User's Guide for *MCestimate*

Matthew Etterson
U.S. Environmental Protection Agency
12 September 2013

Contents

1. Preliminaries

1.1 Disclaimer & distribution

Thank you for your willingness to try **MCestimate**. This is a work in progress, so expect some bugs! If you report them I will do my best to fix them quickly. However, your use of this program is at your own risk.

Matt Etterson
etterson.matthew@epa.gov
218-529-5158

1.2 What is MCestimate and why is it needed?

MCestimate is a tool for estimating probabilities of failure in the presence of competing risks. It was originally developed to facilitate the study of causes of avian nest-failure. However, competing risks arise in virtually all stochastic processes and **MCestimate** can be applied in many contexts outside of avian nest failure. For example, carcass distribution trials are a commonly used technique for understanding either scavenging rates or searcher efficiency to better understand animal mortality as a result of collisions with anthropogenic structures such wind turbines, automobiles, power lines, or

windows. The two fates (scavenged versus discovered) can be viewed as competing risks. Similarly, standard toxicity tests generally include a control treatment in which some animals die due to natural causes or experimental stress that is not related to the toxicity of the test material. These two sources of mortality (intoxication versus control) are competing risks. **MCestimate** can be (indeed is being) used to understand both of these processes. However, for historical reasons and for simplicity of presentation, this user's guide, and many of the menus in **MCestimate** explicitly refer to the analysis of nest survival data.

MCestimate uses nest-monitoring data to estimate daily and overall probabilities of nest-failure due to specific causes (competing risks). It is programmed in MATLAB (Mathworks 2012a) and compiled as a standalone program. However, other good computer programs (White and Burnham 1999, Dinsmore et al. 2002) and code for statistical software (Stanley 2000, 2004a; Shaffer 2004) already exist for estimating daily survival rates for nests when nest outcomes are classified binomially (success/fail). Fewer methods exist for nests with multinomial outcomes (success, depredated, adverse weather, abandonment, etc.). Multinomial methods are slightly trickier than their binomial counterparts because estimating the probabilities of failure due to a given cause, requires that you correctly consider the probabilities of not failing due to all other possible causes. When combined with the need to account for nests not being monitored daily, this gets trickier still. **MCestimate** does this using the Markov chain algorithms described by Etterson et al. (2007a, 2007b). Like other current nest survival methods, **MCestimate** is a generalization of Mayfield (1961, 1975) methods for estimating daily probabilities of failure. The primary generalizations of **MCestimate** over the Mayfield (1961, 1975) methods are:

1. **MCestimate** does not require exact knowledge of the timing of failure events (i.e., daily nest-monitoring is not required).
2. **MCestimate** allows separate estimation of daily probabilities of failure due to multiple causes (competing risks, or cause-specific failure).
3. **MCestimate** allows separate specification of covariates to different fates to explore hypotheses about the influence of environmental conditions on cause-specific failure rates.

MCestimate is also a simple multistate model (see Hougaard 1999 and Pradel 2005 for details on multistate and multievent models). Other software for fitting multistate models is also available (Choquet et al. 2004, 2009), which can accommodate more general models than MCestimate. MCestimate has been designed for easy use with an intuitive graphical user interface.

1.3 Installation

MCestimate installation requires 2 files. The MATLAB Compiler Runtime (**MCRInstaller.exe**) installs a library of MATLAB functions that are used by components built in the MATLAB environment. The MCR is specific to the version of MATLAB in which **MCestimate** was compiled. This build of **MCestimate** used version 7.14.0.739, which is associated with MATLAB 2012a (32-bit). This file can be downloaded from the Mathworks website (http://www.mathworks.com/products/compiler/mcr/). Installing the MCR is a one-time event and should not need to be repeated until **MCestimate** is compiled under a new version of Matlab. Note that to install the MCR you will need administrative rights to your computer. To install, double click on the file **MCRInstaller.exe**. This will call a setup program that you should click through. Once the MCR is installed you should be able to run **MCestimate** by double-clicking the file **MCestimate.exe**. The initial distribution ([insert URL for MED website, (pending)]) consists of the following four files:

1. MCestimate.exe
2. MCestimateUsersGuide.pdf
3. SimData.xls
4. SimDataBad.xlsx

2. Data format

MCestimate is designed to import nest monitoring data from Microsoft® Excel. For importing, your nest data should be arranged with one nest per spreadsheet row. The order in which columns occur is somewhat flexible, but with the following rules: 1) the first (left-most) column is reserved for an alphanumeric nest identifier, and 2) the monitoring data (the actual nest visit dates) must be in the right-most columns.

2.1 Required data

1. Row of unique headers for all columns.
2. Column giving the ultimate fate of each nest (successful, depredated, weather, etc.). You may have as many fates as you wish. However, "unknown" is a treacherous one. Be careful with it!
3. Columns (at least two) giving the dates on which each nest was visited.
   a. The first column gives the first date on which the nest was observed to be active.
   b. The last column gives the date of the last visit to the nest.
   Because there are many ways to code and interpret nest-monitoring data, I recommend the following rules. Note that I define an "observation" on a nest to be a set of two successive nest visits.
      i. For all nests, the first observation should begin with the first visit on which a nest was known to have contents that could fail (eggs or nestlings). The first observation should *not* begin with nest construction.
      ii. For all nests, if a nest was sufficiently developed to have fledged during an observation, then the second visit of the observation should be discarded and the fate of the nest should be recorded as its state on the first visit of the observation (presumably "successful"). To see why this is important, read Stanley (2004b).

Optional data include:
4. Columns giving covariate data (year, treatments, etc.)
5. Column giving the age of a nest on the first visit (where "first visit" is defined as 3.b.i., above).

2.2 Formatting requirements

Two example data sets (**SimData.xls & SimDataBad.xlsx**) are provided in the dropbox foloder (Fig. 1). The data in **SimData.xls** were simulated using fewer covariates than are actually provided in the data. **SimDataBad.xlsx** was created from **SimData.xls** by changing two data values to produce two common

types of errors. Each of the examples in this user's guide uses one of these two files, so you should be able to reproduce the examples exactly. You might also have some fun trying to figure out which covariates I actually used to generate the data (and what the "true" parameter values were).

1. Dates should be given in the following format "01-Apr-09." Deviations may work (e.g. "01/04/09"), but employ them at your own risk!
2. There must be no missing data from any columns used for analysis (except visit dates). For now this is a very strict limitation. Cells with missing data will be assigned to the status "NaN" (not a number). If you import a spreadsheet with missing data, the program will display an error message and you will be unable to perform any analyses until the missing data is supplied or the nests with missing data are removed.
3. For each nest, visits are numbered sequentially beginning with the first active visit to a nest and ending with the last visit. Thus there won't be the same number of visits to each nest. All cells to the right of the last visit to a nest should be left blank.
4. The nest data must be on the first worksheet in your Excel Workbook. Safest thing is to delete all other worksheets.
5. Data may be saved in Excel 1997-2003 Format (*.xls) or Excel 2007 Format (*.xlsx, default).
6. Headers cannot include an asterisk. In general I'd recommend avoiding characters other than numbers and letters. I'd also recommend avoiding having one header that is a proper subset of another (e.g., don't have "F1" and "F12"). It would probably be fine, but it might not…

2.3 Example: SimData.xls

Figure 1. Example data **SimData.xls**

Fig. 1, above, shows portions of the sample data file **SimData.xls**. The sample file has 16 columns of data as described in Table 1.

Table 1. Simulated nest-monitoring data in **SimData.xls**

Column header	Description
nest	Unique identification number for a nest. Can be any combinations of letters, spaces, and numbers.
fate	Ultimate fate of the nest (assumed to be determined on last visit).
yr	Year of study in which nest occurred, to be used as a covariate.
ht	Nest height (m) above ground.
trt	A treatment thought to influence predation rate (hi vs. lo).
age	Age (d) of the nest on the first visit (V1).
V1-V10	Columns of dates of nest-visits to nest. V1 is the first visit on which nest could have failed (i.e. the first visit on which the nest had eggs or nestlings observed).

3. Getting *your* data into *MCestimate*

Start **MCestimate** by double-clicking MCestimate.exe. You should now be looking at the main **MCestimate** window (Fig. 2).

Figure 2. The main **MCestimate** window at startup
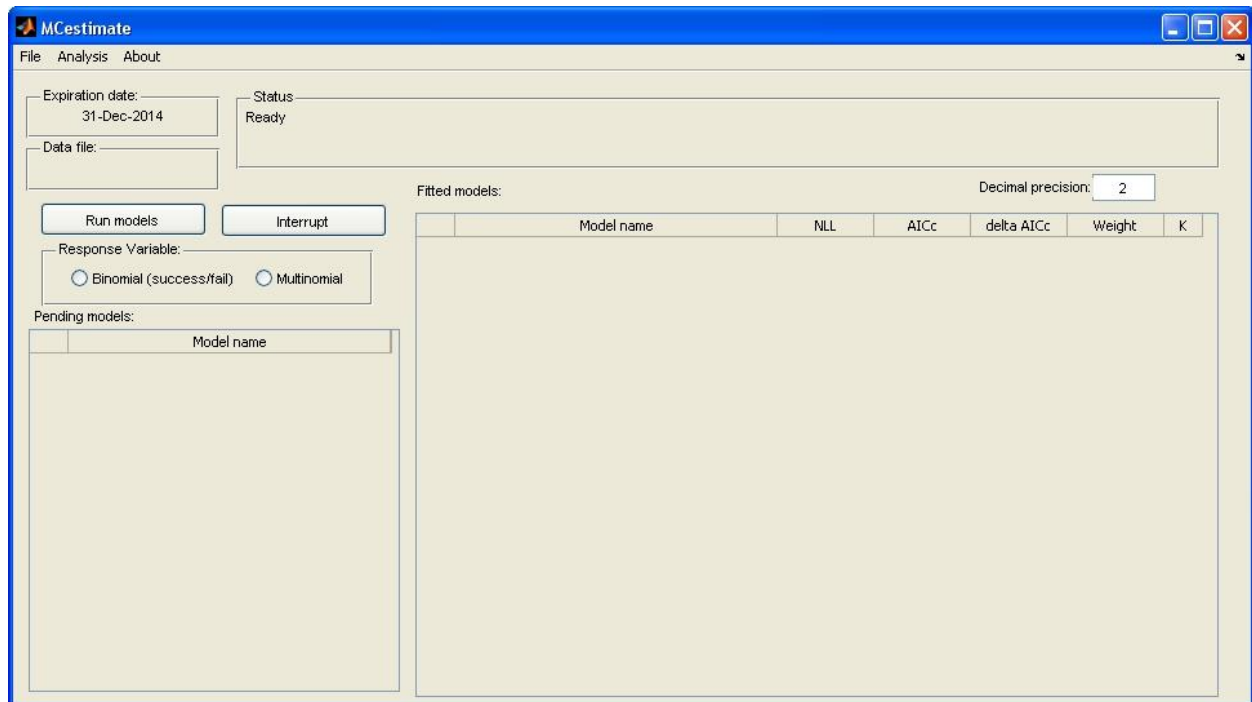
Operations concerned with importing, managing, and formatting data for **MCestimate** are initiated via the **File** menu in the upper left corner (Table 2). The **Analysis** menu is described in Section 4: *Modeling nest failure*. After analysis, most operations are initiated via the **Fitted models** context menu (Section 5). The **About** menu provides information on the build date of the current version. Ultimately it will also provide a recommended citation.

Table 2. **MCestimate** menus

Menu Item	Description
File	Set of commands for importing and formatting data for **MCestimate.**
Analysis	Set of commands for performing analysis and managing output.
About	Information about **MCestimate.**

3.1. The File Menu

The **File** menu in **MCestimate** has 8 choices most of which are designed to assist with importing, formatting , viewing, and exporting data (Table 3, Fig. 3).

Table 3. The **File** menu

Menu Item	Description
Set working directory	Sets the default directory for files for a given project.
Import data	Opens a file browser to select a data file for importing. If a file is chosen, subsequently launches the **DescribeData** GUI (Fig. 4, below).
View file	Displays the **MCestimate** version of the Excel file. This is used for fixing data errors.
Open data file	Opens a properly formatted **MCestimate** file created via the **Import data** option.
Display summary info	Displays summary information about the current data file.
Export MARK file	Exports an encounter history file (*.inp) for analysis in program MARK.
Clear data	Clears all nest data from memory. Do this before opening another nest file.
Exit	Exits **MCestimate.**

Figure 3. The Main *MCestimate* window showing the **File** menu
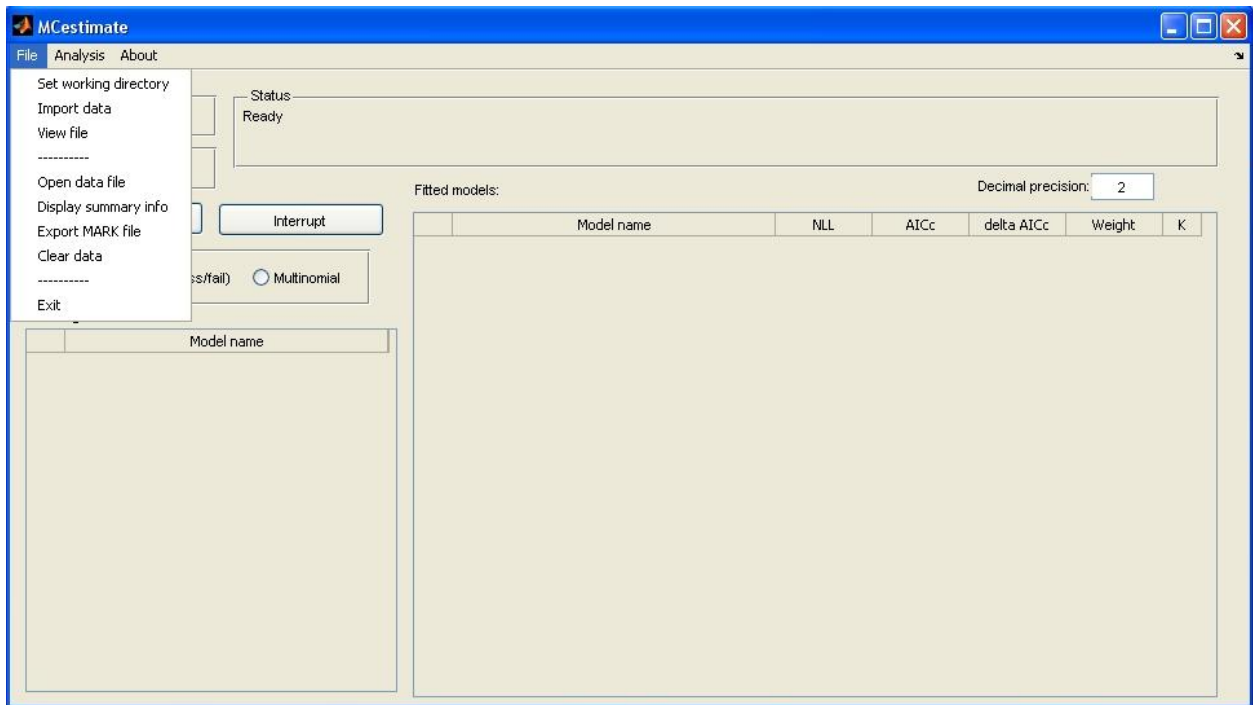
3.2 Importing nest data

To import a spreadsheet into *MCestimate*, choose the **Import data** option from the **File** menu. This will call up a typical windows file browser (Fig. 4) that will display only Excel 2007 (*.xlsx) files. Excel 1997-2003 (*.xls) files can be viewed by changing the type at the bottom of the file browser (Files of type:).

Figure 4. Select your data file browser
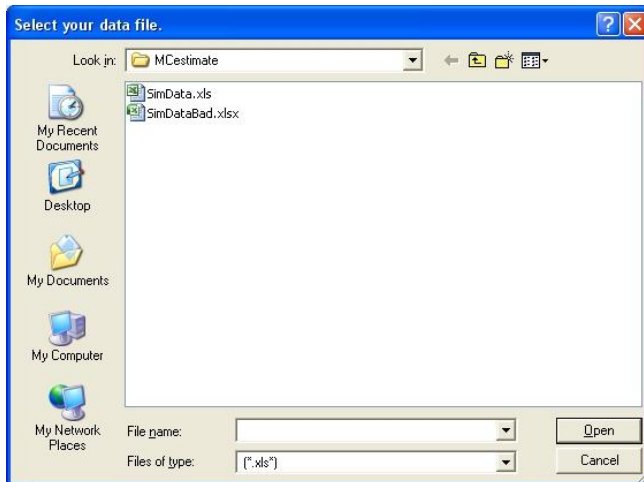
After you choose your file and click **open**, *MCestimate* will automatically open the **DescribeData** window (Fig. 5). If you load **SimData.xls**, your **DescribeData** window should look like Fig. 5.

Figure 5. The **DescribeData** window
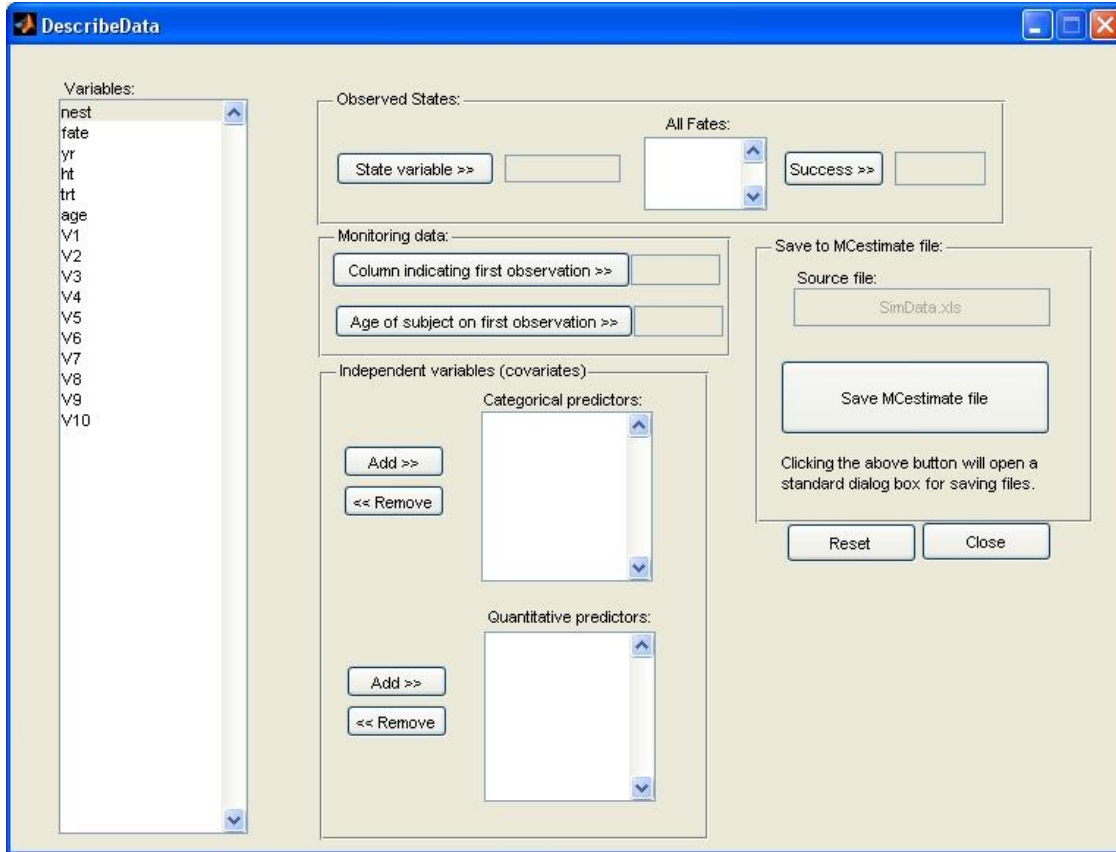
When the **DescribeData** window loads, the data columns of your spreadsheet should be listed in the **Variables** listbox and all other boxes should be empty. Before you can analyze your data, you must identify its structure to *MCestimate*. Not all of the controls on this page must be used.

The required controls are:
1. You must identify the column containing nest outcomes (**State variable**). In **SimData.xls** (Figs. 1, 5) it is the variable labeled **fate** (but you may use any label you wish). Enter it by selecting the variable in the **Variables** box and clicking the **Fate variable >>** button. Once you do so, the **All Fates** listbox will contain a list of all entries in that column of the spreadsheet. If you make a mistake, don't worry. In most cases you can simply highlight the correct variable and click the button again and the incorrect entry will be replaced. If you manage to screw up completely, then you can click the **Reset** button and start over!
2. You must indicate which one of the fates corresponds to a successful nest. Do this by highlighting that fate in the **All Fates** listbox and clicking the **Success >>** button.

3. You must identify the monitoring data by indicating which column contains the first nest visit. This is done by highlighting the proper variable (in **SimData.xls** it is **V1**, though again you may use any headers you wish) and clicking the button **Column indicating first observation**.

4. You must name the data and save it as a *MCestimate* file. Do this by clicking the **Save MCestimate file** button. This will close the **DescribeData** window and you should see a message in the main *MCestimate* window **Status** box "Parsing data …" If you have a lot of data this could take a minute or two. After parsing, the **DescribeData** window will close and a file browser will open to allow you name and save the file. Once the file is saved, a status message will be displayed (Fig. 6). If the parsing algorithm fails, the file browser will not open and a status message will let you know about data formatting problems. Dealing with failed imports is the topic of Sections 3.3 and 3.4 below.

Figure 6. Status message after a successful import
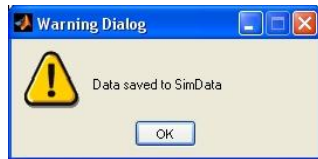
The optional controls are:

5. **Age of subject at first visit:** If this is known it can be entered using the button of the same name.

6. **Categorical predictors:** This listbox is used to identify categorical covariates.

7. **Quantitative predictors:** This listbox is used to identify quantitative covariates. Note: Integer-coded covariates may be entered as either categorical, quantitative, or both. If you code an integer-valued variable in both ways for the same analysis *MCestimate* will either append (c) for categorical, or (q) for quantitative to the name of the variable. For most integer-coded variables, the choice will be obvious about whether to specify that variable as categorical versus quantitative. However, for ordinal variables, in which the codes 1, 2, 3,…, n indicate a meaningful progression (e.g., over time, as in years, or from small to big, as in size classes), you might wish to explore both kinds of effect. The best example I can think of to justify exploring both kinds of effects in a single variable is with a "year" effect. Years may randomly fluctuate in predation pressure or weather events, arguing for a categorical coding. On the other hand, years may follow a trend, arguing for a continuous effect. In the end the choice should best reflect why you want to use the variable as a covariate.

Example: importing SimData.xls

From the **File** menu choose **Import Data** and open **SimData.xls**. Then, using the **DescribeData** window, add variables following the rules above until the **DescribeData** window looks like Fig. 7. Then click **Save MCestimate file**.

Figure 7. The **DescribeData** window for the importing data example
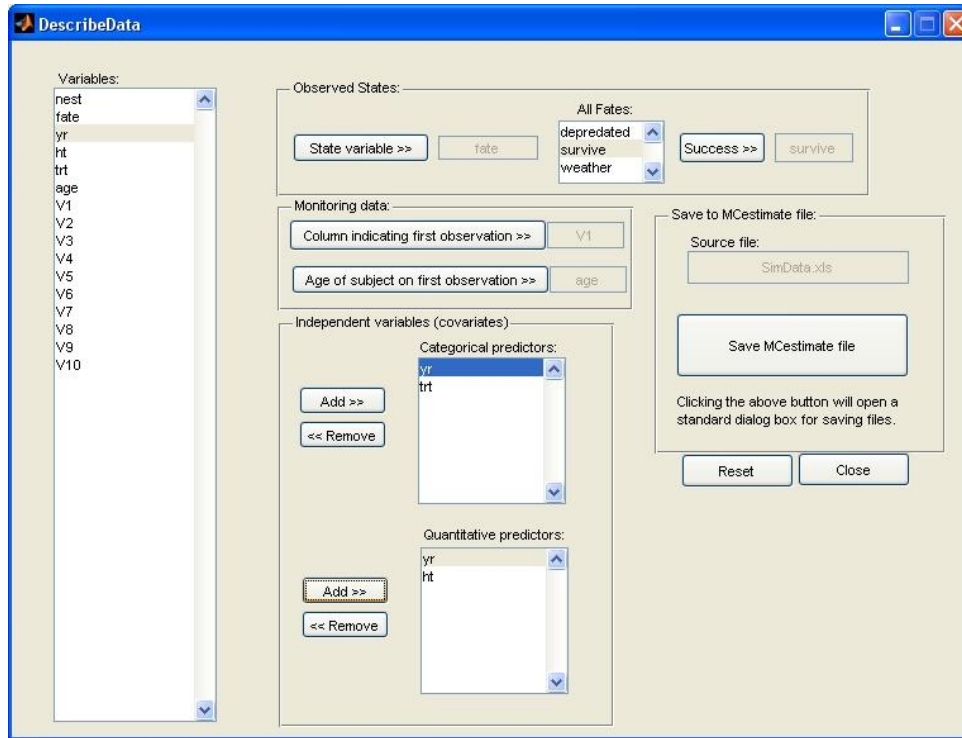
Example: importing SimDataBad.xlsx

This example is designed to illustrate the primary causes for failures in data importing. They are 1) nest observations lasting zero or negative days, and 2) missing data. The parsing algorithm searches for both kinds of errors. If either occurs then the status dialog reports the identification numbers of all nests having one or the other error. For an example, follow the steps above to import data, but this time choose **SimDataBad.xlsx**. After parsing, the status dialog should look like Fig. 8 (below).

Figure 8. Status message after a failed import
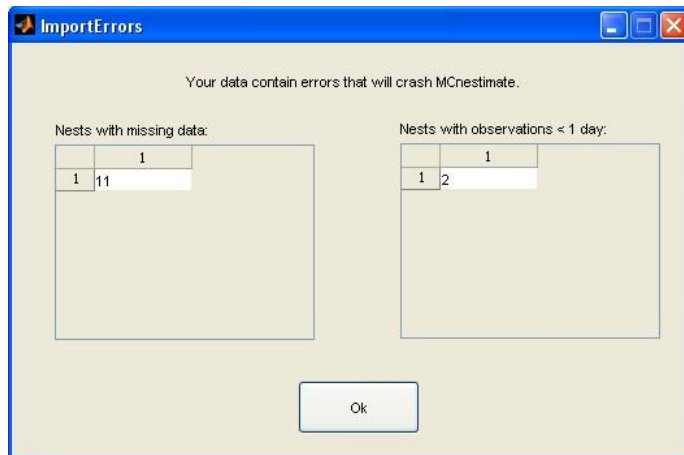
3.3 Viewing import errors

If you have many errors, you will probably want to edit your Excel file and re-import the data. However, if there are only a few errors, or you don't wish to alter your Excel file, you may do some limited editing in **MCestimate**. This is done by choosing **View file** under the **File** menu. Doing so for this example will produce a new window that should resemble Fig. 9.

Figure 9. The imported nest file for **SimDataBad.xlsx**

|	nest	fate	yr	ht	trt	age	V1	V2	V3	V4	V5	V6	V7
	1	2	3	4	5	6	7	8	9	10	11	12	13
1	nest	fate	yr	ht		trt	age	V1	V2	V3	V4	V5	V6	V7
2	1	depredated	1	14.8614	lo	6	5/8/2001	5/12/2001	NaN	NaN	NaN	NaN	
3	2	weather	0	10.9674	lo	9	6/15/2000	6/14/2000	6/21/2000	6/24/2000	NaN	NaN	
4	3	weather	0	9.0823	lo	1	6/15/2000	6/19/2000	6/21/2000	6/25/2000	6/28/2000	7/2/2000	7/4/2000
5	4	survive	0	9.6552	hi	15	7/1/2000	7/4/2000	7/8/2000	NaN	NaN	NaN	
6	5	survive	1	11.1172	hi	11	6/12/2001	6/14/2001	6/17/2001	6/21/2001	6/24/2001	NaN	
7	6	depredated	0	6.4027	lo	2	6/12/2000	6/16/2000	6/18/2000	6/22/2000	NaN	NaN	
8	7	depredated	0	10.6757	hi	2	6/3/2000	6/7/2000	6/9/2000	NaN	NaN	NaN	
9	8	depredated	0	13.4008	hi	11	6/14/2000	6/17/2000	6/21/2000	6/23/2000	6/27/2000	NaN	
10	9	survive	0	4.6667	lo	2	6/3/2000	6/7/2000	6/11/2000	6/13/2000	6/16/2000	6/18/2000	6/20/2000
11	10	depredated	1	14.5165	lo	12	6/28/2001	7/2/2001	7/6/2001	NaN	NaN	NaN	
12	11	survive	0	NaN	hi	14	7/12/2000	7/15/2000	7/19/2000	NaN	NaN	NaN	
13	12	weather	0	11.2101	lo	12	6/9/2000	6/13/2000	6/15/2000	6/17/2000	6/20/2000	NaN	
14	13	survive	0	13.8447	hi	11	7/2/2000	7/6/2000	7/9/2000	7/12/2000	NaN	NaN	
15	14	depredated	0	6.2752	lo	5	6/4/2000	6/8/2000	6/11/2000	6/15/2000	NaN	NaN	
16	15	depredated	1	4.3611	lo	14	6/7/2000	6/9/2001	6/12/2001	NaN	NaN	NaN	
17	16	weather	1	5.6608	hi	3	7/2/2001	7/5/2001	7/8/2001	7/11/2001	NaN	NaN	
18	17	depredated	0	14.0732	lo	4	6/18/2000	6/22/2000	NaN	NaN	NaN	NaN	
19	18	survive	0	14.5480	lo	11	6/24/2000	6/27/2000	6/30/2000	7/2/2000	7/6/2000	NaN	
20	19	weather	1	9.6167	lo	11	6/28/2001	7/1/2001	7/3/2001	7/6/2001	7/10/2001	NaN	
21	20	survive	1	13.8472	lo	5	5/12/2001	5/15/2001	5/17/2001	5/21/2001	5/23/2001	5/27/2001	5/31/2001
22	21	survive	0	5.8240	lo	3	7/2/2000	7/6/2000	7/9/2000	7/13/2000	7/17/2000	7/21/2000	
23	22	survive	1	13.3541	hi	8	6/8/2001	6/12/2001	6/15/2001	6/18/2001	6/22/2001	NaN	
24	23	survive	0	8.4195	hi	8	7/4/2000	7/8/2000	7/12/2000	7/15/2000	7/17/2000	7/20/2000	
25	24	survive	1	13.5138	lo	10	5/14/2001	5/18/2001	5/21/2001	5/23/2001	5/27/2001	NaN	
26	25	depredated	1	14.8732	hi	13	6/8/2001	6/10/2001	6/12/2001	6/16/2001	NaN	NaN	
27	26	depredated	0	13.6630	hi	5	5/16/2000	5/19/2000	NaN	NaN	NaN	NaN	
28	27	survive	1	9.2012	hi	14	7/3/2001	7/5/2001	7/7/2001	7/10/2001	7/12/2001	NaN	

The status message (Fig. 8) reported an observation lasting zero or fewer days for nest 2. Looking across the visit dates for nest 2, you can see that the date in row 3, column 8 is a day earlier than that in row 3, column 7, which will generate an observation lasting -1 days. Similarly, the status message reported missing data for nest 11. **MCestimate** represents missing data as **NaN** (not a number). Looking across row 12, **NaN** appears in column 4 (**ht**). The correct values for these cells are **6/19/2000** and **7.2**, respectively (Fig. 1). These values can be typed directly into the cells. Alternatively you can use the **Edit** menu (Fig. 10). If you simply want to remove all nests with errors, you can use the first two choices. If you want to look at each error individually you can use the third or fourth choices **Delete current cell** and **Delete current row(s).** In particular, try selecting the cell in row 3, column 8 and choose **Delete current cell.** This will remove the value in the selected cell and shift leftward by one cell all cells to the right of the selected cell. Because the incorrect date is nested within a sequence of successful (survived) observations, deleting that particular cell will not affect the estimation results (but this will change in future versions of **MCestimate**!). Choosing **Delete row** will remove the entire row (nest) from the data. Deleting a row will always change the estimation results.

Figure 10. The **Edit** menu in the **NestFile** window
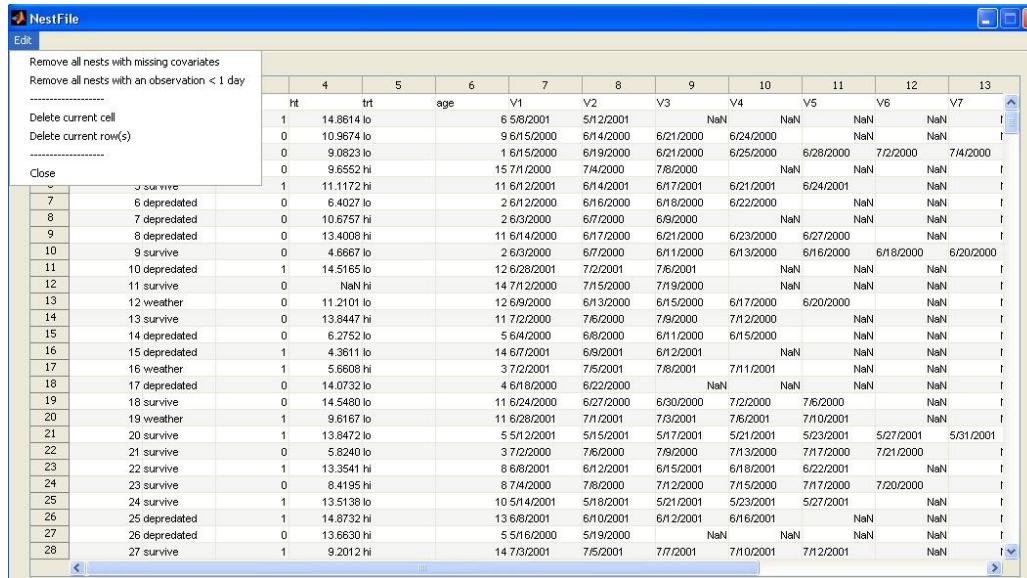
Some final notes about data errors:

 1) *MCestimate* will only look for errors in columns that you specify for analysis. It ignores other columns. For example, had you omitted specifying **ht** as a covariate in the above example, no error would have been generated for the missing value in the cell in row 12, column 4 (though nest 2 would still have generated an error for an observation of less than 1 day). However the displayed nest file would be unchanged (i.e., it would still look like Figs. 9 & 10).

2) Another possible error, having non-numeric data in a column specified for a quantitative covariate, will crash *MCestimate*.

3.4 Re-describing the data after fixing errors

If you make changes to the data file in the **NestFile** window *MCestimate* will automatically try to parse the data again, and if you haven't fixed the errors (or if you've created new ones!) then *MCestimate* will display the same error again. You cannot perform any analyses until you successfully navigate the import screens and get notification of a successful import (e.g., Fig. 6).

3.5 Opening a saved nest file

Once you've successfully described your data, the *MCestimate* window will display a file name under **Data file**. The file has been saved to your hard drive and, in future sessions, can be opened directly in *MCestimate* (without going through the import and describe data processes above). Or you can begin analysis immediately. If you wish to use a previously imported data set choose **Open nest file** from the **File** menu.

3.6. Displaying summary information

If you have a nest file open that has been successfully parsed, then you can quickly view summary information about the file by choosing **Display summary info** from the **File** menu. Doing so for the imported data SimData.xls should look like Fig. 11.

Figure 11. The summary information screen for SimData.mat
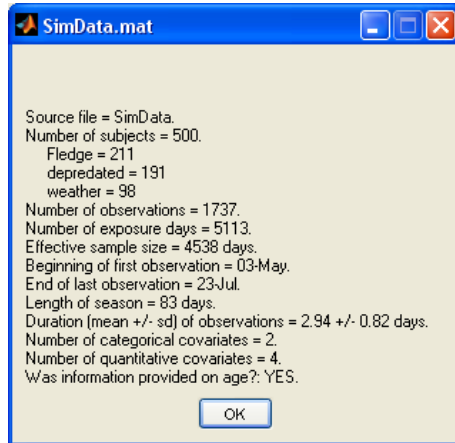
3.7 Exporting a file for analysis in program MARK

You may wish to see whether you get the same answer in *MCestimate* that you get using program MARK (White and Burnham 1999). For binomial fates the two programs should give virtually identical results. I've tried to make this an easy comparison by adding the Export MARK file option. Choosing this option will prompt you for a file name (say **FileName**), then create two files, **FileName.inp** and **FileName.txt**. **FileName.inp** is an encounter history file formatted for immediate import into MARK. **FileName.txt** gives you some supporting information, including the number of encounter occasions (length of the nest season, defined from first active date to last nest-check) and the names of covariates (if you specified any in the **DescribeData** GUI). You will need to input the season length value into the **Encounter occasions** box in MARK. You may wish to add the covariate names under the **Enter Ind. Cov. Names** option in MARK.

4. Modeling nest failure

Analysis is done through the **Analysis** menu, which has five options (Fig. 12, Table 4).

Figure 12. The **Analysis** menu
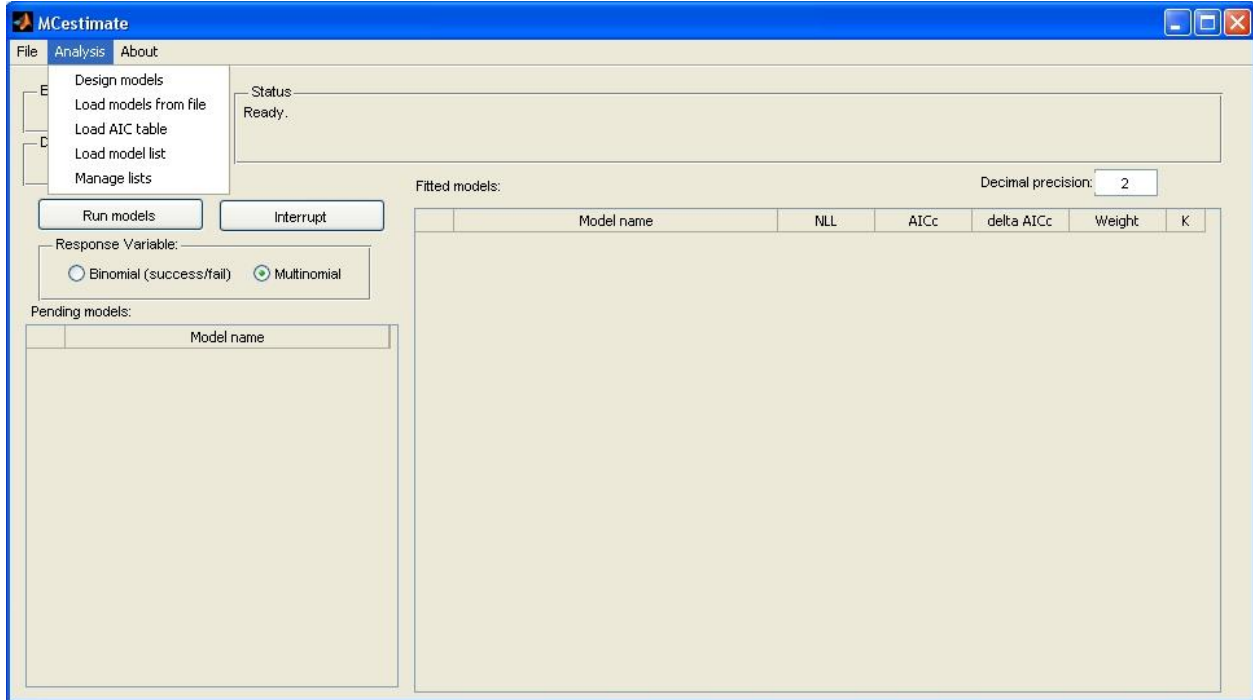
Table 4. The **Analysis** menu

Menu Item	Description
Design models	Opens the **ModelDesign** window (Section 4.2 below).
Load models from file	Imports a list of models from Microsoft Excel (Section 4.3 below).
Load AIC table	Opens a window for displaying previously saved results (Section 6 below).
Load model list	Opens a window for adding a previously specified list of models to run (Section 6 below).
Manage lists	Opens a window displaying both previous results tables and model lists. This choice is for deleting saved results (Section 6 below).

4.1 Specifying the response distribution

Before designing models for analysis you must specify whether you wish to analyze binomial or multinomial responses. The default is multinomial. If your fate variable contained only two fates to begin with then the two options will provide identical results. A note of caution! Changing the response variable during analysis will delete the current pending models list and results table (you will get a

16

warning first). This is to prevent you from inadvertently creating an AIC table comparing multinomial and binomial models.

4.2 Overview: designing models

Primary analysis is done using the **ModelDesign** window, which is accessed using the **Design models** option of the **Analysis** menu. With the **ModelDesign** window open (Fig. 13) you can design multiple models and then run them from the main *MCestimate* window. The default model is the Intercept-only model (no covariates, all failure rates temporally constant). *MCestimate* uses fairly conventional notation to name models. Each failure rate is designated as a basic model parameter and is specified as a function of specific covariates using parentheses. If no covariates are used then the parentheses brackets contain only a dot – indicating that the basic parameter is assumed constant across all nests and over time. This model, the intercept-only model, can be added to the list immediately by clicking the button **Add to pending models**. In the example above this will add the model **weather(.)depredated(.)** to the **Model Set** listbox.

Figure 13. The **ModelDesign** window for SimData.mat as created in Section 3.2
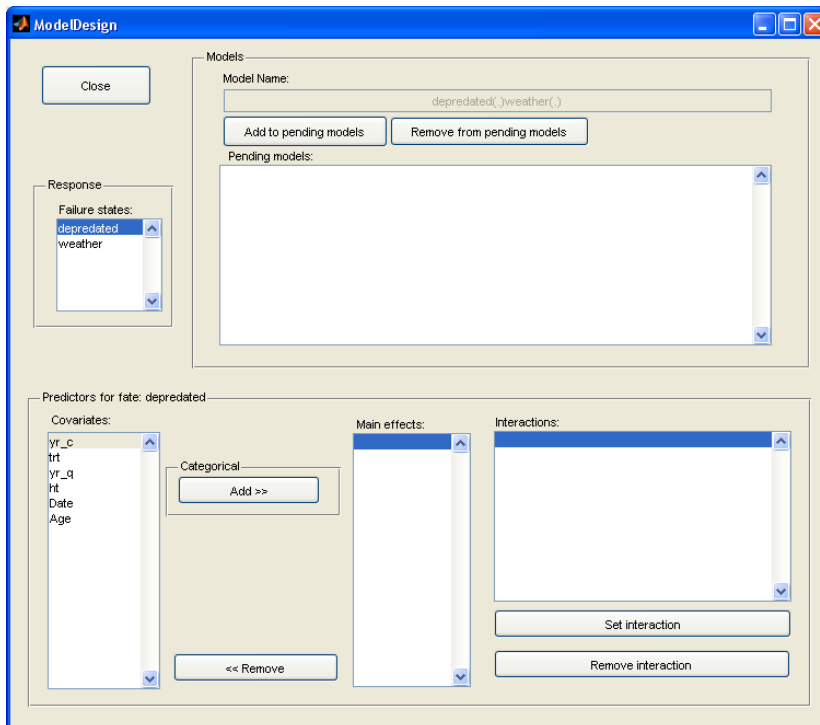
4.3 Building models by specifying covariates

Your specified covariates appear in the **Covariates** listbox of *MCestimate*. Note that there are two covariates (**Date** and **Age**) in that box beyond what you specified in the **DescribeData** step. *MCestimate* creates **Date** automatically and **Age** if an **Age at first visit** column is specified on the **DescribeData** GUI. Fates may be modeled using different covariates. To specify which fate you are currently working on,

click that fate in the **Failure states** listbox. This will set the title of the box below to: **Predictors for fate: [fate]**. Any covariates you specify will always be for the currently selected fate.

The buttons for specifying a covariate to a fate change depending on whether you have selected a categorical or a quantitative covariate. Fig. 13 shows the display for a categorical covariate. If you click on a quantitative covariate in the **Covariates** listbox the display will look like Fig. 14. With the exception of **Date** and **Age** all covariates are assumed to be nest specific (i.e., constant within nests).

Figure 14. The **ModelDesign** window when a quantitative covariate is selected.
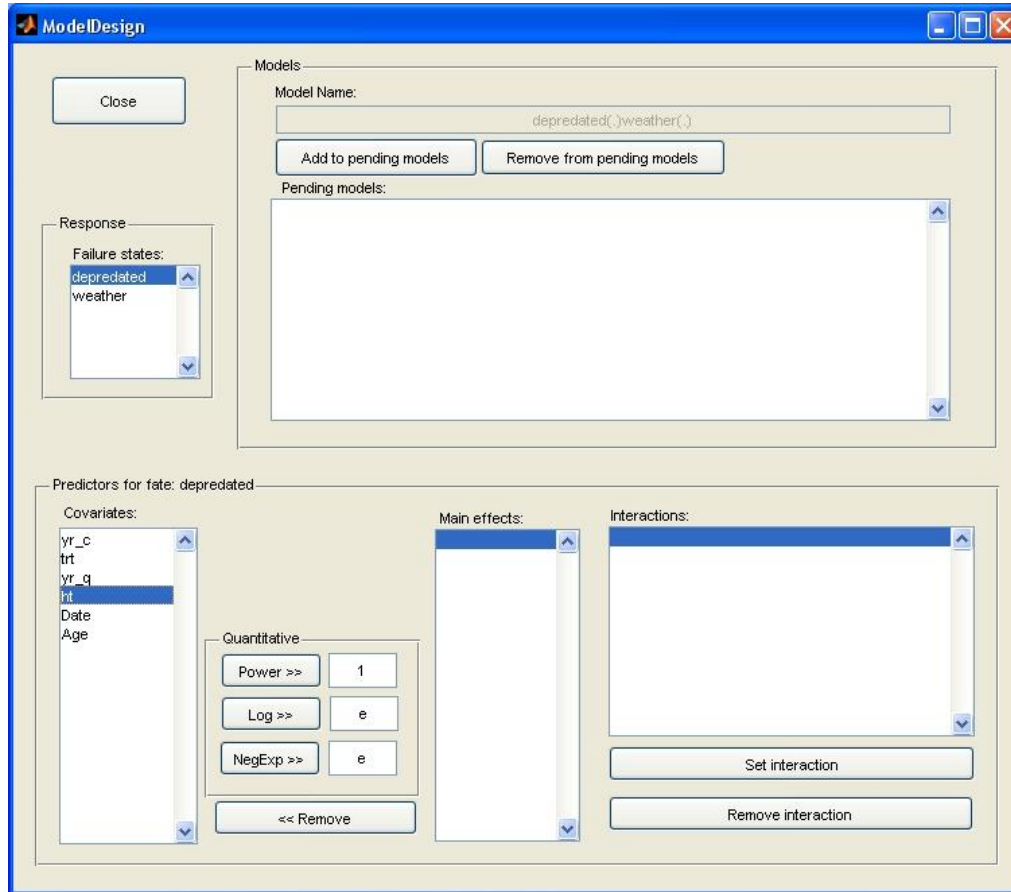
4.3.1 Adding a categorical covariate

Categorical covariates may be added to a model by selecting them in the **Covariates** listbox and clicking **Add >>**.

4.3.2 Adding a quantitative covariate

Quantitative covariates may be added by selecting them in the **Covariates** listbox and clicking the button corresponding to the desired transformation.

Power transformations

To add a power of a quantitative covariate click **Power >>**. *MCestimate* applies the indicated power transformation (edit box to the right of **Power >>**) to the quantitative covariate at run-time. The default is untransformed (power = 1). Any power may be entered. Note, however, that entering zero will simply generate a second intercept. Multiple powers of a given covariate may be entered sequentially (e.g., {height+height^2...}, etc.).

Log transformations

Log transformations work in analogous fashion to the power transformations (described above). The edit box is for entering the desired base. The default is the natural logarithm (**e**). With the exception of **e**, only strictly positive numeric values can be entered in the base edit box. Another restriction is that the base cannot be **1**. I'd recommend sticking to the bases **2**, **e**, and **10**. *MCestimate* will not add an arbitrary constant to zeros to allow log transformations. Instead, if you attempt to log-transform any quantitative covariate that contains data that are not strictly positive, you will get an error message and you will not be allowed to use the log transformation.

Negative exponential transformations

The NegExp transformation is intended to allow you to model thresholds in a quantitative covariate (i.e., a steep slope at small values but negligible additional effect at large values). The edit box is for entering the desired base. The default is the natural logarithm (**e**). The same restrictions apply to the base as with the log transformation. However, there is no restriction on the value of the covariate, except that very large numbers will result in "inf" errors that will crash *MCestimate*.

The Date covariate

MCestimate uses the dates you entered for visits (V1 – V10 in simulated data, Fig. 1) to generate an ordinal design variable (always called **Date**) to enable you to model within-experiment temporal effects. It is the numbered day within the year (i.e., where day 1 = 1 January). Date is a quantitative covariate that varies within nests with every day of the season (whereas other quantitative covariates above are assumed fixed for a given nest). Ordinal dates are assumed to be non-leap year. In creating **Date** *MCestimate* removes the year (i.e., 2 July 2006 and 2 July 2007 would have the same ordinal date: 184). If you want a **year** covariate you must add it as a separate column in your Excel file. You could then ask whether the effect of date differs by year by specifying an interaction between the **Date** and **year** covariates.

The Age covariate

If you specify an age at first visit column, *MCestimate* generates an ordinal **Age** covariate in the same way that it generated **Date**. Thus, like **Date**, the **Age** covariate will change with every day within a nest.

4.3.3 Interactions

Once you've added at least two covariates for a given fate to the **Main effects** listbox, you may specify an interaction. This is done by selecting all covariates you wish to be in the interaction in the **Main effects** listbox and clicking **Set interaction**. If you select more than two (say n) covariates then *MCestimate* builds all lower level interactions when generating the n-way interaction. If you remove a main effect that is specified in an interaction *MCestimate* will crash at run time.

4.3.4 Finalizing a model

To build models you move covariates from the **Covariates** listbox to the **Main effects** listbox using the rules described above and specifying any desired interactions. As you do so, the **Model Name** box in the **Models** frame will keep track of the covariates you are adding and removing. When you are satisfied with the model for the current fate you can select another fate in the **Failure states** listbox and continue. Once you have specified the model you wish for all fates, then you can add it to the set for analysis by clicking the **Add to pending models** button in the **Models** frame. If you change your mind after adding the model to the model set, then simply select it and click **Remove from pending models**. After you select **Add to pending models** the **ModelDesign** window reverts to the opening state, with the default model being the intercept model. Fig. 15 shows the **ModelDesign** window with five models designed and ready for analysis.
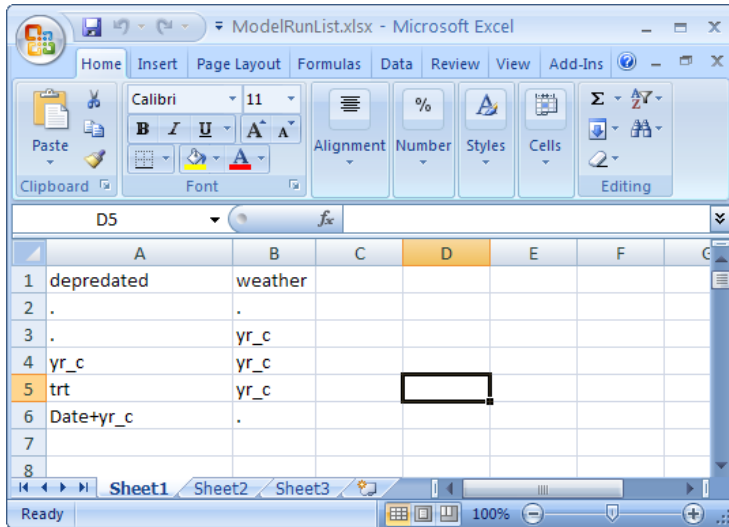
Figure 15. Some models designed in the **ModelDesign** window
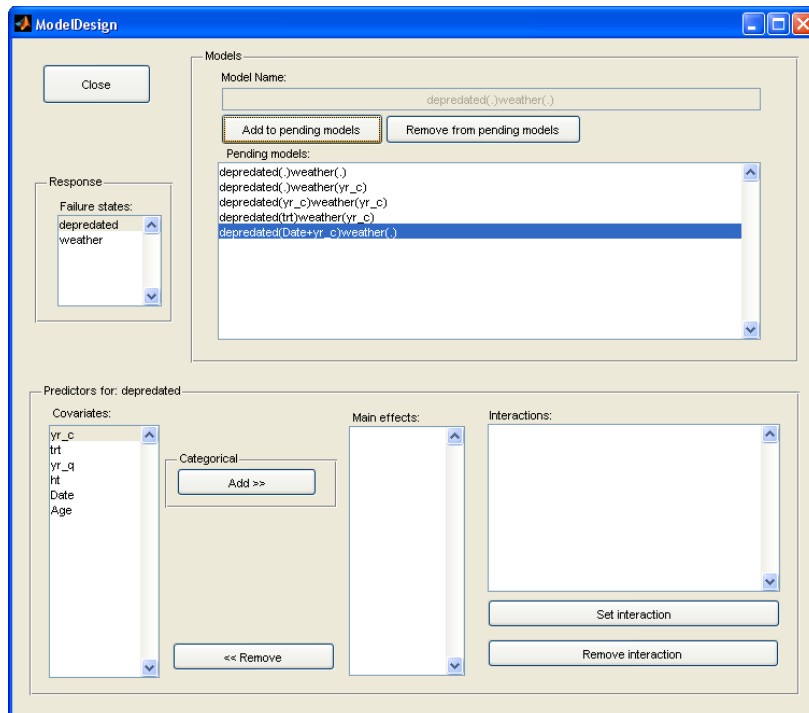
4.4 Importing models from a file

Models may also be typed into Excel and then imported to ***MCestimate***. As an example, the above models have been typed into the file "ModelRunList.xlsx" (Fig. 16). To do so, Choose **Load models from file** from the **Analysis** menu and navigate the desired file. The resulting models will be loaded into the pending models window. In the model list file the fates may be listed in any order, however the syntax must match exactly the syntax that MCestimate uses for building files. Any models that do not conform to this syntax are ignored.

Figure 16. Example model list for import to MCestimate

Probably the easiest way to learn MCestimate model syntax is to build a few models using the model builder (Section 4.3). Here are the basic rules. Main effects are listed by name (yr_c, trt, Date, etc.). Additive models are indicated by separating main effects with a plus sign (+). Interactions are indicated by separating main effects by an asterisk (*). Intercept-only models are indicated with a period (.). Transformations of quantitative covariates are indicated differently, depending on the transformation.

Power transformations of quantitative covariates are indicated with a caret symbol, followed by the desired power (e.g., Date^2). Log transformations differ depending on the desired base. If the natural log is desired, then the "ln" notation is used, whereas if any other base is desired, then the notation logX(covariate) is used, where X is the base (e.g., log10(Date)) indicates a common log of the variable Date). Similarly, the negative exponential transformation notation differs depending on whether the desired base is e or not. When the base is e, then the negative exponential transformation is indicated as exp(-covariate). For example, a negative exponential transformation of the Date variable is indicated as exp(-Date). However, if a different base is desired, the notation used is, for example, base^(-covariate). The resulting transformations can be combined into additive and interactive models as described in the previous paragraph.

4.5 Running the models

Once you've specified as many models as you wish, select **Close** to return to the main ***MCestimate*** window. You might wish to build the models in Fig. 15 to verify that you will get the same results from running ***MCestimate,*** and so you will be able to follow the results menus in the subsequent material. Upon closing the **ModelDesign** window, the main ***MCestimate*** window should look like Fig. 17, with the table of pending models appearing on the left of the window. To run the models click **Run models**. As each model runs, it will be removed from the **Pending models** table to the **Fitted models** table, which is a fairly typical AIC table. If you've specified too many models and want your computer back you can click **Interrupt** to stop the analysis. As each model is added to the AIC table, the table is reordered relative to the model with the lowest AIC value. The AIC values are corrected for sample size, where sample size is defined as the sum of the number of known days survived and the number of failed observations (Rotella et al. 2004).

Figure 17. The main ***MCestimate*** window with models ready to run
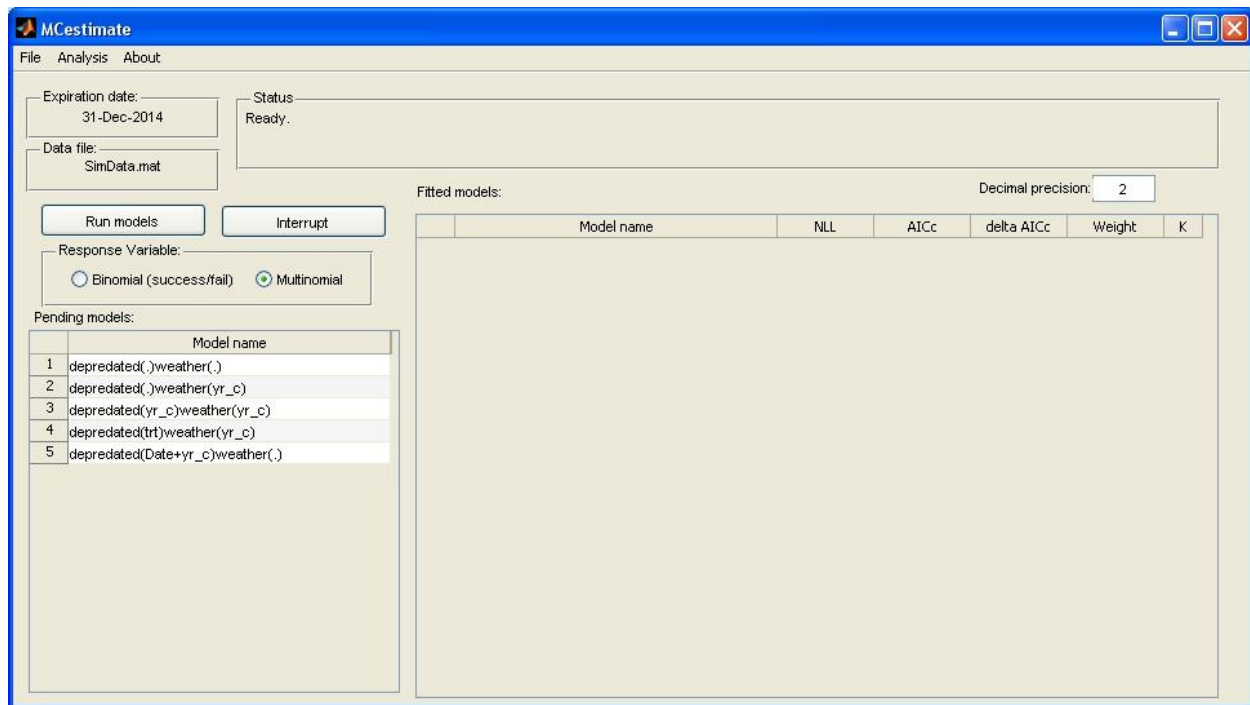
4.6 The Pending models context menu

Before clicking **Run models** or while running is paused by clicking the **Interrupt** button, models can be removed from the **Pending models** table using the pending models context menu. This menu is accessed by selecting a cell in the **Pending models** table and right-clicking, which will bring up four choices (Fig. 18). Choosing **Clear pending models** will delete *all* of the pending models from the table. They cannot be recovered, except by going back to the **DesignModels** GUI and rebuilding them (or reimporting from a file). Choosing **Export model list to Excel** will open a dialog box for saving the Pending models list to an Excel file for later import. This might be useful, for example, if you wish to run

the same set of models on a different data set. Choosing **Save model list** will first prompt you for a name for the model list then save the list for later use (Section 6.2 below) to the same *MCestimate* file in which your data are saved. This option may be removed in future editions as it is now redundant with the ability to save model lists to a distinct Excel file. Finally, the **Delete model** choice will delete all highlighted models in the **Pending models** table.

Figure 18. The **Pending models** context menu
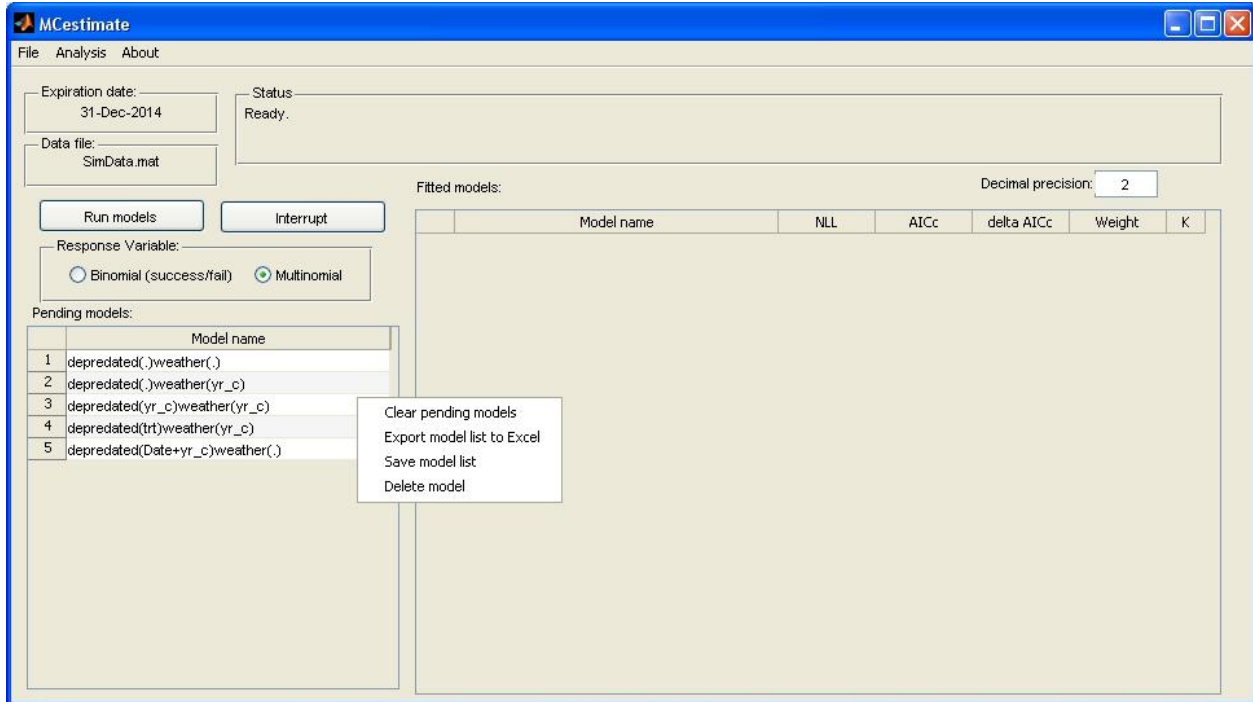
Figure 19. Models after they have been run

	Model name	NLL	AICc	delta AICc	Weight	K
1	depredated(Date+yr_c)weather(.)	962.01	1932.03	0	1	4
2	depredated(trt)weather(yr_c)	968.77	1945.54	13.51	0	4
3	depredated(.)weather(yr_c)	970.46	1946.93	14.9	0	3
4	depredated(yr_c)weather(yr_c)	970.23	1948.47	16.44	0	4
5	depredated(.)weather(.)	973.72	1951.45	19.42	0	2
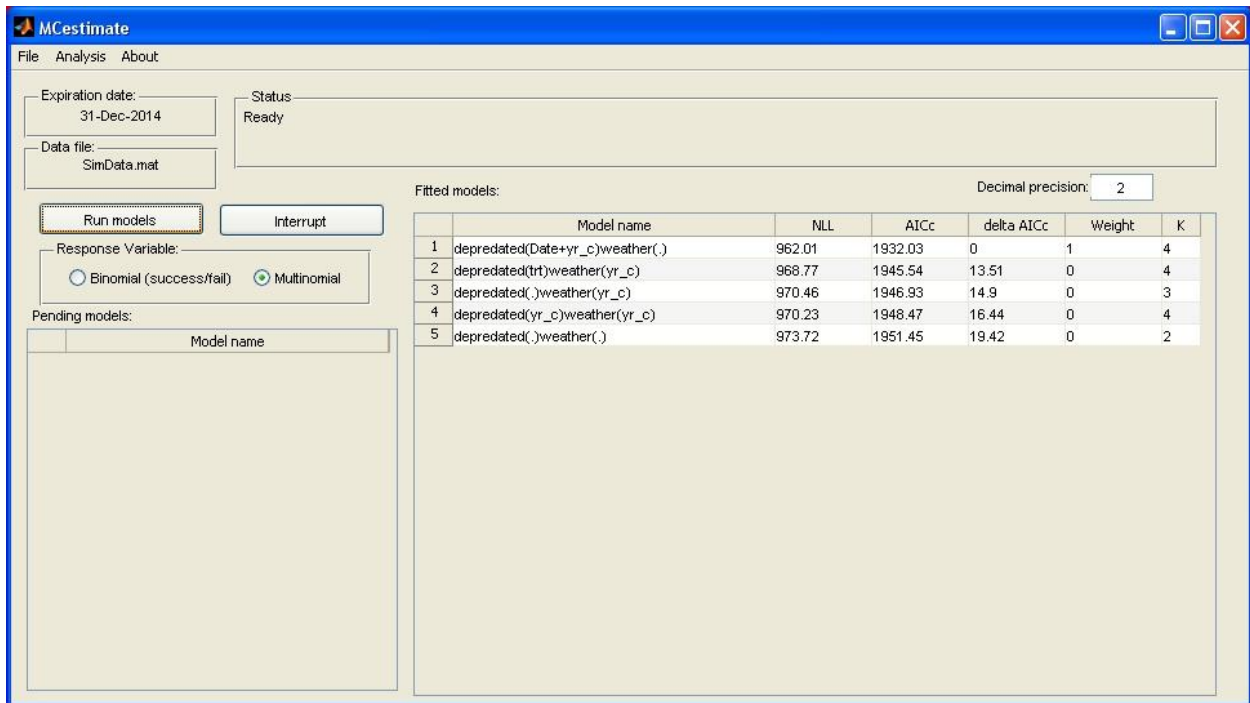
5. Interpreting model results

Once you've run the model, the output table (**Fitted models**) gives you the negative log-likelihood (**NLL**) of the model, the corrected AIC value (**AIC$_c$**), the **delta AIC$_c$** value relative to the lowest AIC$_c$ value in the set, the model **Weight** (calculated following Burnham and Anderson 2002), and the number of estimated parameters (**K**) for the model (Fig. 19). Models are sorted from lowest AIC$_c$ to highest.

The results given in the **Fitted models** table are only a start. Much more information can be gleaned using the context menu. To enable this menu you must select a row (any cell) in the **Fitted models** table and then right-click your mouse. This will bring up a context menu with two choices **Model** (Table 4, Fig. 20) and **Table** (covered in Section 6, below). Most of these options are fairly self-explanatory and can be understood by reading the descriptions in Tables 4 and 5. The **Show parameter estimates** window will probably be the most useful and I describe this option at length below.

Table 4. The **Model** context menu in the **Fitted models** table

Menu Item	Description
Rerun model	This reruns the model with a different (randomly chosen) starting point for the numerical optimization.
Show convergence stats	This brings up a message box giving the convergence statistics for that particular model.
Show parameter estimates	This brings up a new window with each of the estimated parameters (Figs. 20-22 below).
Assess goodness of fit	This brings up a new window that allows you to run a goodness-of-fit test. I have just added four goodness-of-fit algorithms (see Section 5.3).
Export estimates	Exports the maximum likelihood parameter estimates, covariance matrix and the predicted probabilities for each nest *over the period of observation* (not over the life of the nest) to an Excel file.
Delete model	Deletes the models selected in the **Fitted models** table.

Figure 20. The **Model** context menu
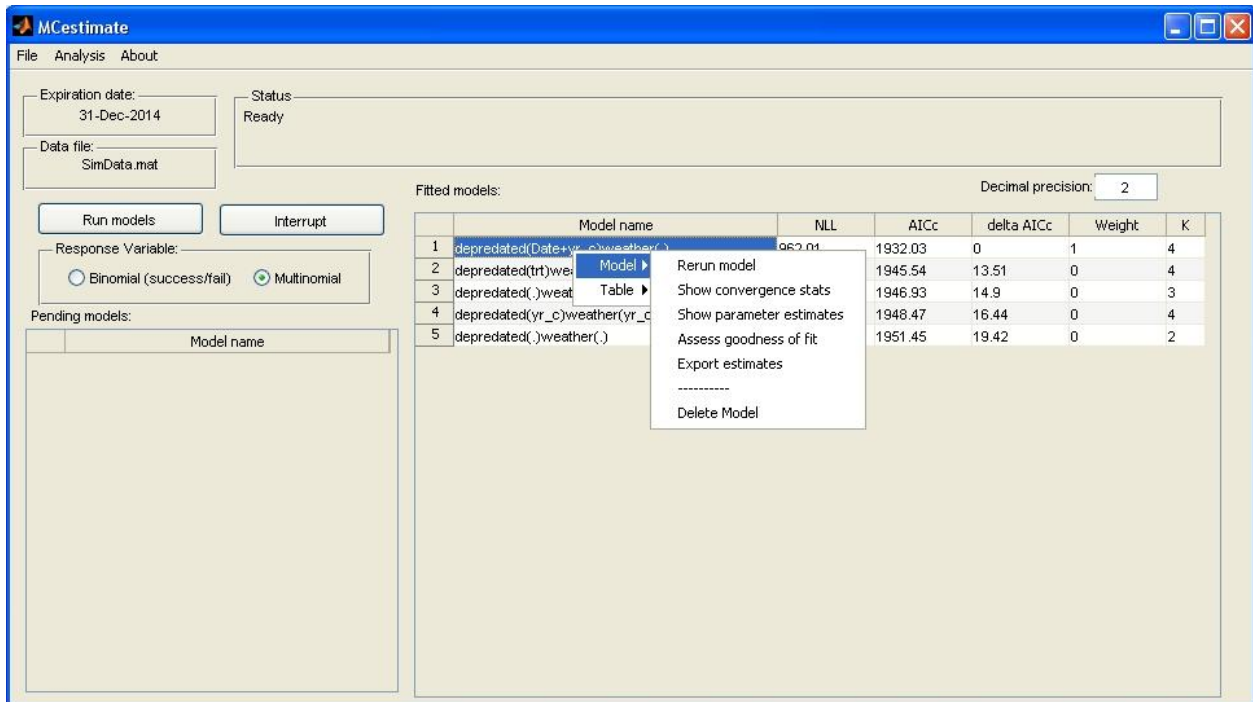
5.1 Model convergence

Occasionally **MCestimate** will fail to converge on a solution. This is more likely to happen the greater the dimensionality (number of parameters) of the model that is designed. Indicators of failed convergence include NaN in the NLL column of the Fitted models table or multinomial logit scale parameters that are equal to 10 (or minus 10), indicating that the optimization ran to a maximum (minimum). The latter is most likely to occur when all nests share the same fate. In particular, if you attempt to optimize a categorical covariate with many levels, it may occur by chance that all nests in that category share the same fate, which may result in the optimization running to a boundary condition (depending on what other covariates are in the model). If the optimization failed to converge due to a poor starting point you may try re-optimizing it from a different starting point. To rerun a model, select that model and chose **model**, **rerun model** in the **Fitted models** context menu. If you have reached a boundary condition due to invariant fates in a given category this will not help. In the latter case, your data do not provide sufficient information for estimation and you should not try to fit that model.

The convergence criteria for each model are also available as a **MCestimate** menu option. To see the convergence statistics for a given model choose **Show convergence stats** under the **model** option in the **Fitted models** context menu.

5.2 Parameter estimates

The most useful context menu choice is the **Show parameter estimates** choice under **Models** (Fig. 20), which loads the **Parms** window (Fig. 21). In the **Parms** window, parameter estimates are given at three scales (Multinomial logit scale, daily probabilities, and overall probabilities).

Figure 21. The Parms window
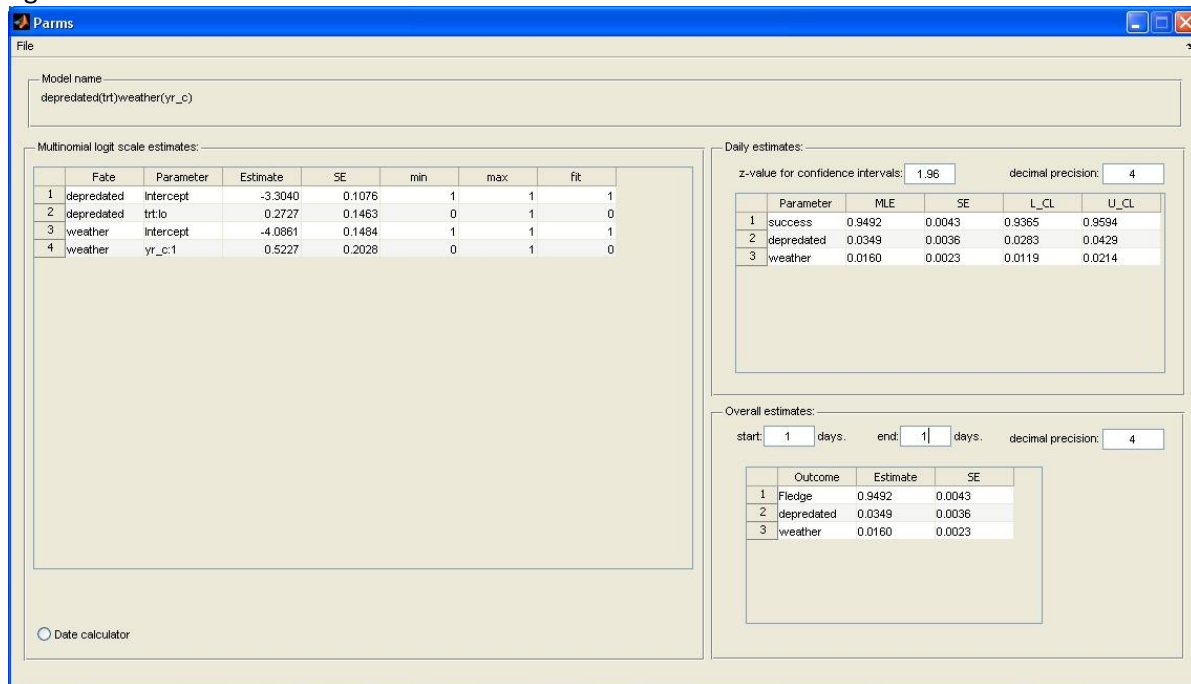
5.2.1 Multinomial logit scale estimates

The logit scale estimates (the "betas") are difficult to convert into survival probabilities because the probabilities depend not only on the betas, but also on particular covariate values. The **fit** column of the first table shows which covariate values are used for fitting.   When the **Parms** window is initially opened, the **fit** column is set to **1** for all intercept parameters, **0** for all categorical covariate levels (representing the reference category(s), which is lumped in with the intercept), and the mean value of a covariate for all quantitative covariates. For interactions, the product of the **fit** values for each main effect in the interaction is taken. These **fit** values are then used to generate the probabilities reported in the **Daily estimates** table, which should contain one row for each fate (including success).

5.2.2 Daily Estimates

Daily survival and failure rates are reported in the **Daily estimates** table. This table has five columns and as many rows as nest outcomes (including success). The column **MLE** gives the fitted probabilities corresponding to the values of covariates in the **fit** column of the **Multinomial logit scale estimates** table. Similarly the subsequent columns give the standard error of the fitted probability and lower and upper confidence limits of the parameter. The standard error is calculated using the delta method (Seber 1982). Confidence limits are calculated following the strategy outlined by Sambamoorthi et al. (1999). They are symmetrical around the linear portion of the model (prior to transformation via the multinomial logit) and become asymmetrical under the transformation. The z-value specifies the desired nominal coverage of the confidence interval. The default is $z = 1.96$ corresponding to an approximate 95% confidence interval. Changes in the **fit** column of the **Multinomial logit scale estimates** table are immediately applied to the **Daily estimates** and **Overall estimates** tables.

5.2.3 Overall Estimates

The third table, **Overall estimates**, extrapolates the daily probabilities for a single nest over the length of a cycle defined by the entries in the **start** and **end** edit boxes. The extrapolation to overall rates is done under the assumption that the particular covariate values in the **fit** column of the **Multinomial logit scale estimates** table pertain to the nest at Age = 1 day. Overall estimates are calculated for the period lasting from Age = **start** through Age = **end**, assuming the nest is still "alive" at age **start**. If you do not specify an Age covariate, the actual values of **start** and **end** are irrelevant and all that matters is their difference (the number of days for which an "Overall" estimate is desired). In this case you should always specify **1** for **start**. The ability to specify **start** is designed to allow you to obtain "Overall" estimates for a given developmental stage (e.g., the nestling phase). Temporal covariates (**Age** and **Date**) are assumed to increment by 1 each day from **start** until **end** is reached. All other covariates are assumed fixed for that particular nest.

5.2.4 Exploring the parameter space

Of course, the mean of quantitative covariates and the reference category of categorical covariates are not necessarily the only portions of the covariate space of interest. Therefore the **fit** column in the first table allows limited interaction. All main effect values can be edited. Intercept values and interactions

cannot be edited. Intercepts must always be "1" and interactions will change with changes in the corresponding main effects.

5.2.5 Design variables for quantitative covariates

All quantitative covariates are rescaled (by dividing by their standard deviation) when they are converted to design variables. **MCestimate** retains the standard deviations in memory so that the covariance matrix for untransformed variables can be estimated from the covariance matrix for the transformed variables.

5.2.6 Design variables for categorical covariates

All categorical covariates are converted to indicator variables (columns of ones and zeros) indicating whether a particular nest is in that category. Each categorical covariate with n levels generates n-1 indicator variables. The other category is lumped into the intercept. **MCestimate** arbitrarily allocates the first level it encounters of any categorical covariate into the intercept. In the example above the level "hi" of the categorical variable **trt** was lumped into the intercept.

5.2.7 Effects of changing covariate values on other covariates in the model

In some cases, changing a covariate value in the **fit** column will automatically change the value of other covariates in the model. When a level of a categorical covariate is set to "1" all other levels of that covariate are set to zero (except the intercept). When a main effect fit value is changed, any interactions involving that main effect will use the new value of that covariate, so the interaction values will change also. Similarly, when multiple powers (or other transformations) of a quantitative covariate are simultaneously included in a model, changing the fit value for one transformation will result in a corresponding change to the fit values of the other (i.e., if "variable" is set to "3" then "variable^2" will automatically change to "9" and vice versa).

Example 1. Interpreting parameter estimates

To get a sense for how to use the **Parms** window, click on a cell in the second row of the output table on the main **MCestimate** window (i.e., highlighting the model: **depredated(trt)weather(yr(c))**), which is the model hypothesizing that the nest-predation rate depends on treatment (a categorical covariate) and that the weather-related failure rate depends on year (a categorical covariate). Then right-click and choose **Model** > **Show parameter estimates**. The **Parms** window should look like Fig. 21 (above).

Suppose you were interested in the predation rate under the treatment condition **lo**. Then click that cell (row 2, column 7) in the **fit** column, type "1" and hit the "Enter" key . The new **Parms** window should look like Fig. 22. Note the changes in daily and overall parameters that have occurred in the second table (compare Fig. 21 to Fig. 22).

Figure 22. Parms window from Fig. 21 with trt:lo set to "1"

Parms

File

Model name
depredated(trt)weather(yr_c)

Multinomial logit scale estimates:

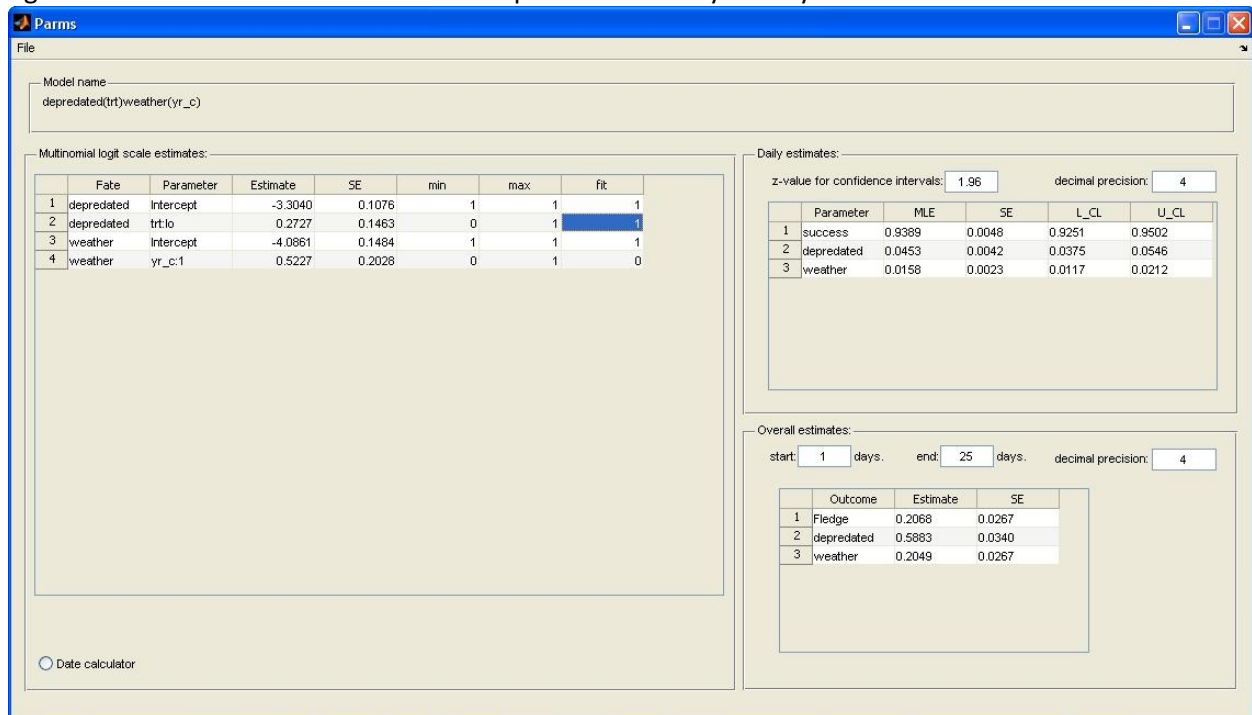|	Fate	Parameter	Estimate	SE	min	max	fit
1	depredated	Intercept	-3.3040	0.1087	1	1	1
2	depredated	trt:lo	0.2727	0.1474	0	1	1
3	weather	Intercept	-4.0861	0.1541	1	1	1
4	weather	yr_c:1	0.5227	0.2087	0	1	0

Daily estimates:

z-value for confidence intervals: 1.96   decimal precision: 4

	Parameter	MLE	SE	L_CL	U_CL
1	success	0.9389	0.0049	0.9247	0.9505
2	depredated	0.0453	0.0043	0.0373	0.0548
3	weather	0.0158	0.0024	0.0116	0.0214

Overall estimates:

start: 1 days.   end: 1 days.   decimal precision: 4

	Outcome	Estimate	SE
1	Fledge	0.9389	0.0049
2	depredated	0.0453	0.0043
3	weather	0.0158	0.0024

○ Date calculator

In Fig. 22, above, the overall rates are equal to the daily estimates because the **start** and **end** edit boxes are set to 1. Setting these values to 1 and 25, respectively (i.e., a 25-day nest cycle, from first egg to fledging) should give you the result displayed in Fig. 23.

Figure 23. Parameter estimates from Example 1 for a 25-day nest cycle

Parms

File

Model name
depredated(trt)weather(yr_c)

Multinomial logit scale estimates:

	Fate	Parameter	Estimate	SE	min	max	fit
1	depredated	Intercept	-3.3040	0.1076	1	1	1
2	depredated	trt:lo	0.2727	0.1463	0	1	1
3	weather	Intercept	-4.0861	0.1484	1	1	1
4	weather	yr_c:1	0.5227	0.2028	0	1	0

Daily estimates:

z-value for confidence intervals: 1.96   decimal precision: 4

	Parameter	MLE	SE	L_CL	U_CL
1	success	0.9389	0.0048	0.9251	0.9502
2	depredated	0.0453	0.0042	0.0375	0.0546
3	weather	0.0158	0.0023	0.0117	0.0212

Overall estimates:

start: 1 days.   end: 25 days.   decimal precision: 4

	Outcome	Estimate	SE
1	Fledge	0.2068	0.0267
2	depredated	0.5883	0.0340
3	weather	0.2049	0.0267

○ Date calculator

Example 2. Interpreting parameter estimates when Date is included

Close the Parms window and go back to the main **MCestimate** screen. Choose the "best" model (**depredated(Date+yr)weather(.)**) and display its parameter estimates. The numbers in the **min**, **max**, and **fit** columns for the **Date** covariate are not easily interpreted because they are ordinal dates. The **Date calculator** allows conversion between calendar dates and ordinal dates (and their transformations). It can be displayed by clicking the **Date calculator** radio button below the **Multinomial logit scale estimates** table (Fig. 24). Typing "167" into the **Ordinal:** edit box should result in "15-Jun" appearing in the **Date (dd-mmm):** edit box. If you've used a transformation of **Date** you can also accommodate for it using the appropriate controls to the right and the transformed design variable will be displayed (on its native scale, not divided by its standard deviation). Changes to any of the three edit boxes (**Date (dd-mmm):**, **Ordinal:**, or **Design variable**:) will carry through to the other two. Note that this calculator cannot currently handle negative exponential transformations.

Figure 24. Using the **Date calculator** to interpret models with "Date" as a covariate
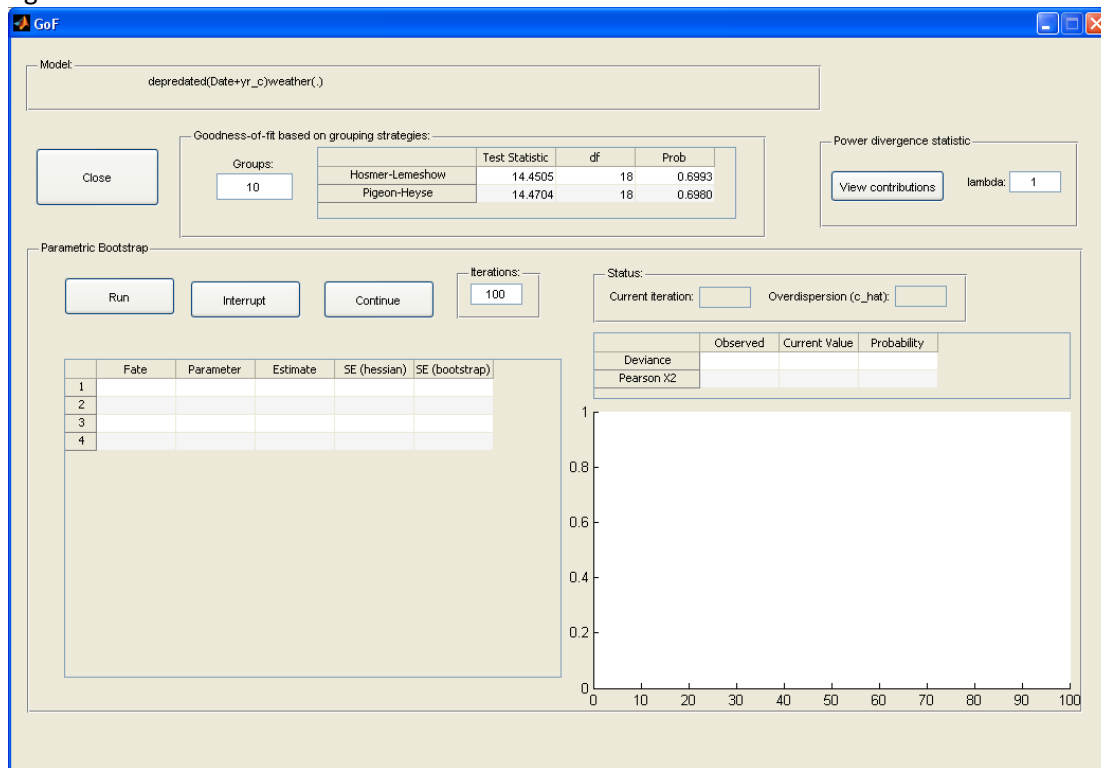

5.3 Goodness-of-fit

Assessing the fit of nest survival models is challenging (Sturdivant et al. 2007). **MCestimate** uses several different algorithms, including a multinomial extension of the Hosmer-Lemeshow test (Hosmer & Lemeshow 2000), a similar method described by Pigeon and Heyse (1999), and two methods based on

parametric bootstrap (White et al. 2001). The latter algorithms rely on repeatedly generating sets of data from the maximum likelihood estimates of the linear coefficients and fitting the same model to the new sets of data. In this way, a null distribution of a given test statistic is generated under the hypothesis that the data arose from the fitted model parameterized with the estimated MLEs. The complement of the proportional rank of the test statistic is taken to be the p-value for that test statistic. Several test statistics are available for use in the parametric bootstrap procedure, including the deviance from a saturated model (Deviance) and various forms of the power-divergence family of statistics (Read & Cressie 1988). The latter family includes the Pearson X^2 statistic and the Freeman-Tukey statistic. I recommend that you use the parametric bootstrap option with caution! It is very slow and often does not exhibit improved performance compared to the Hosmer-Lemeshow and Pigeon-Heyse tests, which are assessed against chi-square distributions with known degrees of freedom (hence do not require lengthy simulations).

To use these algorithms choose **Assess goodness of fit** under the **model** option in the **Fitted models** context menu (Table 4). This will load the **GoF** GUI (Fig. 25).

Figure 25. The **GoF** GUI when first loaded

At the top of the **GoF** GUI the name of the model to which the current goodness-of-fit results pertain is displayed. Immediately below the model name is a panel (**Goodness-of-fit based on grouping strategies**) showing p-values for the Hosmer-Lemeshow test and the Pigeon-Heyse test. The default value for the number of groups to consider is 10 (following guidance from Hosmer and Lemeshow 2000), but you may enter other values. Typically the Hosmer-Lemeshow p-value is very close to that for the

Pigeon-Heyse statistic given the same number of groups. Below these results is the **Parametric Bootstrap** panel. To run a parametric bootstrap test you must choose how many iterations you wish to simulate and click **Run**. Usually 100 iterations is enough to figure out whether you should be concerned about goodness-of-fit, but if so, then you should run at least 500 iterations (better yet 1,000). Plan on having something else to do during that time! Above and to the right of the **Parametric Bootstrap** panel is a small panel labeled **Power divergence statistic**. This box has two uses. First, the parameter lambda determines which member of the **Power divergence family** is used as the test statistic in the **Parametric bootstrap**. A value of lambda = 1 corresponds to Pearson's X^2 statistic. A value of lambda = −0.5 corresponds to the Freeman-Tukey statistic. Reed and Cressie (1988) recommend a value of lambda ≈ ⅔. You don't necessarily need to decide in advance because you can change the value of lambda on the fly (example below). The ability to change the value of lambda is intended to allow you to ensure that your data do not exhibit lack-of-fit under different statistics. It is *NOT* intended to allow you to shop for a statistic under which you get an acceptable p-value! The second use for the **Power divergence statistic** panel is to allow you to view the contributions from each particular observation in your data set to the Power divergence statistic (with parameter lambda). To do so, enter the desired value of lambda and choose **View contributions**. This will bring up a table of contributions (described in example below).

If you run a parametric bootstrap test, in addition to the chosen power divergence statistic, you will also be provided with goodness-of-fit results for the deviance of the current model compared to a saturated model. You will also generate estimates of overdispersion and bootstrap estimates of the standard errors of the linear coefficients. Do not be too alarmed if these disagree with the estimates generated from the Hessian matrix. It can take many iterations for the bootstrap estimates to converge. However, if after many iterations the estimates still do not agree, then the bootstrap estimates are probably more reliable, especially at small sample sizes. Finally, the parametric bootstrap test will also generate a graphical display showing the current value and cumulative probability trajectory of the goodness-of-fit statistics. This is intended to help you decide when enough simulations have been run.

5.4 Example. Assessing Goodness-of-fit

If you have imported SimData.xls as described above and fitted the Models from Section 4, then load the **GoF** GUI so that it looks like Fig. 25. Preliminary indications from the Hosmer-Lemeshow and Pigeon-Heyse tests suggest little evidence for lack-of-fit (Fig. 25). For further verification, a bootstrap goodness-of-fit test could be run. To do so, simply click **Run** within the **Parametric Bootstrap** panel. Fig. 26 shows the bootstrap results after 100 iterations (note this took about 30 minutes to generate).

Figure 26. Bootstrap goodness-of-fit results for the depredated(Date+Yr_c)weather(.) model
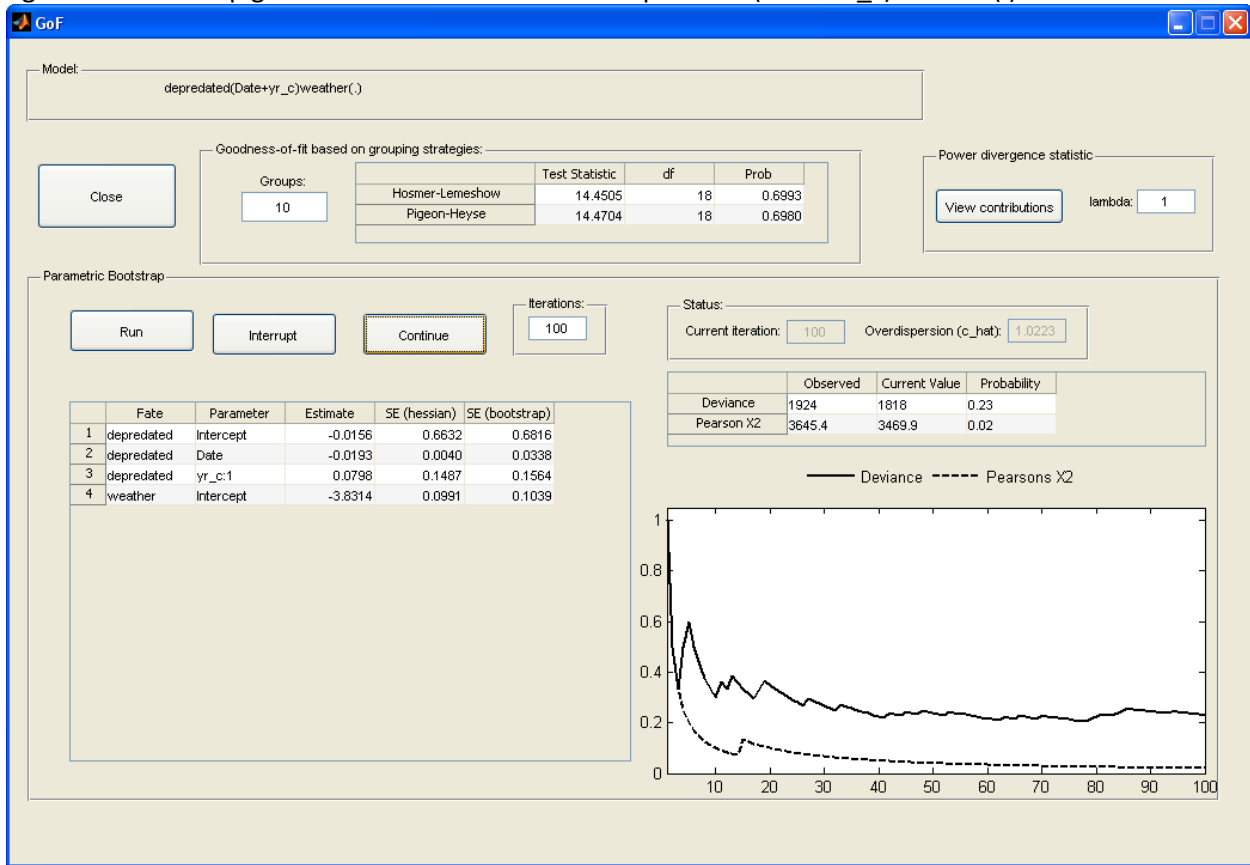
According to the table showing goodness-of-fit statistics (above the panel graph of probabilities), the observed deviance (compared to a saturated model) was 1,924, which was well within the distribution of simulated deviances (Fig. 26). However, the Pearson X^2 statistic was higher than almost all simulated values, indicating lack-of-fit. The probability is 0.02 in this case because the observed value of X^2 from the empirical data is taken to be the first member of the simulation set and one iteration of the simulation (around iteration 16) produced a value of X^2 greater than the observed value. Thus 2 out of 100 values were greater than or equal to the observed value. To see which observations are contributing the most to the test statistic, you can click **View contributions** in the **Power divergence statistic** panel. Doing so in this case, with lambda=1, should produce a table looking like Fig. 27.

Figure 27. Table of observation-specific contributions to the Peason X^2 statistic

	1	2	3	4	5	6	7	8
1	Nest	Date	Length	Outcome	survived	depredated	weather	PDS contrib.
2	39	17-May-2000		2 weather	0.8333	0.1289	0.0378	25.4264
3	201	18-May-2000		2 weather	0.8440	0.1178	0.0382	25.1777
4	105	23-May-2000		2 weather	0.8457	0.1161	0.0383	25.1385
5	166	28-May-2000		2 weather	0.8552	0.1063	0.0386	24.9231
6	450	27-May-2000		2 weather	0.8607	0.1005	0.0388	24.7984
7	64	02-Jun-2000		2 weather	0.8639	0.0972	0.0389	24.7279
8	375	05-Jun-2000		2 weather	0.8688	0.0922	0.0390	24.6197
9	241	06-Jun-2000		2 weather	0.8704	0.0905	0.0391	24.5850
10	142	03-Jun-2000		2 weather	0.8722	0.0887	0.0391	24.5463
11	362	08-Jun-2000		2 weather	0.8735	0.0873	0.0392	24.5176
12	127	09-Jun-2000		2 weather	0.8750	0.0858	0.0392	24.4848
13	41	05-Jun-2000		2 weather	0.8752	0.0855	0.0392	24.4803
14	337	10-Jun-2000		2 weather	0.8765	0.0842	0.0393	24.4528
15	349	13-Jun-2000		2 weather	0.8864	0.0740	0.0396	24.2410
16	381	21-Jun-2000		2 weather	0.8912	0.0690	0.0398	24.1382
17	230	22-Jun-2000		2 weather	0.8924	0.0677	0.0398	24.1128
18	268	23-Jun-2000		2 weather	0.8936	0.0665	0.0399	24.0879
19	361	23-Jun-2000		2 weather	0.8936	0.0665	0.0399	24.0879
20	369	26-Jun-2000		2 weather	0.8971	0.0629	0.0400	24.0161
21	130	29-Jun-2000		2 weather	0.9003	0.0596	0.0401	23.9483
22	231	25-Jun-2000		2 weather	0.9005	0.0594	0.0401	23.9453
23	188	07-Jul-2000		2 weather	0.9083	0.0514	0.0403	23.7859
24	258	04-Jul-2000		2 weather	0.9093	0.0503	0.0404	23.7647
25	442	04-Jul-2000		2 weather	0.9093	0.0503	0.0404	23.7647
26	183	15-Jul-2000		2 weather	0.9151	0.0443	0.0406	23.6468
27	399	11-Jul-2000		2 weather	0.9152	0.0442	0.0406	23.6446
28	455	11-Jul-2000		2 weather	0.9152	0.0442	0.0406	23.6446
29	49	18-Jul-2000		2 weather	0.9175	0.0419	0.0407	23.6000
30	44	22-Jul-2000		2 weather	0.9204	0.0389	0.0407	23.5416
31	126	07-Jul-2000		2 depredated	0.9119	0.0476	0.0405	20.0134
32	100	06-Jul-2000		2 depredated	0.9111	0.0485	0.0404	19.6266
33	408	06-Jul-2000		2 depredated	0.9111	0.0485	0.0404	19.6266
34	412	06-Jul-2000		2 depredated	0.9073	0.0523	0.0403	18.1029
35	224	01-Jul-2000		2 depredated	0.9065	0.0532	0.0403	17.8008
36	400	28-Jun-2000		2 depredated	0.9066	0.0542	0.0402	17.4503

As you can quickly determine by inspection, the greatest contributions to the Peason X^2 statistic in this case are coming from observations on weather failure. Scrolling further down (not shown) you would find that the next largest contributions come from observations on predation events, followed by observations over which nests survived. This is typical, as the occurrence of rare events will dominate the Pearson X^2 statistic when all cells have exactly one observation (hence the need to evaluate the distribution of the test statistic via parametric bootstrap).

Changing the value of lambda in the Power divergence statistic panel will change the test statistic and will retroactively apply the changes to the results of the Parametric Bootstrap. For example, the next figure shows the results for the Freeman-Tukey statistic (Fig. 28).

Figure 28. Parametric bootstrap results for the Freeman-Tukey statistic
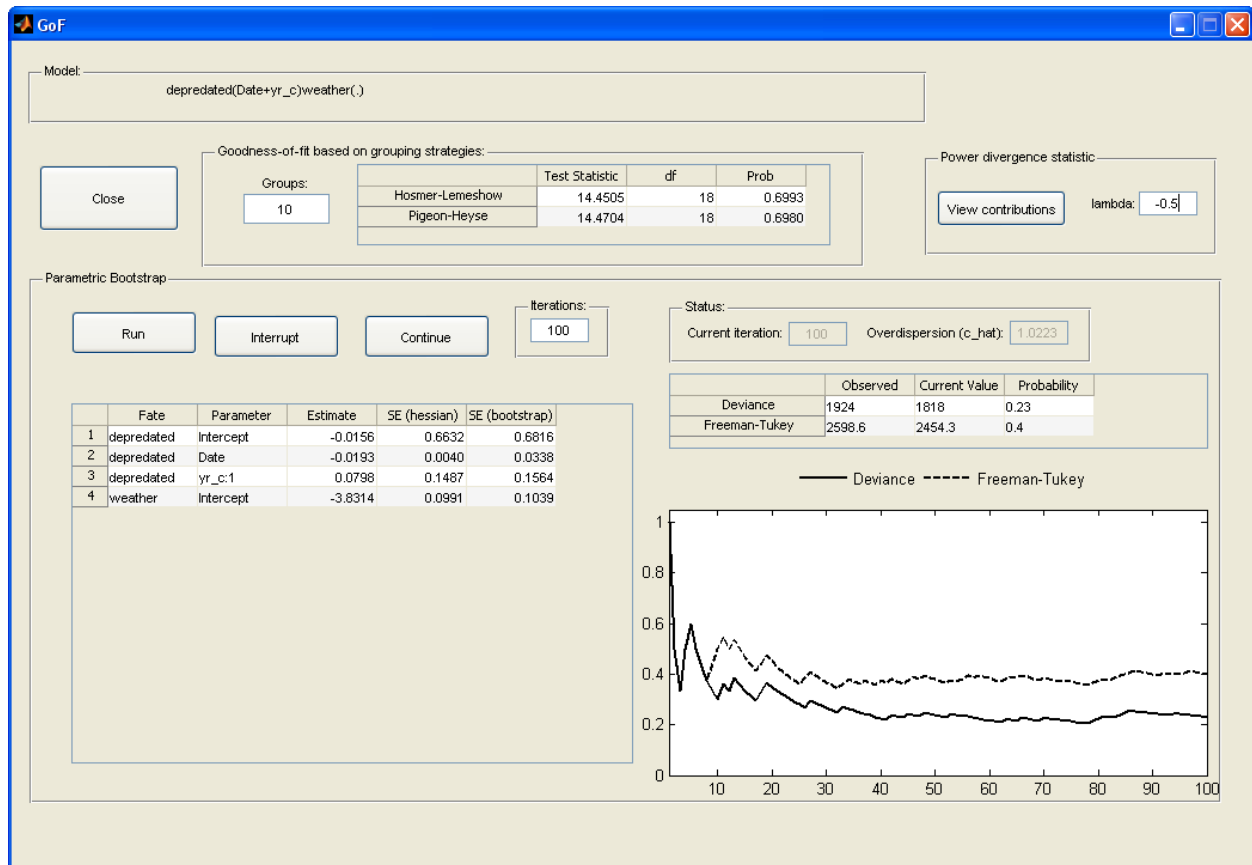
Unlike the Pearson X2 statistic, the Freeman-Tukey statistic does not suggest lack-of-fit (Fig. 28). This should be interpreted with caution! Limited power analyses suggest that the Freeman-Tukey statistic rarely has much power to detect lack-of-fit (for these kinds of data) and that the Pearson X2 statistic tends to reject at the nominal rate when the model fits the data. Thus I do not recommend adoption of the Freeman-Tukey statistic without good *a priori* justification. Fig. 28 also shows that by 100 iterations, the bootstrap estimates of standard errors are fairly close to those obtained from the Hessian (with the exception of the estimate for Date). This happened fairly rapidly in this case and you should not expect such rapid agreement! Finally, note also that overdispersion is small ($\hat{c} = 1.0223$) and can probably be ignored.

5.5 Exporting MCestimate results

All tables in **MCestimate** can be copied and pasted into other programs. To do so, select the cells you wish to copy and press ctrl-c to copy the cells into the windows buffer. They can then be pasted using ctrl-v or the edit-paste menus of most programs.

MCestimate provides two additional ways to export results. The first is **Export estimates** under the **Model** choice of the **Fitted models** context menu. Choosing this option will bring up a file save dialog

(Fig. 29) with a default name being the name of the model under which the estimates are derived (though you may change it).

Figure 29. The file save dialog that appears after choosing **Export estimates.**

This option saves an Excel workbook with three worksheets. **Sheet1** will contain the MLEs (the betas) and the covariance matrix. **Sheet2** will contain fitted probabilities for each nest from the time of discovery to the end of monitoring (note these are *not* fitted probabilities over the life of the nest!). **Sheet3** will contain fitted probabilities for each observation on each nest. I recommend caution interpreting the values on **Sheet2** and **Sheet3**. They are generally equivalent to neither daily nor overall values. This option is included to facilitate a do-it-yourself goodness-of-fit test.

Another way to export results from *MCestimate* is to use the **File** menu on the **Parms** GUI (e.g., Fig. 22, above). To use this function you must first select the name of a quantitative covariate in the **Multinomial logit scale estimates** window. *MCestimate* will then generate 101 values of the selected quantitative covariate evenly spaced between the minimum and maximum value of the covariate. It then uses those 101 values to fit daily and overall probabilities as described above (Sections 5.2.2 and 5.2.3). The resulting fitted probabilities are exported to an Excel file of the same name as the selected covariate (in this case **Date.xls**, Fig. 25). This function is intended to help with graphing the relationship between the fitted probabilities and continuous covariates.

Figure 30. A portion of the file Date.xls, sheet = weather,created using the **Export series** button

	A	B	C	D	E	F	G
1	Date	daily est.	daily SE	L CI	U CI	overall est.	overall SE
2	124	0.0194	0.0020	0.0159	0.0236	0.0194	0.0020
3	124.81	0.0194	0.0020	0.0159	0.0236	0.0194	0.0020
4	125.62	0.0194	0.0020	0.0159	0.0237	0.0194	0.0020
5	126.43	0.0194	0.0020	0.0160	0.0237	0.0194	0.0020
6	127.24	0.0195	0.0020	0.0160	0.0237	0.0195	0.0020
7	128.05	0.0195	0.0020	0.0160	0.0237	0.0195	0.0020
8	128.86	0.0195	0.0020	0.0160	0.0238	0.0195	0.0020
9	129.67	0.0195	0.0020	0.0160	0.0238	0.0195	0.0020
10	130.48	0.0196	0.0020	0.0161	0.0238	0.0196	0.0020
11	131.29	0.0196	0.0020	0.0161	0.0239	0.0196	0.0020
12	132.1	0.0196	0.0020	0.0161	0.0239	0.0196	0.0020
13	132.91	0.0196	0.0020	0.0161	0.0239	0.0196	0.0020
14	133.72	0.0197	0.0020	0.0161	0.0239	0.0197	0.0020
15	134.53	0.0197	0.0020	0.0162	0.0240	0.0197	0.0020
16	135.34	0.0197	0.0020	0.0162	0.0240	0.0197	0.0020
17	136.15	0.0197	0.0020	0.0162	0.0240	0.0197	0.0020
18	136.96	0.0197	0.0020	0.0162	0.0241	0.0197	0.0020
19	137.77	0.0198	0.0020	0.0162	0.0241	0.0198	0.0020
20	138.58	0.0198	0.0020	0.0162	0.0241	0.0198	0.0020
21	139.39	0.0198	0.0020	0.0163	0.0241	0.0198	0.0020
22	140.2	0.0198	0.0020	0.0163	0.0242	0.0198	0.0020
23	141.01	0.0199	0.0020	0.0163	0.0242	0.0199	0.0020
24	141.82	0.0199	0.0020	0.0163	0.0242	0.0199	0.0020
25	142.63	0.0199	0.0020	0.0163	0.0242	0.0199	0.0020
26	143.44	0.0199	0.0020	0.0163	0.0242	0.0199	0.0020

5.6 Visualizing MCestimate results

The **File** menu on the Parms window also contains the option **Plot series**. As above, to use this function you must first select the name of a quantitative covariate in the **Multinomial logit scale estimates** window. *MCestimate* will then generate 101 values of the selected quantitative covariate, as described above and plot them for you with confidence intervals (the latter using specified z-values). For example, with the **Parms** window open for the model **depredated(Date+Yr(c))weather(.)** and the name **Date** highlighted in the **Multinomial logit scale estimates** table, choosing **File/Plot** series will produce three figures that resemble Figs. 31-33, respectively.

Figure 31. Daily survival rate, with Date in the **depredated(Date+Yr(c))weather(.)** model
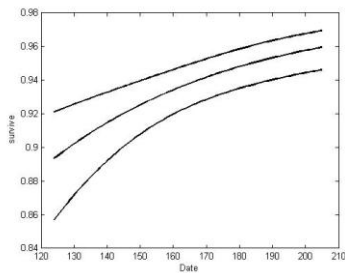
Figure 32. Daily depredation rate, with Date in the **depredated(Date+Yr(c))weather(.)** model
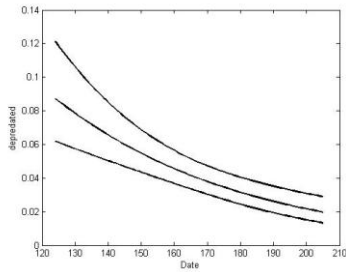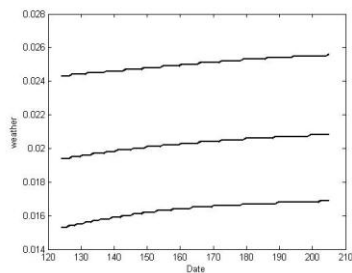

Figure 33. Daily rate of failure due to adverse weather, with Date in the **depredated(Date+Yr(c))weather(.)** model

I would welcome suggestions and thoughts on how to improve these graphing functions!

6. Managing results tables and model lists

Constantly clicking through the creation of models in *MCestimate* can be tedious. Thus model lists and AIC tables can be saved, re-opened, and subsequently modified (add or remove models). The easiest way to do this is through the **Table** options in the context menu of the **Fitted models** table (Table 5, Fig. 34).

Table 5. The **Table** context menu in the **Fitted models** table

Menu Item	Description
Save AIC Table	Saves the AIC table so that it can be loaded later. Saved AIC tables are automatically saved to the data file currently in use.
Save as model list	Saves the models in the AIC table as a list that can be loaded and run at a later time. Saved model lists are automatically saved to the data file currently in use.
Clear AIC Table	Deletes the **Fitted models** table. It cannot be recovered.

Figure 34. The **Table** context menu
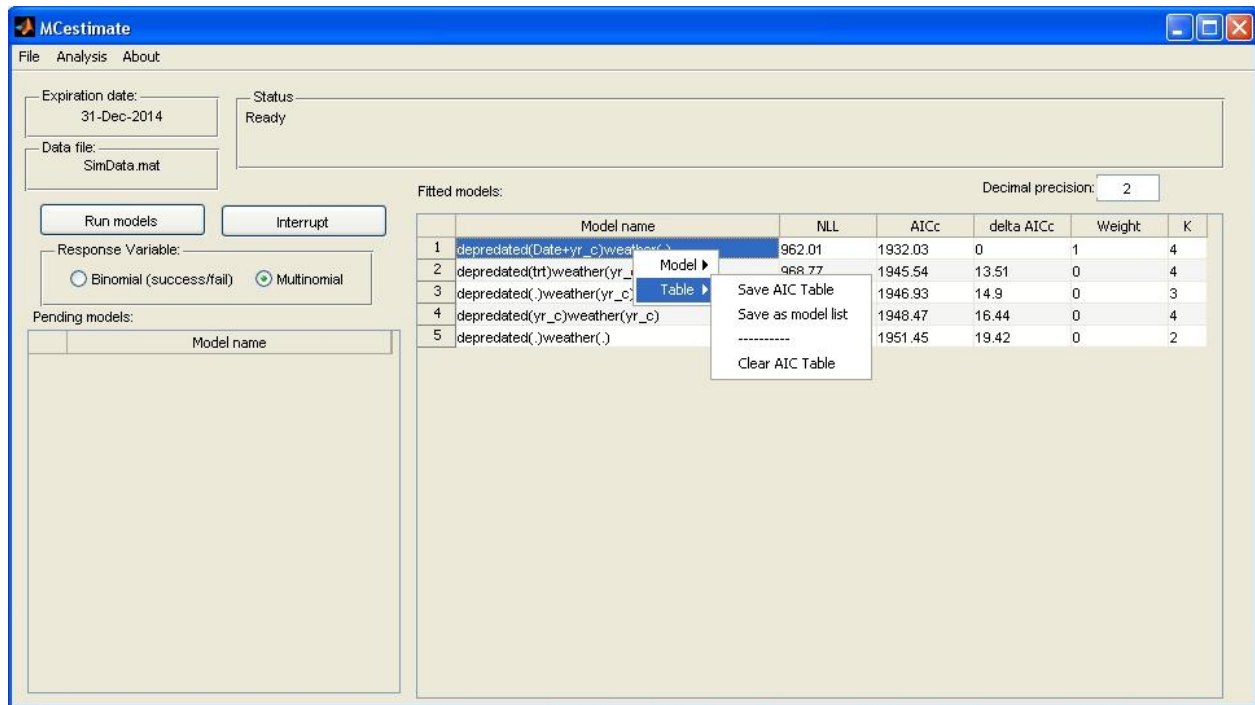
6.1 Saving an AIC Table

Choosing **Save AIC Table** under the **Table** option in the **Fitted models** context menu (Fig. 34) will load a dialog box for specifying a name for the AIC Table. Note: spaces are not allowed in the table name. If you include a space, the table will be saved, but will generate an error when you attempt to load the table subsequently. Names of AIC Tables and model lists (below) should always begin with a letter. When you save an AIC table, *MCestimate* automatically saves it to the current data file on disk.

6.2. Saving a model list

In some cases you may wish to save a model list to run in the future. This can be done in two ways. An existing AIC table can be saved as a model list by choosing **Save as model list** under the **Table** option in the **Fitted models** context menu. As with saving an AIC table, this will bring up a dialog asking you to name the model list. Also as above, the name cannot include spaces. The second way to save a model list is to choose **Save model list** in the **Pending models** context menu.

6.3 Managing lists

Once you've saved an AIC table or a model list it can be subsequently loaded under the **Analysis** menu using the analogous load command (**Load AIC table** or **Load model list**). In either case, the choices you will be offered to load will depend on the response variable that is currently selected in the radio button on the main *MCestimate* page. *MCestimate* will only allow you to load model lists or AIC results that

39

match the selected response variable (binomial or multinomial). As you save model lists and AIC tables you may wish to delete lists that are no longer needed. You can do so under the **Analysis** menu choice **Manage lists**. Saved model lists and AIC tables may be compiled into larger lists by loading them into the main *MCestimate* window. Whenever you load a list or an AIC table you have the option of either appending or replacing the existing table.

Acknowledgments

Literature cited

Burnham, K., and D.R. Anderson. 2002. Model selection and multimodel inference: A practical information-theoretic approach, 2nd Ed. Springer-Verlag, NY.

Choquet, R. A.–M. Reboulet, R. Pradel, O. Gimenez, and J.–D. Lebreton. 2004. M–SURGE: new software specifically designed for multistate capture–recapture models. Animal Biodiversity and Conservation 27.1: 207-215.

Choquet, R., L. Rouan, and R. Pradel. 2009. Program E-SURGE: a software application for fitting multievent models. Environmental and Ecological Statistics 3:8-39.

Dinsmore S.J., G.C. White, and F.L. Knop. 2002. Advanced techniques for modeling avian nest survival. Ecology 83:3476-3488.

Etterson, M.A., L.R. Nagy and T. Rodden-Robinson. 2007a. Partitioning risk among different causes of nest failure. Auk 124:432-443.

Etterson M.A., B.J. Olsen, and R. Greenberg. 2007b. The analysis of covariates in multi-fate Markov chain nest-failure models. Studies in Avian Biology 34:55–64.

Hosmer, D.W. and S. Lemeshow. 2000. Applied Logistic Regression. John Wiley & Sons, Inc. New York, NY.

Hougaard, P. 1999. Multi-state models: A review. Lifetime Data Analysis 5:239-264.

Mathworks. 2009. MATLAB: The language of technical computing. The Mathworks, Inc. Natick, MA.

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

Mayfield, H.F. 1975. Suggestions for calculating nesting success. Wilson Bulletin 87:456-466.

Pigeon, J.G. and J.F. Heyse. 1999. An improved goodness of fit statistic for probability prediction models. Biometrical Journal 41:71-82.

Pradel, R. 2005. Multievent: an extension of multistate capture recapture to uncertain states. Biometrics 61:442-447.

Read, T.R.C., and N.A.C. Cressie. 1988. Goodness-of-Fit Statistics for Discrete Multivariate Data. Springer-Verlag. New York, NY. USA.

Rotella J., S.J. Dinesmore and T.L. Shaffer. 2004. Modeling nest-survival data: A comparison of recently developed methods that can be implemented in MARK and SAS. Animal Biodiversity and Conservation 27:187-205.

Sambamoorthi, N., V.J. Ervin and G. Thomas. 1994. Simultaneous prediction intervals for multinomial logistic regression models. Communications in Statistics – Theory and Methods 23:815-829.

Seber, G.A.F. 1982. The Estimation of Animal Abundance and Related Parameters, 2nd Ed. Blackburn Press. Caldwell, NJ.

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.

Stanley, T.R. 2004a. Estimating stage-specific survival probabilities of nests when nest age is unknown. Auk 121:134:147.

Stanley, T.R. 2004b. When should Mayfield data be discarded? Wilson Bulletin 116:267-269.

Sturdivant, R.X., J.J. Rotella, and R.E. Russell. 2010. A smoothed residual based goodness-of-fit statistic for nest-survival models. Studies in Avian Biology 34:45–54.

White, G.C. and K.P. Burnham. 1999. Program MARK: Survival estimation from populations of marked animals. Bird Study 46:S120-S129. See also: http://warnercnr.colostate.edu/~gwhite/mark/mark.htm#Introduction

White, G.C., K.P. Burnham, and D.R. Anderson. 2001. Advanced features of program MARK. Pages 368-377 in Wildife, Land, and People: Priorities for the 21st Century. Proceedings of the 2nd International Wildlife Management Congress. R. Field, R. J. Warren, H. Okarma, and P. R. Sievert, Eds. The Wildlife Society. Bethesda, MD.