with the Change Data Type menu choice.

91. A new version of the installation program was used that has caused some difficulty, but seems to be working correctly.

Reporting Problems

For questions or to let me know about problems you have encountered, send email. Please try to provide as much documentation as possible to help me duplicate your problem. In particular, I would like to have the input file that caused the problem, and the values you entered for the number of occasions, the number of groups, and the data type. Further, if you have created a results file, please send this via MIME or uuencode. Both the *.DBF and *.FPT files must be forwarded -- both are needed to see the models you have built.

Upcoming Workshops

An alternative to a week-long workshop is to take FW663, Analysis of Vertebrate Populations, a 5-credit graduate course taught by David R. Anderson and Gary C. White in alternate spring semesters at Colorado State. Out-of-state tuition for the course is approximately $2,700, and cheaper for Colorado residents. The class meets MWF from 8-12 from mid-January until the first of April. The class will next be taught spring semester, 2004, beginning mid-January and ending early April.

Another intermediate level workshop is scheduled for June 2-6, 2003, in Fort Collins, Colorado.

Formal Coursework

Individuals desiring a comprehensive treatment of the background material of Program MARK, and gaining a familiarity with using the program, are encouraged to take the course FW663, Sampling and Analysis and of Vertebrate Populations, co-taught by David R. Anderson and Gary C. White. The course meets from mid-January until the first week of April, MWF from 8-12. The class will next be taught spring semester, 2002, beginning January 14 and ending March 29. We strongly encourage students from outside the University to participate in this course.

Known Problems

Some known problems that you should be aware of:

1. The deviance for the closed captures model divided by its degrees of freedom is not a valid estimate of *c*-hat. This problem carries over to the robust design model. I believe the problem is because these models are not in the exponential family, and this type of estimate of c-hat is only valid for models in the exponential family. This problem also extends to the robust design model.
2. The deviance is not a good test of the goodness-of-fit of the model for sparse data. Sparse data may result from few releases, but even with lots of releases and a large number of encounter occasions, the data will be sparse. Sparse data result in small expected values for lots of the capture histories. To have observed a capture history, at least 1 animals must have been observed with this history. For a small expected value, the contribution to the deviance is large, i.e., (1 - Expected)^2 / Expected is large. A pooling algorithm does not seem possible that will fix this problem. To obtain a valid test of the goodness-of-fit of the model, and an appropriate estimate of c-hat, a parametric bootstrap procedure has been implemented in the Tests menu choice.
3. When you run several numerical estimation jobs at once, the program often aborts after you've appended one of these jobs to the Results Database. You can still retrieve the results from the remaining jobs by using the Output | Append menu item to open the MRK????Y.TMP temporary files. This problem is caused by bugs in the Visual Objects compiler having to do with the dynamic allocation of memory for multiple threads.
4. Program MARK is not particularly stable, mainly because of limits on the amount of available memory. Suggestions for improving the stability of the program are to maximize the amount of memory (RAM) available to the program by closing other jobs, closing PIM windows within MARK when through with them, and occasionally restarting the program to free any memory tied up in memory leaks. The more memory you have, the more stable the program is.

Literature Cited

Barker, R. J. 1997. Joint modeling of live-recapture, tag-resight, and tag-recovery data. Biometrics 53:666-677.

Brownie, C., D. R. Anderson, K. P. Burnham, and D. S. Robson. 1985.

Statistical inference from band recovery data - a handbook, 2nd ed. U.S. Fish Wildl. Serv. Res. Publ. Num. 131, Washington, D.C. 305pp.

Brownie, C., J. E. Hines, J. D. Nichols, K. H. Pollock, and J. B. Hestbeck. 1993. Capture-recapture studies for multiple strata including non-Markovian transitions. Biometrics 49:1173-1187.

Burnham, K. P., D. R. Anderson, G. C. White, C. Brownie, and K. H. Pollock. 1987. Design and analysis methods for fish survival experiments based on release-recapture. Amer. Fish. Soc. Monograph 5:1-437.

Burnham, K. P. 1993. A theory for combined analysis of ring recovery and recapture data. Pages 199-213 in J.-D. Lebreton and P. M. North, editors. Marked individuals in the study of bird population. Birkhauser Verlag, Basel, Switzerland.

Hestbeck, J. B., J. D. Nichols, and R. A. Malecki. 1991. Estimates of movement and site fidelity using mark-resight data of wintering canada geese. Ecology 72:523-533.

Huggins, R. M. 1989. On the statistical analysis of capture-recapture experiments. Biometrika 76:133-140.

Huggins, R. M. 1991. Some practical aspects of a conditional likelihood approach to captureexperiments. Biometrics 47:725-732.

Kendall, W. L., J. D. Nichols, and J. E. Hines. 1997. Estimating temporary emigration using capture-recapture data with Pollock's robust design. Ecology 78:563-578.

Kendall, W. L., and J. D. Nichols. 1995. On the use of secondary capture-recapture samples to estimate temporary emigration and breeding proportions. J. Applied Statistics 22:751-762.

Kendall, W. L., K. H. Pollock, and C. Brownie. 1995. A likelihood-based approach to capture-recapture estimation of demographic parameters under the robust design. Biometrics 51:293-308.

Jolly, G. M. 1965. Explicit estimates from capture-recapture data with both death and immigration stochastic model. Biometrika 52:225-247.

Lebreton, J.-D., K. P. Burnham, J. Clobert, and D. R. Anderson. 1992. Modeling survival and testing biological hypotheses using marked animals: case studies and recent advances. Ecol. Monogr. 62:67-118.

MacKenzie, D. I., J. D. Nichols, G. B. Lachman, S. Droege, J. A. Royle, and C. A. Langtimm. 2002. Estimating site occupancy when detection probabilities are less than one. Ecology 83:2248-2255.

Pollock, K. H., J. D. Nichols, C. Brownie, and J. E. Hines. 1990. Statistical inference for capture-recapture experiments. Wildlife Monographs 107. 97pp.

Pradel, R. and J.-D. Lebreton. 1993. User's manual for program SURGE. Version 4.2. Centre D'Ecologie Fonctionnelle et Evolutive, C.N.R.S., Montpellier-Cedex, France. 33pp.

Pradel, R. 1996. Utilization of capture-mark-recapture for the study of recruitment and population growth rate. Biometrics 52:703-709.

Schwarz, C. J., and A. N. Arnason. 1996. A general methodology for the analysis of capture-recapture experiments in open populations. Biometrics 52:860-873.

Seber, G. A. F. 1965. A note on the multiple recapture census. Biometrika 52:249-259.

Seber, G. A. F. 1970. Estimating time-specific survival and reporting rates for adult birds from band returns. Biometrika 57:313-318.

Seber, G. A. F. 1982. The estimation of animal abundance and related parameters. 2nd ed. Macmillan, New York, USA. 654pp.

Seber, G. A. F. 1986. A review of estimating animal abundance. Biometrics 42:267-292.

Seber, G. A. F. 1992. A review of estimating animal abundance II. Reviews of the International Statistics Institute 60:129-166.

My colleagues have lost consciousness over MARK.

If you don't use WordPerfect, you may not have the WPMathA fonts used in some of the PDF documents provided above (which may result in various equations being unreadable). You can install the needed fonts from Corel by going to http://www.corel.com/6763/downloads/WordPerfect/wpwin/9/wpfonts.exe , which should solve the problem.

The Program MARK web page has been accessed times since August 16, 2000.

Last updated