USGS
science for a changing world

Prepared in cooperation with Bureau of Land Management and the
National Renewable Energy Laboratory

GenEst User Guide—Software for a Generalized Estimator of Mortality

Chapter 19 of
Section C, Computer Programs
Book 7, Automated Data Processing and Computations

Techniques and Methods 7-C19

U.S. Department of the Interior
U.S. Geological Survey

GenEst User Guide—Software for a Generalized Estimator of Mortality

By Juniper Simonis, Daniel Dalthorp, Manuela Huso, Jeffrey Mintz, Lisa Madsen, Paul Rabie, and Jared Studyvin

Chapter 19 of
Section C, Computer Programs
Book 7, Automated Data Processing and Computations

Prepared in cooperation with Bureau of Land Management and the National Renewable Energy Laboratory

Techniques and Methods 7-C19

U.S. Department of the Interior
U.S. Geological Survey

U.S. Department of the Interior
RYAN K. ZINKE, Secretary

U.S. Geological Survey
James F. Reilly II, Director

U.S. Geological Survey, Reston, Virginia: 2018

For more information on the USGS—the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment—visit https://www.usgs.gov/ or call 1–888–ASK–USGS (1–888–275–8747).

For an overview of USGS information products, including maps, imagery, and publications, visit https:/store.usgs.gov.

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Although this information product, for the most part, is in the public domain, it also may contain copyrighted materials as noted in the text. Permission to reproduce copyrighted items must be secured from the copyright owner.

Suggested citation:
Simonis, J., Dalthorp, D., Huso, M., Mintz, J., Madsen, L., Rabie, P., and Studyvin, J., 2018, GenEst user guide—Software for a generalized estimator of mortality: U.S. Geological Survey Techniques and Methods, book 7, chap. C19, 72 p., https://doi.org/10.3133/tm7C19.

ISSN 2328-7055 (online)

Contents

Abstract 1
1. Introduction 2
1.1. Description of Product 2
1.1.1. Software Modules 2
1.1.2. Comparison to Other Methods 3
1.2. Installation Instructions 3
1.2.1. Requirements 3
1.2.2. Installation 4
1.2.2.1. R Software, Version 3.5.0 or Higher 4
1.2.2.2. GenEst Package 4
1.3. Getting Started 5
1.4. Working with `shiny` 5
1.5. Layout 6
2. Data Input 7
2.1. Data Files 7
3. Analyses 10
3.1. General Inputs 10
3.2. Searcher Efficiency 11
3.2.1. Model Inputs 11
3.2.2. Model Comparison 13
3.2.2.1. Selected Data 13
3.2.2.2. Figures 14
3.2.2.3. Estimates 15
3.2.2.4. Model Selection 16
3.3. Carcass Persistence 17
3.3.1. Model Inputs 17
3.3.2. Table & Figure Selection and Model Selection 18
3.3.2.1. Selected Data 18
3.3.2.2. Model Comparison 18
3.3.2.3. Figures 20
3.3.2.4. Estimates 20
3.3.2.5. Model Selection 21
3.4. Mortality Estimation 23
3.4.1. Model Inputs 23
3.4.1.1. Splitting Mortality Estimates 25
3.4.1.2. Scaling Mortality Estimates 27
3.5. Detection Probability 28
3.5.1. Model Input for Estimating Generic Detection Probabilities 28
3.5.1.1. Search Schedule 28
3.5.2. Table and Figure Selection 30
3.5.2.1. Summary 30
3.5.2.2. Figure 31

4. Worked Examples 32
4.1. Example 1—Wind Facility: Bats, Road and Pad Surveys of All Turbines 32
4.1.1. Data Loading 32
4.1.2. General Inputs 33
4.1.3. Searcher Efficiency 34
4.1.4. Carcass Persistence 38
4.1.5. Mortality Estimation 42
4.2. Example 2—Solar Photovoltaic Facility, All Birds 48
4.2.1. Data Loading 48
4.2.2. Modeling Search Efficiency 48
4.2.3. Modeling Carcass Persistence 51
4.2.4. Mortality Estimation 57
4.2.5. Detection Probability 60
Acknowledgments 63
About the Authors and Collaborators 63
References Cited 64
5. Appendix 66
5.1. Akaike Information Criterion (AIC) and Model Selection 66
5.2. Number Format and .csv Files 66
5.3. Dates 66
5.4. File Names 66
5.5. Variables and Column Names 67
5.6. Some Terminology Used in Searcher Efficiency and Carcass Persistence Modeling 67
5.7. Searcher Efficiency 68
5.8. Carcass Persistence 70
5.9. Search Schedule 71
5.10. Density-Weighted Proportion (DWP) 71
5.11. Carcass Observations 72

Figures

Figure 1. "Getting Started" in GenEst 6
Figure 2. Downloading data 7
Figure 3. Data Input tab with preview of the uploaded Searcher Efficiency data 9
Figure 4. Required and optional general inputs 10
Figure 5. Searcher efficiency model inputs 12
Figure 6. Options to explore results of searcher efficiency model fitting 13
Figure 7. Example plots of estimated and observed searcher efficiency by search event 14
Figure 8. Estimates and confidence limits of p and k for each cell 15
Figure 9. Model selection screen indicating which searcher efficiency model to use for each size class of trial carcasses 16
Figure 10. Carcass persistence model inputs 17
Figure 11. Options to explore results of carcass persistence model fitting 19
Figure 12. Example plots of estimated and observed carcass persistence by time for each combination of levels of predictor variables 20

Figure 13. Estimates and confidence limits of location and scale parameters for each cell for selected model and size class 21
Figure 14. Model selection screen indicating which carcass persistence model to use for each size class of trial carcasses 22
Figure 15. Model inputs screen requesting information on sampling fraction and identifying the column in the carcass observations file that represents the date on which the carcass was found by searchers 24
Figure 16. Estimated total mortality at the site based on user input 25
Figure 17. Estimated mortality at the site in each combination of size and ground cover 26
Figure 18. Estimated mortality at the site in each combination of size and ground cover 27
Figure 19. Model inputs screen requesting information on search schedule 29
Figure 20. Summary of detection probability and confidence limits for relevant covariate classes for the selected size class 30
Figure 21. Boxplot of estimated detection probability for relevant covariate classes for the selected size class 31
Figure 22. "Data Input" tab with "Carcass Observations" displayed after uploading 32
Figure 23. Starting the analysis with "General Inputs". 33
Figure 24. Model inputs for searcher efficiency analysis 34
Figure 25. Searcher efficiency model comparison 35
Figure 26. Searcher efficiency plots 36
Figure 28. Carcass persistence model building 38
Figure 29. Carcass persistence model comparison tab 39
Figure 30. Fitted carcass persistence models 40
Figure 31. Carcass persistence model selection 41
Figure 32. Mortality estimation tab 42
Figure 33. Total mortality and options for mortality splits 43
Figure 34. Mortality estimates by season 44
Figure 35. Mortality estimates by species 45
Figure 36. Seasonal mortality by species 46
Figure 37. Bat mortality estimates by species and season (transpose of season by species) 47
Figure 38. Data for searcher efficiency analysis 48
Figure 39. Searcher efficiency model comparison 49
Figure 40. Diagnostic searcher efficiency plots for large carcasses at a solar PV project 50
Figure 41. Selecting the SE models for each size class for the solar PV example 51
Figure 42. Fitting carcass persistence models for solar PV dataset 52
Figure 43. Comparison of carcass persistence models for solar PV example 53
Figure 44. Fitted carcass persistence distributions for the $\mathrm{l} \sim \mathrm{constant}; \mathrm{s} \sim \mathrm{constant}$ models 54
Figure 45. Fitted carcass persistence distributions for the $\mathrm{l} \sim \mathrm{Season}; \mathrm{s} \sim \mathrm{constant}$ models 55
Figure 46. Selection of carcass persistence models for solar PV example 56
Figure 47. Total mortality in the solar PV example 57
Figure 48. Mortality by species for solar PV example 58
Figure 49. Mortality by species and season for the solar PV example 59
Figure 50. Summary statistics for all mortality estimates are available in tabular format under the "Summary" tab 60
Figure 51. Average search schedule for solar PV example data 61
Figure 52. Detection probability by season for small birds (solar PV example) 62

Tables

Table 1. Data needed for each analysis 8
Table 2. Format for searcher efficiency data 69
Table 3. Carcass Persistence Data Format 70
Table 4. Search Schedule Format 71
Table 5. DWP data example 72

Abbreviations

AICc	Akaike information criterion corrected for finite sample size
ΔAICc	difference between AICc value for a given model and the lowest AICc value among models tested
.csv	comma-separated values (format for some input files)
CI	confidence interval
CO	Carcass Observations (usually referring to the data file with results from carcass surveys)
CP	Carcass Persistence (often referring to the data file with results from carcass persistence field trials)
CRAN	Comprehensive R Archive Network (for downloading R statistical software and supplemental packages)
IQR	Inter-quartile range = 25^{th} and 75^{th} quantiles of a dataset
DWP	Density Weighted Proportion; proportion of total mortality expected to fall within searched areas (often referring to the data file with DWPs for each size class at each surveyed unit)
GUI	Graphical User Interface, referring to the system of menus, tabs, textboxes, and the like for managing data, calling R functions, and showing results
SE	Searcher Efficiency (often referring to the data file with results from carcass persistence field trials)
SS	Search Schedule (often referring to the data file that lists the search schedules followed at each unit, with optional covariate columns)

Variables and Parameters

α	significance level
$1-\alpha$	confidence level
$\mathbf{A}$	$X \times nsim$ matrix of simulated arrival intervals
dwp	density weighted proportion; fraction of the total carcasses that arrive in a searched area
f	sampling fraction; proportion of potential sampling units that were searched
g	overall detection probability within the searched area, within the period of study
I	search interval; number of days between searches of the same area/unit
I_r	assumed search interval for reporting an estimate of r
k	factor by which searcher efficiency decreases with each successive search
M	mortality or number of fatalities
$\widehat{M}$	estimated mortality or number of fatalities
p	probability of observing a carcass in the first search after carcass arrival (assuming the carcass is present at the time of search)
r	probability that a carcass persists until the first search after arrival
X	number of carcasses observed in searches

This page left intentionally blank

GenEst User Guide—Software for a Generalized Estimator of Mortality

By Juniper Simonis[1], Daniel Dalthorp[2], Manuela Huso[2], Jeffrey Mintz[2], Lisa Madsen[3], Paul Rabie[4], and Jared Studyvin[4]

Abstract

GenEst (Generalized Estimator) is a software tool for estimating the total number of individuals arriving in an area during a specific time period when their detection probability is unknown but estimable. Its development was motivated by the need to accurately estimate the total number of bird and bat fatalities occurring at wind and solar energy facilities, but it is applicable in various other contexts as well. Simple counts of carcasses are not an accurate measure of the true number of fatalities because some carcasses are inevitably missed in carcass searches. Furthermore, simple carcass counts do not allow comparison among locations or years because carcasses may be detected at different rates. This software uses data collected during carcass searches and estimates of detection rates to accurately estimate the number of fatalities and to provide a measure of precision associated with the estimate. These estimates are fundamental to understanding acute and cumulative effects of renewable energy development on wildlife populations. The software package is available with a user-friendly graphic interface as well as a flexible and powerful command-line implementation. GenEst includes tools for estimating searcher efficiency, carcass persistence, and other detection probability parameters from experimental field trials. The software includes example datasets for analyses, standard R package help files, this user guide, and vignettes detailing use at the command-line.

[1]DAPPER Stats.
[2]U.S. Geological Survey.
[3]Oregon State University.
[4]Western EcoSystems Technology, Inc.

1. Introduction

1.1. Description of Product

GenEst (Generalized Estimator) is a software package designed for use by anyone analyzing data associated with estimating bird or bat fatalities at renewable-energy facilities, such as wind and solar facilities, but it has applicability in many other situations, as well. It is designed to addresses the general problem of estimating the size of a population when not all animals are present on all survey occasions—or a *superpopulation* (Williams and others, 2011)—when the probability of detection is generally less than one. The population is not closed, and the probability of detection can vary among individuals due to physical characteristics of the individual, such as size, or on the environmental conditions in which the individual exists, such as vegetation or season. To estimate the number of fatalities, carcasses are usually collected during distinct searches repeated at (generally) constant intervals through time and counts of carcasses are adjusted for imperfect detection. Imperfect detection may be due to any of several possible detection biases, for example—(1) search teams fail to find carcasses that are present in the searched area at the time of the search, (2) scavengers remove carcasses before searches are conducted, (3) carcasses fall outside the searched area, or (4) fatalities occur outside the monitored period. In parallel with the search process, investigators typically conduct field trials to estimate the effects of the first two components, and use observed locations of carcasses as well as knowledge regarding the sampling fraction to estimate the third. The fourth is often a matter of educated guess. Accurate estimation of the detection biases is critical to accurate estimation of total mortality. Because of imperfect detection, the simple count of observed carcasses does not accurately represent the actual population of animals killed by turbines and cannot be used as an index of mortality because it strongly depends on the details of the search protocol, which can vary markedly from site to site and from year to year. GenEst includes several modules that use data provided by the user to estimate searcher efficiency, carcass persistence, and detection probability parameters as well as estimates of total mortality for groupings of carcasses of interest to the user.

GenEst is written in R (R Core Team, 2018) and C++ and is available as a fully documented R package, including both a user-friendly Graphic User Interface (GUI) coded using `shiny` (Chang and others, 2018) and command-line interface for savvy R users who wish to have more flexibility and power to do custom analyses that cannot be done in the GUI. This User Guide focuses on the GUI version of GenEst. Vignettes within the R package detail the command-line interface and can be accessed by entering `browseVignettes("GenEst")` in R.

1.1.1. Software Modules

GenEst software includes three primary modules (or tabs), each with submodules:

1. Data input—Upload the data input files (Searcher Efficiency, Carcass Persistence, Search Schedule, Density Weighted Proportion, and Observed Fatalities), all in .csv format.
2. Analyses—Determine the best searcher efficiency and carcass persistence models and combines input data from several search classes and/or time periods into an estimate of total mortality.
3. Help—Find help, download files, and read disclaimers.

1.1.2. Comparison to Other Methods

There are currently several estimators in use that were created specifically to estimate mortality at wind power facilities (Shoenfeld, 2004; Huso and others, 2012; Peron and others, 2013; Korner-Nievergelt and others, 2015; Wolpert, 2015; Dalthorp and others, 2017). GenEst's accuracy, flexibility, and ease of use are unrivaled. The primary differences between GenEst and other estimators are that GenEst is in most respects much more flexible than any of the others and provides more accurate estimates with respect to bias and coverage probabilities for confidence intervals because fewer constraining assumptions are made and more accurate techniques for incorporating estimator uncertainties are used. It is a semi-periodic fatality estimator, closely related to Wolpert's (2015) estimator, and accounts for estimation uncertainties in a novel way that allows for:

1. straightforward analysis of complex datasets that may include multiple carcass size classes, detection probabilities that depend on environmental covariates (for example, visibility, season, search team, carcass type, area), variable search schedules, search coverage that varies with search unit (for example, size of search plot may vary by turbine), and
2. easy analysis and summary of results by user-defined subcategories such as species, unit, time interval, species group (raptor, passerine, waterfowl, migratory bat, hibernating bats), turbine type, etc.

GenEst provides flexibility in modeling carcass persistence times and can accommodate a decrease in searcher efficiency with repeated searches for the same carcass. It allows users to model searcher efficiency and carcass persistence as functions of environmental covariates and to compare models using Akaike information criterion (AICc) and graphs.

1.2. Installation Instructions

1.2.1. Requirements

R software, version 3.5.0 or later

R packages:

`corpus`	Perry (2017)
`cbinom`	Dalthorp)2018)
`DT`	Xie (2018)
`Gsl`	Hankin (2006)
`Gtools`	Warnes and others (2015)
`htmltools`	RStudio, Inc. (2017)
`lubridate`	Grolemund and Wickham (2011)
`matrixStats`	Bengtsson (2018)
`mvtnorm`	Genz and others (2018)
`Rcpp`	Eddelbuettel and Francois (2011)
`shiny`	Chang and others (2018)
`shinyjs`	Attali (2018)
`sticky`	Brown (2017)

1.2.2. Installation

There are several steps to installing the package and getting up and running the first time. Do not skip any steps.

1.2.2.1. R Software, Version 3.5.0 or Higher

GenEst requires R statistical computing software (R Core Team, 2018), which is free and open source. With version 3.5.0, R introduced deep changes in internal structure, and R version 3.5.0 or later (released in April 2018) is required for GenEst. To install, use your Internet browser to navigate to https://www.r-project.org/ and click following links: "CRAN mirror" (choose the site nearest to your location) and "Download R for ...." If you are installing on a computer with Microsoft Windows®, click "base", and "Download R x.x.x for Windows". You may need to be logged in with administrative privileges to install R. If so, be sure to log back in using your standard working credentials before continuing. If installing on a computer with Mac® or a UNIX®-like operating system, follow the instructions provided at the https://www.r-project.org/ website.

R packages that have been downloaded and installed with an earlier version of R will not be available in the updated version unless they are either reinstalled or copied from the library folder of the previous installation to the library folder of the new installation and updated in the new version of R using `update.packages(checkBuilt = TRUE, ask = FALSE)`.

1.2.2.2. GenEst Package

Make sure you are logged into your computer in the way you normally work. For most corporate and government users, that will be without administrative privileges. GenEst software is bundled as an R package, and official copies of each version of GenEst are stored at https://code.usgs.gov/ecosystems/GenEst. To install, you will first need to install several, required, third-party packages. All are free and open source and available from CRAN. The easiest way to install them is to run the following commands in R (with guidance concerning potential dialog boxes given below the commands):

```
package_new <- c("corpus", "cbinom", "DT", "gsl", "gtools", "htmltools", "lubridate",
"matrixStats", "mvtnorm", "Rcpp", "shiny", "shinyjs", "sticky", "survival")

package_new <- package_new[!(package_new %in% installed.packages()[,"Package"])]
if(length(package_new) > 0) install.packages(package_new)
```

- If asked about a "CRAN mirror", choose the nearest location.
- If asked whether you want to use a "personal library", choose "Yes"
- If you are on Windows and are asked whether you want to install packages and their dependencies "from source", choose "No" (unless you are ready to go to lunch, in which case, you can select "Yes" and the installation may well be done by the time you get back).

To install GenEst itself, navigate to https://code.usgs.gov/ecosystems/GenEst/tags and click the link to the latest version, 1.x.x. Download the software (GenEst_1.x.x.zip for Windows or GenEst_1.x.x.tar.gz for other operating systems) and do not unzip. Install the package from the local copy. If you are working from the simple R Gui, select "Install package(s) from local files..." from the "Packages" menu. If you are working in R Studio, select "Packages" from the "Tools" menu, and then in the "Install from: " drop-down menu, select "Package Archive File (.zip, .tar.gz)" and browse for the downloaded GenEst package.

1.3. Getting Started

Initializing the GenEst GUI is as simple as running the following commands in R:

```
library(GenEst)
runGenEst()
```

This will launch the application in your default web browser.

1.4. Working with `shiny`

GenEst is an R package that uses an add-on package called `shiny` for a user-friendly graphical interface to the underlying R functions for data management and analysis. GenEst's system of menus, tabs, buttons, and data input forms are convenient and intuitive. Knowing a few, simple `shiny` conventions will smooth the process of learning to use GenEst.

- ***Takeover of the R session***. Opening the application ties up the active R session. As a result, if you would like to use an active R session alongside the GUI, you will need to initiate a second, independent R session. GenEst does not automatically save data and results to a working directory in R, but, if desired, results can be downloaded and saved or copied into the clipboard and pasted into other programs for further analysis, writing reports, or storage.
- ***Gray screen of death***. GenEst does have extensive error-checking, but not every error is caught. When `shiny` encounters an unexpected error, the browser window turns gray, indicating that the link between R and the GUI has been broken and cannot be recovered and marking the end of that session. Users may still navigate among tabs in the application and view current data and models, although changes to the data and analyses can no longer be made. Control can be restored to the R command-line by clicking on the R window and pressing the Esc key. After pressing Esc, notes about where and how the crash occurred will be displayed. These notes will be cryptic to most users but may be invaluable for helping the developers track down the source of the error. Users may then begin a new GenEst `shiny` session by entering `runGenEst()`.
- ***Drop-down menus.*** Users upload data files to GenEst and then use drop-down menus, text boxes, check boxes, and action buttons to communicate to GenEst about the data structure and desired analyses. GenEst has fairly sophisticated checks on data formatting and will not offer users the choice to use improperly formatted data in the drop-down menus. For example, in a drop-down menu asking users to identify which column contains carcass discovery dates, the only choices offered will be those that are properly formatted as dates (see 5. Appendix). Users then select the desired column(s). Where applicable, users may select more than one item from a given drop-down menu by clicking the choices in succession. For example, in modeling searcher efficiency, a user may first select season and then visibility as two covariates. Selections can be erased by clicking in the textbox and pressing Backspace key.
- ***Updating of calculations***. When a user makes changes in one part of an analysis, associated calculations are automatically updated. For example, if results for one searcher efficiency model are displayed in the application and a user selects a different searcher efficiency model to consider, the searcher efficiency graphs and tables are automatically updated to match the user's interest. It should be noted, though, that "downstream" results (in particular, estimates of mortality and detection probabilities) are erased rather than automatically updated when users select new datasets or new searcher efficiency or carcass persistence models.

1.5. Layout

The GUI is laid out using tabs: "Data Input", "Analyses", and "Help". The application opens to a "Getting Started" page under the "Help" tab (fig. 1), which describes the general layout and work flow for the software. The Data Input and Analyses tabs allow active user interaction for analyses.

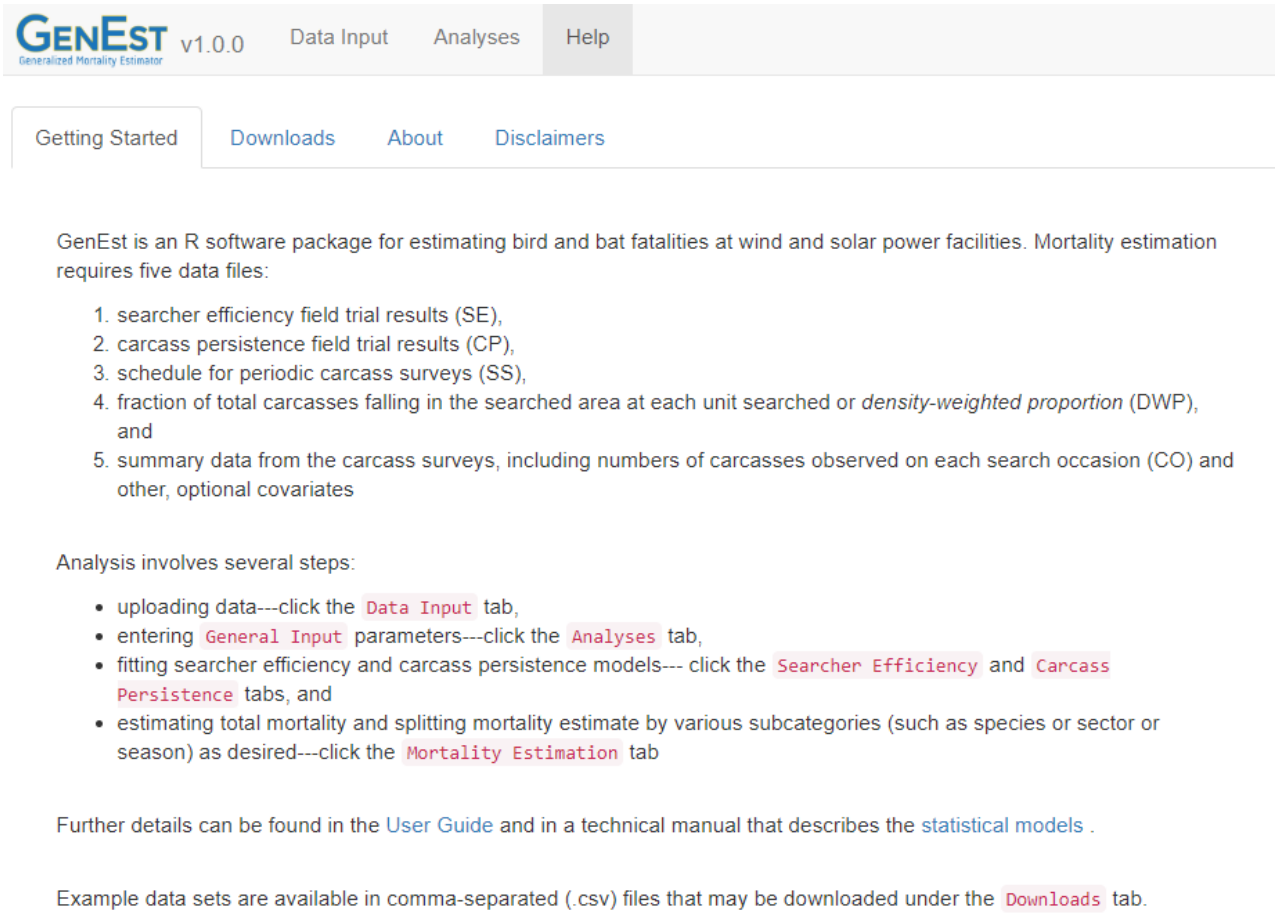

Figure 1. "Getting Started" in GenEst.

2. Data Input

2.1. Data Files

Several example datasets that can be used within the GUI are provided in the R package. The example datasets may be downloaded under the "Downloads" tab. The datasets can be examined, edited, used as templates for formatting new datasets, or uploaded to the appropriate pages in the GUI for analysis. Begin by downloading the "mock" dataset (fig. 2), which is a full, properly formatted dataset. The data are largely random and have no particular significance, but the example is useful for a quick introduction to how the software works.

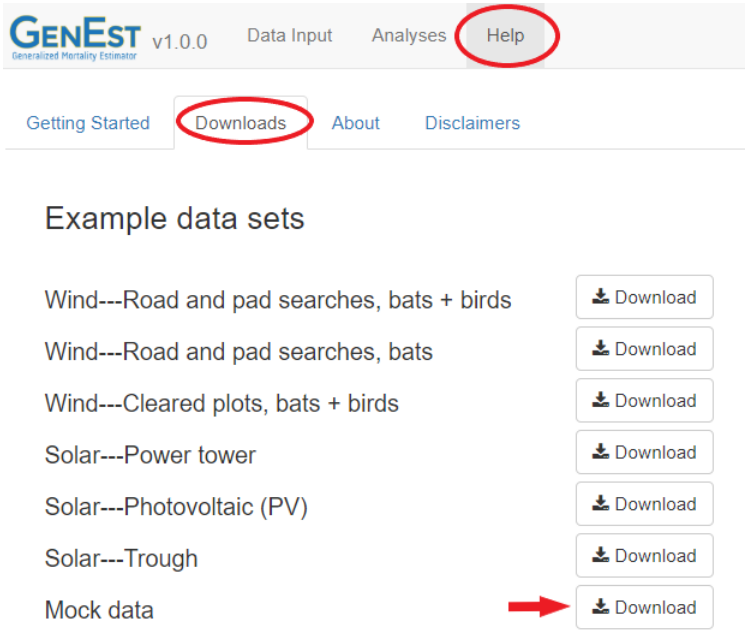

Figure 2. Downloading data.

Mortality estimation in GenEst requires five data files—Searcher Efficiency (SE), Carcass Persistence (CP), Search Schedule (SS), Density-Weighted Proportion (DWP), and Carcass Observations (CO). The files should be formatted with fields separated by commas, either as a comma-separated value (.csv) or a comma-delimited regular text (.txt) file (as is common in North America, Australia, South Africa, and most of Asia), or, alternatively, fields may be separated by semi-colons (;), as is common in Europe and South America. Appendix (section 5) details the necessary and available formats for each of the files, and example data files included in the software can serve as templates for formatting. The files that are required for an analysis depend on a user's specific desires (table 1). Upon upload, each data file becomes immediately available for inspection (fig. 3) and is available for analysis throughout the GUI.

Table 1. Data needed for each analysis.

Analysis	Searcher Efficiency	Carcass Persistence	Search Schedule	Density-Weighted Proportion	Carcass Observations
Search Efficiency	x				
Carcass Persistence		x			
Detection Probability	x	x	x		
Mortality Estimation	x	x	x	x	x

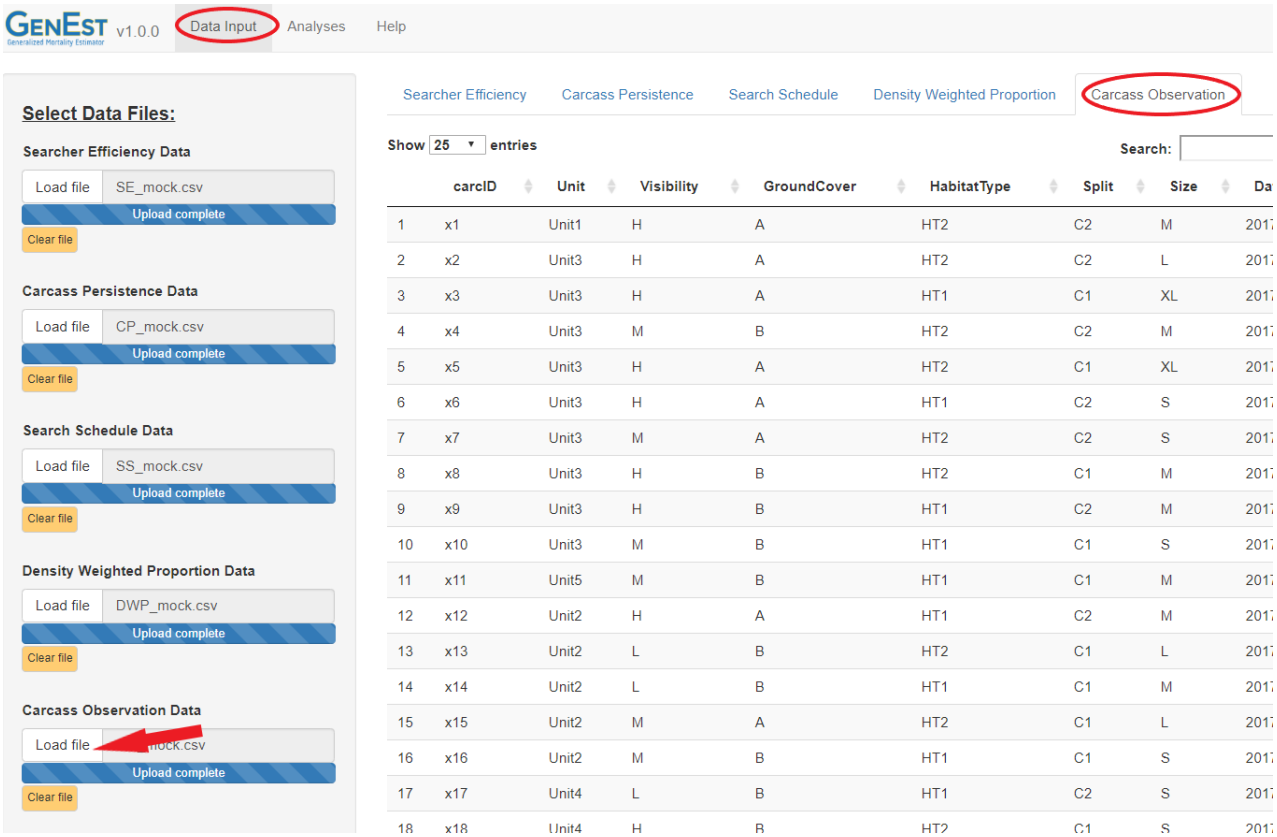

Figure 3. Data Input tab with preview of the uploaded Searcher Efficiency data.

3. Analyses

Following upload of the data files, the user should navigate to the "Analyses" tab.

3.1. General Inputs

Under the "General Inputs" subtab, the user provides two values: "Number of iterations" and "Confidence Level" (fig. 4) and may supply a third, optional value which is the name of the carcass size class column in the datasets. "Number of iterations" controls the number of parametric bootstrap iterations used to build confidence intervals for parameters. Accuracy improves asymptotically as the number of iterations increases, but the run time also increases. The number of iterations must be an integer and defaults to 1000. Values up to 10000 may be used. The "Confidence Level" defines the probability level used to describe the spread of distributions (for parameter estimates and state variables). The default is 0.90 (or 90%), but users may override this and use any value between 0 and 1. In addition, the user may provide a "Size Class Column (optional)." In the `mock` dataset, the size class column is called `Size`, which we select for this analysis. Detection probabilities often depend strongly on carcass size, and GenEst fits separate searcher efficiency and carcass persistence models for distinct size classes if a size class column is provided. Sizes are entered into the input files as categorical variables like `small` and `large`. The software offers as options for size class only those columns that occur in all three of Searcher Efficiency, Carcass Persistence, and Carcass Observations input files. [Note that sizes and other covariate values are case-sensitive. For example, `Small`, `small`, and `SMALL` are treated as different sizes by GenEst.]

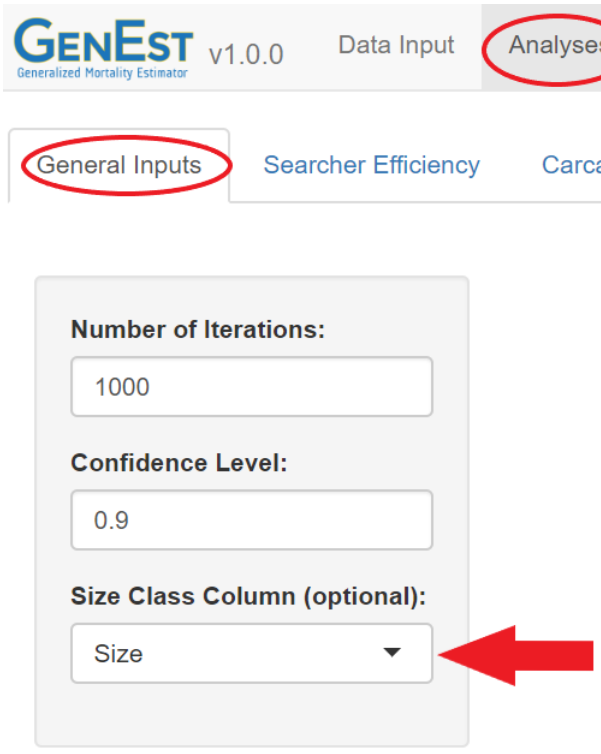

Figure 4. Required and optional general inputs.

3.2. Searcher Efficiency

Searcher efficiency is the probability that a searcher will observe a carcass that is present in the searched area at the time of the search. Searcher efficiency is typically estimated through field trials. Marked carcasses are placed in the field for possible discovery by search teams during routine carcass surveys. Trial carcasses, representing as closely as possible the species composition of turbine-caused fatalities at the facility, are placed within the searched area at various times during the monitoring period at locations known only to the trial conductor and not the searchers. After each search, if a trial carcass is not reported as found by the searchers, the trial conductor will visit the site to determine whether the carcass was present but truly missed or no longer present and hence not eligible for inclusion in the searcher efficiency trial data. If trial conductors allow missed carcasses to remain in the field for potential discovery on a later search, the data may provide information about how searcher efficiency for an individual carcass changes over time. Searcher efficiency typically decreases as carcasses age and as the relatively easy-to-find carcasses are preferentially removed in the first searches after carcass arrival. Searcher efficiency often varies with characteristics of the individual carcasses or conditions of its search, for example, size of the carcass, carcass coloration, vegetation density, season, or individual searcher.

GenEst uses single- or multiple-search carcass detection surveys to estimate searcher efficiency. Searcher efficiency is modeled with two parameters: $p =$ the probability that a carcass that is present is found during the first search after it arrived; and $k =$ proportional change in searcher efficiency with each successive search. A value of $k = 0$ asserts that carcasses that are missed on the first search after carcass arrival cannot be discovered in a later search, and $k = 1$ means that searcher efficiency remains constant regardless of carcass age and the number of times a carcass has been missed in previous searches. GenEst does not allow k to exceed 1, which asserts that searcher efficiency does not increase with carcass age.

Searcher efficiency parameters can be estimated for carcasses grouped into size classes and any combination of categorical variables (for example, season and vegetation class) that occur as columns in the searcher efficiency data that is input by the user. The searcher efficiency parameters are estimated from the survey data in each "cell" (that is, each combination of covariate levels and size class) using maximum likelihood methods, as implemented with custom-written functions for the GenEst package. For full details, see *GenEst Statistical Models* (Dalthorp and others, 2018). Users have the option to allow GenEst to estimate k from the data or to fix k at any acceptable level ($k \in [0, 1]$). If data from only the first attempt to find a trial carcass is recorded, k cannot be estimated and must be set by the user. In addition, if no trial carcasses are discovered after being missed in the first search, k cannot be estimated and users must enter a fixed k.

3.2.1. Model Inputs

Searcher efficiency model inputs are selected on the left-hand side panel of the "Searcher Efficiency" tab, "Selected Data" sub-tab (fig. 5).

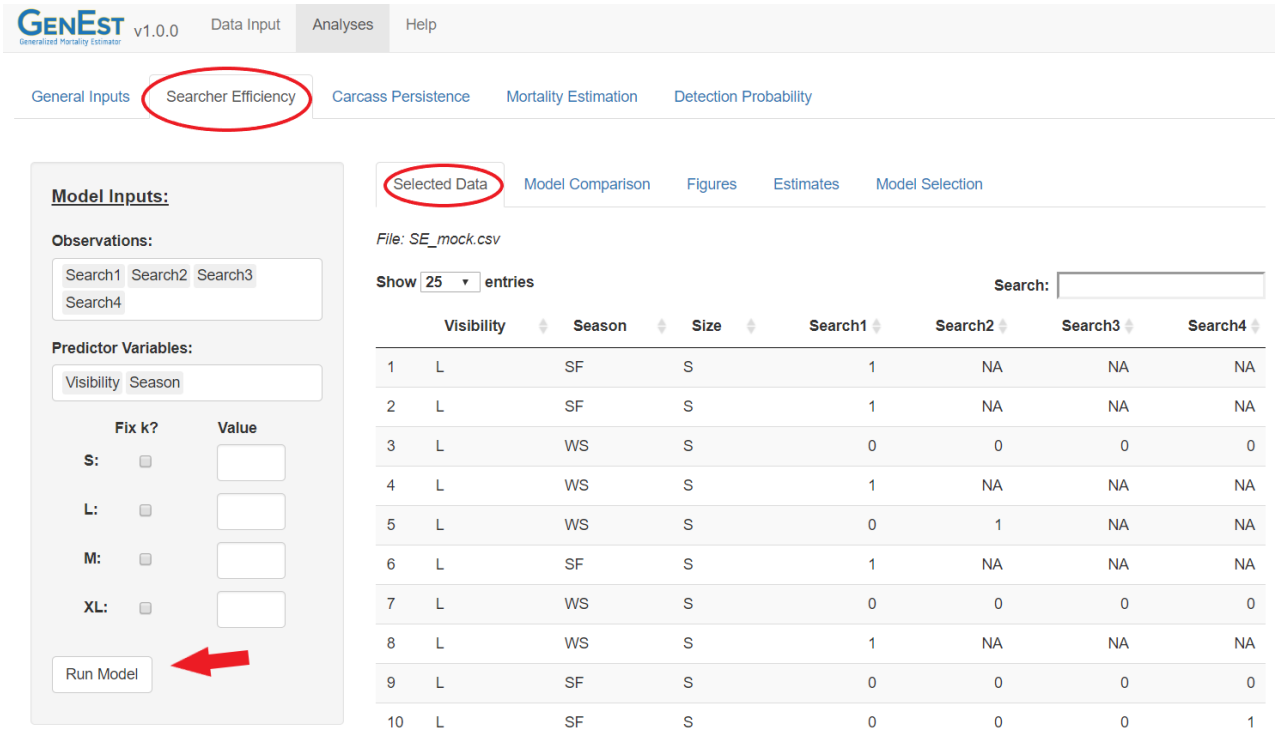

Figure 5. Searcher efficiency model inputs.

- "Observations:"
  - The user is required to select the columns in the input data set that correspond to the search efficiency trial results for one or more searches after carcasses were placed (1 = found, 0 = missed, NA = not searched or carcass no longer available).
  - The columns must be selected in order of search occasion.
  - At least one column must be selected.
- "Predictor Variables:"
  - This input is optional.
  - Users may select up to two categorical variables that they believe might affect searcher efficiency and are represented in the field trial data
  - If the searcher efficiency model is to be used in mortality estimation (as opposed to a stand-alone analysis of searcher efficiency) the selected predictor columns must also be included in the carcass observations data.
  - The predictors selected in this analysis can be, but do not need to be, used in the carcass persistence analysis.
- "Fix k?"
  - If k cannot be estimated from the data (either because carcasses were only available for one search or because the data are otherwise insufficient), the user must enter by hand a fixed value for k between 0 and 1 (inclusive) if the searcher efficiency model is to be used for mortality estimation. If the data do allow estimation of k (that is, data from repeated attempts to find the same carcasses are included in the SE file), the user has the option of estimating k or fixing k at a particular value for all predictor variable and size combinations.

3.2.2. Model Comparison

Once all input data are selected, the user clicks on “Run Model.” GenEst fits all combinations of fittable models for both p and k that involve the user-selected predictors. For example, if user selects A and B as predictors, then GenEst will attempt to fit the constant, A, B, $A + B$, and $A * B$ models for both p and k (see appendix 5.6 for discussion of model forms). For some datasets, not all models will be fittable. For example, if covariates include `Visibility` and `Season` but there is no trial data available for, say, the easy visibility class in spring, then the `Visibility * Season` model cannot be fit and will not be included in the list of fitted models.

When the model-fitting calculations are complete, the “Model Comparison” tab is automatically opened (fig. 6), showing a table of models sorted by AICc scores (appendix 5.1). If a size class column was entered under "General Inputs", by default the model comparison table lists model fit for the first size class encountered in the SE file. The models for other size classes can be viewed by selecting the desired size from the "Size Class" drop-down menu on the left sidebar. Several output sub-tabs for evaluating model fit are available: "Selected Data", "Model Comparison", "Figures", "Estimates", and "Model Selection".

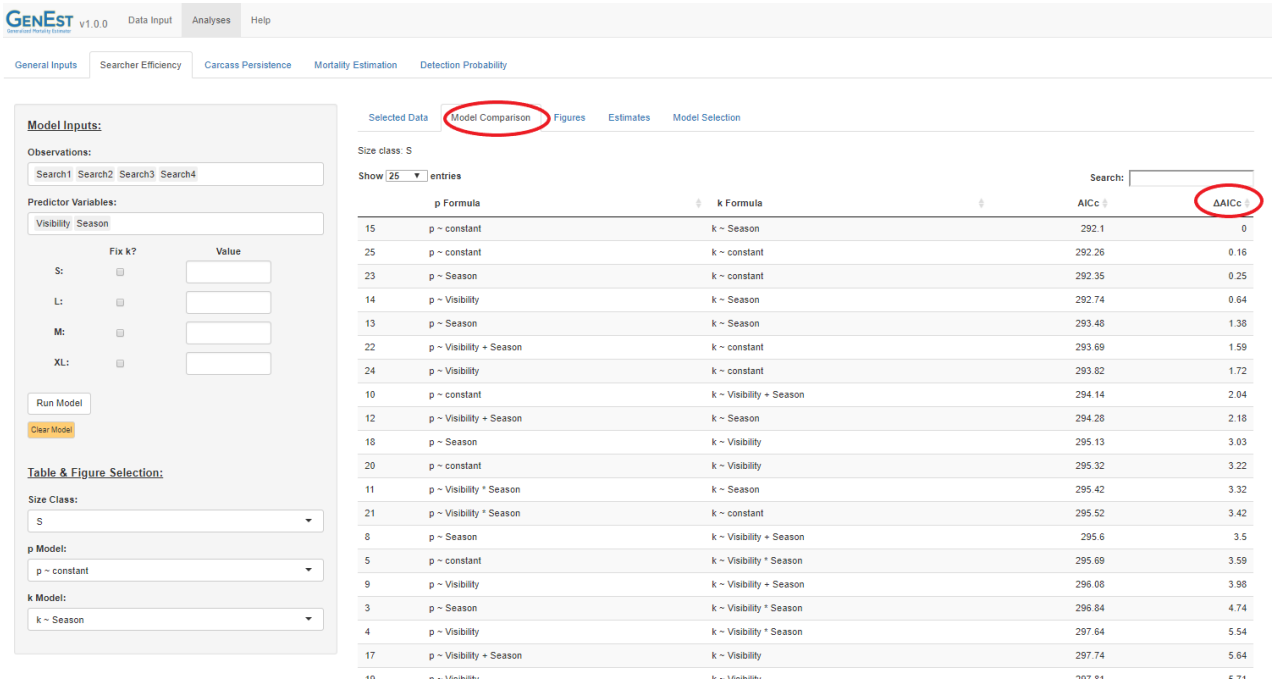

Figure 6. Options to explore results of searcher efficiency model fitting.

3.2.2.1. Selected Data

Data used to fit the searcher efficiency models for the selected size class are displayed under the “Selected Data”, which shows the first 10 rows of the .csv file that was uploaded earlier in the Data Input process. Additional rows may be viewed by following the page navigation links under the lower right corner of the data table.

3.2.2.2. Figures

Searcher efficiency estimates from the selected model are plotted against search occasion for each possible combination of class levels of the input predictor variables (fig. 7). The initial set of figures is for the model with the lowest AICc score (appendix 5.1) for the size class that occurs first in the SE data (if a size class column was entered under “General Inputs”). To produce figures for other size classes and other models, select the desired elements from the drop-down menus in the lower left section of the "Figures" window (fig. 7). The user-selected model is plotted along with a reference model, which is the model that provides the closest fit to the data as measured by maximum likelihood. Although the reference model provides the closest fit to the data, it typically has large error bars associated with it and is often overfit and not the best or most desirable model. A better model would be one with medians that do not deviate far from those of the reference model but with less uncertainty as reflected in shorter boxes (fig. 7).

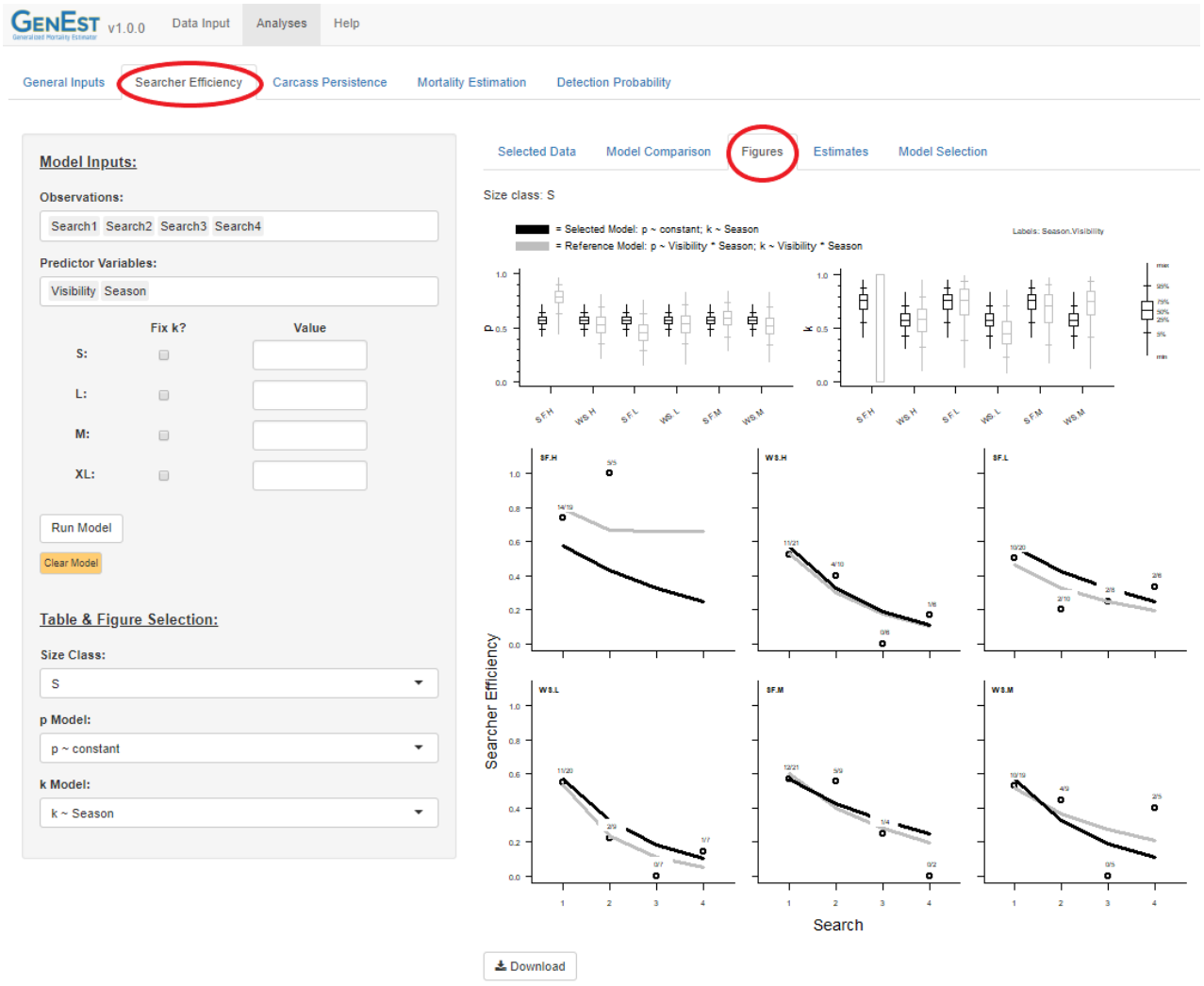

Figure 7. Example plots of estimated and observed searcher efficiency by search event. Dark lines represent selected model, gray lines represent the most complex model, that is, separate estimates for each class combination (when possible). Open circles represent observed searcher efficiency and ratios above each open circle represent observed data used to estimate the point, that is, found/available. Boxplots at the top represent estimated p and k for selected model (dark lines) compared to the reference model (gray lines).

3.2.2.3. Estimates

Estimates and confidence level (as set by user in “General Inputs”) of both p and k for each cell for the selected model and size class are given in a table (fig. 8), In addition, the median persistence time and the proportion of carcasses expected to persist 1, 3, 7, 14, or 28 days are given. For all but the exponential model, persistence probability changes with time since arrival. Because exact time of death is not known, the probability of persisting through an interval is calculated as the integral of the persistence function over the specified interval divided by the interval length. For full details, see the "GenEst Statistical Models" document linked under the "Help" tab in GenEst. For example, if the predictors are `Visibility` (`H`, `M`, `L`) and `Season` (`SF`, `WS`), the "cells" would be the combinations of predictor levels (`H.SF`, `L.SF`, `M.SF`, `H.WS`, `L.WS`, `M.WS`) with the confidence interval (CI) given for estimated p and k. The table can be saved to a .csv file by clicking the “Download” button.

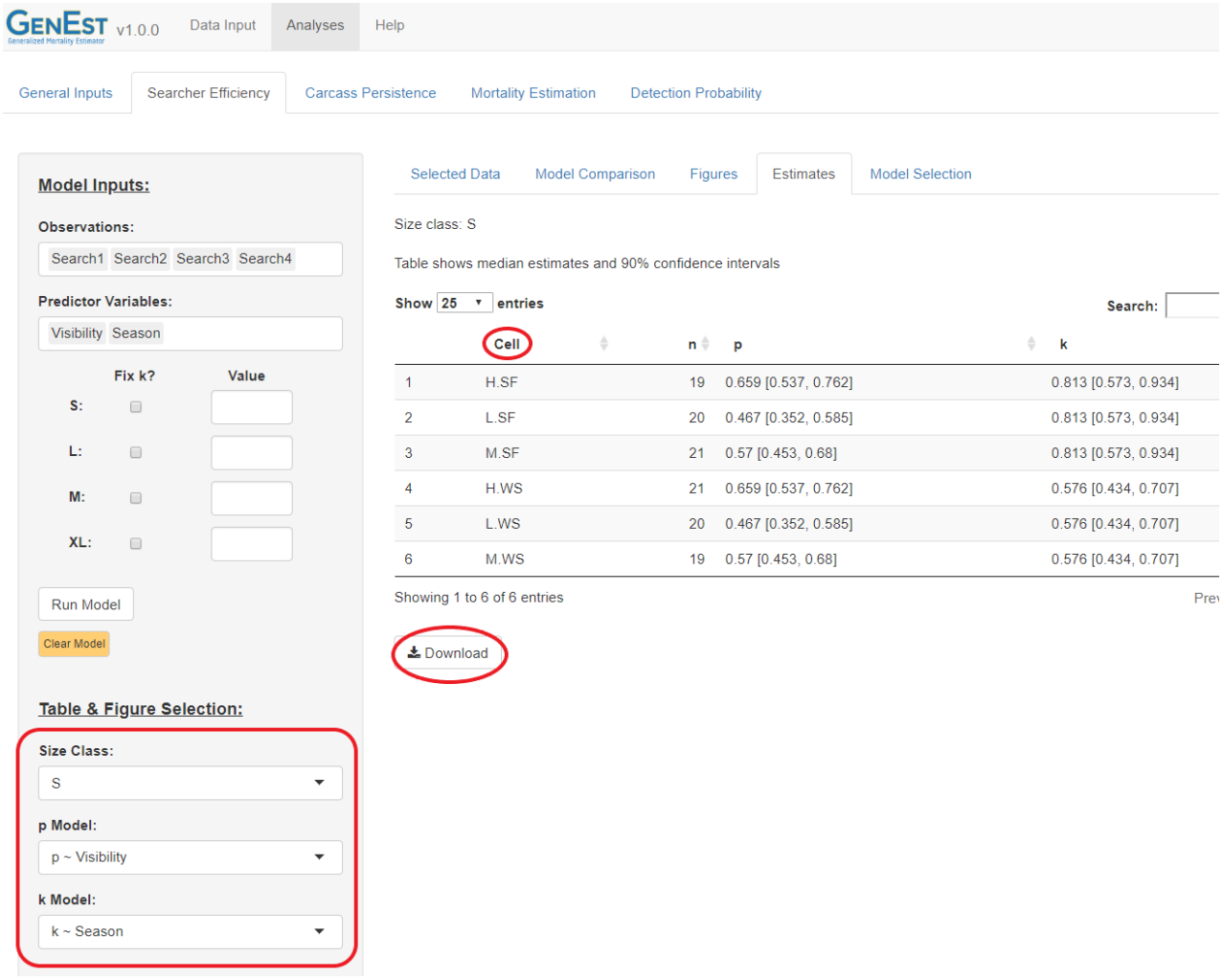

Figure 8. Estimates and confidence limits (set by user in “General Inputs”) of p and k for each cell (or combination of covariate levels) for selected model.

3.2.2.4. Model Selection

After comparing the models by AICc (appendix 5.1) under the "Model Comparison" tab and diagnostic plots ("Figures" tab), user must click on the “Model Selection” tab and explicitly choose an SE model to use in further analyses of mortality estimation or detection probability. Available models are listed in drop-down menus by ascending order of ΔAICc for each size class (fig. 9). There is no need to “Run Model” again. The selections will be held in memory for later use in mortality estimation.

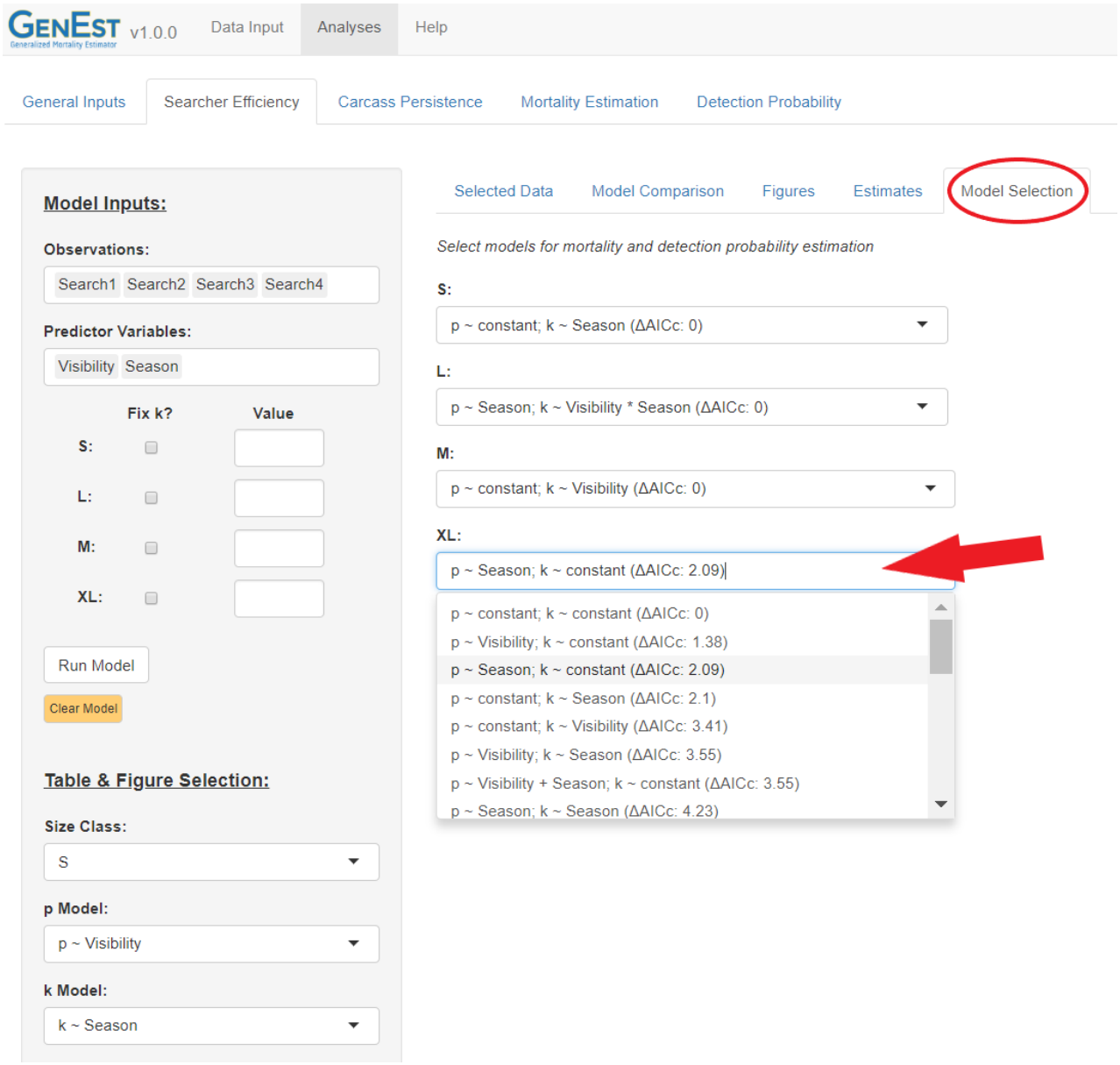

Figure 9. Model selection screen indicating which searcher efficiency model to use for each size class of trial carcasses.

3.3. Carcass Persistence

"Carcass persistence" is the probability that a carcass arriving at time 0 will continue to persist until a time *t* days later. Trial carcasses representing as closely as possible the species composition of turbine-caused fatalities at the facility are placed at various times throughout the monitored period (not necessarily immediately preceding a search) within the searched area. The exact date and time of placement are recorded. The trial conductor then revisits each carcass regularly, typically more frequently soon after placement than later, recording the date and time of visit and the presence or absence of the carcass. Carcass persistence often varies with characteristics of the individual carcasses or conditions of its search, for example, size of the carcass, season, vegetation density, etc.

GenEst allows the user to fit up to four persistence distributions to the carcass persistence data: exponential, Weibull, lognormal, and loglogistic. Carcass persistence parameters (l = location and s = scale; Kalbfleisch and Prentice, 2002) can be estimated for carcasses grouped into size classes and any combination of up to two categorical predictors (for example, season and vegetation class) that occur as columns in the carcass persistence data input by the user. The carcass persistence parameters are estimated from the trial data in predictor cell by size combination using maximum likelihood.

3.3.1. Model Inputs

Carcass persistence model inputs are selected on the left-hand side panel of the "Carcass Persistence" tab and the "Selected Data" tab (fig. 10).

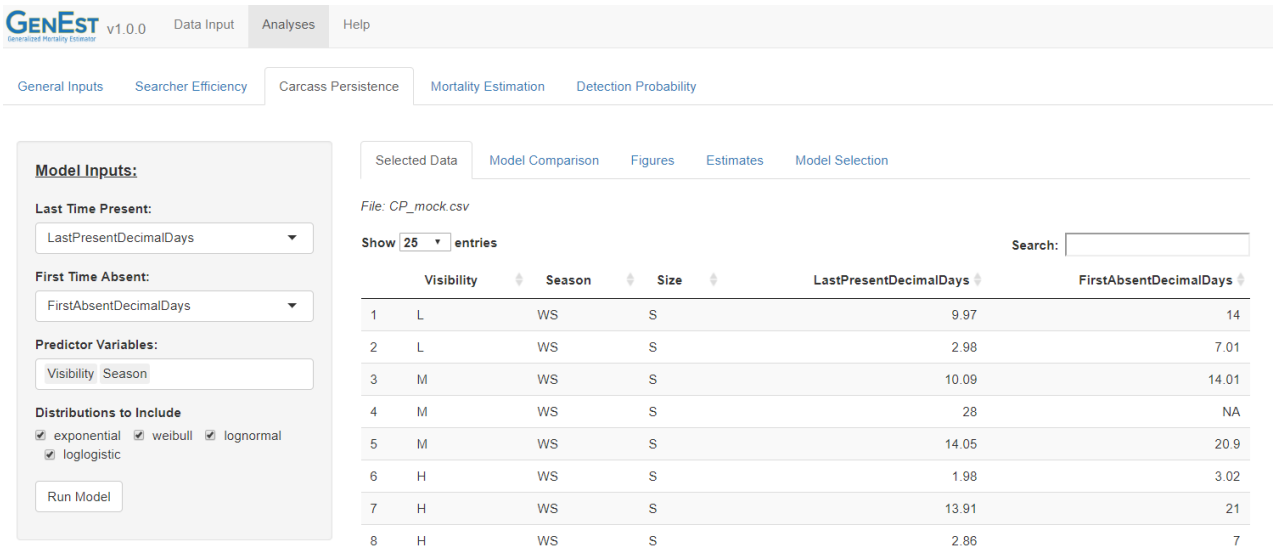

Figure 10. Carcass persistence model inputs.

- "Last Time Present:"
  - The user is required to select the column in the input dataset that corresponds to the length of time (in decimal fraction of days, starting at time = 0) between when a carcass was placed in the field until the last time it was observed.
- "First Time Absent:"
  - The user is required to select the column in the input dataset that corresponds to the first time (in decimal days since placement at time $t = 0$) that the persistence trial carcass was determined to be absent. For carcasses that were never removed during the trial, enter NA or Inf.
- "Predictor Variables:"
  - This input is optional.
  - The user can select up to two categorical predictors that they believe might affect carcass persistence and are represented in the carcass persistence file.
  - If the carcass persistence model is to be used in mortality estimation (as opposed to a stand-alone analysis carcass persistence), the selected predictor columns must also be included in the carcass observations data.
  - The predictors selected in this analysis can be, but do not need to be, used in the searcher efficiency analysis.

3.3.2. Table and Figure Selection and Model Selection

Several alternatives are available for viewing the data and fitted models: Selected Data, Model Comparison, Figures, Estimates, and Model Selection.

3.3.2.1. Selected Data

Data used to fit the carcass persistence models are displayed under the "Selected Data Tab." This is a subset of the .csv file input earlier in the Data Input process.

3.3.2.2. Model Comparison

The model comparison tab provides the AICc and ΔAICc (appendix 5.1) for all models fit using the covariates selected under "Model Inputs" (fig. 10).

Once all input data are selected, the user clicks on "Run Model," and the "Model Comparison" sub-tab is automatically opened when all the candidate models have been fit (fig. 11).

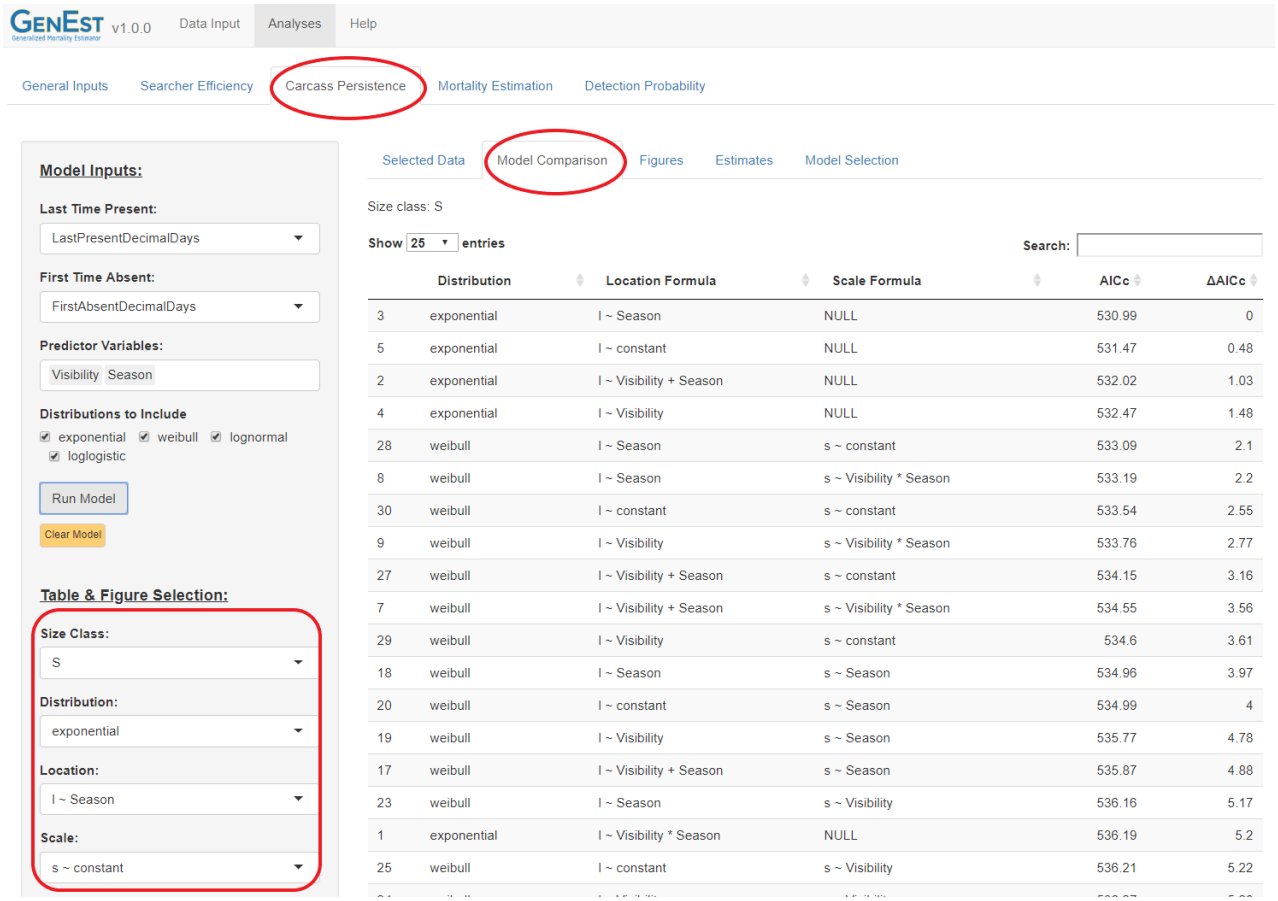

Figure 11. Options to explore results of carcass persistence model fitting.

3.3.2.3. Figures

Carcass persistence estimates from the selected model are plotted against time for each possible combination of class levels of the input predictor variables (fig. 12).

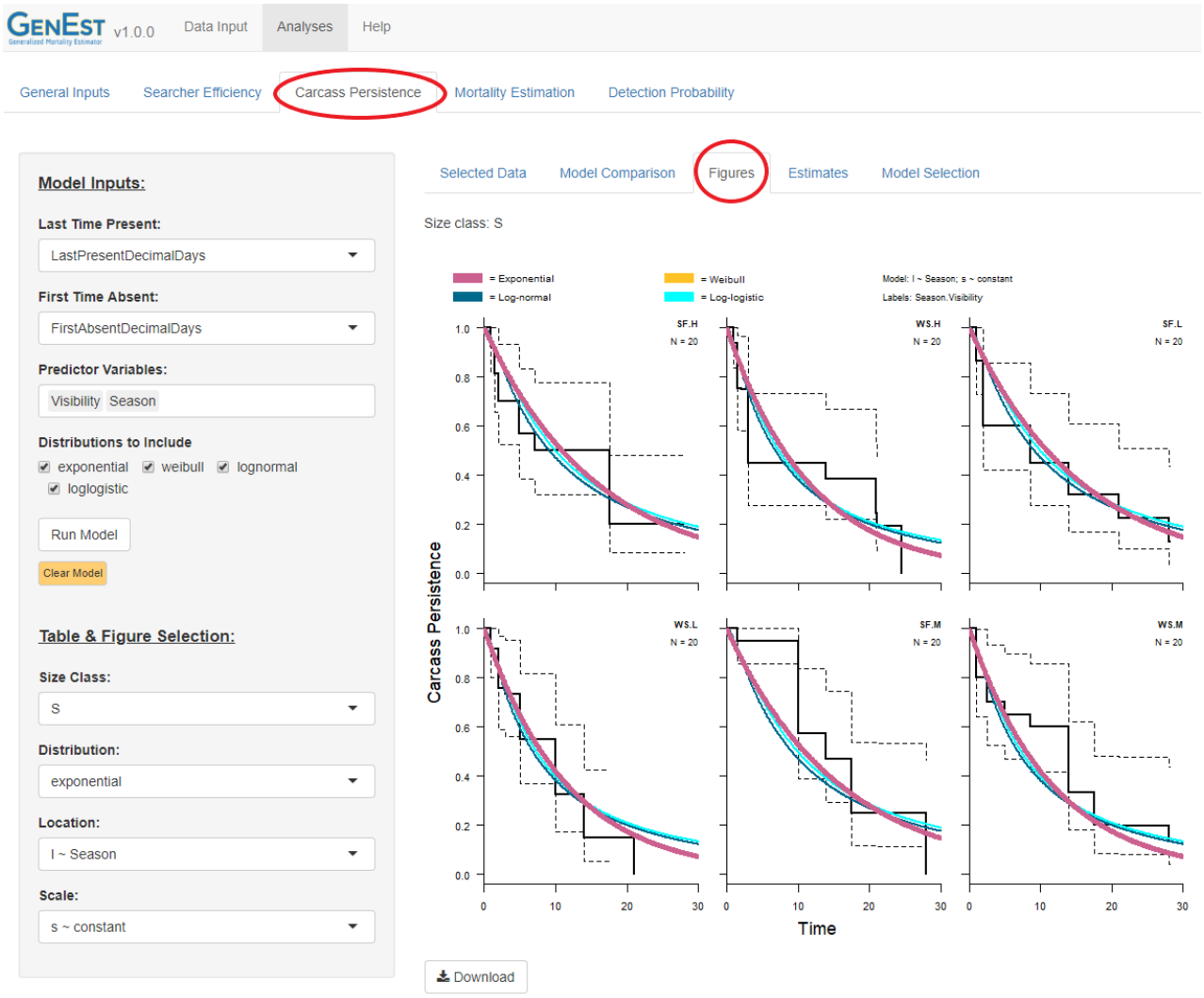

Figure 12. Example plots of estimated and observed carcass persistence by time for each combination of levels of predictor variables. Stair-step solid and dotted lines are Kaplan-Meier plots representing observed data and empirical confidence limits. All fitted persistence distributions are represented by solid curves, but the width of selected persistence distribution will be thicker than others.

3.3.2.4. Estimates

Estimates and confidence limits (as set by the user in “General Inputs”) of the location and scale parameters for each class combination of covariates for the selected model are given in a table (fig. 13). The table can be saved to a csv by clicking the “Download” button. (Note that the exponential distribution has only one parameter, l. For each "cell", or combination of covariate levels for model predictors, the number of trial carcasses (n); the median and confidence interval for estimated location and scale parameters (with confidence level entered under "General Inputs", the median carcass persistence time; and a series of r statistics, with r_I = estimated probability that a carcass that arrives at a uniform random time in an interval of I days persists until the end of the interval.

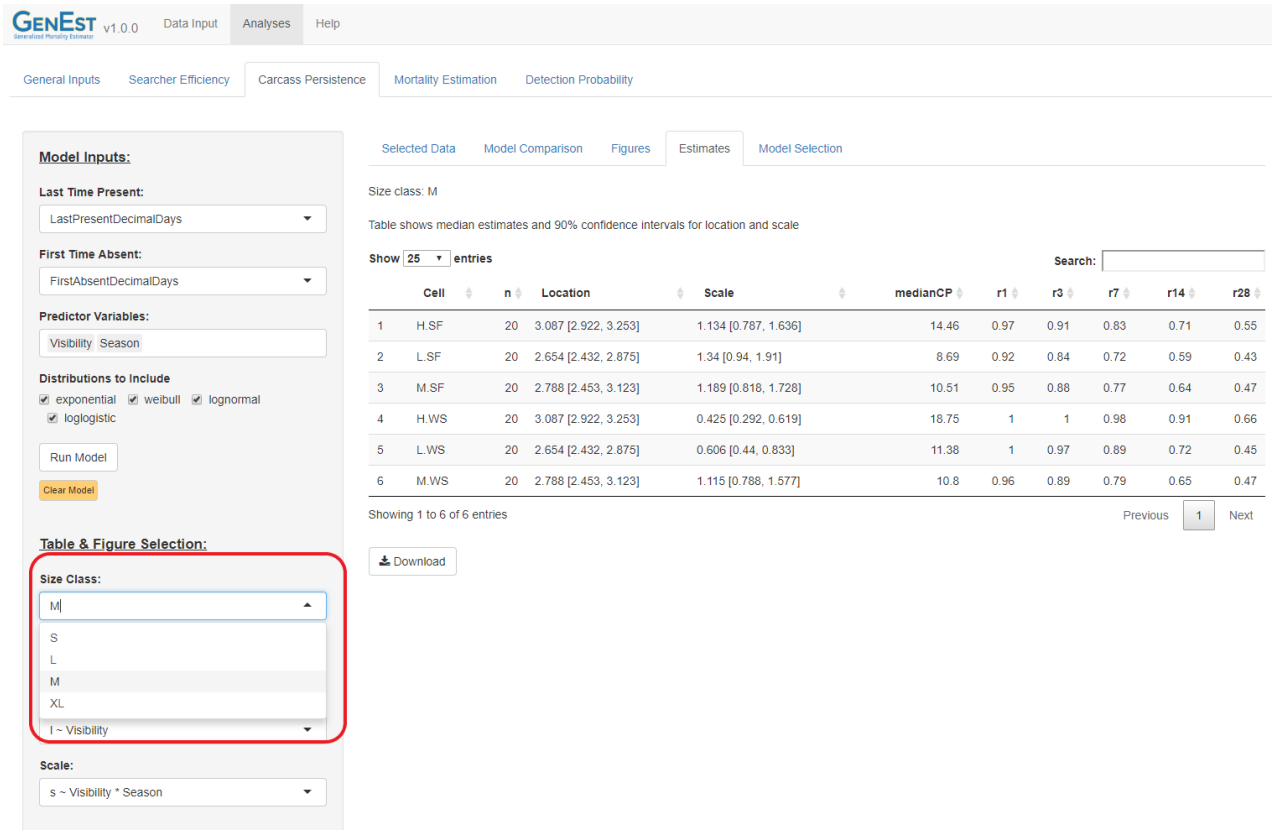

Figure 13. Estimates and confidence limits (as set by user in "General Inputs") of location and scale parameters for each cell (or combination of covariates) for selected model and size class.

3.3.2.5. Model Selection

Before continuing on to estimate mortality or detection probability, user must click on the "Model Selection" tab and explicitly choose a CP model to use in further analyses. Available models are listed in ascending order of ΔAICc (fig. 14). User may choose models by using the drop-down menus. GenEst automatically registers the selected models as the user's choices, and there is no need to confirm the choice or mechanism to do so. Also, there is no need to "Run Model" again. The selections will be held in memory for later use in mortality estimation.

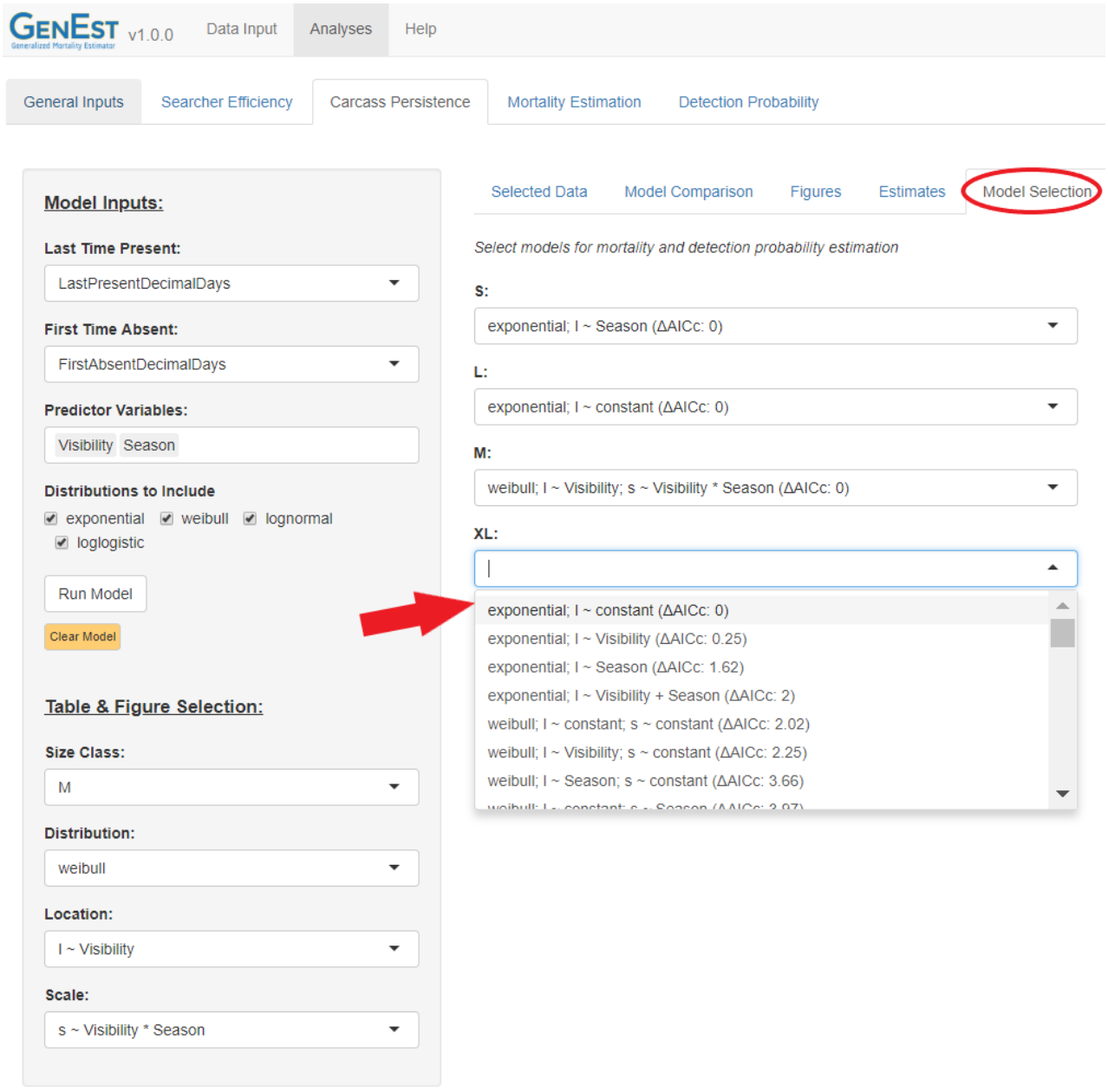

Figure 14. Model selection screen indicating which carcass persistence model to use for each size class of trial carcasses.

3.4. Mortality Estimation

Once the user has selected the desired searcher efficiency and carcass persistence models for each size class (if applicable), the user proceeds to either “Mortality Estimation” or “Detection Probability”. Explicit selection of searcher efficiency and carcass persistence models is required for mortality estimation and can be accomplished by clicking on the “Model Selection” tabs in the “Searcher Efficiency” and “Carcass Persistence” sections of the “Analyses” pages.

3.4.1. Model Inputs

Under the "Mortality Estimation" tab, users are asked to provide the “Fraction of Facility Surveyed” (fig. 15). If the facility is a wind farm or solar facility with distinct and comparable individual sources of mortality or "units" (for example, individual turbines or rows of PV panels), then a number between 0 and 1 representing the fraction of units included in the searched sample is entered. For example, if 23 of 100 turbines are searched at a wind facility, the “Fraction of Facility Sampled” would be 0.23; or if a solar facility is divided into 100 similarly sized groups of PV panels and 23 of the groups are surveyed, then the fraction of the facility sampled would likewise be 0.23.

By contrast, DWP is used to account for the fraction of a unit's carcasses falling into the area searched at the unit and may vary with unit and size class. For example, in road & pad monitoring at a wind facility, the sampling unit may be a turbine, and the "Fraction of Facility Surveyed" would be the fraction of turbines that are searched at a site. The fraction of carcasses falling in the searched area beneath a turbine may vary from turbine to turbine and would be entered in the density-weighted proportion (DWP) data file (appendix 5.10). Alternatively, for example, a solar power tower site might be divided into sampling units comprising the inner circle and each ring of reflectors. If 100% of the area within the inner circle is searched and 20% of the area within each ring of reflectors surrounding the power tower in the outer circle is sampled, the “Fraction of Facility Sampled” would be 1 because all units were sampled. In the density-weighted proportion file, the DWP for the inner circle would be 1 and each individual sampling unit in the rings in the outer circle would have a DWP = 0.2 to account for carcasses falling outside the sampled plots within a ring. It is not recommended that a completely random sample of units within the outer circle be taken as carcass density may change with distance from the tower (Huso and others, 2016), but a random sample of the same proportion within each ring will result in an unbiased sample of density.

The user is also asked to identify the column in the input Carcass Observation file (input under “Data Input”) that represents the date on which a carcass was found. This information will be combined with the information in the Search Schedule file to estimate the probability of persisting after arrival.

Before the "Estimate" button is activated and the mortality estimation be performed, the user must fit searcher efficiency and carcass persistence models and select which models to use in the mortality estimation. Once the SE and CP models have been fit, user must click on the "Model Selection" tabs under both the "Searcher Efficiency" and "Carcass Persistence" tabs to register model choices by selecting from the drop-down menus.

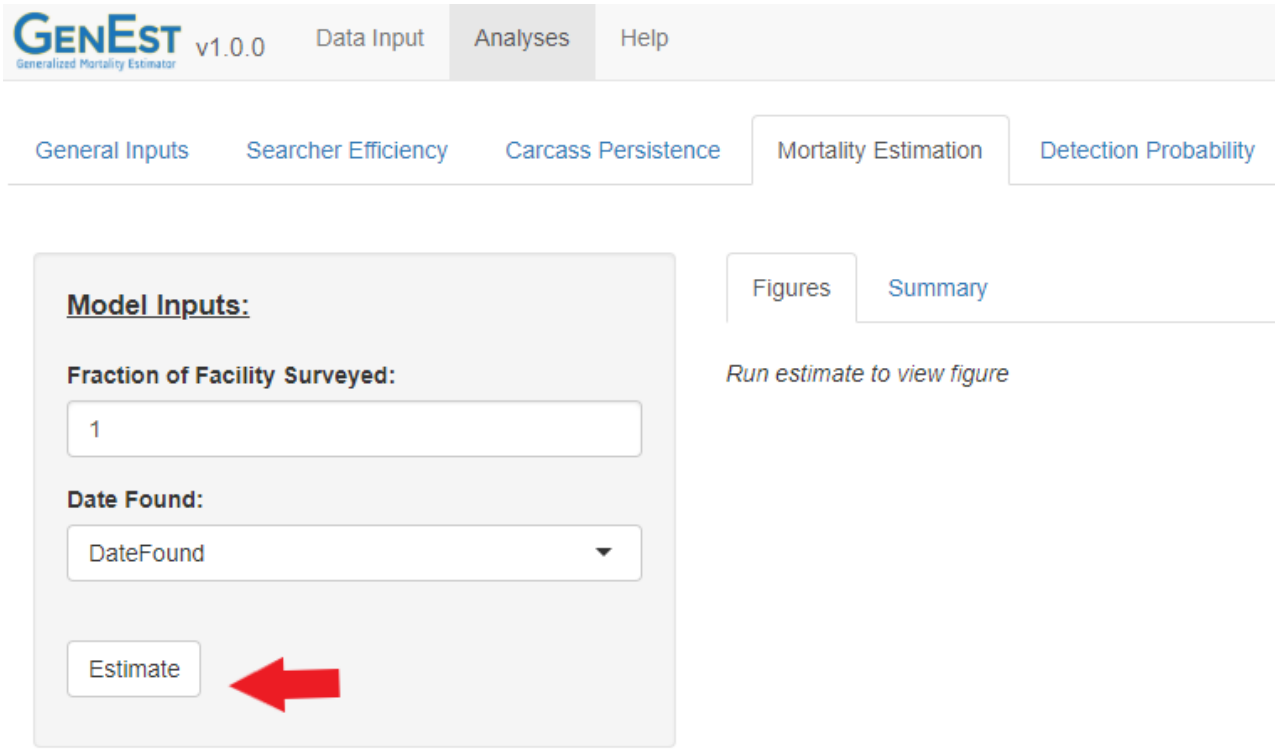

Figure 15. Model inputs screen requesting information on sampling fraction and identifying the column in the carcass observations file that represents the date on which the carcass was found by searchers.

Once the requested user input is provided, clicking “Estimate” will generate mortality estimates and confidence intervals. The calculations may take several seconds to complete or even longer if a large number of iterations was entered in the "General Inputs" page and a large number of carcasses were found. The program automatically generates a figure representing total mortality of all species at the site, a number that may or may not be of interest to the user (fig. 16).

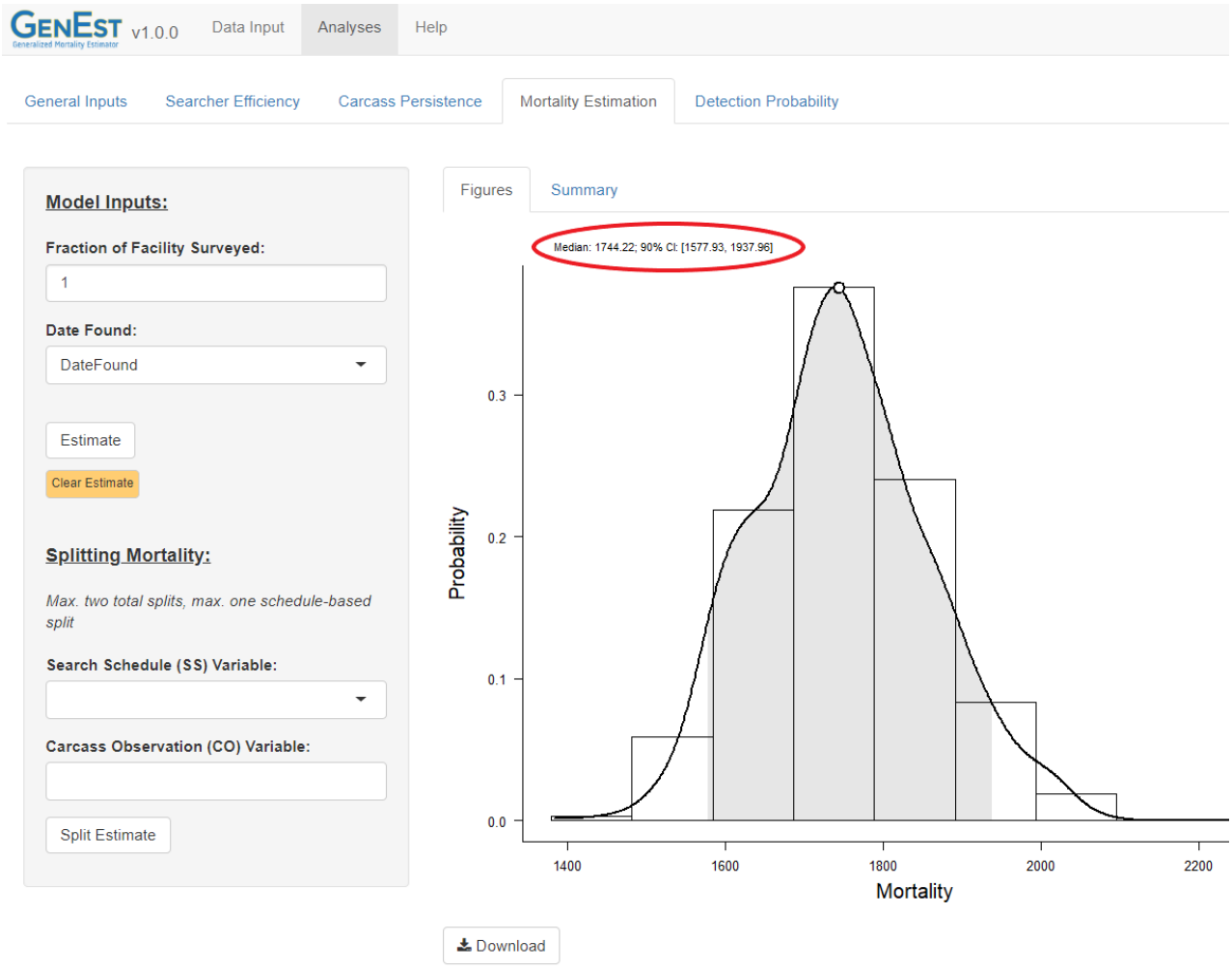

Figure 16. Estimated total mortality at the site based on user input.

3.4.1.1. Splitting Mortality Estimates

Mortality estimates may be split into subcategories as defined by selected columns in the search schedule (SS) and carcass observation (CO) files. SS splits generally refer to temporal subsets of the monitoring period (for example, season or search date), and CO splits generally refer to characteristics of the carcasses (for example, carcass size or species group) or the conditions under which they were found (for example, visibility class or turbine type). GenEst can accommodate up to two splits at a time, at most one of which can be a temporal split. Each split can have many levels. Clicking "Split Estimate" will generate a graph and a summary table of the selected split combinations (fig. 17).

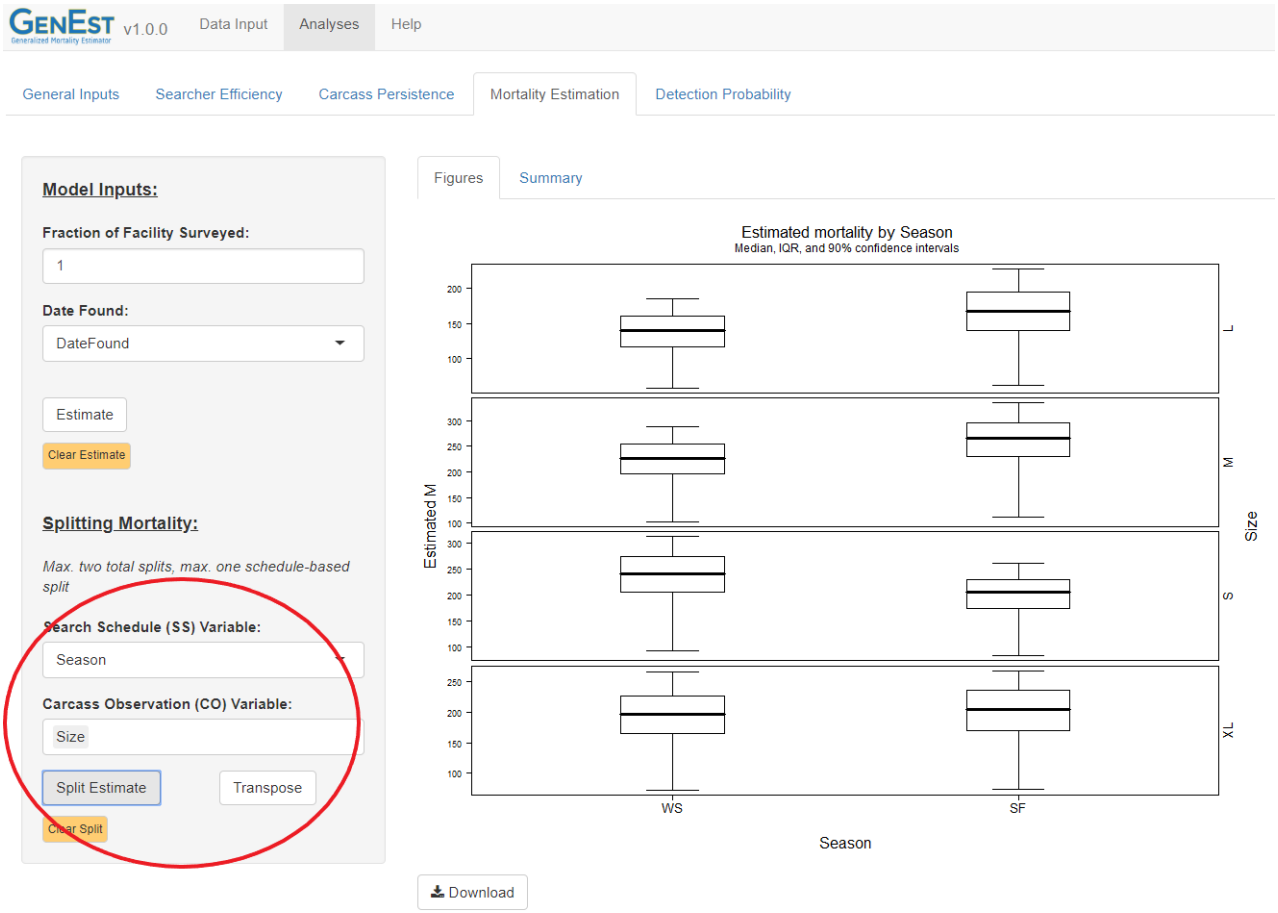

Figure 17. Estimated mortality at the site in each combination of size and ground cover (graph has been truncated to preserve space).

Selecting "Summary" will open a table of estimates and selected quantiles for each split combination (fig. 18). The table shows the number of carcasses that arrived in each cell (X) and a 5-number statistical summary (median, IQR, and a confidence interval) of the estimated mortality in each cell. Note that, when splitting on a temporal variable that can be found in the SS file, X need not be integer-valued. This is because the arrival times of carcasses are unknown and must be estimated. For example, a carcass that was found at the beginning of summer may have arrived in either the summer or spring, and GenEst allocates some fraction of carcass to summer and the rest to spring, according to a probabilistic model (Dalthorp and others, 2018). The table can be downloaded and saved as a .csv file (appendix 5.2) for use in reports and publications.

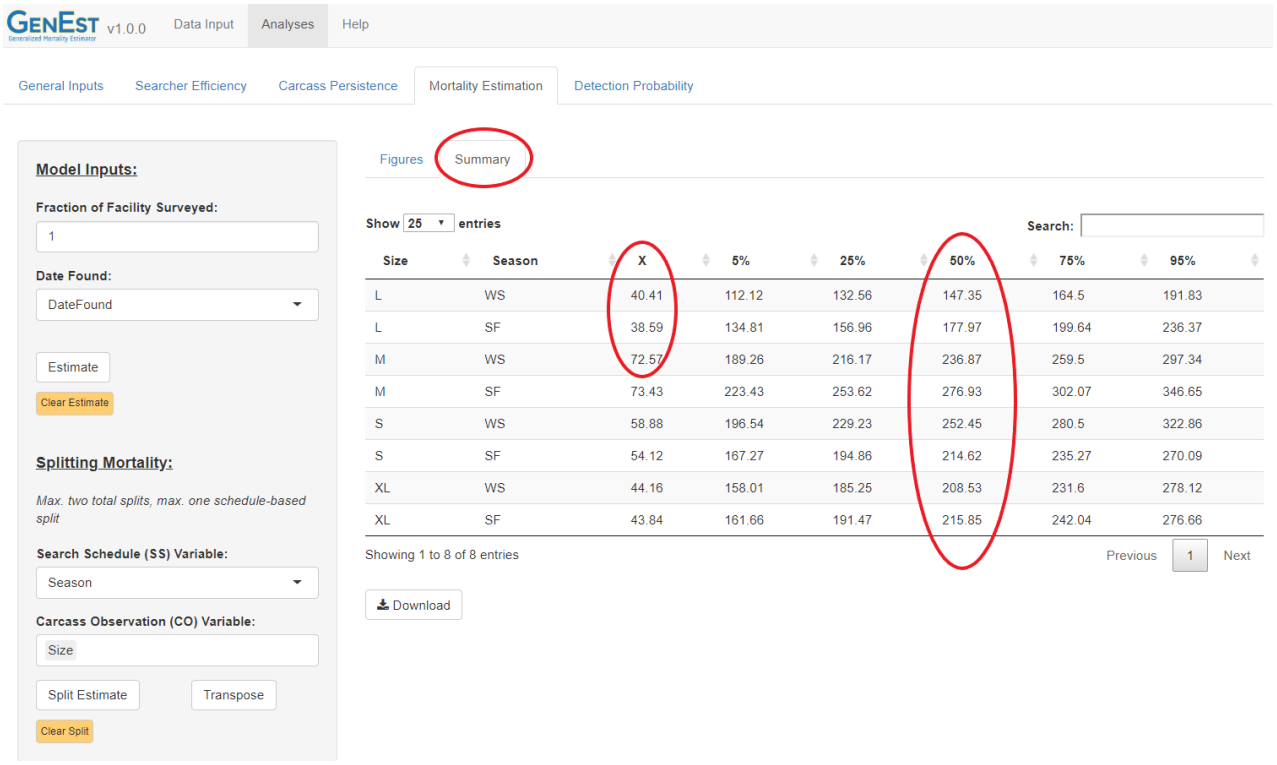

Figure 18. Estimated mortality at the site in each combination of size and ground cover. Table of quantiles used to produce associated graph.

3.4.1.2. Scaling Mortality Estimates

GenEst provides raw mortality estimates for the whole facility or for user-specified splits of the facility (for example by sector, turbine type, visibility class, habitat type, or other factor) for the period of time spanned by the monitoring. Users may scale the mortality estimates to represent mortality on a per megawatt, per turbine, per PV-array, per-unit-area or other basis, by multiplying the point estimate and confidence bounds of mortality by the appropriate scaling factor to represent a convenient unit for reporting a mortality rate. For example, if mortality was estimated to be 67 with a CI of [41, 93] at a 50 MW site, the scaled mortality would be 1.34 with a CI of [0.82, 1.86]. It must be cautioned that such a scaled estimate is a convenience for comparison to other wind facilities, but that same overall per-MW metric might not be well-suited for making adaptive management decisions if there is variability among turbines within the facility based on habitat context or turbine type.

Extrapolation beyond the monitored period can be done via adjusting the "Sampling Fraction" parameter. For example, if it is assumed that at this site summer time represents 80% of the total bat mortality, scaling to represent the entire year is accomplished by entering 0.8 as the sampling fraction and re-running the analysis (or entering, for example, 0.4 = 0.8 * 0.5 if the sampling fraction was 0.5 in the original analysis). This extrapolation should be interpreted with caution because it assumes that the fraction of mortality occurring outside the monitored period is known with certainty. There are some additional complications as well. For example, a 3-month sampling interval may represent 80% of the bat mortality but only 25% of the mortality for a resident, winter-active species of bird.

3.5. Detection Probability

Detection probability can vary with carcass size, species, visibility class, season, search schedule (for example, detection probabilities are likely to be higher for daily searches than for weekly searches), carcass arrival time (for example, carcasses arriving early in the monitoring season may be available for discovery for several searches but carcasses arriving near the end of the monitoring season may be available for only one or two searches), and other factors. Defining an overall detection probability that applies to all carcasses or even defining several detection probabilities that apply to various subclasses of carcasses (for example, large carcasses in a moderate visibility class in summer) would typically be an oversimplification and potentially misleading if used for estimating mortality. Instead, GenEst uses a sophisticated, carcass-specific detection probability for mortality estimation (Dalthorp and others, 2018), which is difficult to summarize in an informative way. However, rough summaries of detection probabilities by covariate levels (for example, easy, moderate, and difficult visibility; or small and large carcasses) and search schedules can be useful for design and planning purposes, and GenEst can provide such summaries even though they are not used in the mortality estimation itself.

3.5.1. Model Input for Estimating Generic Detection Probabilities

Explicit selection of searcher efficiency and carcass persistence models is required for estimation of detection probabilities and can be accomplished by clicking on the "Model Selection" tabs in the "Searcher Efficiency" and "Carcass Persistence" sections of the "Analyses" pages.

3.5.1.1. Search Schedule

Detection probability depends on search schedule, with more frequent searches being associated with greater detection probabilities. However, search schedules may vary markedly among turbines, so it would be unusual for there to be a single, definitive search schedule that applies to an entire site. Therefore, GenEst asks users to define an idealized search schedule and then estimates generic detection probabilities for relevant covariate classes based on the idealized search schedule input by the user. GenEst can calculate a search schedule based on the average search interval length in the Search Schedule data. This is calculated and represented on the right when user clicks "Create Average Schedule." Alternatively, user can provide the search interval and final search day and use "Create Custom Schedule" to generate generic detection probability estimates for the custom schedule (fig. 19).

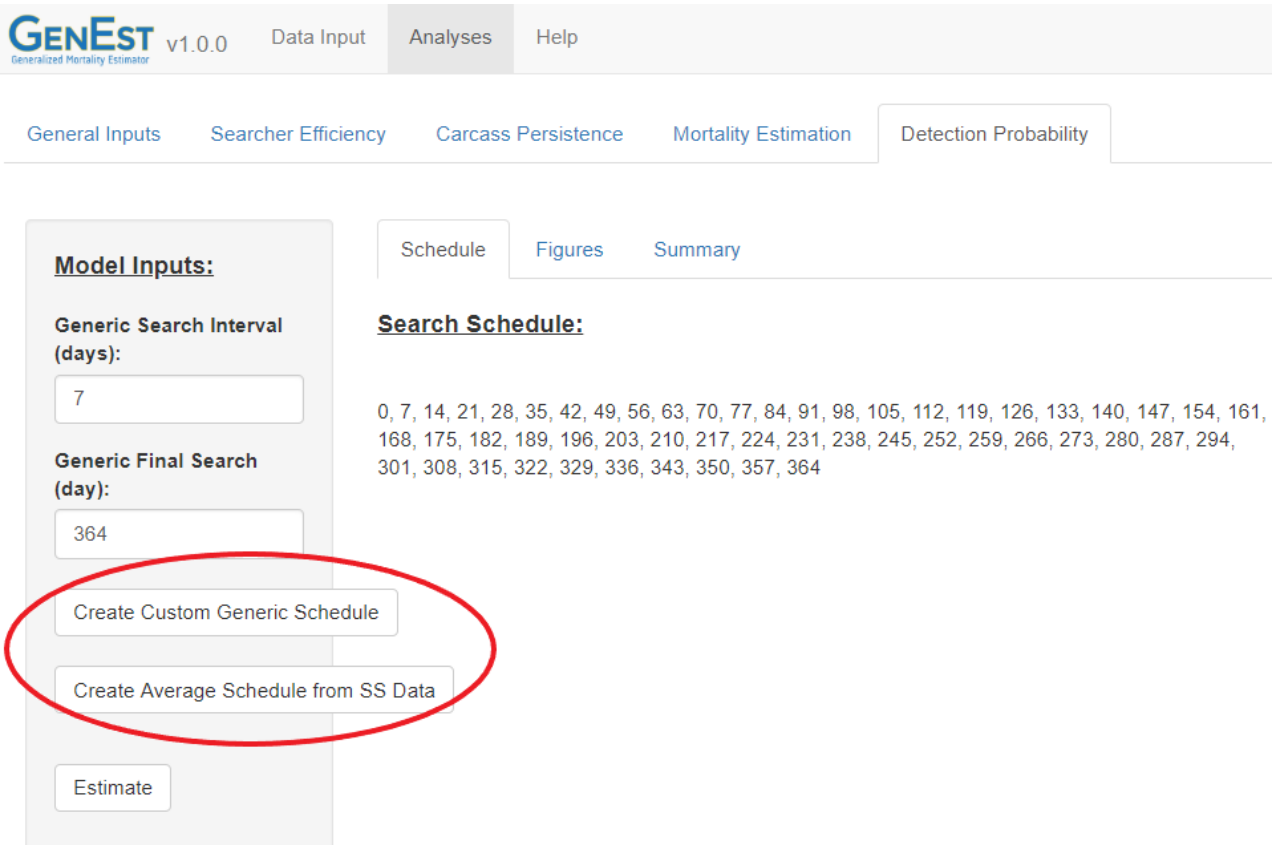

Figure 19. Model inputs screen requesting information on search schedule. Alternatively, user can input regular search interval over a set period.

3.5.2. Table and Figure Selection

3.5.2.1. Summary

After the user clicks “Estimate”, GenEst will calculate estimates of detection probabilities for the given search schedule for the cells represented in the fitted SE and CP models (for example, easy, moderate, and difficult visibility in spring and fall). When the calculations are complete, the "Summary" tab will open to show the medians and confidence intervals in each model cell for the selected size class. User may choose a different size class to view by selecting it from the menu in the "Table & Figure Selection" section. The table can be downloaded and saved as a .csv for use in reports and publications (fig. 20).

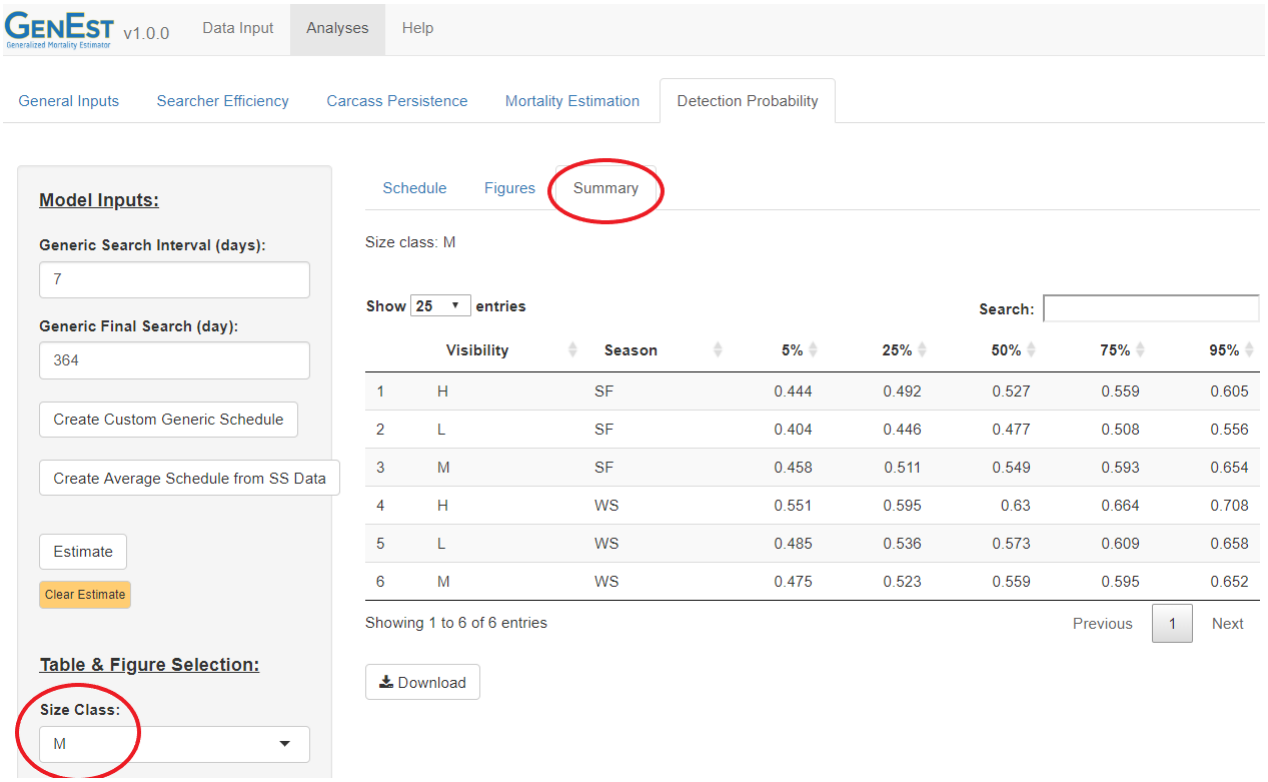

Figure 20. Summary of detection probability and confidence limits for relevant covariate classes (covariates that were included in the selected carcass persistence or searcher efficiency models) for the selected size class.

3.5.2.2. Figure

Selecting the “Figure” tab opens a graphic representation of the data in the “Summary” table for the selected size class (fig. 21).

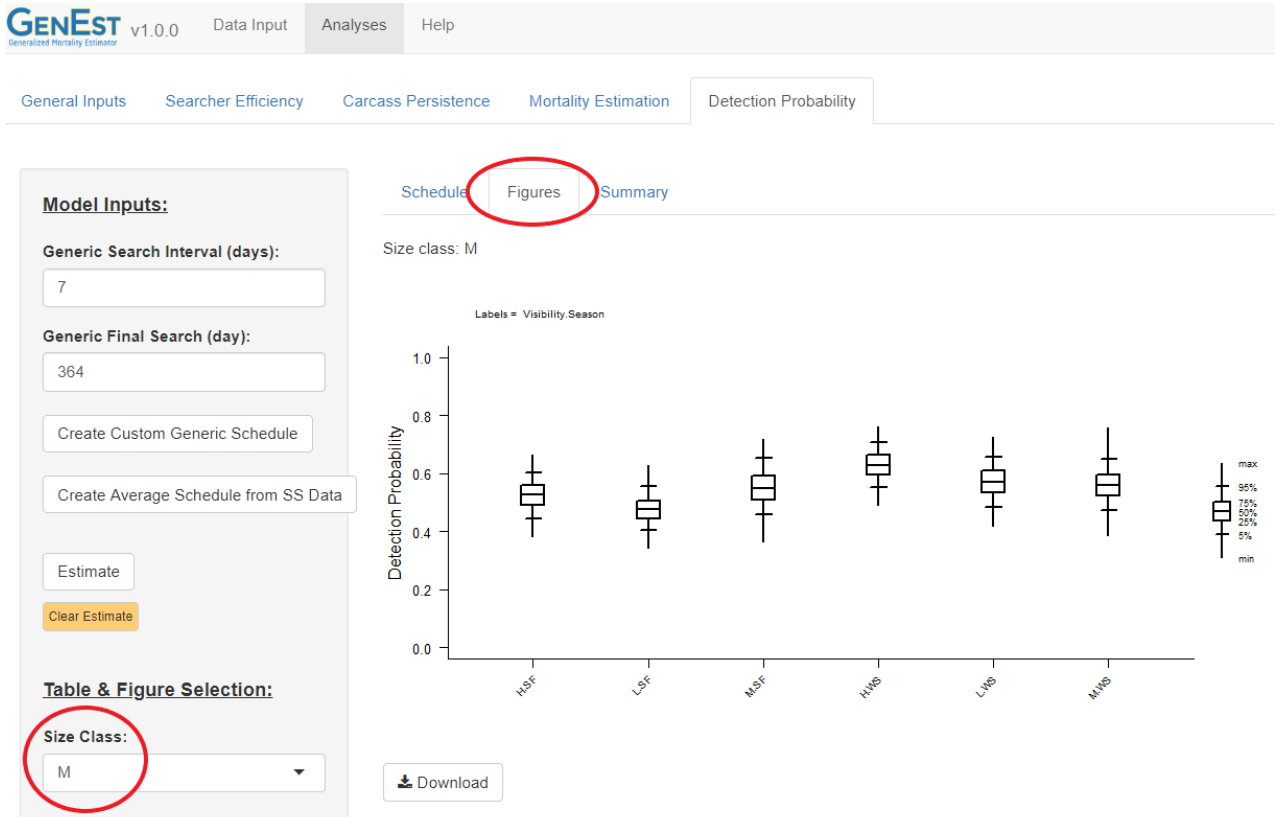

Figure 21. Boxplot of estimated detection probability for relevant covariate classes (covariates that were included in the selected carcass persistence or searcher efficiency models) for the selected size class. Graphic representation of the data under the “Summary" tab.

4. Worked Examples

4.1. Example 1—Wind Facility: Bats, Road and Pad Surveys of All Turbines

In this example, we consider a wind farm at which searches for bats have been conducted from spring to fall at all 100 turbines at a site, on roads and pads only. Our first goal is to estimate with 90% confidence overall bat mortality at the farm over the monitoring period. We are also interested in the mortality by season and species.

4.1.1. Data Loading

Click "Browse" and locate appropriate files for each data type: Searcher Efficiency (SE), Carcass Persistence (CP), Search Schedule (SS), Density-Weighted Proportion (DWP), and Carcass Observation (CO). The files used for this example are located in the `extdata\wind_RPbat` folder within the GenEst library directory or may be downloaded through the "Help/Resources" tab. As each dataset is uploaded, it is displayed in the "Data Input" window (fig. 22).

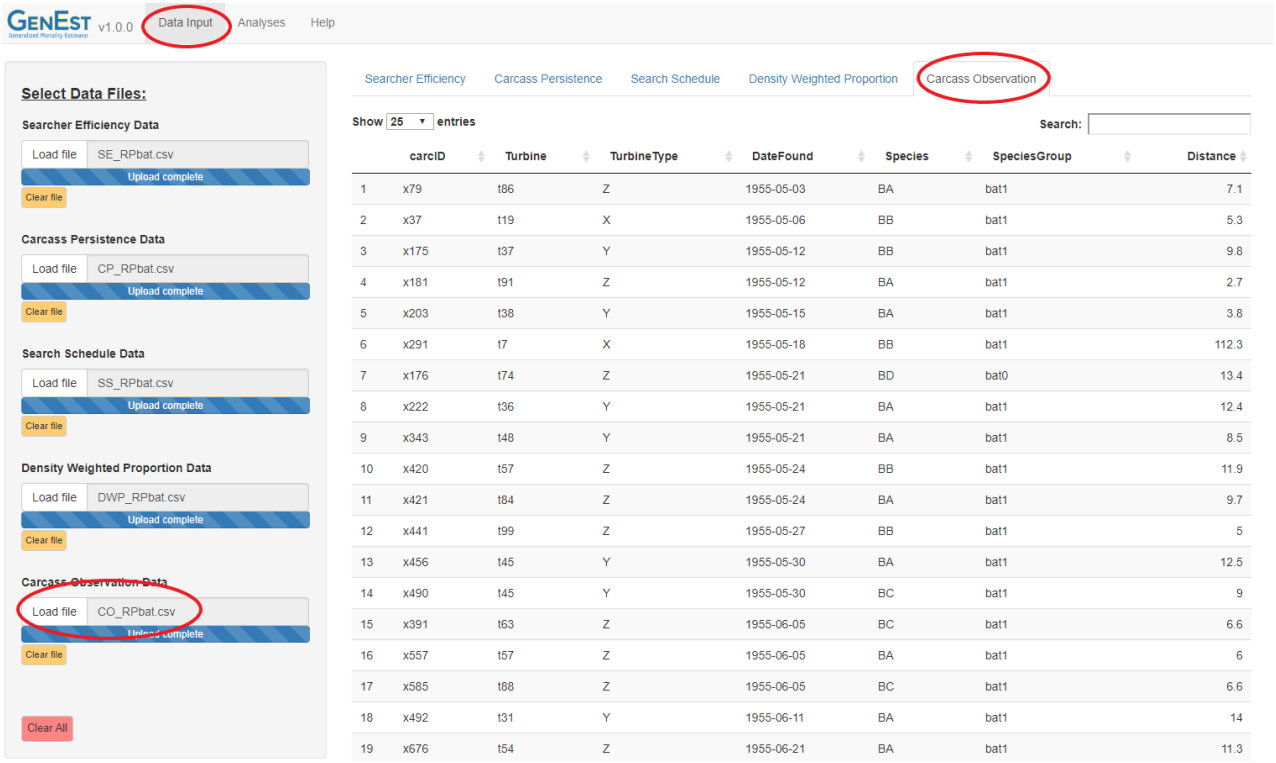

Figure 22. "Data Input" tab with "Carcass Observations" displayed after uploading.

4.1.2. General Inputs

Analysis begins with setting the "General Inputs", which are overall governing parameters for the analysis and can be found under the main "Analyses" tab (fig. 23).

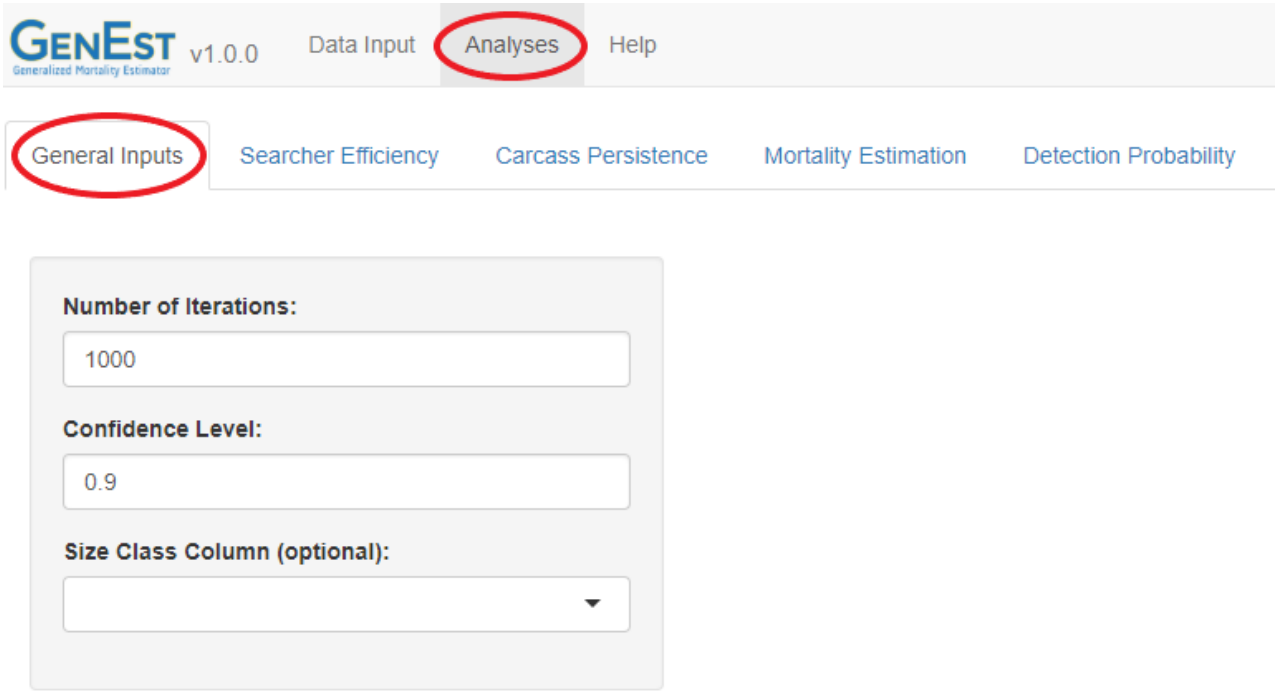

Figure 23. Starting the analysis with "General Inputs".

Confidence intervals are constructed using parametric bootstrapping, which is a simulation technique that accounts for uncertainty in estimation. Increasing the number of simulation iterations increases the accuracy of the confidence intervals, but it also increases the calculation time. The default number of iterations is 1000, but a smaller number may be desired for exploratory analysis or a larger number (up to 10000) for final analyses.

Users define the desired "Confidence Level" for the analyses. A higher confidence level gives greater assurance that the true parameter (for example, mortality or searcher efficiency) is in the resulting confidence interval (CI). The greater assurance is accomplished simply by appropriately increasing the width of the interval. For example, for a given analysis a 50% CI might be [90, 110], while a 95% CI would be more like [70, 130].

Even though carcasses of different sizes may have much different searcher efficiency and carcass persistence parameters, GenEst allows straightforward analyses of data sets that include carcasses belonging to several different size classes. Selecting a size class is optional; doing so allows separate SE and CP models to be fit for each size class. The dataset for this example consists of only bat observations, and we will not select a size class.

4.1.3. Searcher Efficiency

Estimation of searcher efficiency parameters is required for mortality estimation and can be performed using the "Searcher Efficiency" tools under the "Analyses" tab (fig. 24).

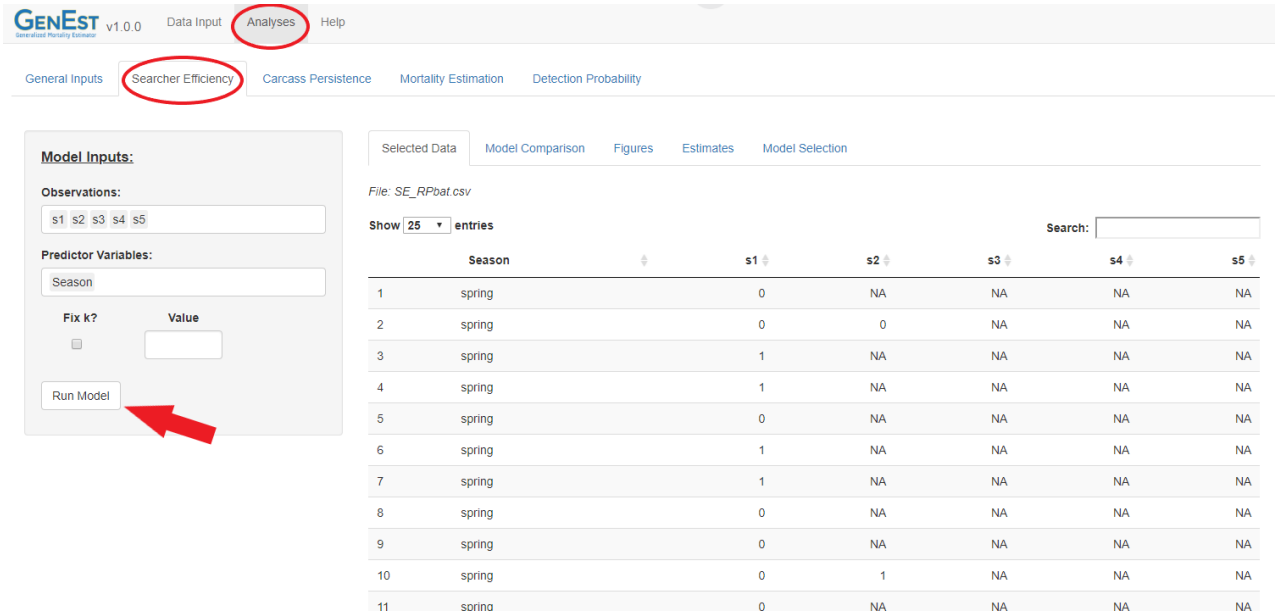

Figure 24. Model inputs for searcher efficiency analysis.

Searcher efficiency is modeled as a function of two parameters, p = probability of finding a carcass on the first search after it arrives in the field (assuming that it is present at the time of the search) and k = proportional change in searcher efficiency with each successive search. Estimation of k requires that field trial carcasses that are missed in the first search after carcass placement are left in the field to allow searchers at least one more opportunity to find them. From the "Observations" drop-down menu, the user selects the names of the columns in the searcher efficiency data file that contain the results for each search occasion. If data from only one search occasion for each carcass is included in the searcher efficiency file, user must specify "Fix k?" as "Yes" and enter a number between 0 and 1 (inclusive) for k. In this example, trial carcasses were checked several times after initial placement in the field, so it is possible (and desirable) to estimate k rather than assume a value for a fixed k.

Searcher efficiency can be modeled as a function of environmental covariates (such as season or ground visibility). Users specify which covariates (if any) they wish to consider in models of searcher efficiency in the drop-down menu for "Predictor Variables". In this example, `Season` is selected as a covariate. There are four fitted models, which include combinations of p and k as a function of `Season` or constant (indicated by "`~ constant`"). After clicking "Run Model" (fig. 24), users will see the "Model Comparison" table (fig. 25), which shows all fitted models in order of AICc (appendix 5.1). The table shows that in this example the best model in terms of AICc is the one with constant p and constant k.

To estimate mortality, users must first choose a single model from the list of models fit. GenEst provides several tools for guiding the selection. First, under the "Model Comparison" tab, the models are ordered by AICc (appendix 5.1).

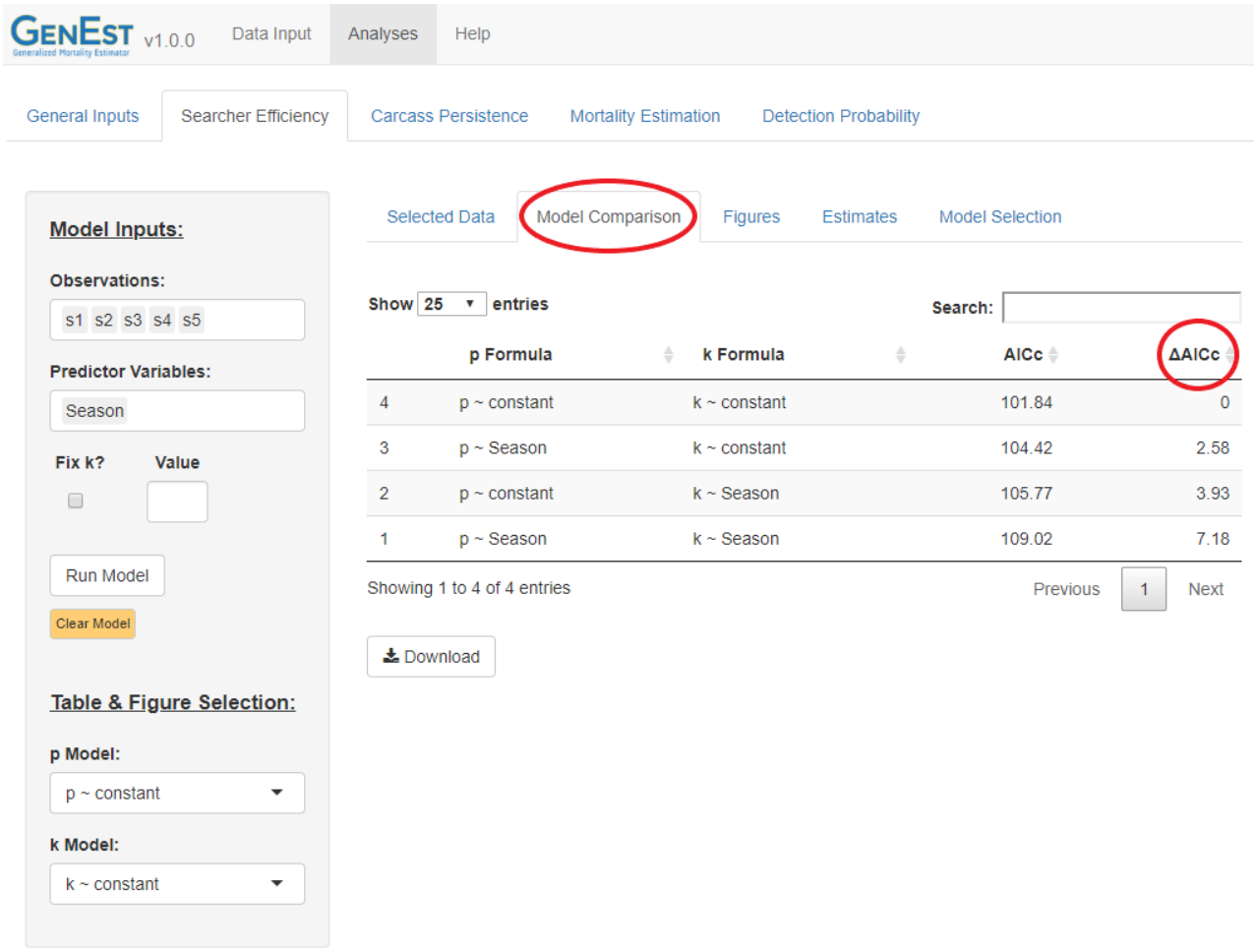

Figure 25. Searcher efficiency model comparison.

After fitting searcher efficiency models, the user may select the "Figures" tab to see diagnostic plots to identify potential problems with model fits and to help distinguish between models with similar AICc scores (fig. 26). The default is to show the model with the best AICc score in comparison with the "Reference Model", which is the most saturated (that is, the model with the most terms) fittable model. Because it includes as many terms as possible, the reference model has point estimates that most closely match the data and provides a good reference to compare with other models. However, the reference model may overfit the data—especially when the number of trial carcasses is small—which leads to relatively wide confidence intervals and introduces some instability.

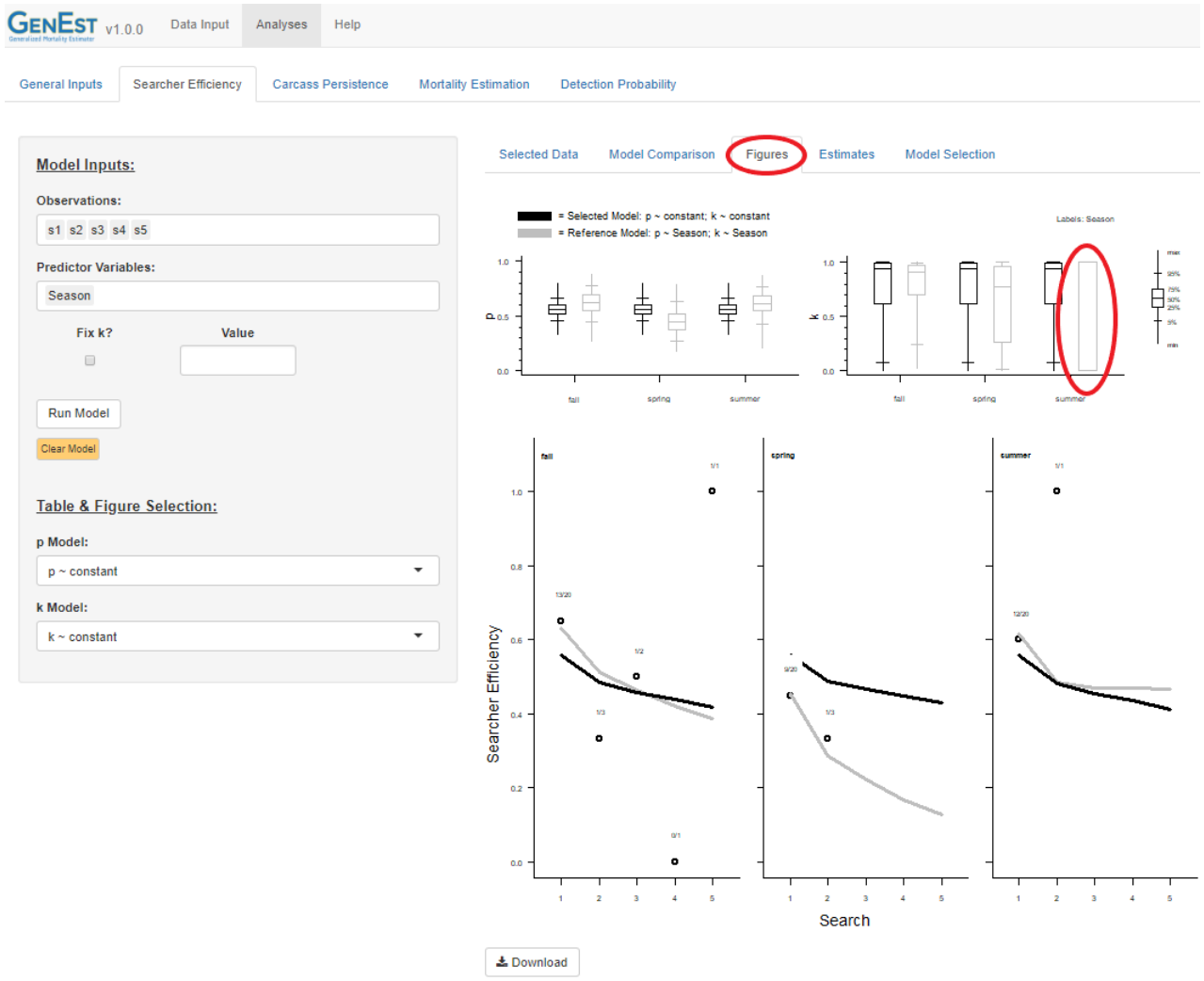

Figure 26. Searcher efficiency plots.

The boxplots in the top section of figures for the SE diagnostic plots (fig. 26) show the fitted p and k parameters for each "cell" or combination of factor levels among all the covariates. The default is to show results for the model with the best AICc score, but users may select other models via the model drop down menus on the left side-bar. In this example, the boxplots show a fairly good match between the p estimates (left panel of boxplots) for the selected model and the reference model, although the reference model appears to estimate slightly lower p than the selected model in spring and slightly higher p's in fall and summer. The variances appear to be smaller (as reflected in the smaller vertical spread of the boxes) for the selected model.

The boxplots for the estimated k values (top, right panel in fig. 26) reveal a problem with the fits for the reference model. In particular for `summer`, the gray box for the reference model extends from 0 to 1, indicating that, as far as the model can tell, k is somewhere between 0 and 1 with a median of either 0 or 1. Often, that 0-1 pattern identifies a model that includes too many parameters given the number of observations. That is not a cause for concern in the reference model (gray), but users should choose a model that avoids the 0-1 pattern in the diagnostic boxplots for the selected model (black). If all models with fitted k exhibit that pattern, users should consider fixing $k = 1$ if some carcasses were discovered after the first search occasion or $k = 0$ if no carcasses were discovered after the first search occasion.

The line plots (fig. 26, lower panels) show how searcher efficiency changes with each successive search according to the data (dots) and the fitted models (gray and black lines).

A final choice must be made for which model to use. In this example, the model with the lowest AICc is $p \sim$ constant, $k \sim$ constant, and it appears to fit the data reasonably well according to the boxplots and line plots, and there does not appear to be a problem of overfitting. We select that model from among the choices presented in the drop-down menu under the "Model Selection" tab (fig. 27).

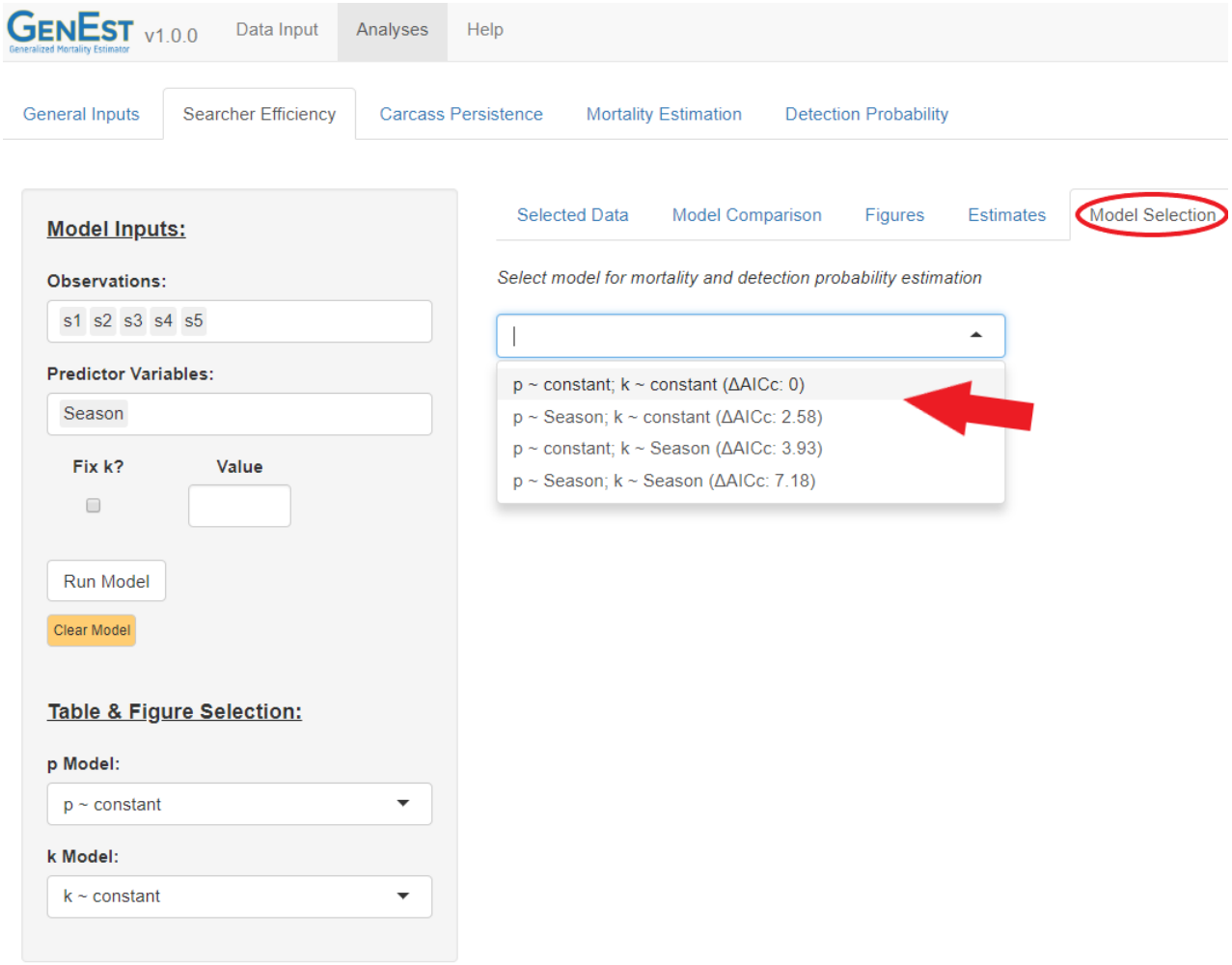

Figure 27. Searcher efficiency model selection.

This finishes the Searcher Efficiency stage.

4.1.4. Carcass Persistence

Suppose we wish to consider carcass persistence models based on exponential, Weibull, lognormal, or loglogistic distributions with location and/or scale dependent on the season. Our CP file contains a column for the `Season` in which each carcass trial was performed, plus columns `LastPresent` and `FirstAbsent`, indicating the left and right endpoints of the interval over which the carcass went missing. Navigate to "Analyses" and "Carcass Persistence" (fig. 28). Select `LastPresent` for "Last Time Present" and `FirstAbsent` for "First Time Absent" and select `Season` as our predictor. To consider our model classes of interest, tick the boxes for "Distributions to Include" exponential, Weibull, lognormal, and loglogistic.

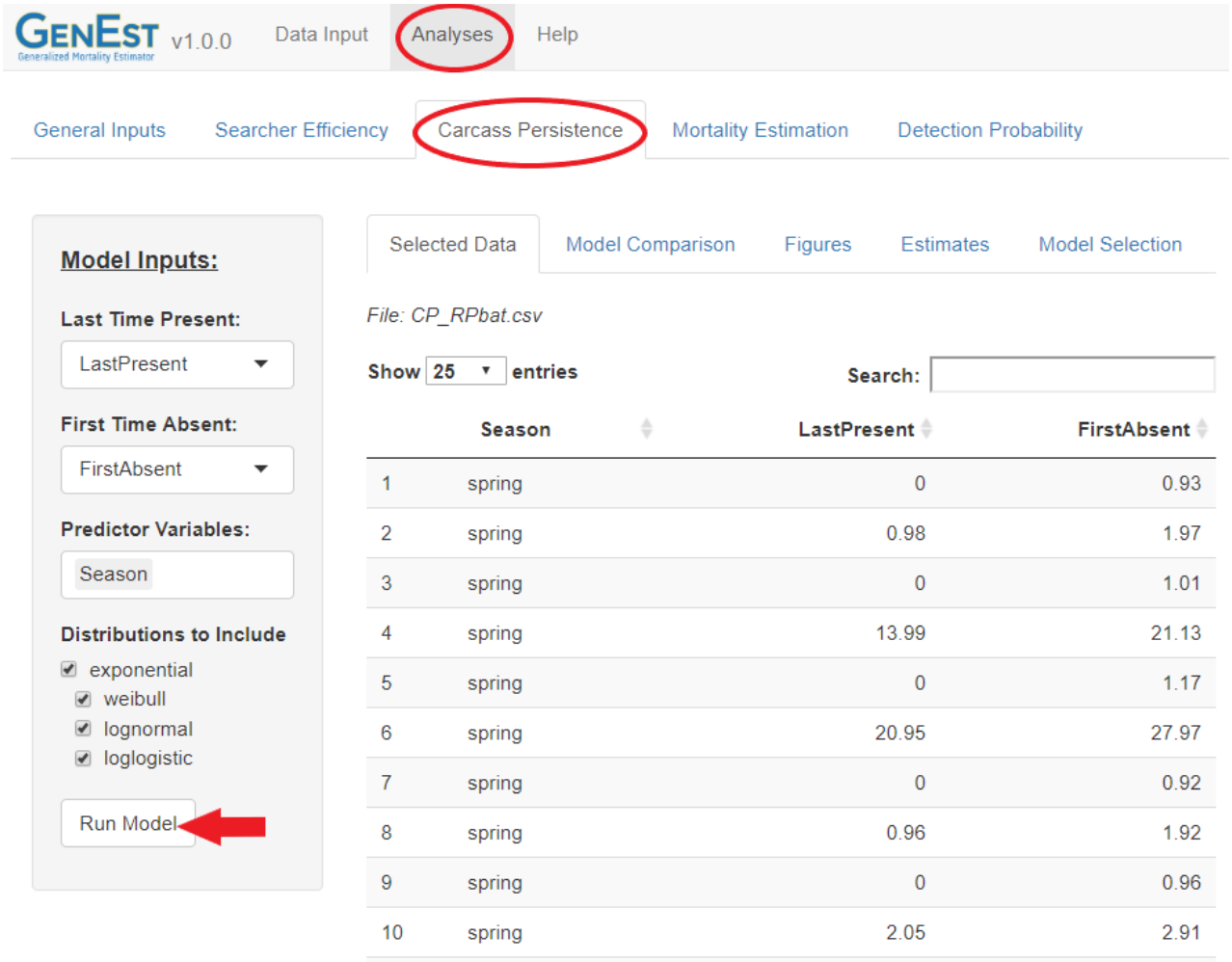

Figure 28. Carcass persistence model building.

Finally, click "Run Model" (fig. 28). A collection of models will be generated, ordered by ascending AICc, and displayed under the "Model Comparison" tab (fig. 29). Models with lower AICc scores are considered to have higher quality fits (appendix 5.1), although differences of less than about 3-4 are generally considered to be indistinguishable by this measure, in which case other criteria should be used for model selection. In this example, the three best fits are for the 2-parameter models (Weibull, lognormal, and loglogistic) with both location and scale constant across seasons ($l \sim \text{constant}; s \sim \text{constant}$). No exponential models are among the top ten best fits and none has $\Delta\text{AICc} < 15$, which is not a surprise because the exponential model has only one parameter and lacks the flexibility that the 2-parameter models provide.

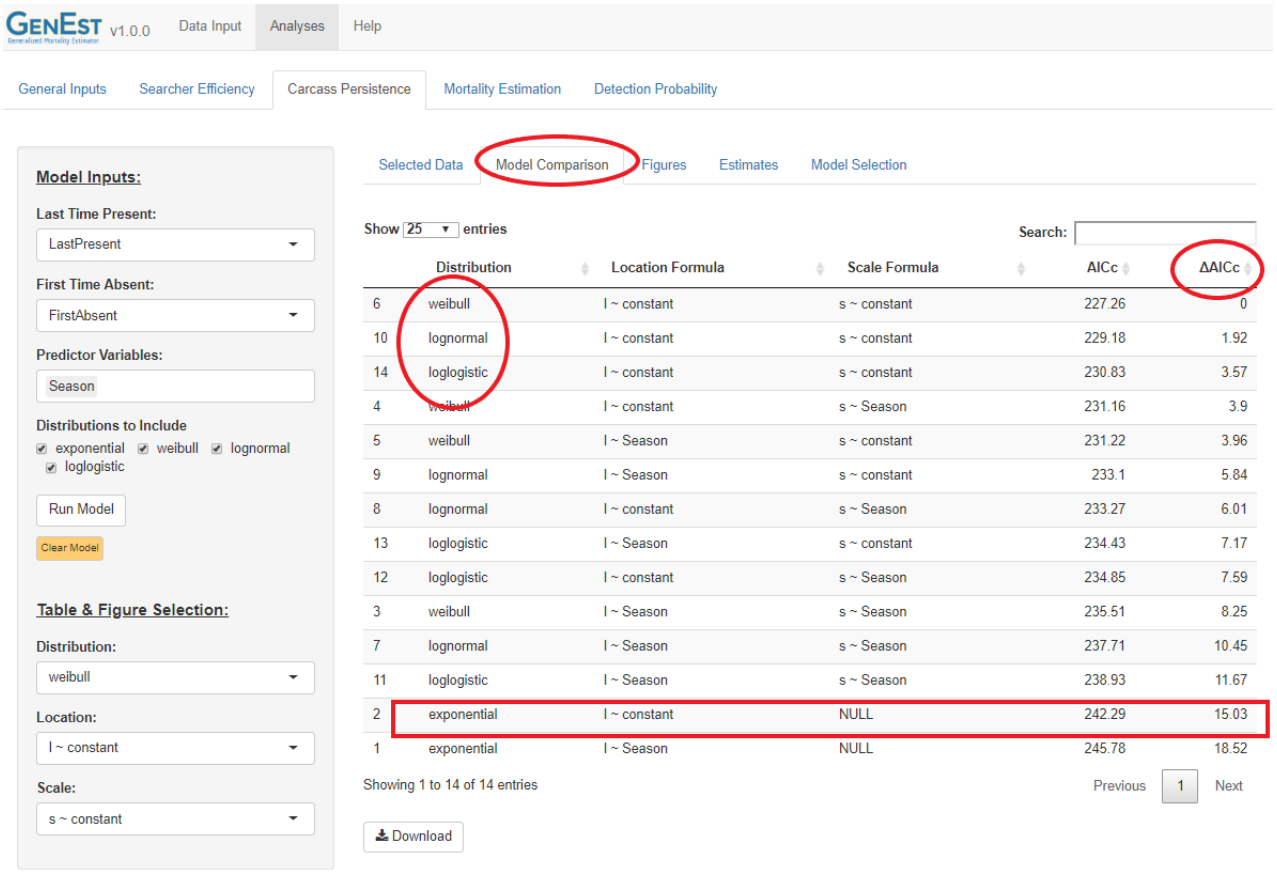

Figure 29. Carcass persistence model comparison tab.

Users may click the "Figures" tab to view fits for selected carcass persistence models (fig. 30). The figures show fitted line plots superimposed onto Kaplan-Meier curves that show the empirical data along with a 95% confidence interval. Select a model that fits the empirical curve reasonably well for all cells. Note that the curves for the 2-parameter models are largely indistinguishable whereas the curve for the exponential model is distinct from the others.

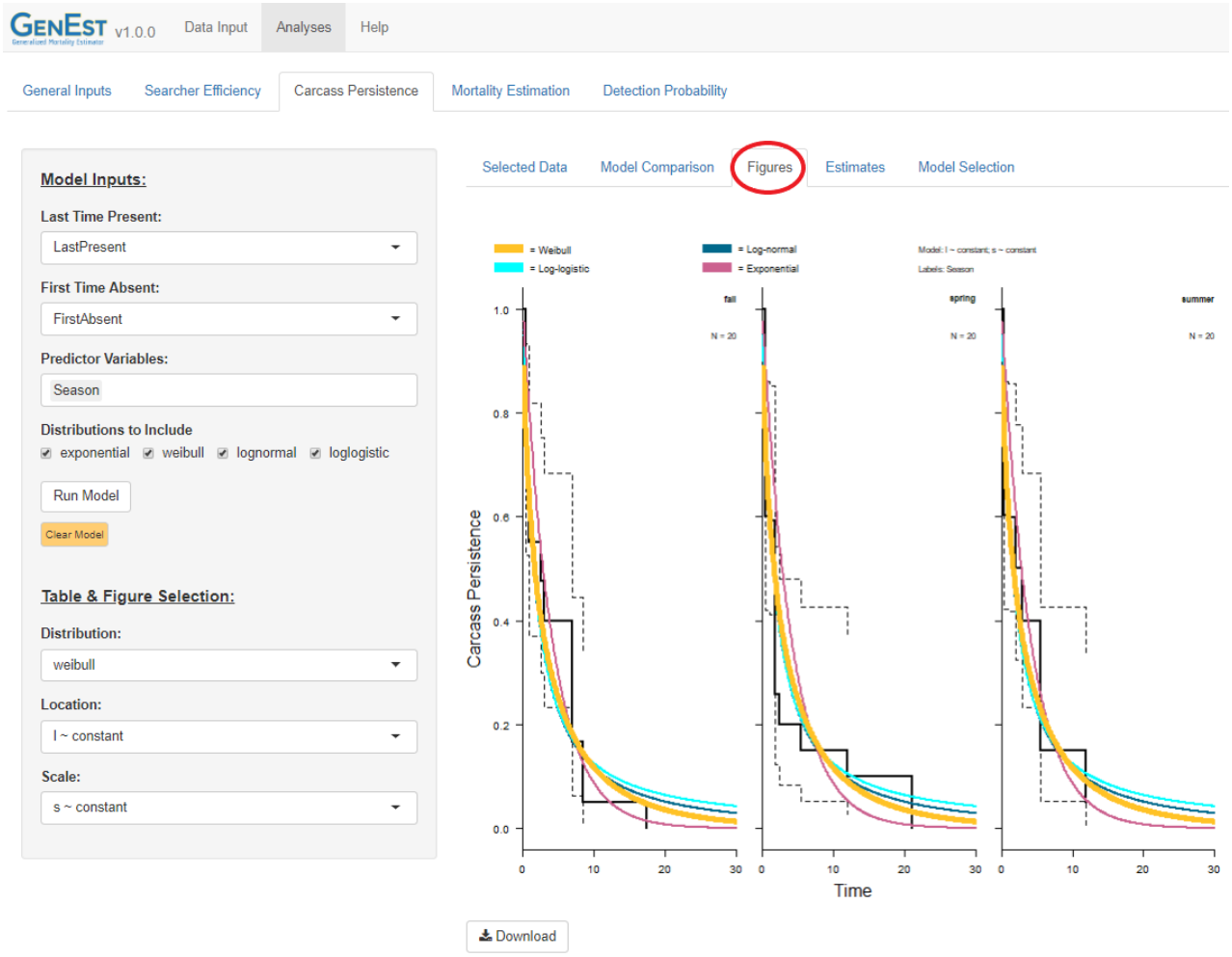

Figure 30. Fitted carcass persistence models.

To see further details about the parameter estimates, navigate to the "Analyses" > "Carcass Persistence" > "Estimates" tab. Finally, user must choose a model from the drop-down menu on the "Model Selection" tab (fig. 31). For this example, we will choose the model with the lowest AICc, which happens to be Weibull with both location and scale constant (or $l \sim$ constant and $s \sim$ constant).

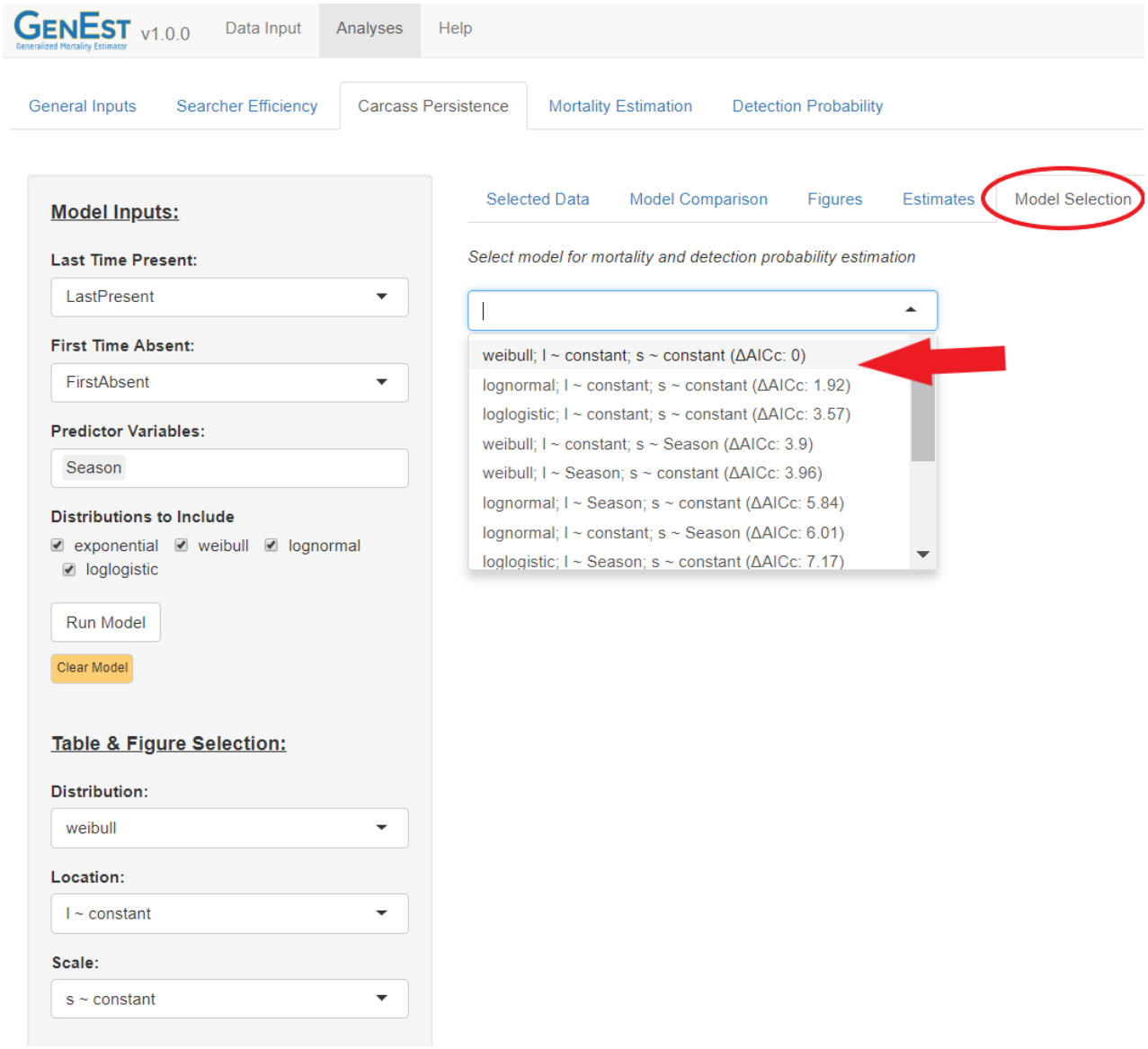

Figure 31. Carcass persistence model selection.

This finishes the carcass persistence phase.

4.1.5. Mortality Estimation

After searcher efficiency and carcass persistence models have been fitted and selected, the number of fatalities can be estimated from the "Mortality Estimation" tab (fig. 32). In this example, all turbines at the site are surveyed and inference is limited to the dates spanned by the carcass surveys, so the "Fraction of Facility Surveyed" is set to 1. Select "DateFound" for the "Date Found" column. Select "bat" as the DWP column and click "Estimate" to estimate total mortality and bring up additional options for splitting mortality by other variables (fig. 33).

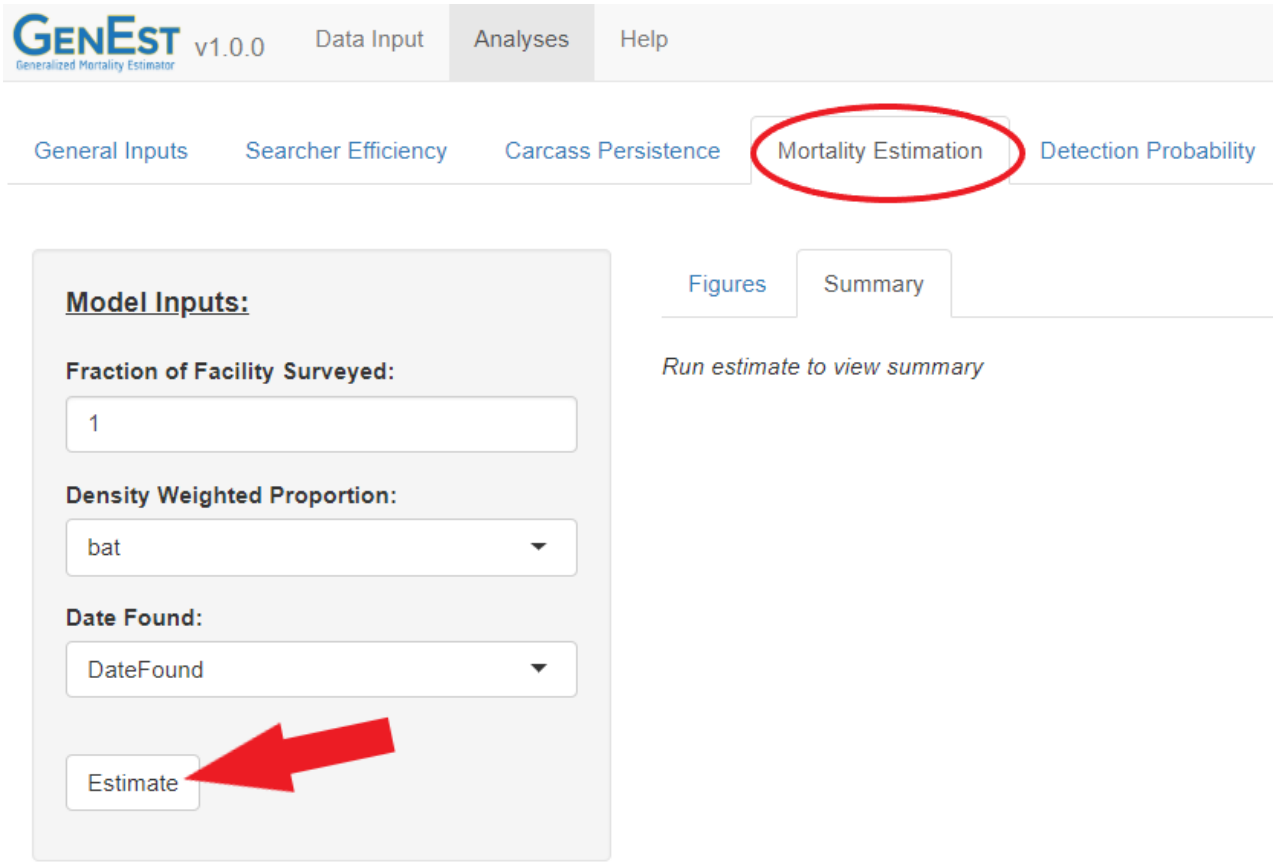

Figure 32. Mortality estimation tab.

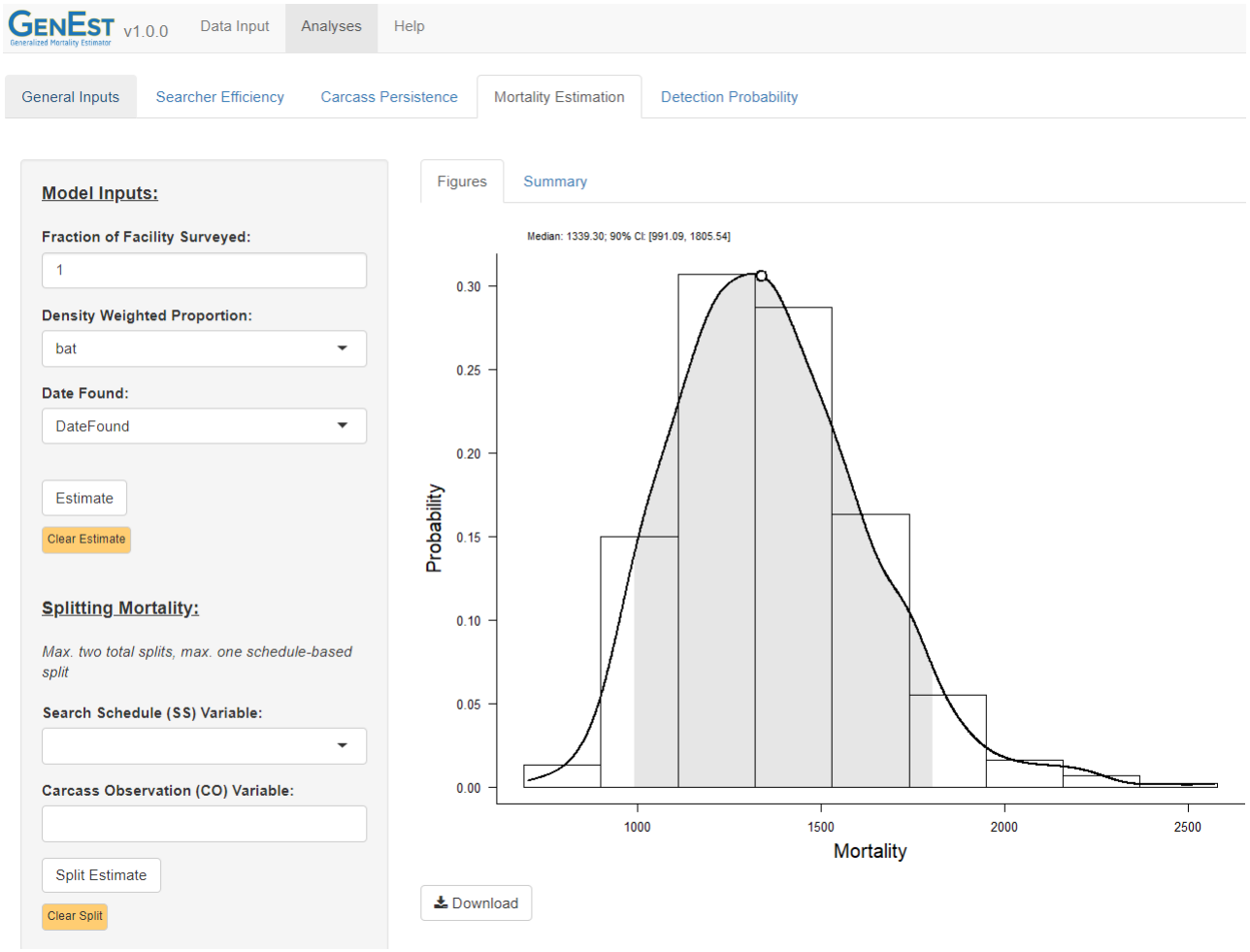

Figure 33. Total mortality and options for mortality splits.

The figure for total mortality represents the estimated mortality as a histogram of probable values, with the central confidence interval highlighted with gray shading (fig. 33). The median and confidence interval (90% in this example because that is the confidence level that was entered under the "General Inputs" tab) are given above the figure. Note that the distribution is not symmetric but is right-skewed instead, which is typical. The mean is a misleading statistic for skewed distributions because it depends strongly on the degree of skewness. That is why GenEst reports medians rather than means. Other statistics of the estimated mortality can be viewed under the "Summary" tab.

Many practitioners summarize fatality analyses at wind energy facilities by reporting a per-turbine or per-MW fatality rate. Such a rate is useful as a quick index of the relative impact of a facility, though it may obscure biologically meaningful relationships between mortality and turbine type or mortality and landscape context. A per-unit index of fatality can be obtained by dividing the median estimate and confidence interval through by the number of units in the facility. In this example there were 100 turbines, so the per-turbine fatality estimate with 90% confidence interval would be 13.4 (9.2, 20.0).

To obtain estimates for levels of temporal covariates (such as season) that can be found in the search schedule file, we select them as "Search Schedule (SS) Variable" splits and click "Split Estimate" to show the estimated numbers of fatalities by season (fig. 34). The statistics can be viewed in a table under the "Summary" tab.

Figure 34. Mortality estimates by season.

To estimate mortality by species, first remove the season covariate by clicking in the "Search Schedule (SS) Variable" drop-down menu and pressing the backspace key. Add a "Carcass Observation (CO) Variable" split for `Species` and click "Split Estimate" (fig. 35). New boxplots and table of statistics representing mortality by species are created.

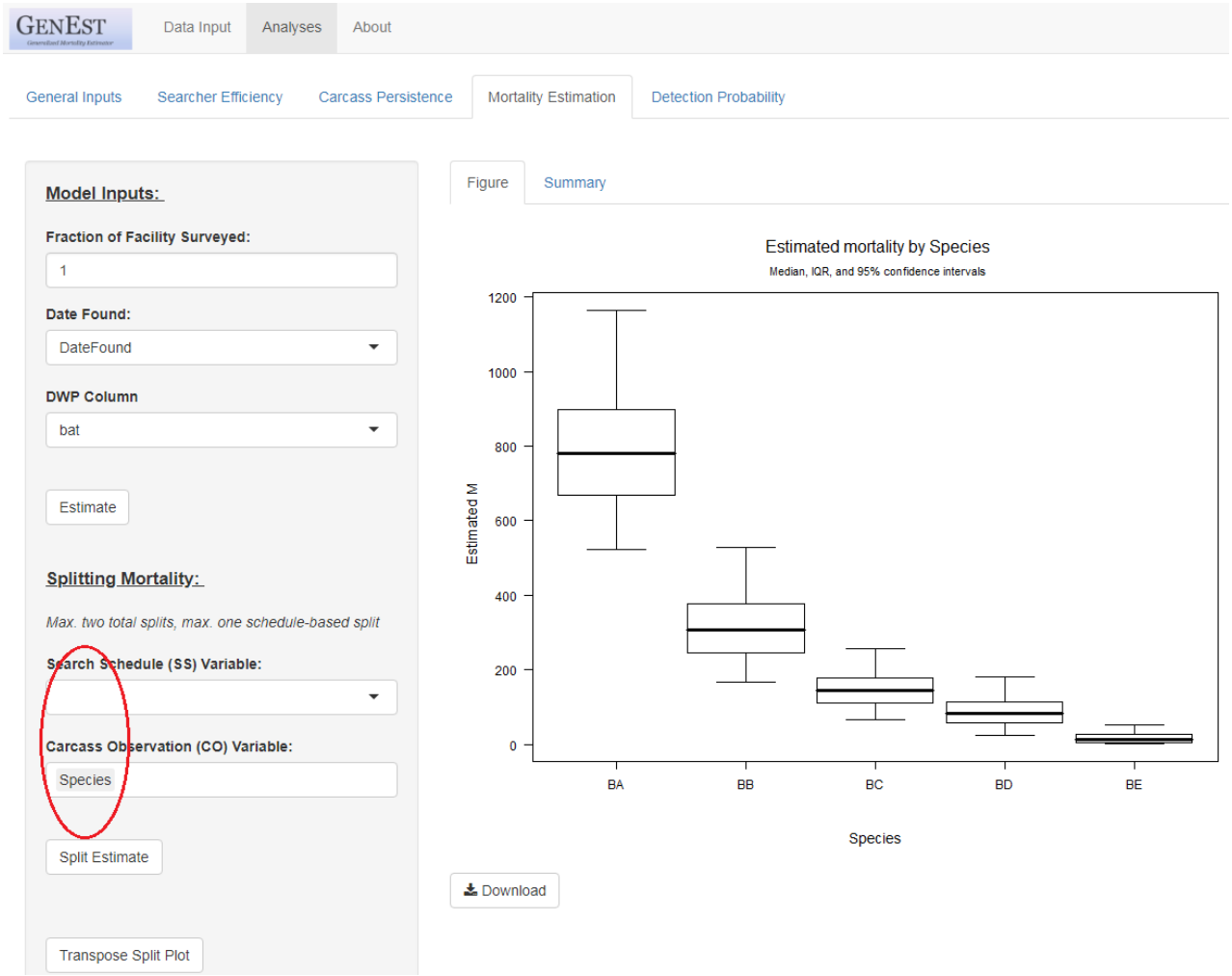

Figure 35. Mortality estimates by species.

After seeing these results, perhaps we are curious about how Species BA observations were distributed among seasons. We can find out by splitting using both an SS and a CO variable. Add back the `Season` covariate to the SS split variables and click "Split Estimate" once more. We now end up with five panels, one for each species (fig. 36), with each panel showing the estimated mortality by season for the given species. If we wish instead to view species along the *x*-axis with separate panels for each season, we can click "Transpose Split Plot" for an alternative view (fig. 37).

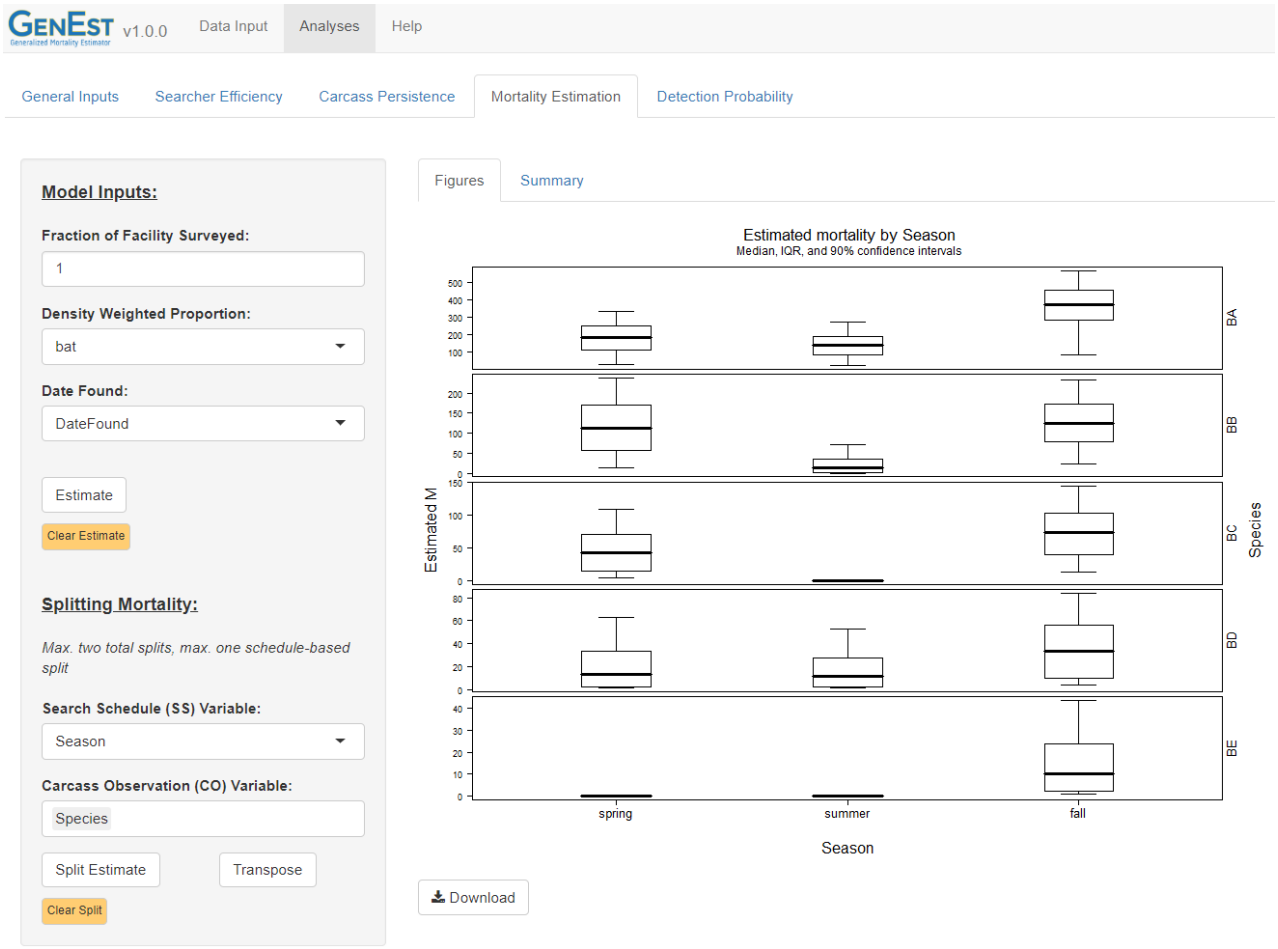

Figure 36. Seasonal mortality by species.

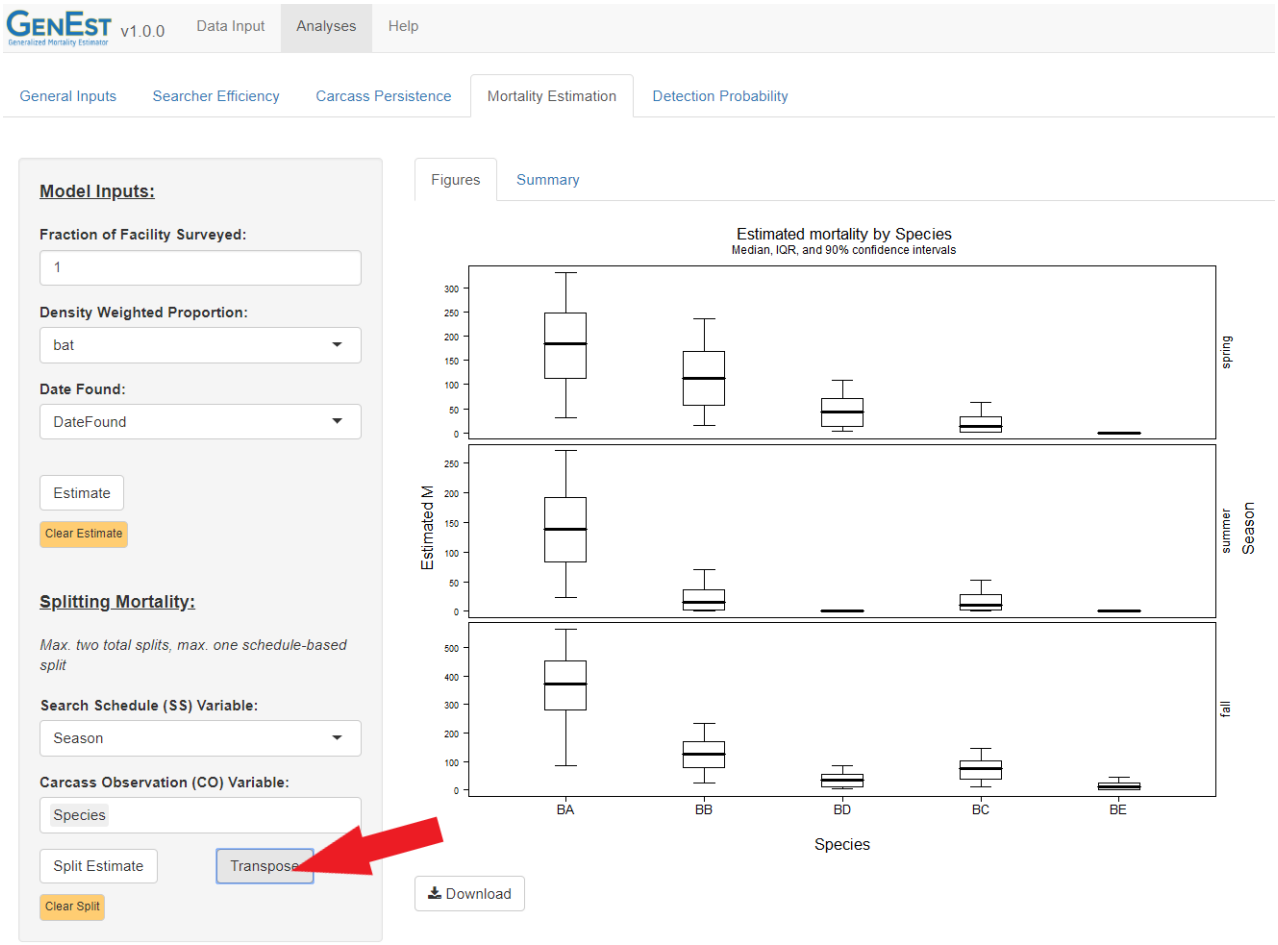

Figure 37. Bat mortality estimates by species and season (transpose of season by species).

We now have an idea of how many bats were killed at the site during the monitoring period and how fatalities were distributed among seasons, including how our five species of bats found were affected by season.

4.2. Example 2—Solar Photovoltaic Facility, All Birds

In this example, we will show how GenEst could be used to estimate mortality at a solar photovoltaic generating facility. Suppose we are interested in creating a 95% confidence interval for overall mortality, assessing whether mortality varies by season, whether it is constant throughout the facility, and how different species or groups of species are affected.

4.2.1. Data Loading

To begin, load the required Searcher Efficiency (SE), Carcass Persistence (CP), Search Schedule (SS), Density Weighted Proportion (DWP), and Carcass Observation (CO) data files by clicking on the "Browse" buttons for each file type under the "Data Input" tab. The files used for this example are located in the `extdat\solar_PV` folder within the GenEst library directory.

Navigate to the "Analyses" > "General Inputs" tab and verify that the confidence level matches the desired 90% for our confidence interval. Select `Size` from the "Size Class Column" drop-down menu. Our dataset contains observations on three classes of birds: small, medium and large. Using size class in our models allows each bird size to be evaluated independently for the purposes of estimating searcher efficiency and carcass persistence. Leaving size out would build a single SE model and a single CP covering all animals, which would normally not be appropriate because SE and CP often strongly depend on size.

4.2.2. Modeling Search Efficiency

Navigate to the "Searcher Efficiency" tab under "Analyses" and select the "Observations" columns from the SE file representing the carcass searches (fig. 38). There were four searches, labeled `Search1`…`Search4`. In the "Predictor Variables" drop-down, select `Season`. Because trial carcasses were left in the field after an initial search so that searchers had more than one opportunity to find the carcasses, k can be estimated, and user is not forced to assume a value for a fixed k. Select “No” for “Fix k?”.

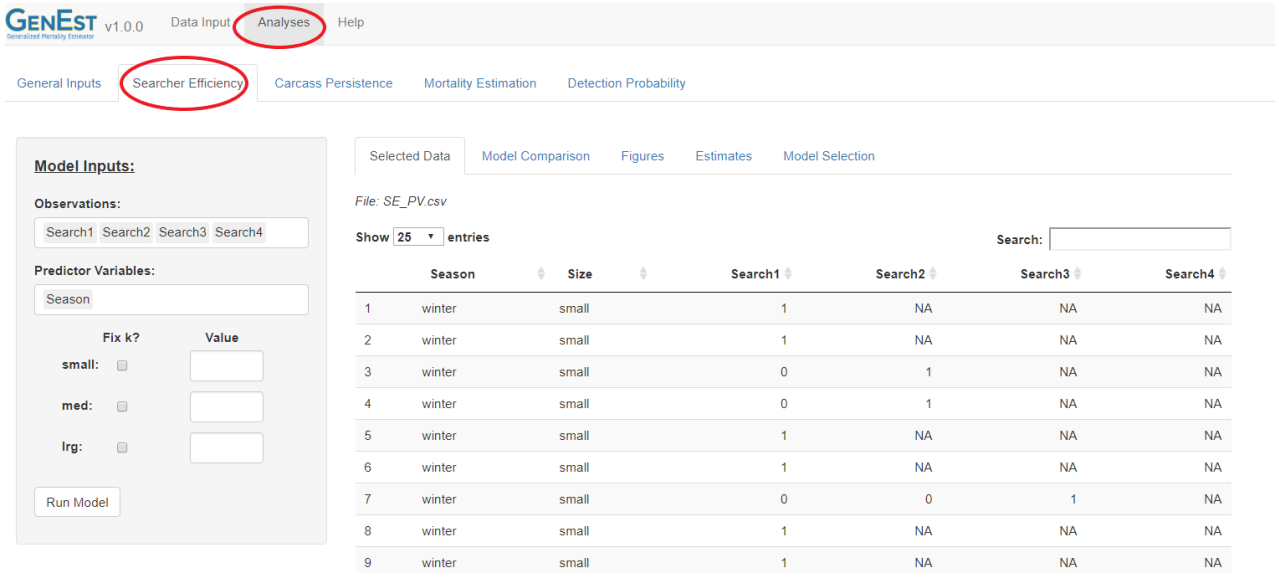

Figure 38. Data for searcher efficiency analysis.

Run the searcher efficiency model. The screen will switch to a model comparison table, showing the various models which have been tried. Models are ordered by their AICc (fig. 39). As a rule of thumb, a lower AICc indicates a higher quality model, but models with ΔAICc (the difference between each model's AICc and the AICc of the "best") less than about 3-4 are generally considered indistinguishable by this measure (appendix 5.1). In this example, with the large carcasses, the $p \sim$ Season; $k \sim$ constant and $p \sim$ constant; $k \sim$ constant models have AICc's that differ by only 0.812, so other criteria should be used to choose between these two models.

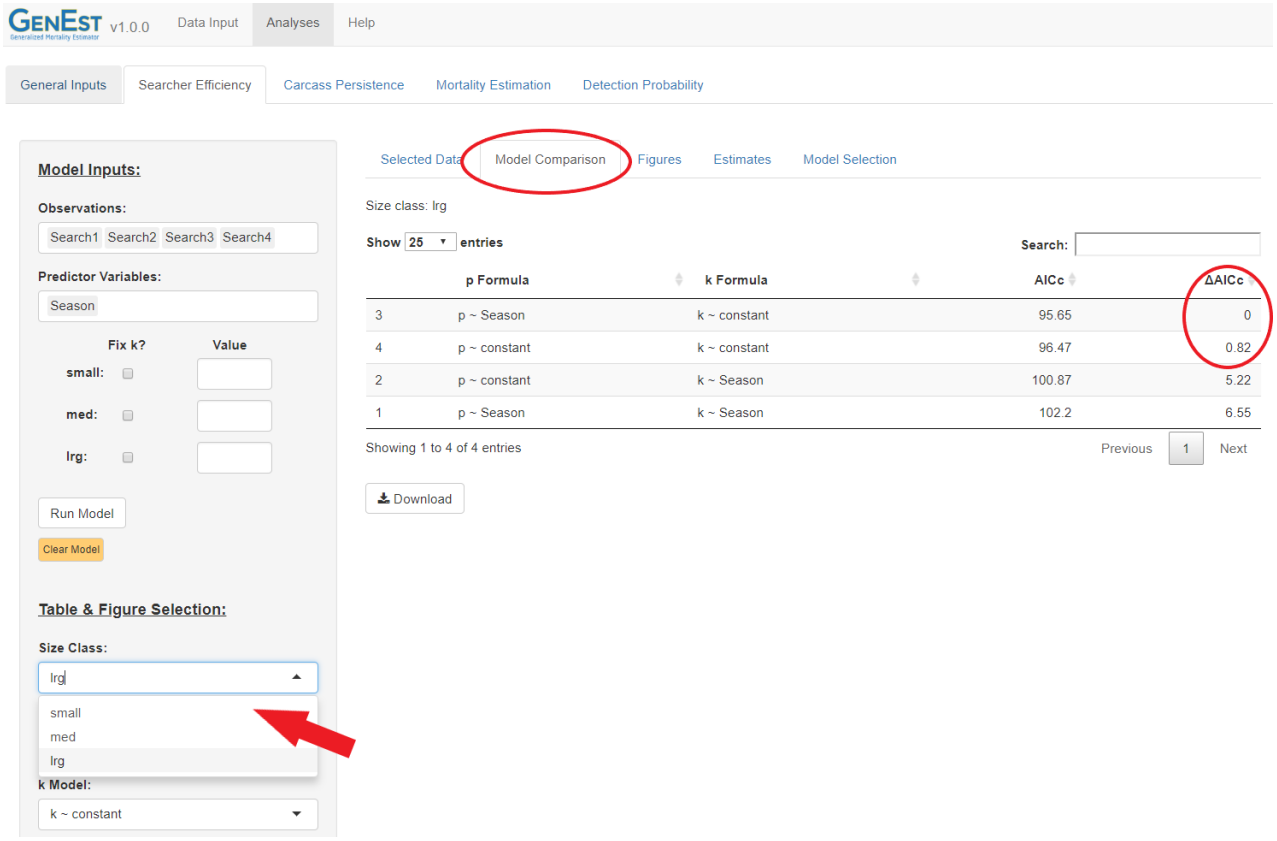

Figure 39. Searcher efficiency model comparison.

The plots (fig. 40) reveal a problem with the $p \sim$ Season; $k \sim$ constant model. In particular, the boxplots for estimated k for the selected model (fig. 40, upper right panel, black boxes) extend from 0 to 1 with no visible median line. This means that the model was unable to find an adequate estimate of k, so a different model should be used. The $p \sim$ constant; $k \sim$ constant model does not have that problem with k and is preferable. In cases where all fitted models have unstable estimates of k, users may wish to choose a fixed k.

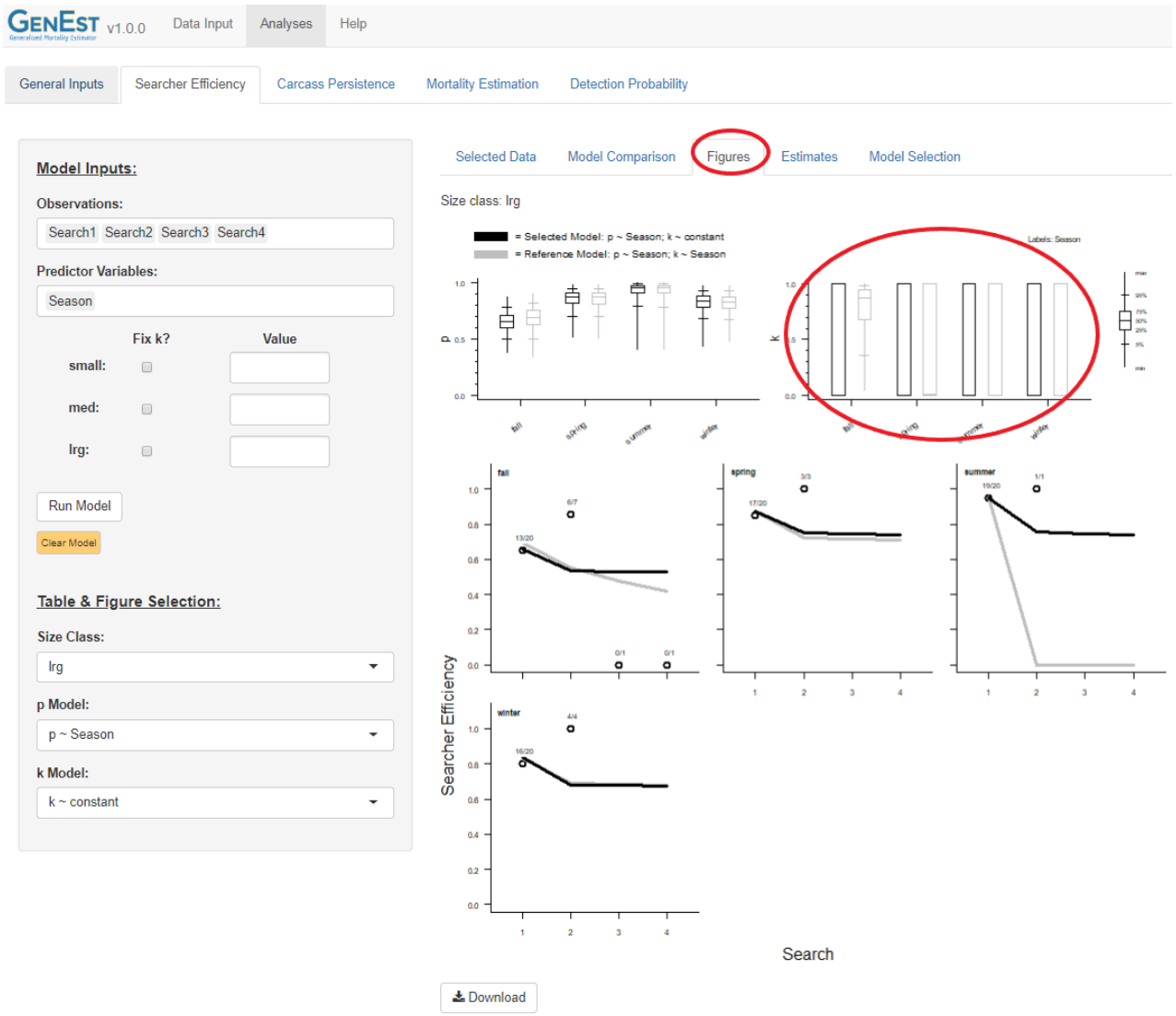

Figure 40. Diagnostic searcher efficiency plots for large carcasses at a solar PV project.

To continue with mortality estimation, users must choose a single search efficiency model for each size class. Navigate to "Analyses" > "Searcher Efficiency" > "Figures" to review a selection of plots to help choose among SE models. Initially the models for the small bird size class are shown. To view other size classes, select them from Size Class menu at left. Tables containing estimates of p and k parameters by model for the size class selected are available in the "Estimates" tab. To tell GenEst which models to use, click on the "Model Selection" tab. For this example, we select the best AICc model (defaults) for small and medium birds and the $p \sim$ constant; $k \sim$ constant model for large birds (because the top AICc model had the unstable k estimate in spring) (fig. 41).

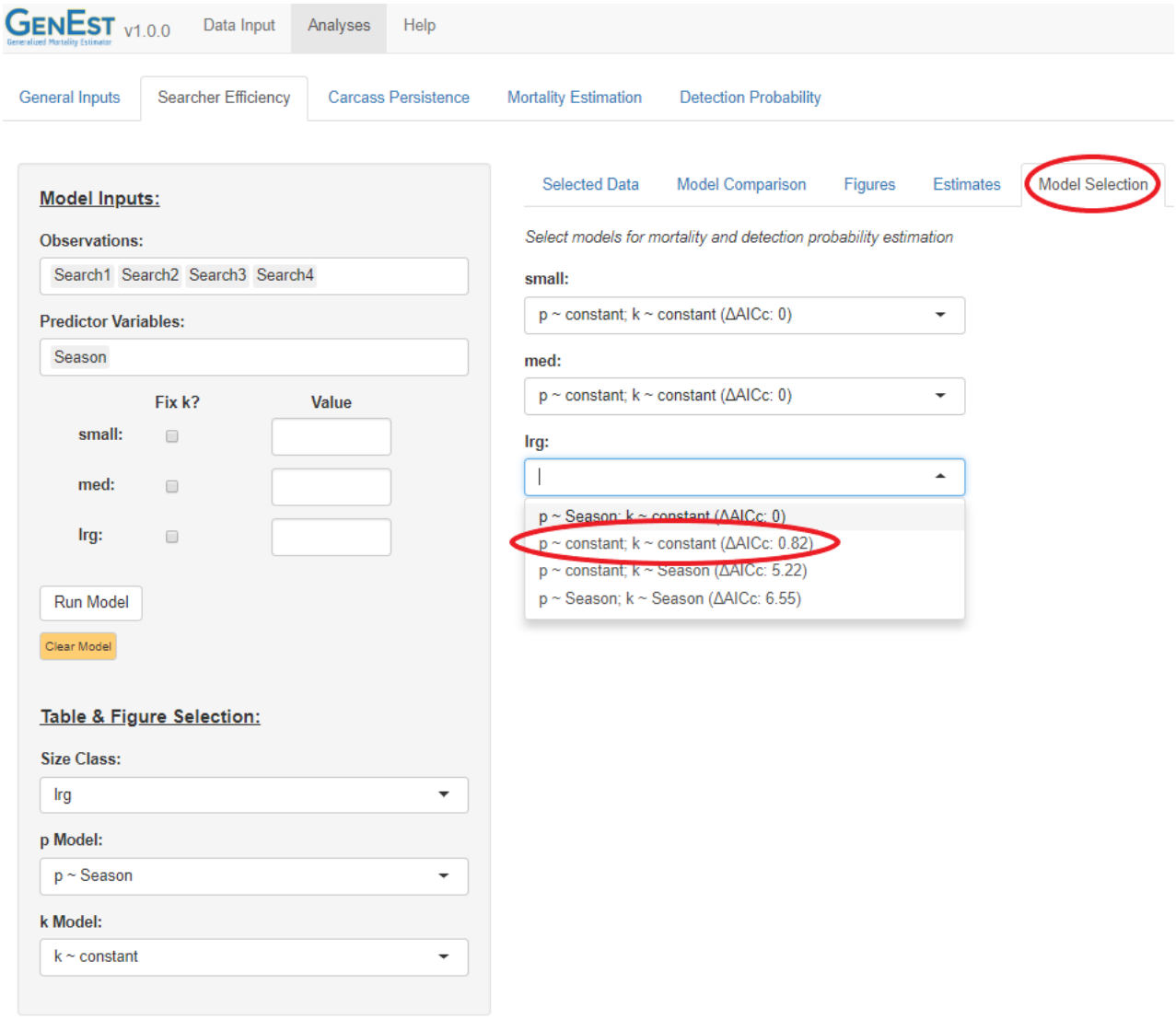

Figure 41. Selecting the SE models for each size class for the solar PV example.

This finishes the Searcher Efficiency stage.

4.2.3. Modeling Carcass Persistence

We wish to consider carcass persistence models with location and scale parameters potentially dependent on season. Our CP file contains a column for `Season` in which each carcass trial was performed, plus columns `LastPresent` and `FirstAbsent`, indicating the left and right endpoints of

the interval during which the carcass went missing. Navigate to the "Carcass Persistence" tab. Select `LastPresent` for "Last Time Present", `FirstAbsent` for "First Time Absent", and `Season` as "Predictor Variables". To consider our model classes of interest, tick the boxes for exponential, Weibull, lognormal, and loglogistic models. Finally, press "Run Model" (fig. 42).

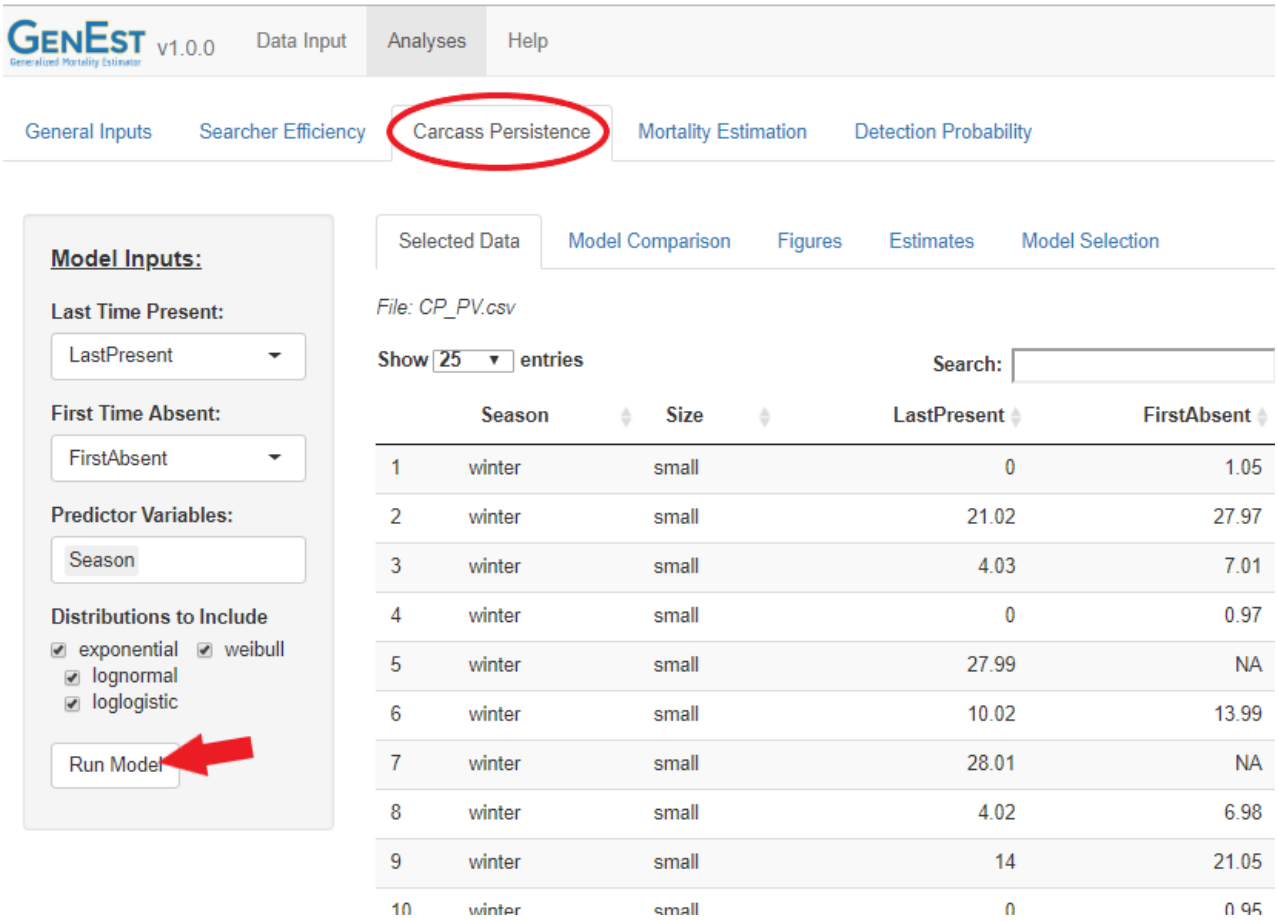

Figure 42. Fitting carcass persistence models for solar PV dataset.

For each size class, a collection of models is generated and ordered by ascending AICc.

For the small carcasses, the best models in terms of AICc are the 2-parameter models (Weibull, lognormal, loglogistic), and the exponential models rank far lower (fig. 43). Many of the 2-parameter models have AICc scores comparable to the "best" (that is, with ΔAICc less than 3 or 4), and other criteria should be used for distinguishing among these. In particular, the diagnostic plots (viewable under the "Figures" tab) can be helpful. The models are generally split into groups according to location and scale formulas, with the Weibull best among the distributions for each particular model form. The figures for the $l \sim$ constant; $s \sim$ constant models (fig. 44) show (1) the 2-parameter models are quite similar, while the exponential model gives a substantially worse fit, and (2) persistence times in the spring seem to be shorter than at other times of the year, so the constant models may not be adequate.

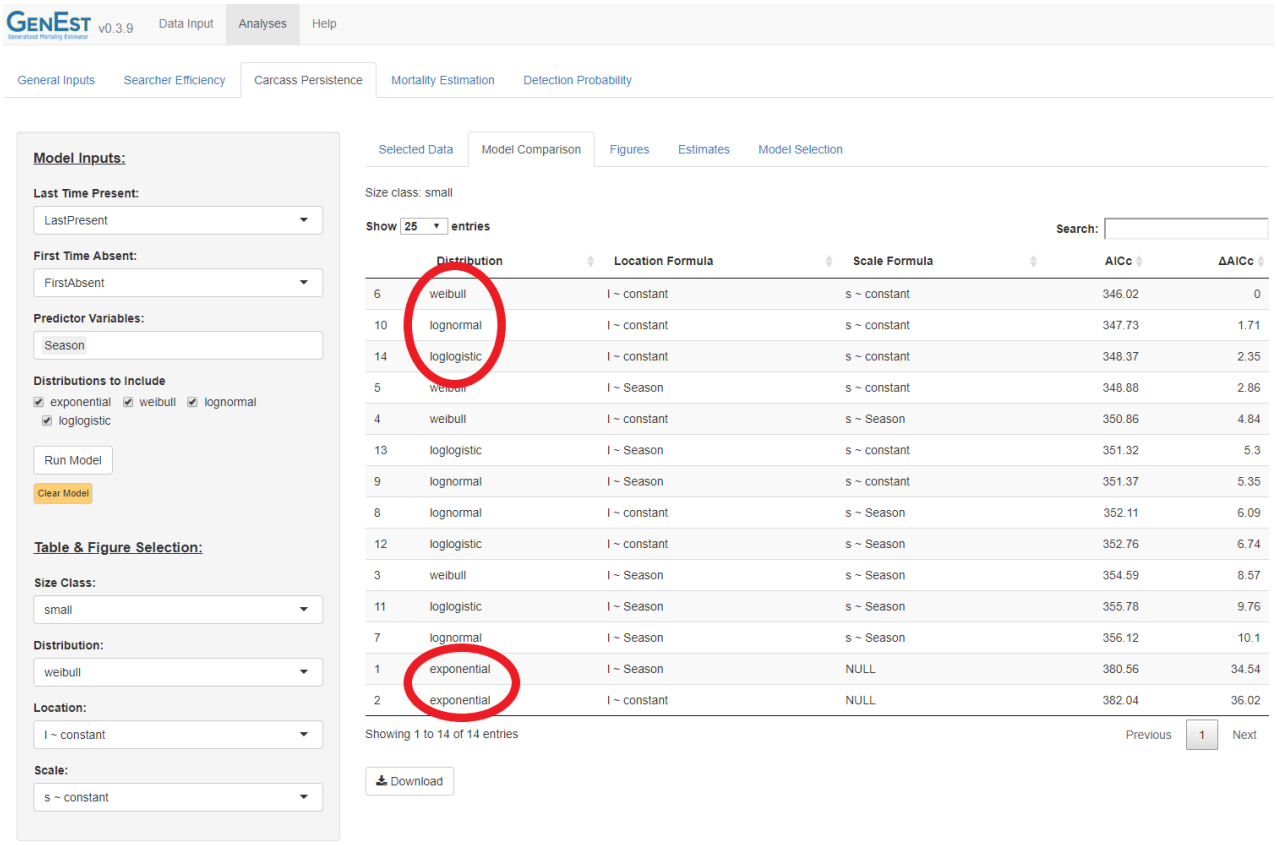

Figure 43. Comparison of carcass persistence models for solar PV example.

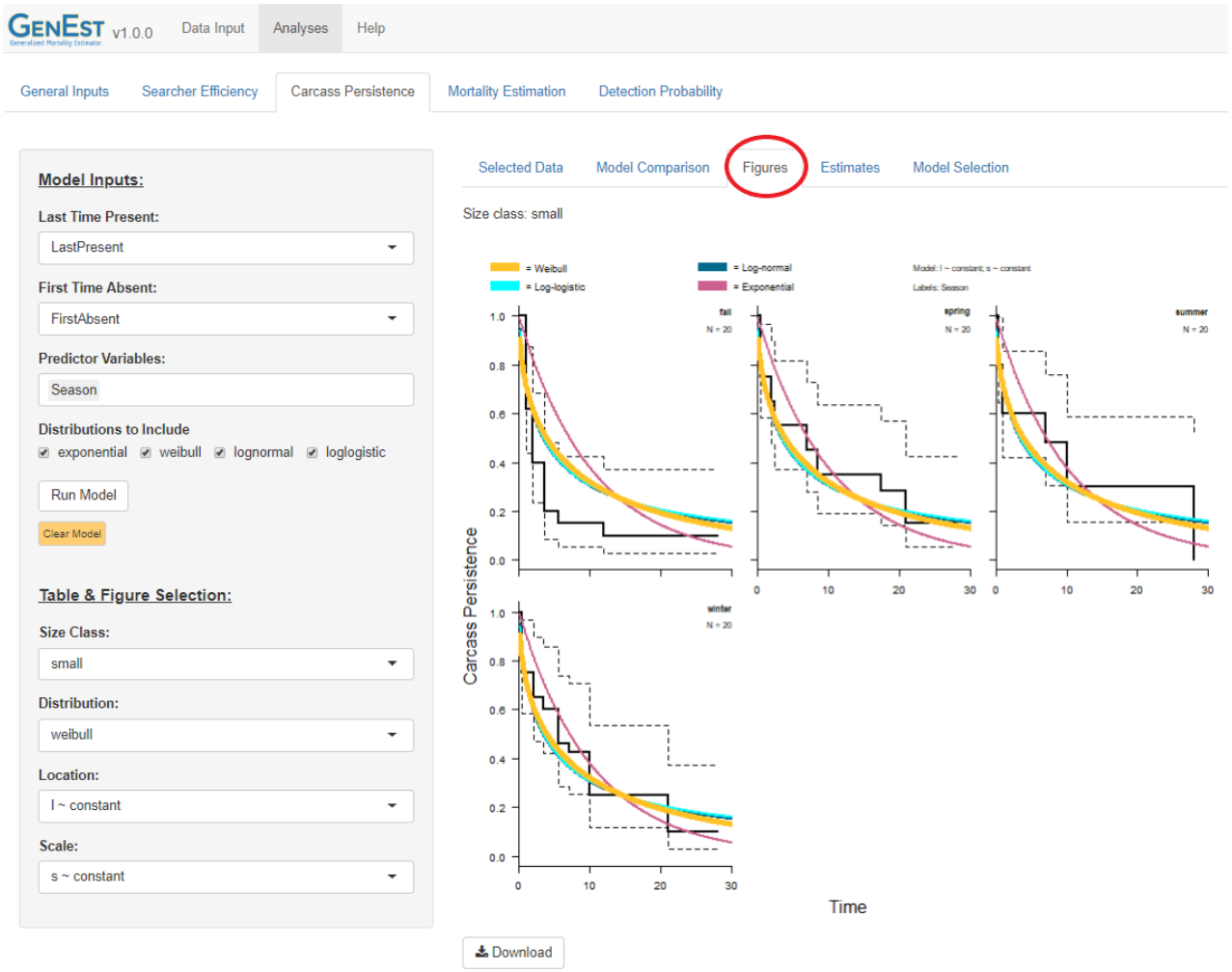

Figure 44. Fitted carcass persistence distributions for the $l \sim \text{constant}; s \sim \text{constant}$ models.

The $l \sim$ Season; $s \sim$ constant models (fig. 45) are comparable to the simpler $l \sim$ constant; $s \sim$ constant models in terms of AICc, but they appear to more accurately reflect differences in CP among seasons.

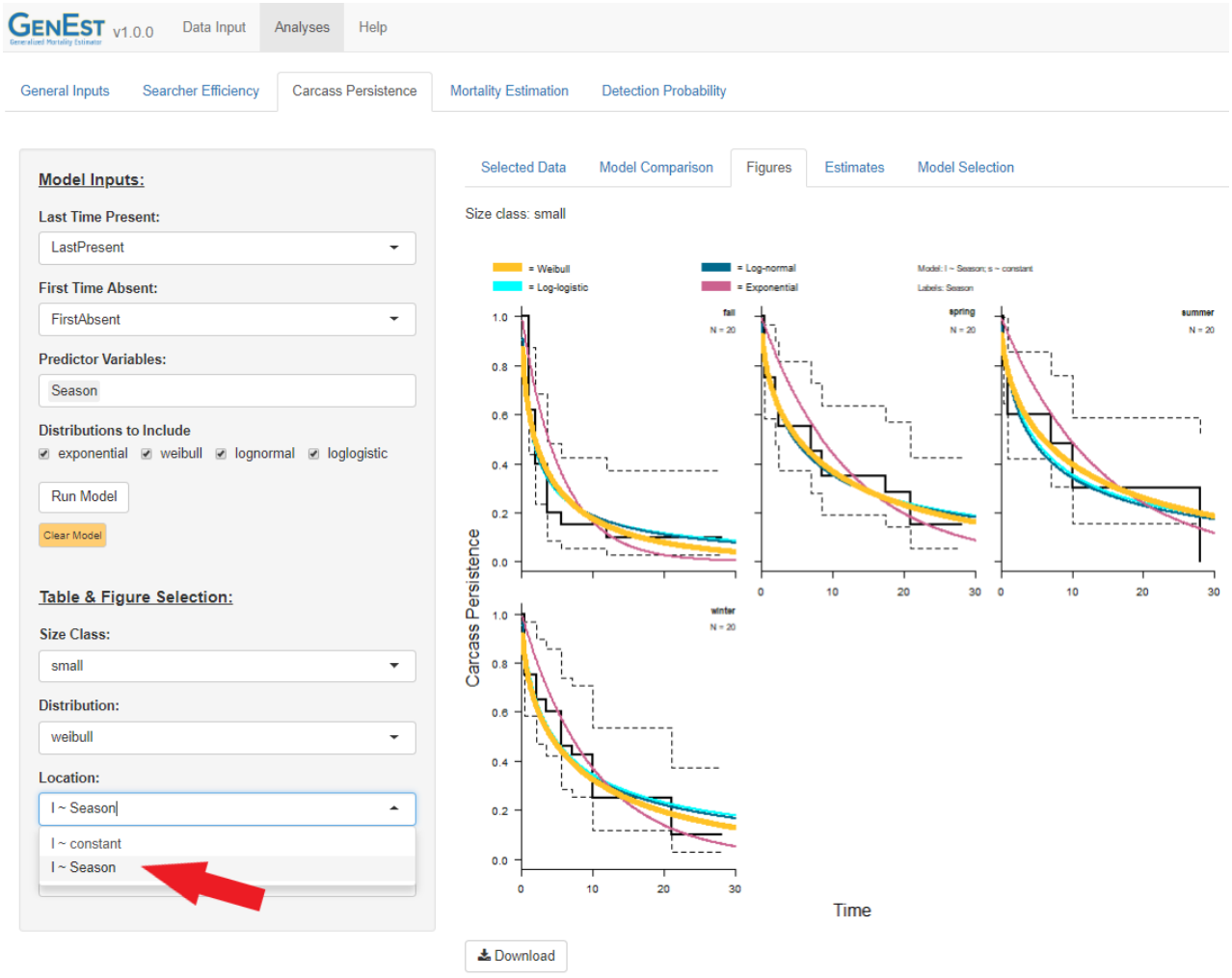

Figure 45. Fitted carcass persistence distributions for the $l \sim$ Season; $s \sim$ constant models.

To view the fit of other models, and for other side classes, use the Tables & Figure Selection drop-down boxes at left to display to the desired figure panel. Final selections on which models to use must be made in the Analyses > Carcass Persistence > Model Selection tab. For the small carcasses, we select the $l \sim$ Season; $s \sim$ constant model, which has ΔAICc = 2.84 (fig. 46) and for medium and large carcasses, we select the model with the lowest AICc score.

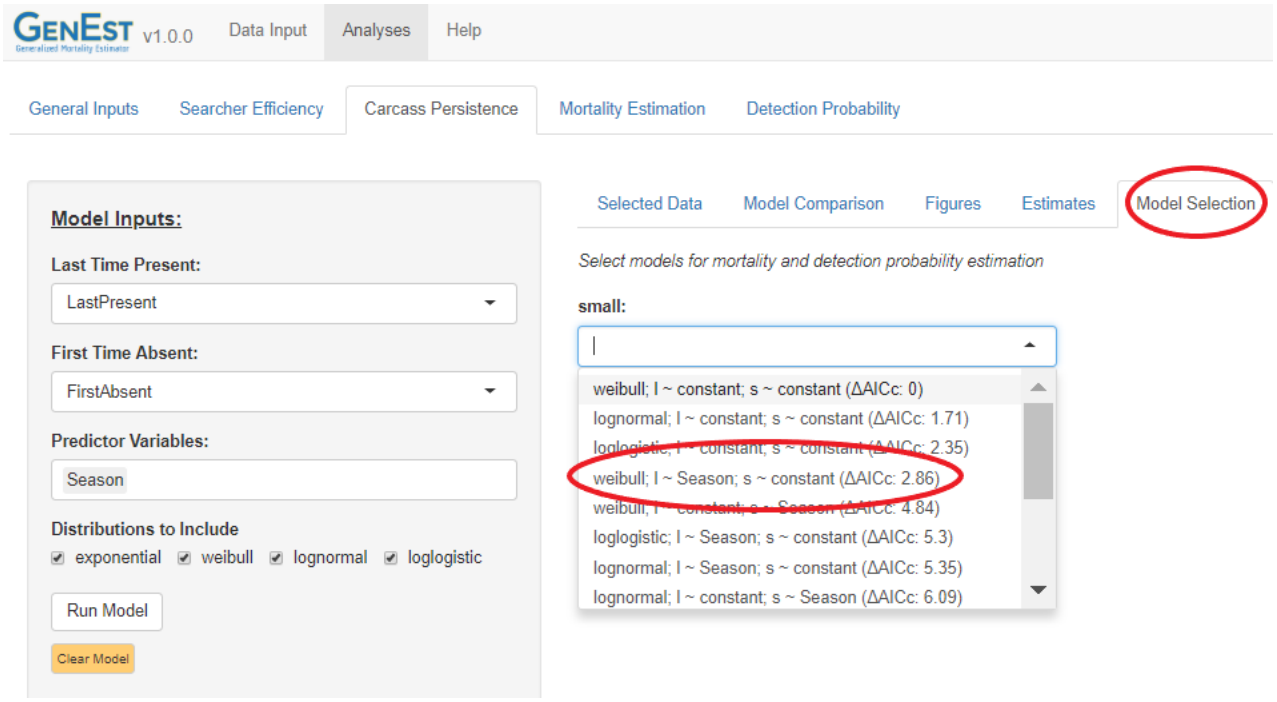

Figure 46. Selection of carcass persistence models for solar PV example.

This finishes the Carcass Persistence stage.

4.2.4. Mortality Estimation

To complete the estimation of mortality, navigate to the "Mortality Estimation" tab. In this example, the whole facility was surveyed, so enter 1 for "Fraction of Facility Surveyed" and select "DateFound" for the "Date Found" column. After estimation is completed, a histogram with overlaid density is shown illustrating mortality in the "Figures" tab (fig. 47). This is the total combined number of fatalities across all carcass size classes.

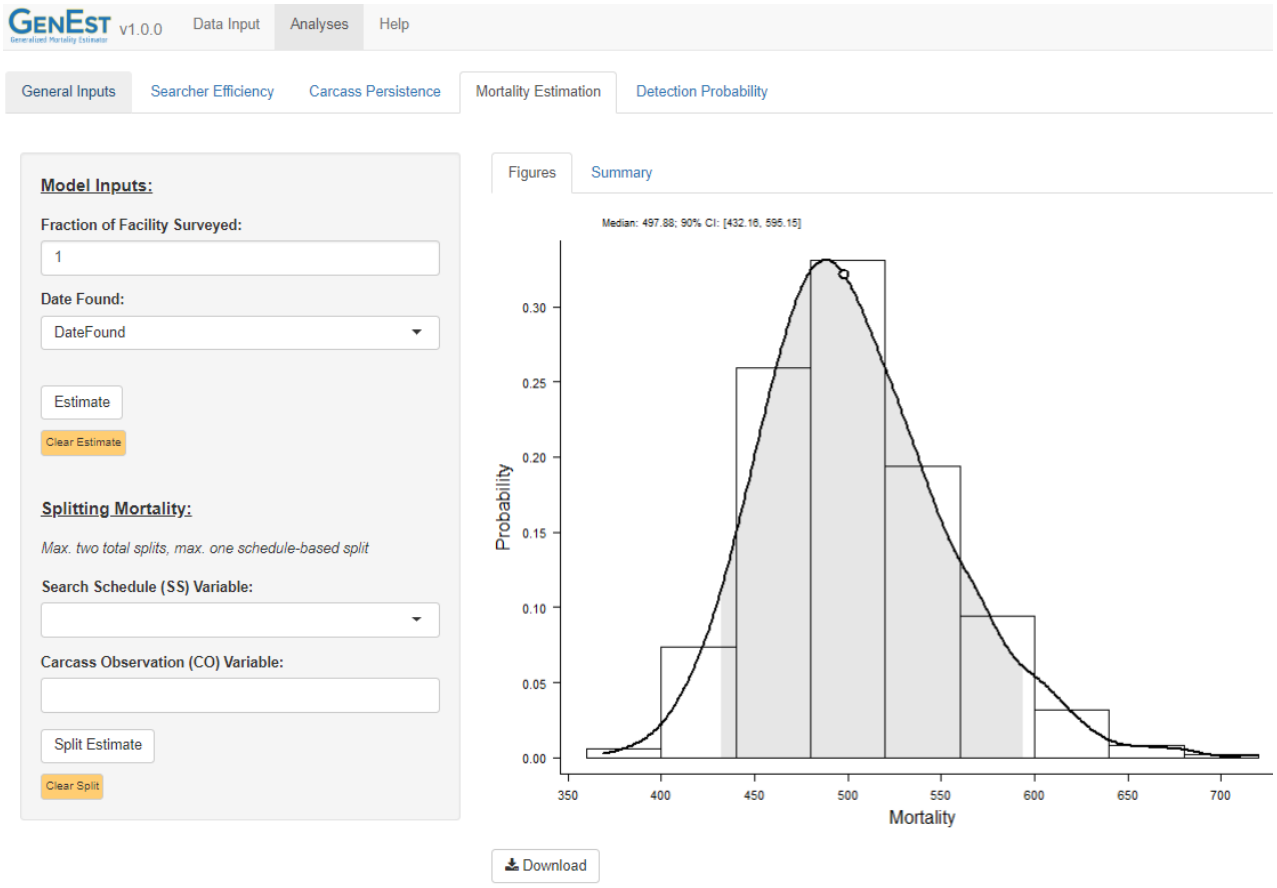

Figure 47. Total mortality in the solar PV example.

Mortality can be split into categories according to: (1) characteristics of the carcasses or where they were found ("Carcass Observation (CO) Variable"), or (2) when carcasses were found ("Search Schedule (SS) Variable"). For example, to see mortality estimates by species, select "Species" as the "Carcass Observation (CO) Variable" and click "Split Estimate" (fig. 48).

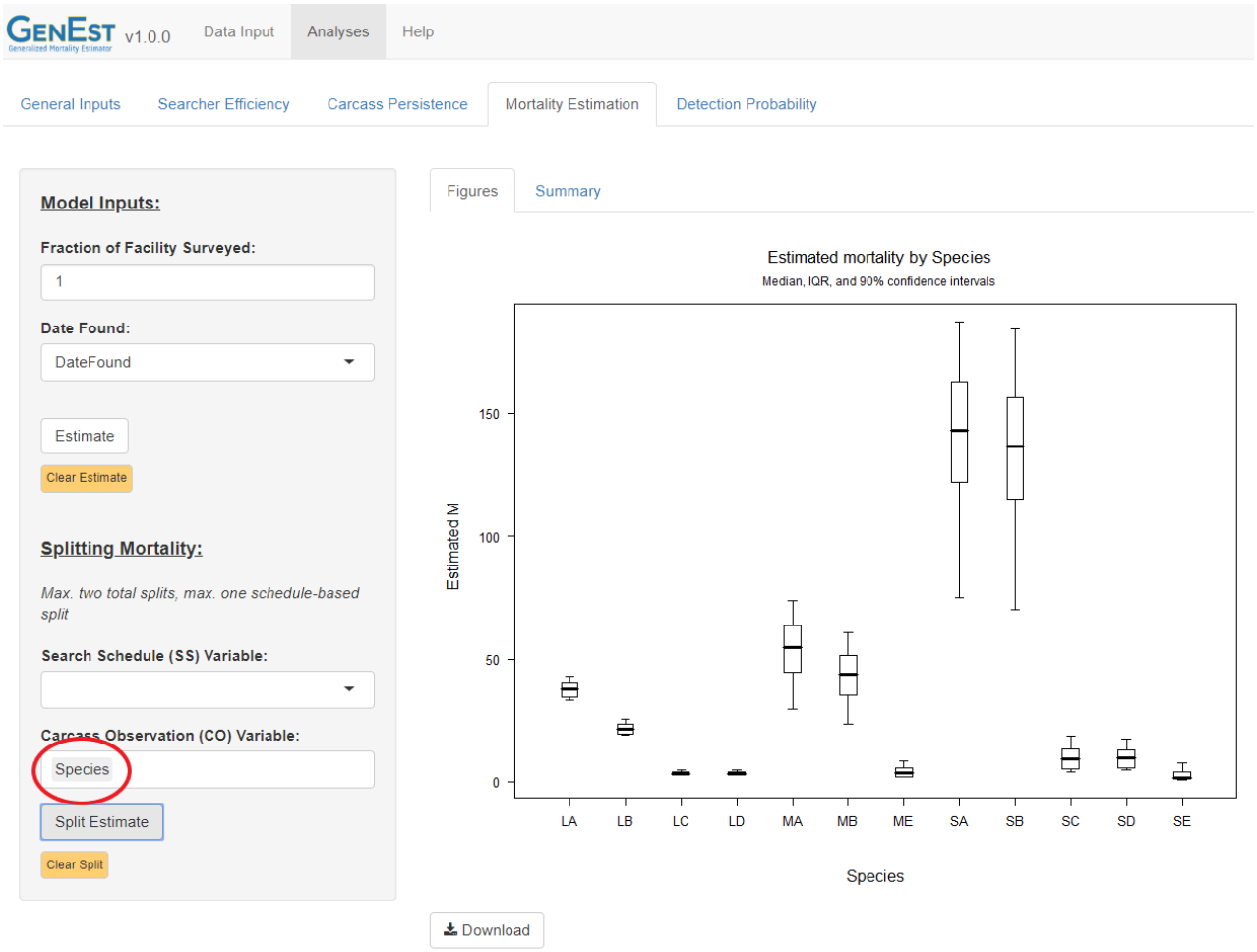

Figure 48. Mortality by species for solar PV example.

To view mortality by species and season, select "Season" as the schedule variable and "Species" as the observation variable and click "Split Estimate" (fig. 49). The figures are too crowded vertically to see well. If we wish instead to view species along the *x*-axis with separate panels for each season, we can click "Transpose Split Plot" for an alternative view (fig. 50).

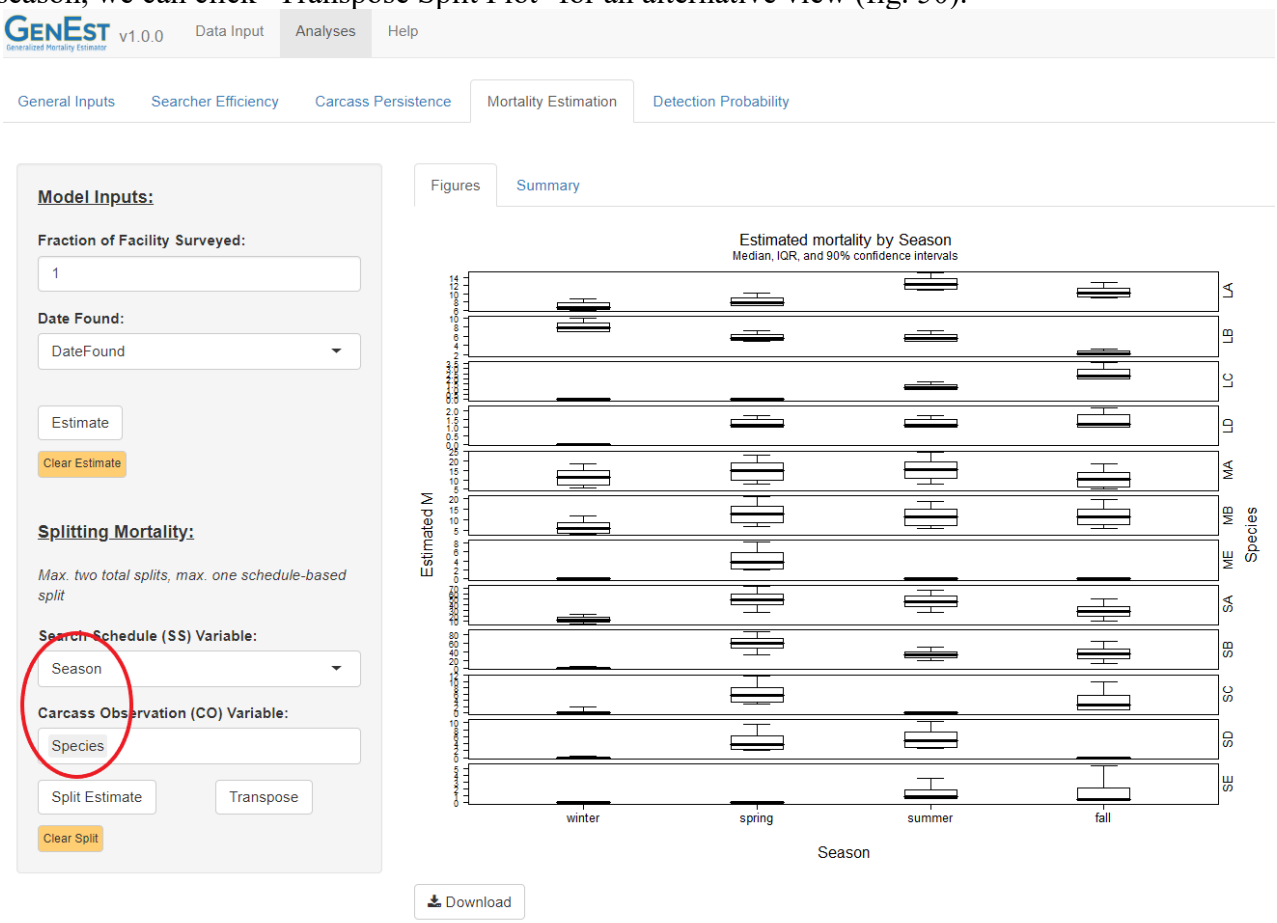

Figure 49. Mortality by species and season for the solar PV example. NOTE: The full figure is quite large and has been truncated to save space in the User Guide. It can be viewed in its entirety in the GUI.

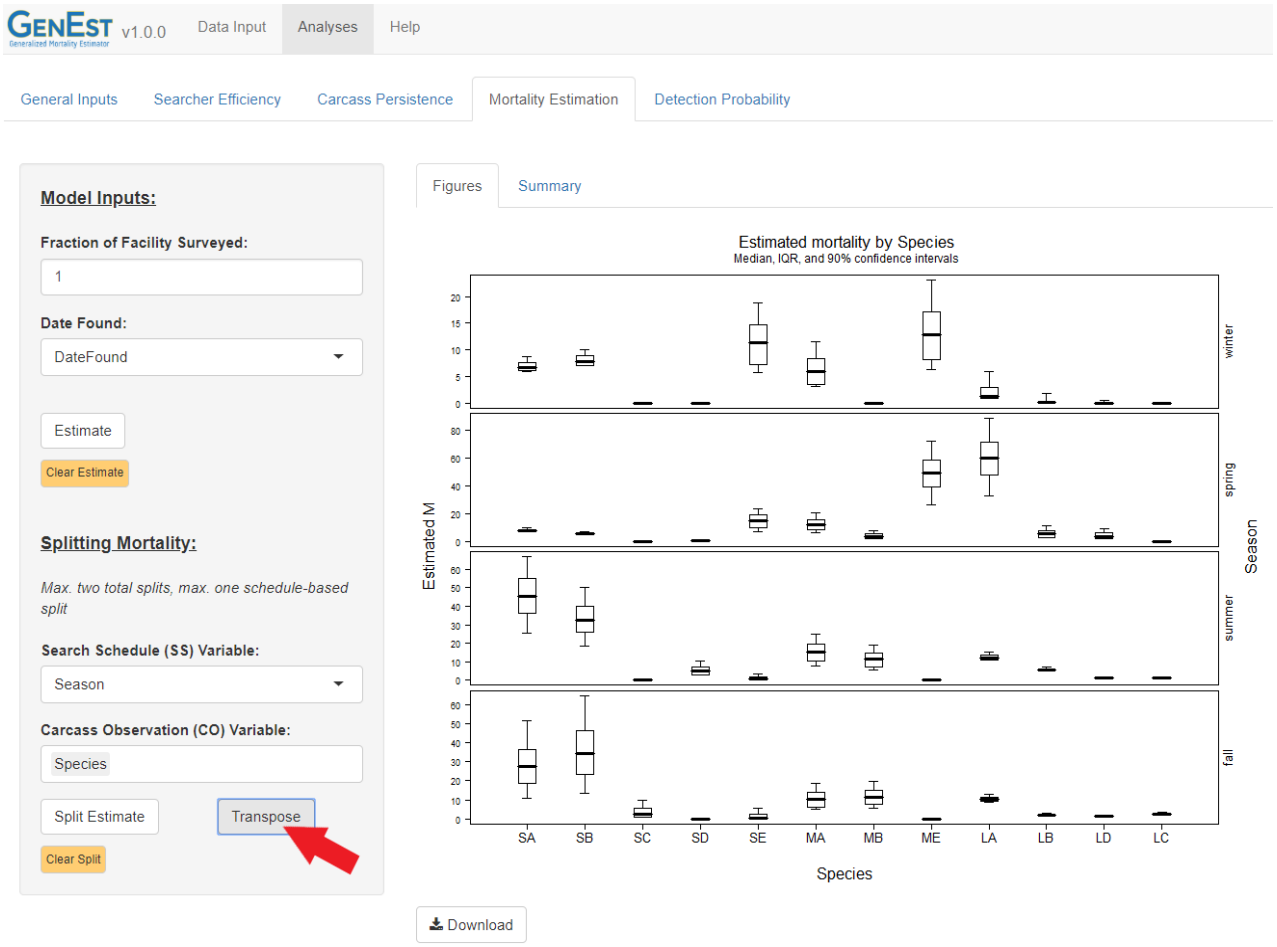

Figure 50. Summary statistics for all mortality estimates are available in tabular format under the "Summary" tab.

4.2.5. Detection Probability

GenEst can estimate detection probability as a function of the covariates selected in the searcher efficiency and carcass persistence sections of the analysis. However, detection probability depends not only on the covariates (carcass size, season, visibility class, etc.) but also on the search schedule (which may vary with unit surveyed) and on carcass arrival time. Thus, it is difficult to calculate simple, meaningful estimates of detection probability. As a convenience, GenEst calculates a generic estimate of detection probability as a function of covariates and search schedule as entered by the user. This estimate of detection probability is not used in the mortality estimates but provides a useful check on what variables may be influencing detection probability and gives a reasonable approximation of what the detection probability was for various covariate combinations.

For example, for small birds, we selected a model that included a seasonal effect for carcass persistence, so detection probability depends on season. To inspect how estimated detection probability varies by season, click on the "Detection Probability" tab (fig. 51). Detection probability depends on search schedule, and users have two options for entering a desired search schedule. The first option is to use the "average" search schedule for site. For this option, click the upper "Create Schedule" button and view the resulting schedule to the right on the same page.

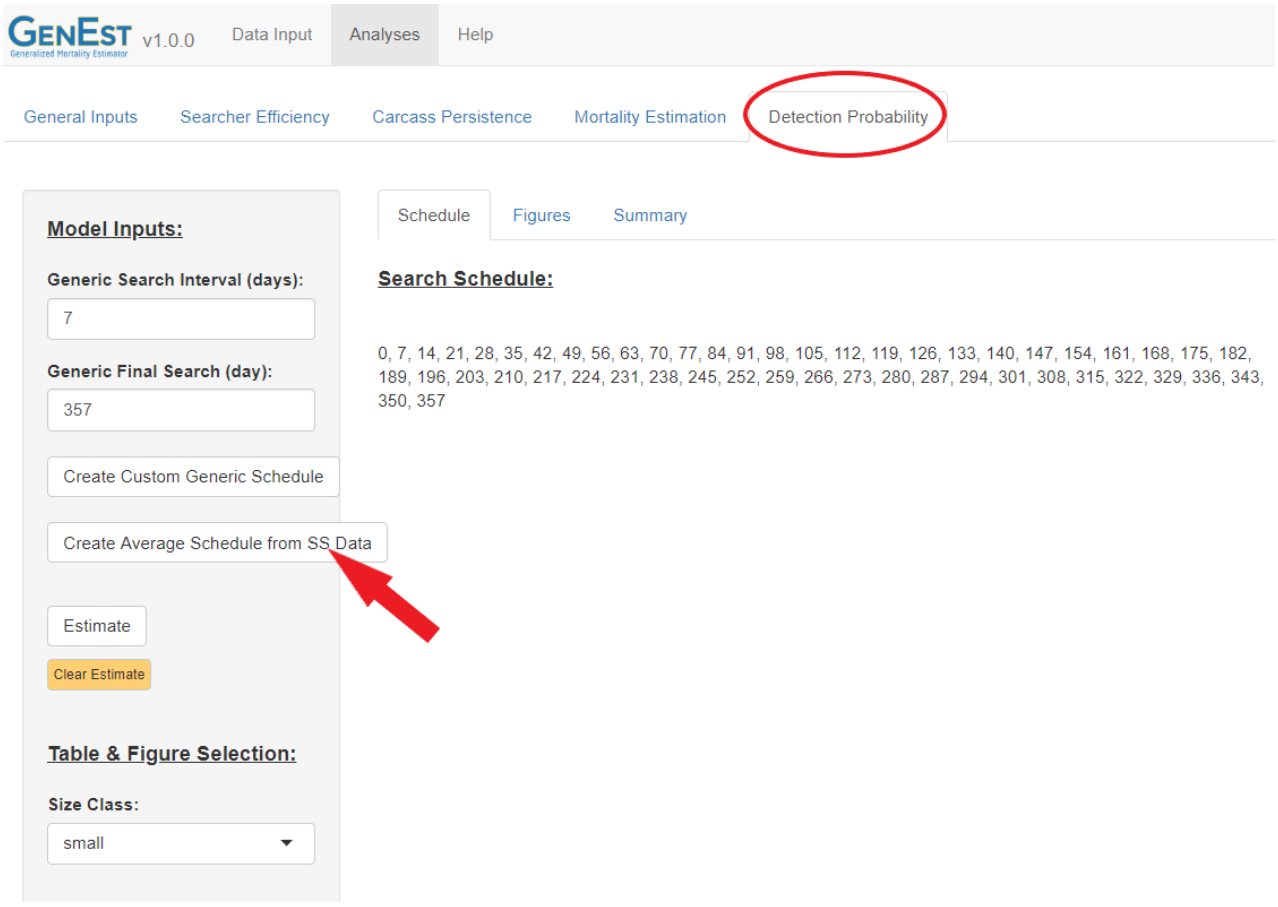

Figure 51. Average search schedule for solar PV example data.

After clicking "Estimate", a table of summary statistics for the first size class is shown under the "Summary" tab (fig. 52). User may choose a different size class to view by selecting it from the menu in the "Table & Figure Selection" section, and the table will be automatically updated. The table can be downloaded and saved as a .csv for use in reports and publications. Graphical representations of summary statistics are provided under the "Figure" tab.

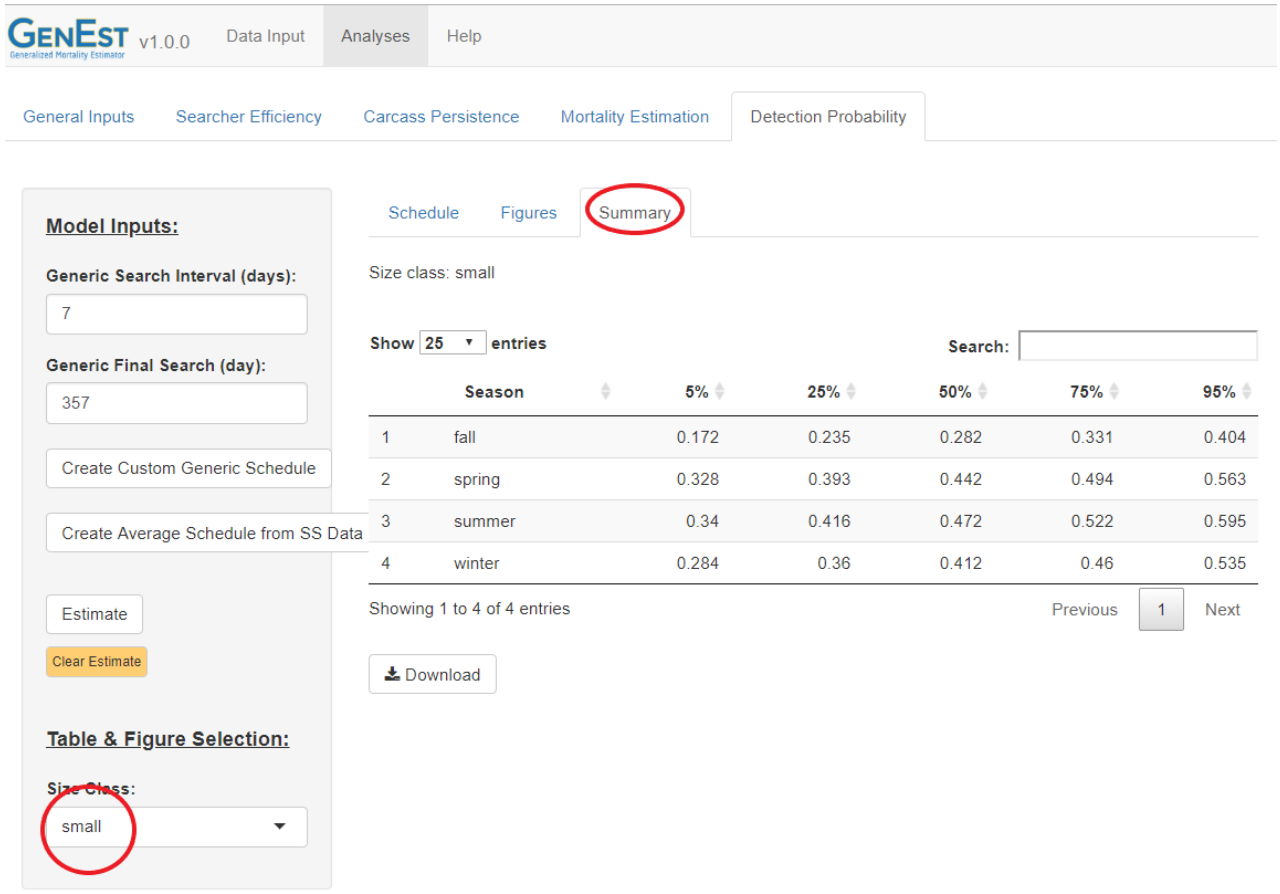

Figure 52. Detection probability by season for small birds (solar PV example).

Acknowledgments

We would like to thank the GenEst steering committee for their tireless support of this project: Cris Hein (Bat Conservation International), Taber Allison (American Wind Wildlife Institute), Jerry Roppe (Avangrid Renewables), Danna Small (Pattern Energy Group),
Rachel London (United States Fish and Wildlife Service), and Dave Cowan (TerraForm Power) Special thanks go to Fränzi Korner Nievergelt of the Swiss Ornithological Institute and Robert Wolpert of Duke University for their contribution to the conceptual development of estimators and consultations on implementation of GenEst throughout our project. We thank the generous reviewers for their insightful evaluation of our statistical approaches and for critical feedback that served to make this software and its user guide more user-friendly, including Gillian Vallejo, David Tidhar, Guillaume Marchais, Christy Humphrey, Andrew Ryckman, Mitch Craig, Sara Weaver, Daniel Maynard, Tyler Stuart, Fränzi Korner-Nievergelt, Janine Aschwanden,

Werner Stefan, Mark Miller, Linda Schueck, and Anonymous. Funding for this project was provided by the U.S. Bureau of Land Management, the Ecosystems Mission Area Wildlife Program of the U.S. Geological Survey, the U.S. Department of Energy, National Renewable Energy Laboratory, WEST Inc. LLC, and Avangrid Renewables for which we are very grateful.

About the Authors and Collaborators

Numerous mortality estimators have been developed to estimate fatalities of birds and bats at wind energy facilities, but failure to meet their inherent assumptions can lead to radically different estimates of mortality, resulting in confusion and poor inferential capacity. The statisticians who developed and/or collaborated on GenEst are the authors of several of the estimators in current use. Recognizing the commonalities and differences among our approaches, we realized they could (and should) all be incorporated under a single umbrella approach. The result is GenEst, or the Generalized Estimator (GenEst).

Juniper Simonis is a data scientist and programmer with DAPPER Stats. They consult on mortality estimation at renewable power facilities, fish passage, and endangered species management.

Dan Dalthorp is a statistician with the U.S. Geological Survey, working primarily on statistical modeling and software development for assessing bird and bat fatalities at wind and solar facilities.

Manuela Huso is a research statistician with the U.S. Geological Survey. She works primarily on issues relating to bird and bat fatalities at wind and solar facilities.

Jeff Mintz is a statistician and recently earned his MS in statistics from the Oregon State University.

Lisa Madsen is associate professor of statistics at Oregon State University. Her research focuses on count data and simulation methods with applications to ecology and the environment.

Paul Rabie is a statistician with Western Ecosystems Technology, Inc. Paul has extensive experience estimating impacts to wildlife due to renewable energy facilities.

Jared Studyvin is a statistician with Western Ecosystems Technology, Inc. He has done lots of consulting work in the mortality estimation of renewable sites for both common and rare species.

References Cited

Attali, D., 2018, shinyjs—Easily improve the user experience of your shiny apps in seconds—R package, version 1.0: Web page, accessed October 29, 2018, at https://CRAN.R-project.org/package=shinyjs.

Bengtsson, H., 2018, matrixStats—Functions that apply to rows and columns of matrices (and to vectors)—R package, version 0.53.1: Web page, accessed October 19, 2018, at https://CRAN.R-project.org/package=matrixStats.

Brown, C., 2017, sticky—Persist attributes across data operations—R package, version 0.5.2: Web page, accessed October 19, 2018, at https://CRAN.R-project.org/package=sticky.

Burnham, K.P., and Anderson, D.R., 2004, Multimodel inference—Understanding AIC and BIC in model selection: Sociological Methods and Research, v. 33, p. 261–304.

Chang, W., Cheng, J., Allaire, J.J., Xie, Y., and McPherson, J., 2018, shiny—Web application framework for R—R package, version 1.1.0: Web page, accessed October 19, 2018, at https://CRAN.R-project.org/package=shiny.

Dalthorp, D., 2018, cbinom—Continuous analog of a binomial distribution—R package, version 1.3: Web page, accessed October 19, 2018, at https://rdrr.io/cran/cbinom/.

Dalthorp, D., Huso, M., and Dail, D., 2017, Evidence of absence (v2.0) software user guide: U.S. Geological Survey Data Series 1055, 109 p., https://doi.org/10.3133/ds1055.

Dalthorp, D., Madsen, L., Huso, M., Rabie, R., Wolpert, R., Studyvin, J, Simonis, J., and Mintz, J., 2018, GenEst statistical models—A generalized estimator of mortality: U.S. Geological Survey Techniques and Methods, book 7, chap. A2, 13 p.

Eddelbuettel, D., and Francois, R., 2011, Rcpp—Seamless R and C++ integration: Journal of Statistical Software, v. 40, no. 8, p. 1–18, http://www.jstatsoft.org/v40/i08/.

Genz, A. Bretz, B., Miwa, T., Mi, X., Leisch, F., Scheipl, F., and Hothorn, T., 2018, mvtnorm—Multivariate normal and t distributions—R package, version 1.0-8: Web page, accessed October 19, 2018, at http://CRAN.R-project.org/package=mvtnorm.

Grolemund, G., and Wickham, H., 2011, Dates and times made easy with lubridate: Journal of Statistical Software, v. 40, no. 3, p. 1–25, http://www.jstatsoft.org/v40/i03/.

Hankin, R.K.S., 2006, Special functions in R—Introducing the gsl package: R News, v. 6, no. 4.

Huso, M., Dietcsch, T., and Nicolai, C., 2016, Mortality monitoring design for utility-scale solar power facilities: U.S. Geological Survey Open-File Report 2016-1087, 4 p., http://dx.doi.org/10.3133/ofr20161087.

Huso, M.M.P., Som, N., and Ladd, L., 2012, Fatality estimator user's guide: U.S. Geological Survey Data Series 729, 22 p., https://dx.doi.org/10.3133/ds729.

Kalbfleisch, J.D., and Prentice, R. L., 2002, The statistical analysis of failure time data: Hoboken, New Jersey, Wiley, 462 p.

Korner-Nievergelt, F., Behr, O., Brinkmann, R., Etterson, M.A., Huso, M.M.P., Dalthorp, D., Korner-Nievergelt, P., Roth, T., and Niermann, I., 2015, Mortality estimation from carcass searches using the R-package carcass—A tutorial: Wildlife Biology, v. 21, no. 1, p. 30–43.

Peron, G., Hines, J.E., Nichols, J.D., Kendall, W.L., Peters, K.A., and Mizrahi, D.S., 2013, Estimation of bird and bat mortality at wind-power farms with superpopulation models: Journal of Applied Ecology, v. 50, no. 4, p. 902–911, https://doi: 10.1111/1365-2664.12100.

Perry, P.O., 2017, corpus—Text corpus analysis—R package, version 0.10.0: Web page, accessed October 19, 2018, at https//CRAN.R-project.org/package=corpus.

R Core Team, 2018, R—A language and environment for statistical computing, v3.5.1: Vienna, Austria, R Foundation for Statistical Computing web page, accessed October 19, 2018, at https://www.R-project.org/.

RStudio, Inc., 2017, htmltools—Tools for HTM—R package, version 0.3.6: Web page, accessed October 19, 2018, at https://CRAN.R-project.org/package=htmltools.
Shoenfeld, P., 2004, Suggestions regarding avian mortality estimation: Report prepared for the West Virginia Highlands Conservancy, West Virginia, by Mountaineer Wind Energy Center Technical Review Committee, 6 p.
Warnes, G.R., Bolker, B., and Lumley, T., 2015, gtools—Various R programming tools—R package, version 3.5.0: Web page, accessed October 19, 2018, at https://CRAN.R-project.org/package=gtools.
Williams, K.A., Frederick, P.C., and Nichols, J.D., 2011, Use of the superpopulation approach to estimate breeding population size—An example in asynchronously breeding birds: Ecology, v. 92, no. 4, p. 821–828, https://doi.org/10.1890/10-0137.1.
Wolpert, R., 2015, ACME—A partially periodic estimator of avian and chiropteran mortality at wind turbines: Cornell University Library, https://arxiv.org/abs/1507.00749.
Xie, Y., 2018, DT—A wrapper of the JavaScript library 'DataTables'—R package, version 0.4: Web page, accessed October 19, 2018, at https://CRAN.R-project.org/package=DT.

5. Appendix

5.1. Akaike Information Criterion (AIC) and Model Selection

Akaike information criterion (AIC) is a statistical score for the quality of a model fit (Burnham and Anderson, 2004). Its companion, AICc, is similar but provides a small-sample bias correction. Complicated models that use many parameters may fit the data more closely than a simpler model but are penalized for their complexity and relative instability. AICc scores have meaning only when compared with the AICc's of other models that are fit using the same data. ΔAICc (the difference between each model's AICc and the AICc of the "best") provides a rough but useful guide for model selection but should not be relied upon as definitive. Its utility is in identifying relatively poor models and in narrowing the choice of plausible models to a manageable number. Preference should normally be given to models with ΔAICc less than 6 or 7. Models with differences of less than 3 or 4 are largely considered indistinguishable by this measure. Choices among such models should be based on other criteria, such as diagnostic plots and an understanding of the search process that generated the data.

5.2. Number Format and .csv Files

Input data files must be comma-separated files (`.csv`) or simple text files (`.txt`). Data fields may be separated by either commas (`,`) or semicolons (`;`). For comma-separated data, it is assumed that the decimal separator is a period or radix (`.`). For example, 3 + 47/100 would be represented as 3.47, as is common in most of North America and Asia. For semicolon-separated data, it is assumed that the decimal separator is a comma (`,`), with 3 + 47/100 represented as 3,47, as is common in most of Europe and South America.

5.3. Dates

Dates should be formatted as integers representing the year, month, and day (respectively), separated by a hyphen (yyyy-mm-dd) or as integers representing month, day, and year (respectively), in format yyyy/mm/dd, mm/dd/yyyy, or dd/mm/yyyy. The latter two are default formats for Excel dates, but care must be taken when using this format because GenEst will not accept ambiguous dates. For example if a column of dates contains exclusively dates that are ambiguous such as 05/04/2012 (which could be May 4 or April 5), GenEst will not allow the column to be used for dates. Dates in a column must all be in the same format.

5.4. File Names

In general, input files can have any names that are allowed by the operating system you are working in. In Windows, file names may be any length but may not contain characters ASCII control characters (`\:/*?"<>`). In Unix-like operating systems, file names may not include the forward slash character (`/`). In a Mac OS the colon (`:`) and forward slash (`/`) are not allowed, and file names may not begin with a period (`.`).

A convenient naming convention, which we have adopted in the example data sets, is to include SE, CP, DWP, SS, and CO in the names of the searcher efficiency, carcass persistence, density-weighted proportion, search schedule, and carcass observation files, respectively. This convention is not required, but it does facilitate the management of data files. Another useful convention is to avoid using spaces in file names and use underscore (_) or camel case instead (that is, using capital letters rather than spaces to delineate words; for example, `DataFile.csv` instead of `data file.csv`).

5.5. Variables and Column Names

In general, naming of columns and factor levels should be consistent across files. Users have great latitude in assigning variable and column names, but there are some requirements. Column names may contain combinations of letters, numbers, underscores (_), and periods (.), but names must not begin with a number or with a period immediately followed by a number. Names are case-sensitive, so "season" would be different from "Season". Column names and data values must not contain hyphens (-), and predictor levels should not contain periods (.). In addition, unit identifiers and names of columns for units must not contain spaces. For example, "`Turbine`" would be fine but "`Turbine Number`" would not be. Similarly, "`t1`" (for turbine #1) would be acceptable, but "`t 1`" or "`t-1`" would not be.

All column names and data values are case sensitive, so "`turbine`" would be distinct from "`Turbine`", and "`Large`" would be different from "`large`". Be careful! This is a common source of error.

Carcass identifier (ID) columns are required in the SE, CP, and CO data files. Each carcass should be assigned a unique ID within each file. (Note that if the first column in a spreadsheet saved as a `.csv` file is named `ID`, Microsoft Excel® may mis-interpret the file as an SYLK file and will not be able to read it.)

If size class is to be used as a variable, identically-named size class columns must be included in the SE, CP, and CO files. In addition, each carcass size represented in the size class column of the CO data must have a corresponding column in the DWP file. For example, if one `large` and three `small` carcasses are found in the carcasses surveys, the DWP file must contain columns named `large` and `small` to enable proper DWP adjustments for each of the carcasses.

Variables used in the searcher efficiency (SE) and carcass persistence (CP) models must have corresponding variables in the carcass survey data (CO). For example, if `season` is used as a covariate for carcass persistence or searcher efficiency, then the CO file must also have a `season` column with levels (for example, `spring`) that can be found in the corresponding SE or CP file.

5.6. Some Terminology Used in Searcher Efficiency and Carcass Persistence Modeling

Formatting and terminology for searcher efficiency (SE) and carcass persistence (CP) models closely follow standard R formatting of models that use the `formula` convention, but GenEst has some additional conventions. In GenEst, size classes and covariates for SE and CP are assumed to be categorical, so, for example, sizes can be `sml` and `lrg` but not 12.61 and 974.3. Should a user include a numeric covariate in these models, each unique value would be taken as a separate factor level rather than numeric values on a continuum and would likely result in an error.

Suppose `A` and `B` are selected as predictors. Then, GenEst would fit five different models for each parameter to be estimated (p and k for SE, l and s for CP). The simplest model is a constant model which makes no distinctions among covariate levels, so, for example, searcher efficiency would be assumed to be the same regardless of season, visibility, search team, or any other factor. For the p parameter, the form of the constant model would be $p \sim \mathrm{constant}$ (pronounced "p is constant"). Although p is constant in this model, its value is unknown. GenEst will use the searcher efficiency data to estimate its value as $\hat{p}$, with the "hat" distinguishing the estimate ($\hat{p}$) from what it is estimating (p). In all its estimates, GenEst accounts for estimator uncertainty, giving confidence intervals for what the data and model tell us about the (unknown) value of the parameter(s). The next simplest models are the single-factor models, for example, $p \sim A$ (pronounced "p is a function of A") or $p \sim B$. The most complicated model is the interaction or full-cell model, $p \sim A * B$. In this model, p may take on a

different value in each *cell* or combination of covariate levels, like easy visibility in spring, moderate visibility in fall, and easy visibility in winter. A compromise between the single-factor models and the full-cell model is the additive model: $p \sim A + B$, where A and B both influence p but the effect of one covariate does not depend on the level of the other.

The term *cells* appears frequently in the analyses and discussions. It refers to combinations of covariate levels. For example, if two covariates are `season` with `sp`, `su`, and `fa` as possible values (or *levels*) and `visibility` with levels `easy` and `moderate`, then there would be 6 cells: `sp.easy`, `sp.moderate`, `su.easy`, `su.moderate`, `fa.easy`, and `fa.moderate`.

5.7. Searcher Efficiency

The first few rows of data from a hypothetical SE field trial are shown in table 2. Each row represents the fate of a single carcass in the searcher efficiency trials. Required columns include, at a minimum, a column with unique carcass identifiers[5] and at least one column indicating whether carcasses that were placed as part of the field trials were discovered or not discovered in later carcass surveys. Columns `s1, s2,` … show the fate of carcass `pkID` on the 1st, 2nd, … searches after the carcass was placed. A value of 1 indicates that the carcass was discovered, a 0 indicates that the carcass was present but not discovered, and NA indicates that the carcass was not present (usually because it was discovered in a previous search or removed by scavengers). Values other than 0 or 1 are ignored and can be used to indicate carcasses that were not present at the time of search (perhaps because they had already been discovered in a previous search or been removed by scavengers) or that were not included in the area searched during that specific carcass survey (perhaps because the carcass was at a unit that was not scheduled for a search or because some aberrant conditions prevented a search where the carcass was placed).

[5] When using Microsoft Excel to open a `.csv` file whose first column is named "ID", note that Excel may misinterpret the file as a `.slk` file and show some warning dialog boxes. It should be able to open the file properly anyway, but to avoid the hassle of the warnings, name the first column something besides "ID". For example, "I.D." or "SE_ID" or "carcID" would work fine.

If the *k* parameter (section 3.2.1) is to be estimated from searcher efficiency trials, carcasses that are not discovered in the first search (`s1` in table 2) should be left in the field for discovery in later searches. Carcass fates in each subsequent search are represented in additional, optional columns (`s2-s5` in table 2). For example, carcass `pk1` was present but not discovered (0) on the first search (`s1`) after being placed in the field and was either not present or its location was not searched (NA) on subsequent searches (`s2-s5`). After a carcass is found (1), it is considered unavailable for discovery subsequent searches (NA). Carcasses in this example were left in the field for up to 5 searches after carcass placement.

Table 2. Format for searcher efficiency data.

pkID	Size	Season	s1	s2	s3	s4	s5
pk1	bat	spring	0	NA	NA	NA	NA
pk2	bat	spring	0	0	1	NA	NA
pk3	bat	fall	1	NA	NA	NA	NA
pk4	bat	fall	1	NA	NA	NA	NA
pk5	sml	spring	0	NA	NA	NA	NA
pk6	sml	spring	1	NA	NA	NA	NA
pk7	sml	fall	1	NA	NA	NA	NA
pk8	sml	fall	0	NA	NA	NA	NA
pk9	lrg	spring	0	NA	NA	NA	NA
pk10	lrg	spring	0	1	NA	NA	NA
pk11	lrg	fall	0	1	NA	NA	NA
pk12	lrg	fall	0	1	NA	NA	NA
...							

If search data are provided for one search only, *k* cannot be estimated from the data, and users must provide an assumed, fixed value of *k* if an estimation of mortality or detection probability is desired (section 3.2.2).

The SE file may contain optional columns including one for carcass size class (`Size` in this example) and any number of covariates (`Season` in this example). Users can specify in the GUI which column (if any) to use for carcass size class and which columns (if any) to use as covariates.

Users may choose whatever column names they desire, but the following conventions are expected when estimating mortality or detection probabilities:

1. if mortality estimates are to be distinguished by class, the SE, CP, and CO files must each have a carcass class column and the column name must be the same in each file. In addition, the DWP file must have separate columns for each level of carcass class (for example `sml`, `med`, `lrg`) represented in the CO file;
2. covariate columns that are to be included in the SE model must be found in both the SE file and the CO file.

5.8. Carcass Persistence

The first few rows of a hypothetical CP field trial are presented in table 3.

Table 3. Carcass Persistence Data Format.

cpID	Season	LastPresent	FirstAbsent
cp1	spring	10.03	14.09
cp2	spring	3.98	7.05
cp3	spring	20.95	28.05
cp4	fall	0	2.84
cp5	fall	27.41	Inf
cp6	fall	17.06	17.06
...			

Required columns:

1. A column containing `LastPresent`: a numeric value from 0 to the value of `FirstAbsent`.
2. A column containing `FirstAbsent`: a numeric value from the value of `LastPresent` to `Inf`.

In this example, fresh carcasses were placed in the field over the course of the entire monitoring period, evenly divided among seasons (spring, summer, fall). Carcasses were checked (approximately) 1, 2, 3, 4, 7, 10, 14, 21, and 28 days after placement in the field (exact times were entered as decimal fractions of days after placement).

Exact scavenging times are not known, but a carcass that was present at one check and absent at the next check is assumed to have been scavenged at some point in the interval. The left endpoint of the interval is entered as `LastPresent` and the right endpoint as `FirstAbsent`. For carcasses not scavenged by the end of the study, `LastPresent` is the time of the last check and `FirstAbsent` is `Inf`. For carcasses whose scavenging time is known exactly (for example, scavenging was recorded by camera), `LastPresent = FirstAbsent`.

The CP file may contain optional columns including one for carcass size class (not used in this example) and any number of covariates (`Season` in this example). Users can specify in the GUI which column (if any) to use for carcass size class and which columns (if any) to use as predictors.

5.9. Search Schedule

The first few rows of a hypothetical search schedule at a site with 8 units is shown in table 4.

Table 4. Search Schedule Format.

SearchDate	Season	t1	t2	t3	t4	t5	t6	t7	t8
1955-04-15	spring	1	1	1	1	1	1	1	1
1955-04-18	spring	1	0	0	0	1	0	0	0
1955-04-21	spring	0	1	0	0	0	1	0	0
...									

Required Columns:

1. A column for search date.
2. A column for each search unit. Values of either 1 or 0, indicating whether the unit (column) was searched or not on the given date (row).

For this example, carcass searches were conducted on roads and pads within a 120 m radius from all 100 turbines at our fictitious wind power facility. Monitoring began on 1955-04-15 and continued through 1955-11-01. Searches spanned 3 seasons: spring, summer, fall. Search intervals varied by turbine and by time of year, ranging from daily searches at some turbines in the fall and searches once every 12 days in the spring at some other turbines. Search schedules for all turbines are stored in `SS_RPbat.csv`, with a column for search dates (including all dates that any turbine was searched), a column of 0s and 1s for each turbine, indicating whether it was searched on the given date, and zero or more optional columns giving additional information about the date (for example, season).

An assumption of the model is that no carcasses that arrived prior to the beginning of the monitored period at a site are included in the sample. Typically, this assumption is enforced by performing a "clean-out" or "clear-out" search to mark the beginning of the monitoring season. All carcasses discovered in that initial search are excluded from the survey data, and the date of the clean-out search is taken as the start of the period of inference.

5.10. Density-Weighted Proportion (DWP)

The density-weighted proportion (DWP)—or the expected proportion of total carcasses that arrive in the searched area wtihin each unit—is entered into a `.csv` file by search unit and by size class (table 5). A column to designate search units is required. At wind facilities, individual turbines are often taken as the search units. Because carcasses of different sizes likely have different fall patterns and hence proportions landing within a fixed distance from a turbine, DWP for each size class at a turbine can differ. A solar PV facility might be differentiated into units according to the type of area (for example, pond, fence line, road, solar panels, operations building) or into different blocks of PV panels. Carcasses may be assigned to distinct classes to reflect likely differences in detection probability. For example, detection probabilities for eagles would likely differ from detection probability for sparrows and would be put into a different carcass size class. Likewise, at a solar power tower facility, singed carcasses may be easier for scavengers to find than unsinged carcasses are and, as a result, may have a lower detection probability, so carcasses could be classed as singed or unsinged. All carcass classes appearing in the carcass observation file must be represented as columns in the DWP file.

Table 5. DWP data example.

unit	bat	sml	med	lrg
t1	0.183	0.054	0.107	0.055
t2	0.192	0.07	0.119	0.07
t3	0.205	0.059	0.122	0.063
t4	0.194	0.05	0.113	0.057
t5	0.179	0.048	0.103	0.051
t6	0.171	0.043	0.098	0.054
...				

5.11. Carcass Observations

The example carcass observations are stored in the `CO_RPbat.csv` file.

carcID	Turbine	TurbineType	DateFound	Species	SpeciesGroup	Distance
x315	t8	X	1955-04-27	BC	bat1	16.2
x382	t57	Z	1955-04-30	BC	bat1	10.6
x431	t72	Z	1955-05-03	BA	bat1	8.4
x496	t44	Y	1955-05-03	BC	bat1	3.3
x537	t43	Y	1955-05-06	BC	bat1	13.8
x651	t79	Z	1955-05-06	BB	bat1	26.4
...						

Required columns:

1. A column identifying the search unit (`Turbine` in this example): This column must have the same name as the column holding search units in the DWP file. Additionally, levels of the search units (`t8`, `t57`, `t72`, ... in this example) used in this column need to be valid column names within the search schedule file (SS). In particular, hyphens (`-`), strings beginning with a number or a dot (`.`) immediately followed by a number, and several other special characters are not allowed.
2. A column for date found: Column for date on which a carcass was observed. Format should be yyyy-mm-dd or mm/dd/yyyy.

In our simulated example, 264 carcasses were observed, representing 4 species of bats. Optional columns were added noting unique carcass ID, species, species group, distance from the turbine, and the type or status of the turbine. These optional covariates can be used to create summaries of interest (called splits) once mortality has been estimated at the site.

Publishing support provided by the U.S. Geological Survey
Science Publishing Network, Tacoma Publishing Service Center

For more information concerning the research in this report, contact the
Director, Forest and Rangeland Ecosystem Science Center
U.S. Geological Survey
777 NW 9th St., Suite 400
Corvallis, Oregon 97330
https://www.usgs.gov/centers/fresc/

ISSN 2328-7055 (online)
https://doi.org/10.3133/tm7C19