GenEst (v0.1.0) Software User Guide

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

Techniques and Methods XXX

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 Reilly, 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.

Disclaimer: The findings and conclusions in this article are those of the author(s) and do not necessarily represent the views of the U.S. Fish and Wildlife Service or the U.S. Bureau of Land Management.

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: Juniper Simonis, Manuela Huso, Daniel Dalthorp, Jeffrey Mintz, Lisa Madsen, Paul Rabie, and Jared Studyvin, 2018, GenEst (v1.0.0) software user guide: U.S. Geological Techniques and Methods XXX, xx p., https://doi.org/10.3133/tmXXX.

ISSN 2327-638X (online)

Contents

Contents 1
Figures 3
Tables 4
Abbreviations 4
Variables and Parameters 5
GenEst (v1.0.0) Software User Guide 7
Abstract 7
1. Introduction 8
1.1 Description of Product 8
1.1.1 Software Modules 9
1.1.2 Comparison to Other Methods 9
1.2 Installation Instructions 10
1.2.1 Requirements 10
1.2.2 Installation 11
1.3 Getting Started 13
1.4 Layout 13
2. Data Input 14
2.1 Data Files 14
3. Analyses 16
3.1 General Inputs 16
3.2 Searcher Efficiency 17
3.2.1 Model Inputs 18
3.2.2 Table & Figure Selection and Model Selection: 19
3.2.2.1 Selected Data 20
3.2.2.2 Figures 21
3.2.2.3 Estimates 21
3.2.2.4 Model Comparison 22
3.2.2.5 Model Selection 23
3.3 Carcass Persistence 23
3.3.1 Model Inputs 25
3.3.2 Table & Figure Selection and Model Selection 26
3.3.2.1 Selected Data 27
3.3.2.2 Figures 27
3.3.2.3 Estimates 28
3.3.2.4 Model Comparison 29
3.3.2.5 Model Selection 30
3.4 Mortality Estimation 30
3.4.1 Model Inputs 31
3.4.1.1 Splitting Mortality Estimates 33
3.5 Detection Probability 35
3.5.1 Model Input 35
3.5.1.1 Search Schedule 35
3.5.2 Table and Figure Selection 37
3.5.2.1 Summary 37
3.5.2.2 Figure 38

4. Worked Examples 39
4.1 Example 1 – Wind facility, bats, road and pad surveys of all turbines 39
4.1.1 Data Loading 40
4.1.2 General Inputs 40
4.1.3 Searcher Efficiency 42
4.1.4 Carcass Persistence 47
4.1.5 Mortality Estimation 51
4.2 Example 2 – Solar photovoltaic facility, all birds 57
4.2.1 Data Loading 58
4.2.2 Modeling Search Efficiency 60
4.2.3 Modeling Carcass Persistence 64
4.2.4 Mortality Estimation 68
4.2.5 Detection Probability 71
Acknowledgments 74
References Cited 76
A Appendix. Input File Formats 78
A.1 Searcher Efficiency 78
A.2 Carcass Persistence 79
A.3 Search Schedule 80
A.4 Carcass Observations 81

Figures

Figure 1. Welcome screen disclaimer for GenEst, as opened in Chrome (v67.0.3396.87). 14
Figure 2. Visible uploaded "Mock" searcher efficiency data in the GenEst GUI. 15
Figure 3. Required and optional general inputs. 16
Figure 4. Searcher efficiency model inputs 18
Figure 5. Options to explore results of searcher efficiency model fitting. 20
Figure 6. Example plots of estimated and observed searcher efficiency by search event. 21
Figure 7. Estimates and confidence limits (α set by user in "General Inputs") of p and k for each class combination of covariates for selected model. 22
Figure 8. Model selection screen indicating which searcher efficiency model to use for each size class of trial carcasses. 23
Figure 9. Carcass persistence model inputs. 25
Figure 10. Options to explore results of carcass persistence model fitting. 27
Figure 11. Example plots of estimated and observed carcass persistence by time for each combination of levels of predictor variables. 28
Figure 12. Estimates and confidence limits (α set by user in "General Inputs") of location and scale parameters for each class combination of covariates for selected model. 29
Figure 13. Model selection screen indicating which carcass persistence model to use for each size class of trial carcasses. 30
Figure 14. 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. 32
Figure 15. Estimated total mortality at the site based on user input. 33
Figure 16. Estimated mortality at the site in each combination of size and ground cover. 34
Figure 17. Estimated mortality at the site in each combination of size and ground cover. Table of quantiles used to produce associated graph. 35
Figure 18. Model inputs screen requesting information on search schedule. 37
Figure 19. Summary of detection probability and confidence limits for relevant covariate classes. 38
Figure 20. Box plot of estimated detection probability for relevant covariate classes 39
Figure 21. "Data Input" tab with "Carcass Observations" displayed after uploading. 40
Figure 22. Starting the analysis with "General Inputs". 41
Figure 23. Model inputs for searcher efficiency analysis 42
Figure 24. Searcher efficiency model comparison. 44
Figure 25. Searcher efficiency diagnostic plots. 45
Figure 26. Searcher efficiency model selection 47
Figure 27. Carcass persistence model building. 48
Figure 28. Carcass persistence model comparison tab. 49
Figure 29. Fitted carcass persistence models. 50
Figure 30. Carcass persistence model selection 51
Figure 31. Mortality estimation tab. 52
Figure 32. Total mortality and options for mortality splits. 53
Figure 33. Mortality estimates by season. 54
Figure 34. Mortality estimates by species. 55
Figure 35. Seasonal mortality by species. 56
Figure 36. Bat mortality estimates by species and season (transpose of season by species). 57
Figure 37. Data import 59
Figure 38. General inputs 59

Figure 39. Data for searcher efficiency analysis. 61
Figure 40. Searcher efficiency model comparison. 61
Figure 41. Diagnostic searcher efficiency plots for large carcasses at a solar PV project..... 63
Figure 42. Selecting the SE models for each size class for the solar PV example. 64
Figure 43. Fitting carcass persistence models for solar PV data set..... 65
Figure 44. Fitted carcass persistence distributions for the $l \sim 1; s \sim 1$ models. 66
Figure 45. Fitted carcass persistence distributions for the $l \sim \text{Season}; s \sim 1$ models..... 67
Figure 46. Selection of carcass persistence models for solar PV example 68
Figure 47. Total mortality in the solar PV example..... 69
Figure 48. Mortality by species for solar PV example 70
Figure 49. Mortality by species and season for the solar PV example. 71
Figure 50. Average search schedule for solar PV example data. 73
Figure 51. Detection probability by season for small birds (solar PV example)..... 74

Tables

Table 1. Files needed for each analysis 15
Table 1. Example format of carcass persistence data. ..... **Error! Bookmark not defined.**

Abbreviations

AICc	Akaike information criterion corrected for finite sample size
ΔAIC	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)
CDF	cumulative distribution function
CI	confidence interval
CP	carcass persistence
CRAN	Comprehensive R Archive Network (for downloading R statistical software and supplemental packages)
DWP	Density Weighted Proportion; proportion of total mortality expected to fall within searched areas
IQR	interquartile range; bounds of the central 50% of a distribution; the 0.25 and 0.75 quantiles of a distribution
Left	left-hand component of interval in which a carcass was removed; the latest time (in decimal days since placement) that a persistence trial carcass is noted to be

	still present (see Right, below)
LRT	likelihood ratio test
PDF	probability density function
PI	prediction interval
.rds	serialized R data object (format for some input files)
Right	right-hand component of interval in which a carcass was removed; the earliest time (in decimal days since placement) that a persistence trial carcass is known to be missing (see Left, above)
SE	searcher efficiency
superpopulation	the total number of animals that ever enter the sampled population during the period of inference

Variables and Parameters

α	significance level (distinguished from shape parameter by context)
$1-\alpha$	confidence level
α, β	shape and scale parameters for persistence distributions
a	spatial coverage, fraction of carcasses falling into searched area; see *dwp*
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
λ	mortality rate over a set time period
M	mortality or number of fatalities
$\widehat{M}$	estimated mortality or number of fatalities
nsim	number of simulation draws

p	searcher efficiency, probability of observing a carcass that is present in the searched area at the time of the search
r	probability that a carcass persists until the first search after arrival
span	the length of the monitoring period
v	temporal coverage, fraction of total mortality expected to occur during the monitored period
X	number of carcasses observed in searches

GenEst (v1.0.0) Software User Guide

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. Simple counts of carcasses are not an accuratemeasure of the true number of fatalities because 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. Included in the software are 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..

1.0 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 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 et al. 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. In tracking 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, nor can it be used as an index of mortality because is not linearly related to the population it is intended to represent. GenEst is comprised of several modules that use data provided by the user to estimate searcher efficiency (**Searcher Efficiency**), carcass persistence (**Carcass Persistence**), and detection probability (**Detection Probability**) parameters

as well as estimates of total mortality (**Mortality Estimation**) for groupings of carcasses of interest to the user.

GenEst is written in R (R Core Team 2017) and available as a fully documented R package, including a command-line interface and as a user-friendly Graphic User Interface (GUI) coded using `shiny` (Chang et al. 2017). This User Guide focuses on the GUI version of GenEst. Vignettes within the R package detail the command-line interface.

1.1.1 Software Modules

GenEst software includes two primary modules, each with submodules:

1) Data input—Points the software to the folder containing the data input files (Searcher Efficiency, Carcass Persistence, Search Schedule, Density Weighted Proportion, and Observed Fatalities), all in .csv format.

2) Analyses—Determines 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.

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 (Dalthorp et al. 2017, Wolpert 2015, Peron et al. 2013, Huso et al. 2012, Korner-Nievergelt et al. 2015, Shoenfeld, 2004). GenEst's accuracy, flexibility, and ease of use are unrivaled. The primary differences are that GenEst is generally much more flexible than any of the others and will provide more accurate estimates both with respect to bias and coverage probabilities for confidence intervals because fewer constraining assumptions are made. 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 data sets that may include multiple carcass size classes, detection probabilities that depend on envirnomental covariatess (e.g., visibility, season, search team, carcass type, area), variable search schedules, search coverage that varies with search unit (e.g. turbine at a wind project or ring at solar power tower facility), 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 includes modules to allow the user to model searcher efficiency and carcass persistence as function of environmental covariates and to compare models using AICc and graphs.

1.2 Installation Instructions

1.2.1 Requirements

- R software, version 3.5.0 or later
- R packages:

`cbinom`	(Dalthorp 2018)
`devtools`	(Wickham et al. 2018)
`digest`	(Eddelbuettel et al. 2018)
`DT`	(Xie 2018)
`gsl`	(Hankin 2006)
`gtools`	(Warnes et al. 2015)
`htmltools`	(RStudio and Inc. 2017)
`lubridate`	(Grolemund and Wickham 2011)
`matrixStats`	(Bengtsson 2018)
`mvtnorm`	(Genz et al. 2018),
`Rcpp`	(Eddelbuettel and Francois 2011)
`shiny` (Chang et al. 2018)	
`shinydashboard`	(Chang and Ribeiro 2018)
`sticky`	(Brown 2017)
`survival`	(Therneau 2017)

`testthat` (Wickham 2011)

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.

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 nearest site) 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 `r-project.org` website.

If you are updating your copy of R from a previously installed version, the new version will be installed alongside the old version rather than on top of it, and both versions will be operational. Unless you know the old version will be needed for some specific purpose in future, uninstalling it is recommended. This can be done using the Control Panel in Windows®. R packages that the user has previous 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)`.

Third-Party R Packages

Several third-party R packages are required. 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. The packages can be installed from within R by running the following commands from the R command line (by copy-and-paste):

```
list.of.packages <- c("cbinom", "devtools", "digest", "DT", "gsl", "gtools",
"htmltools", "lubridate", "matrixStats", "mvtnorm", "Rcpp", "shiny",
"shinydashboard", "sticky", "survival", "testthat")

new.packages <- list.of.packages[
    !(list.of.packages %in% installed.packages()[,"Package"])]

if(length(new.packages) > 0) install.packages(new.packages)
```

Depending on the current status of your R installation, it could take anywhere from a fraction of a second to several minutes to install and/or check the required packages. If new packages need to be installed, a dialog box may ask you to select a CRAN mirror. Choose the nearest location. Next, you might be asked if you want to use a personal library instead (or something to that effect). Choose "Yes." Once these packages have been installed, you will not need to go through the wait again. If you are working in Windows and are asked whether you wish to install packages or dependencies "from source", choose "No" (unless you are prepared to wait for what will seem like an extraordinarily long time).

GenEst software is bundled as an R package. To complete the installation for Windows®, download `GenEst_0.1.0.zip` from the `user` folder at `https://code.usgs.gov/ecosystems/internal/GenEst` to a convenient location but do not unzip. Then run `install.packages(file.choose())` in R and choose the downloaded .zip file. If installing to Mac® or UNIX®-like operating system, download `GenEst_0.1.0.tar.gz` from the `user` folder and run `install.packages(file.choose()).` R may prompt Mac® users to install "Command Line Tools" for `Xcode`, which are necessary for installing packages from source.

GenEst will eventually be available from the R package repository (http://cran.r-project.org), at which point, the stable version of the software will be able to be downloaded and installed locally simply by running `install.packages("GenEst")` without the need to install the third-party packages by hand and to first download GenEst before installing.

1.3 Getting Started

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

```
GenEst::runGenEst()
```

which will launch the application in your default web browser. The welcome screen for the GUI has a disclaimer that must be acknowledged before proceeding. Note that opening the app 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 an additional R session.

1.4 Layout

The GUI is laid out using tabs, as shown in the top left corner of **Figure 1**: Home, Data Input, Analyses, and About. After acknowledging the disclaimer, users may navigate between the tabs. The Data Input and Analyses tabs allow active user interaction for analyses.

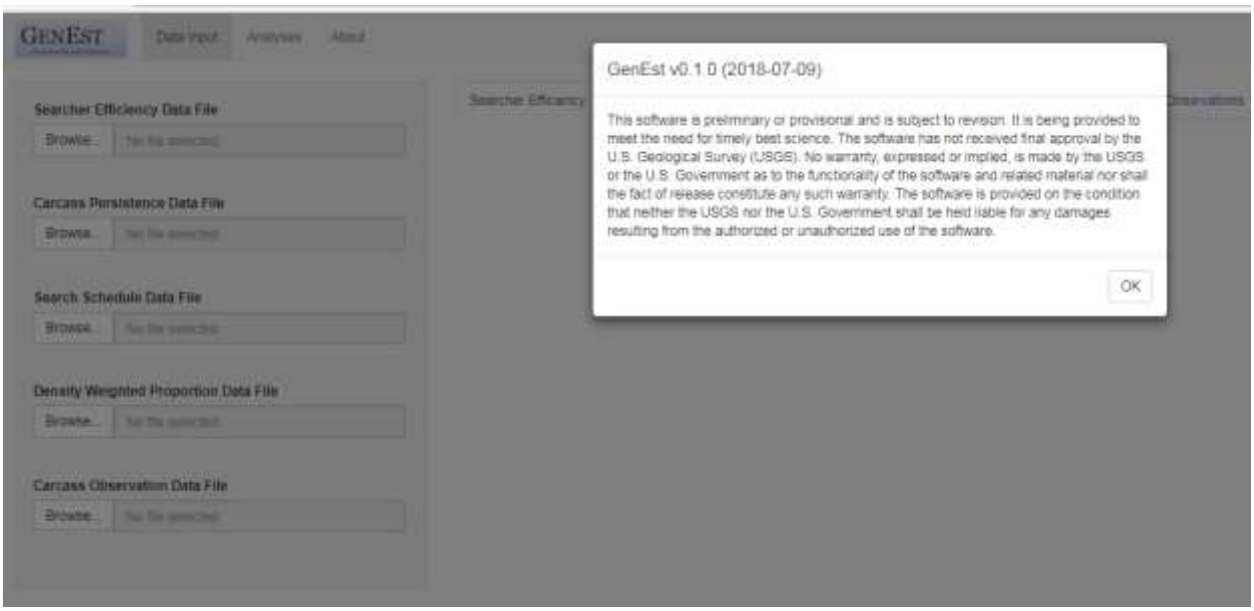

Figure 1. Welcome screen disclaimer for GenEst, as opened in Chrome (v67.0.3396.87).

2.0 Data Input

2.1 Data Files

Having acknowledged the disclaimer, the GUI is now able to accept data for use. Some example data files are provided in the R package that can be used within the GUI (in the `extdata` subfolder in the `GenEst` package folder, which is located with other R package folders on the user's computer). Clicking on the "Browse…" button allows the user to navigate to the desired files for upload. Data for example analyses include a generic data set with random values (`mock`), data from three fictitious solar projects (photo-voltaic, parabolic trough, and power tower), and three data sets from a fictitious wind power project with 100 turbines (roads and pads searched for bats on all 100 turbines, roads and pads searched for birds and bats on all 100 turbines, cleared plots searched out to 90 m on 23 turbines).

The GUI takes five types of data file: Searcher Efficiency, Carcass Persistence, Search Schedule, Density Weighted Proportion, and Carcass Observations. The files should be formatted with comma separation (either a comma separated value (.csv) or a comma delimited regular text (.txt) file). **Appendix A** details the necessary and available formats for each of the files. The files required to upload depend on the user's specific desires (**table 2**). Upon upload, each data file becomes available for inspection (Figure 2) as well as use throughout the GUI.

Table 2. Files needed for each analysis

Analysis	Search Efficiency	Carcass Persistence	Search Schedule	Carcass Observations
Search Efficiency	x			
Carcass Persistence		x		
Detection Probability	x	x	x	
Mortality Estimation	x	x	x	x

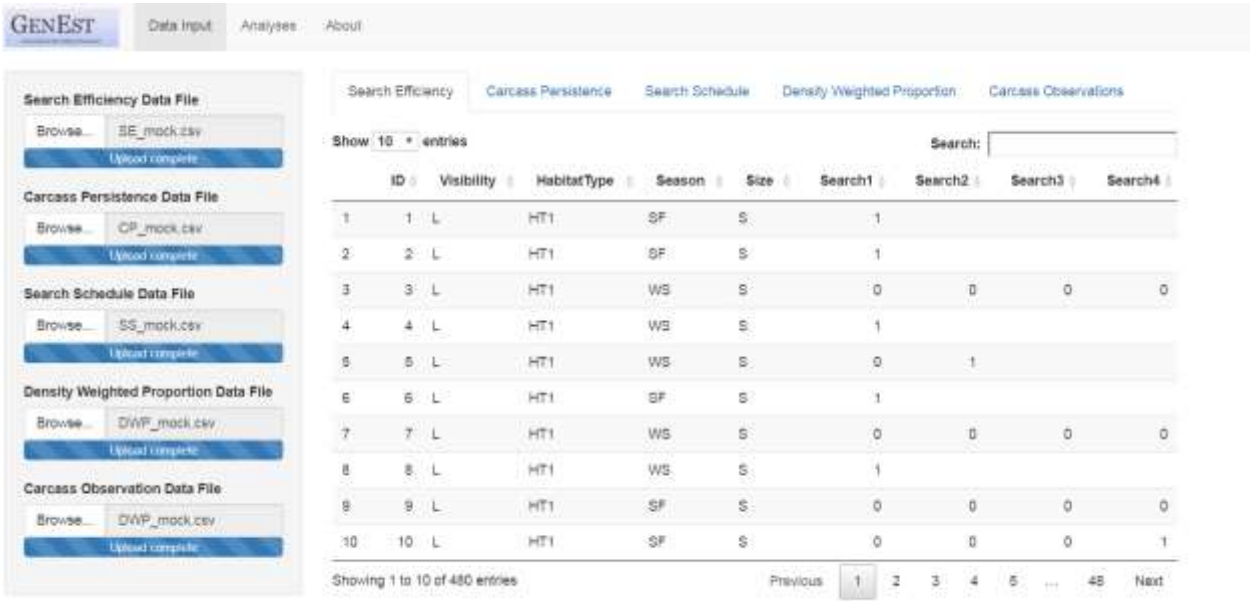

Figure 2. Visible uploaded "Mock" searcher efficiency data in the GenEst GUI.

3.0 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" (**Figure 3**) and may supply a third (optional) value. The "Number of iterations" controls the number of replicates used to estimate parameters, defaults to 1000, and must be an integer value 1 or larger, with current support up to 10000. The "Confidence Level" defines the probability used to describe the spread of distributions (for parameters and state variables), defaults to 0.95 (a.k.a. 95%) and must be between 0 and 1. In addition, the user is asked to choose a "Size Class Column (optional)." The software offers as options those columns that occur in all three of Searcher Efficiency, Carcass Persistence and Carcass Observations input files.

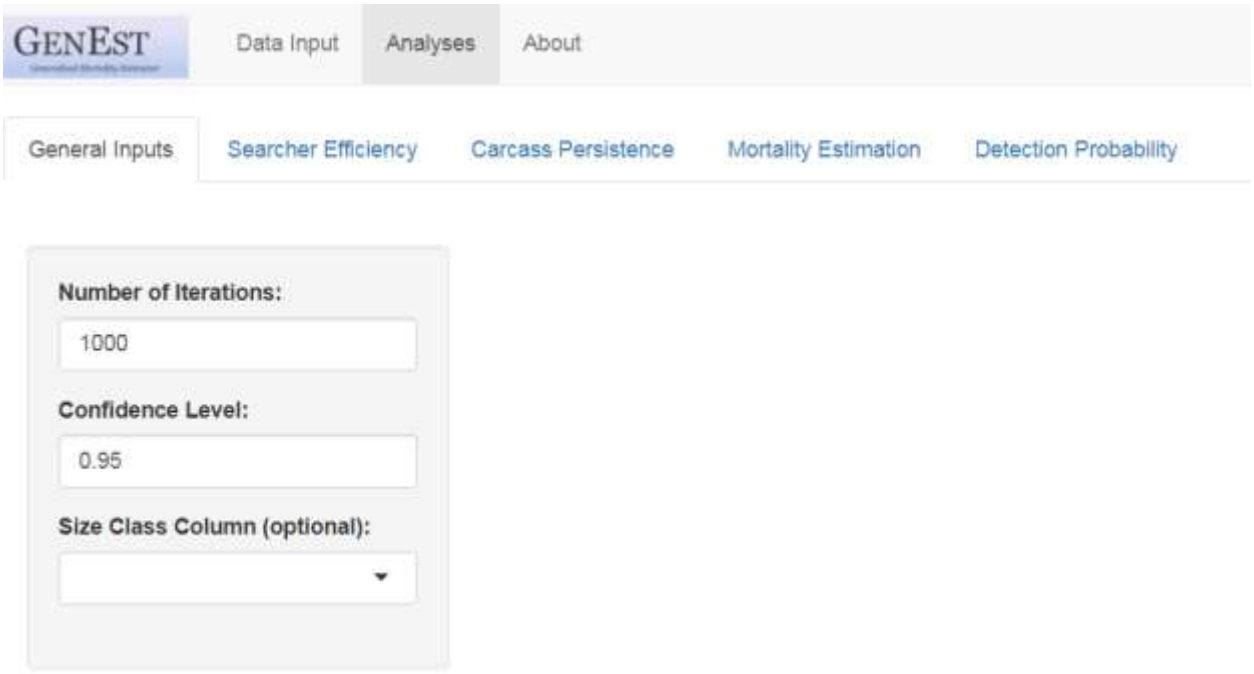

Figure 3. Required and optional general inputs.

3.2 Searcher Efficiency

Searcher efficiency is the probability that a searcher will observe a carcass that is present in the search area when the search is conducted. 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 at regular intervals (not necessarily immediately preceding a search) within the searched area, at locations known only to the trial conductor, 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 ascertain that indeed the carcass was present, but truly missed or no longer present and hence not eligible for inclusion in the searcher efficiency trial data. Trial conductors are encouraged not to remove carcasses when found so they can be monitored for persistence. Allowing missed carcasses to remain in the field for potential discovery on a later search provides information about how searcher efficiency for an individual carcass changes over time, usually decreasing as carcasses age and as the 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, e.g., size of the carcass, carcass coloration, vegetation density, individual searcher, etc.

GenEst uses single or repeat 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, the proportional reduction in the probability that a present carcass is found on each subsequent search. A value of $k = 0$ implies 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. Searcher efficiency parameters can be estimated for carcasses grouped into size classes and any combination of up to two categorical

variables (e.g., season and vegetation class) that occur as columns in the searcher efficiency data input by the user. The searcher efficiency parameters are estimated from the survey data in each class × size combination using maximum likelihood methods, as implemented with custom-written functions for the GenEst package. For full details, see Dalthorp et al. 2018 (in review) . 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.

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 (**Figure 4**).

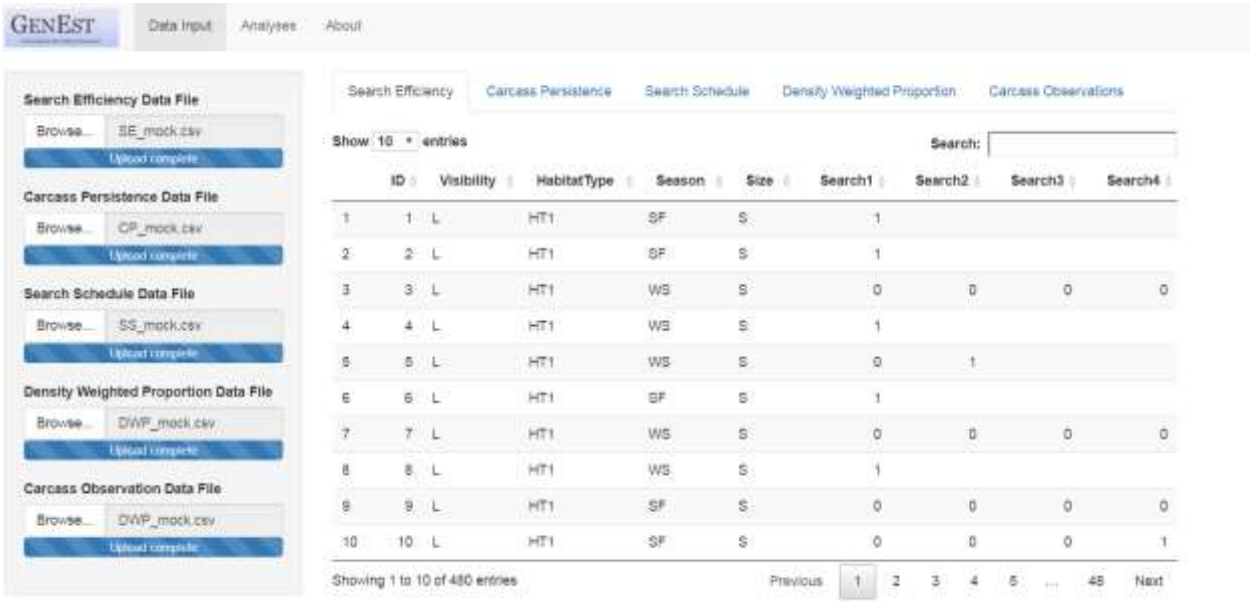

Figure 4. Searcher efficiency model inputs

- "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 (found = 1, missed = 0, 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.
    - The user can 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), 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, repeated attempts to find the same carcasses), 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 Table & Figure Selection and Model Selection:

Once all input data are selected, the user clicks on "Run Model." A section titled "Table & Figure Selection:" will appear below "Model Inputs:" on the left and once calculations are

completed the “Model Comparison” sub-tab is automatically opened and the “Table & Figures” selections are automatically populated with the size class that occurs first in the CO data (if Size column was entered under “General Inputs”) and the model for p and k with the lowest AICc value (**Figure 5**). Several output sub-tabs for evaluating model fit are available: Selected Data, Figures, Estimates, Model Comparison, and Model Selection (fig. 5).

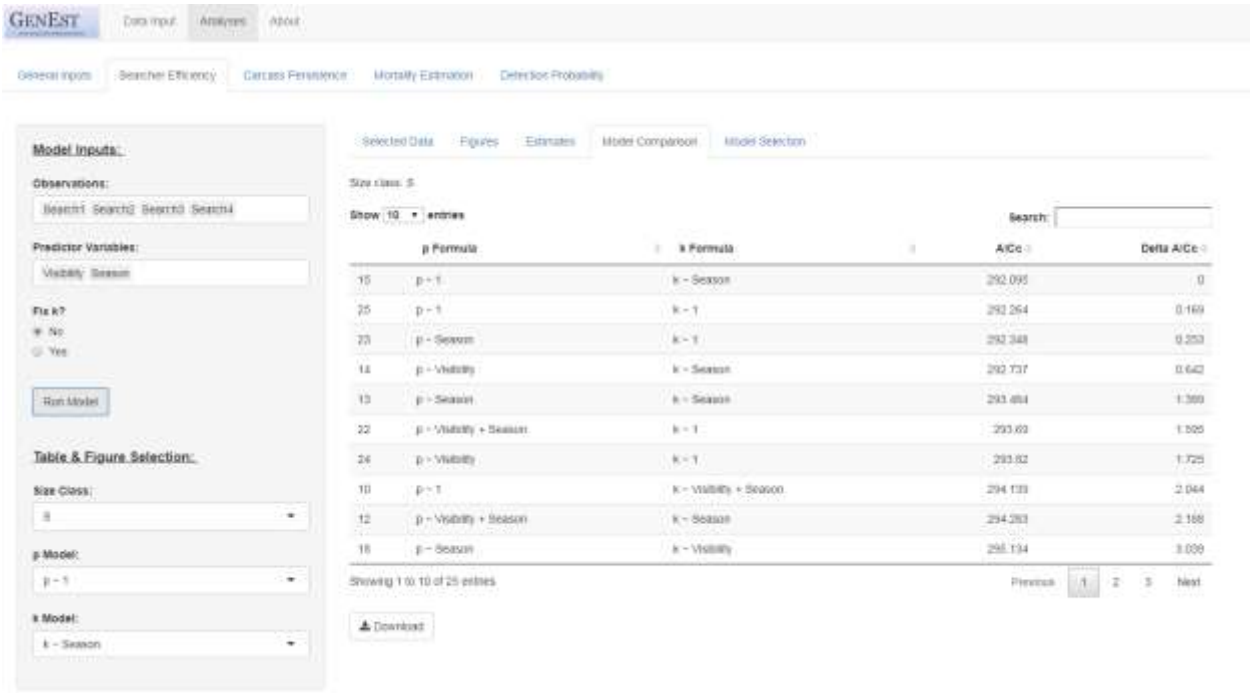

Figure 5. Options to explore results of searcher efficiency model fitting.

3.2.2.1 Selected Data

Data used to fit the searcher efficiency models for the selected size class (selected under Table & Figures Selection on the left screen) are displayed under the “Selected Data Tab.” This is a subset of the .csv file input earlier in the Data Input process.

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 (**Figure 6**).

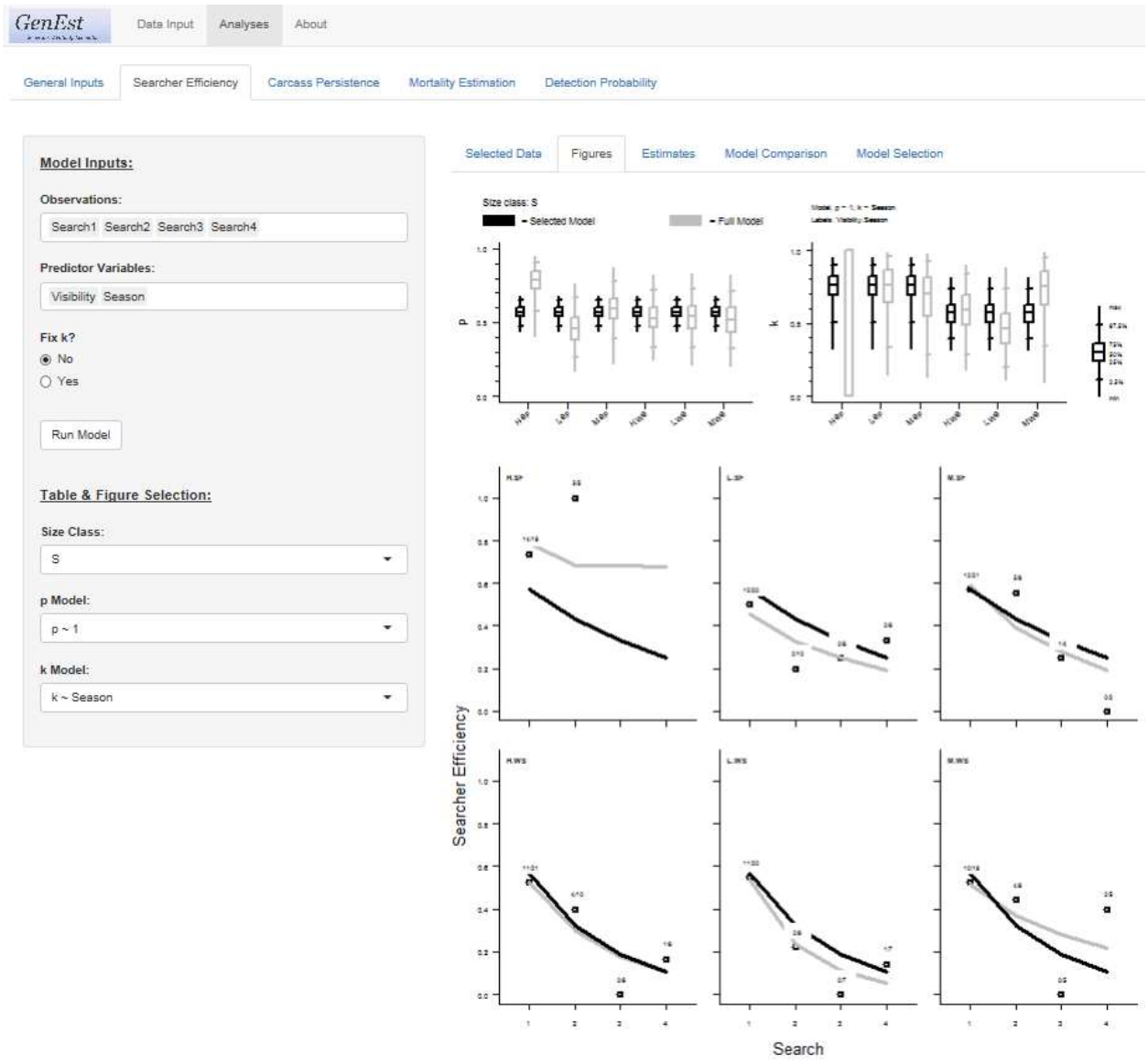

Figure 6. Example plots of estimated and observed searcher efficiency by search event. Dark lines represent selected model, gray lines represent the most complex model, i.e., separate estimates for each class combination (when possible). Open dots represent observed searcher efficiency and ratios above each open dot represent observed data used to estimate the point, i.e., found/available. Box plots at the top represent estimated p and k for selected model (dark lines) compared to most complex model (grey lines).

3.2.2.3 Estimates

Estimates and confidence limits (α set by user in "General Inputs") of p and k for each class combination of covariates for selected model are given in table format (Figure 7). The table

can be saved to a csv by clicking the “Download” button.

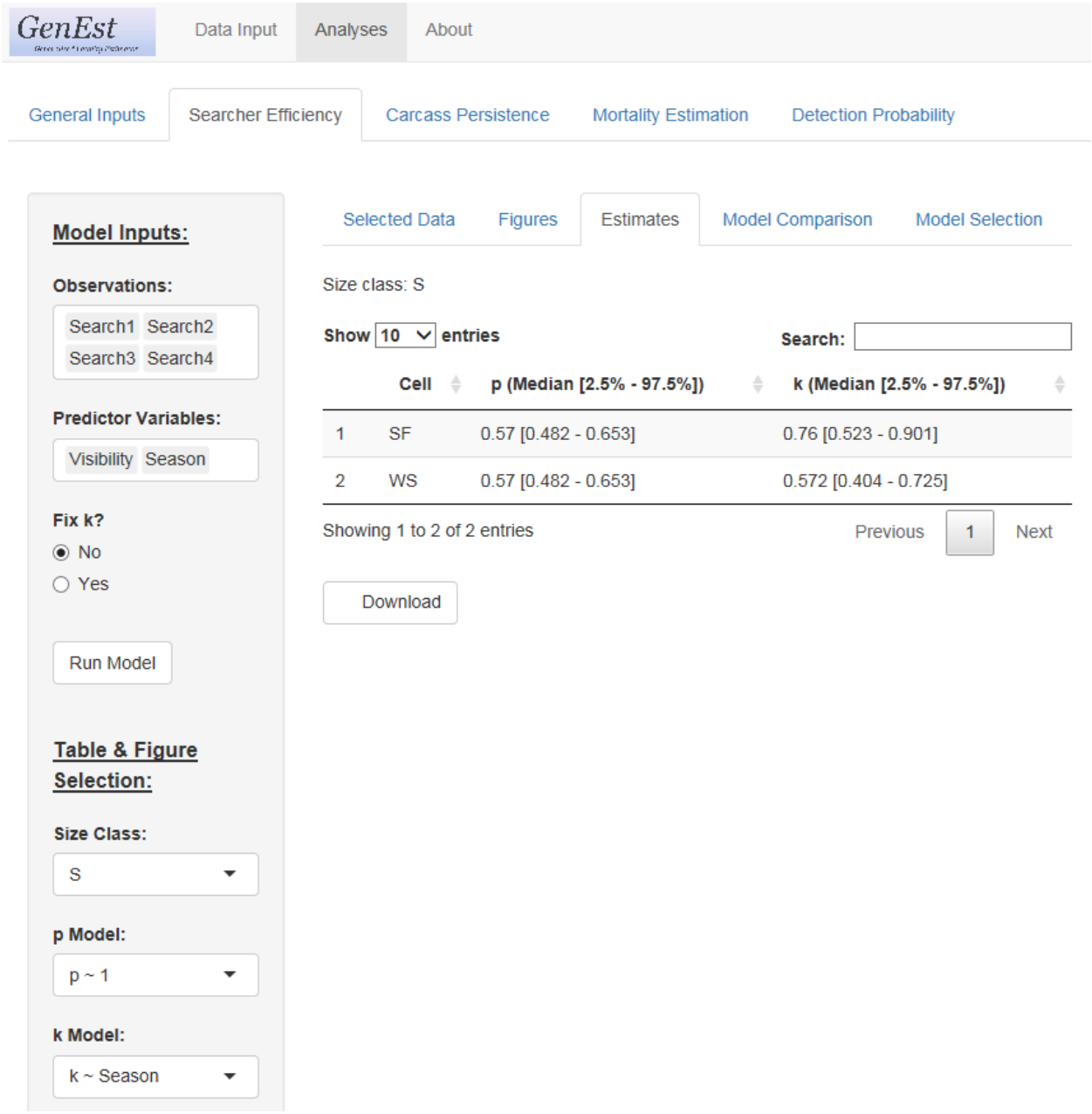

Figure 7. Estimates and confidence limits (α set by user in “General Inputs”) of p and k for each class combination of covariates for selected model.

3.2.2.4 Model Comparison

The model comparison tab provides user with AICc and ΔAICc for all models fit using the covariates selected under “Model Inputs” (Figure 5).

3.2.2.5 Model Selection

The user selects the model to be used in calculating detection probability and mortality. Available models are listed in ascending order of ΔAICc. The default selected models are those with delta AICc = 0, i.e., the "best" model with respect to AICc (**Figure 8**). There is no need to "Run Model" again. The selections will be held in memory for later use in mortality estimation.

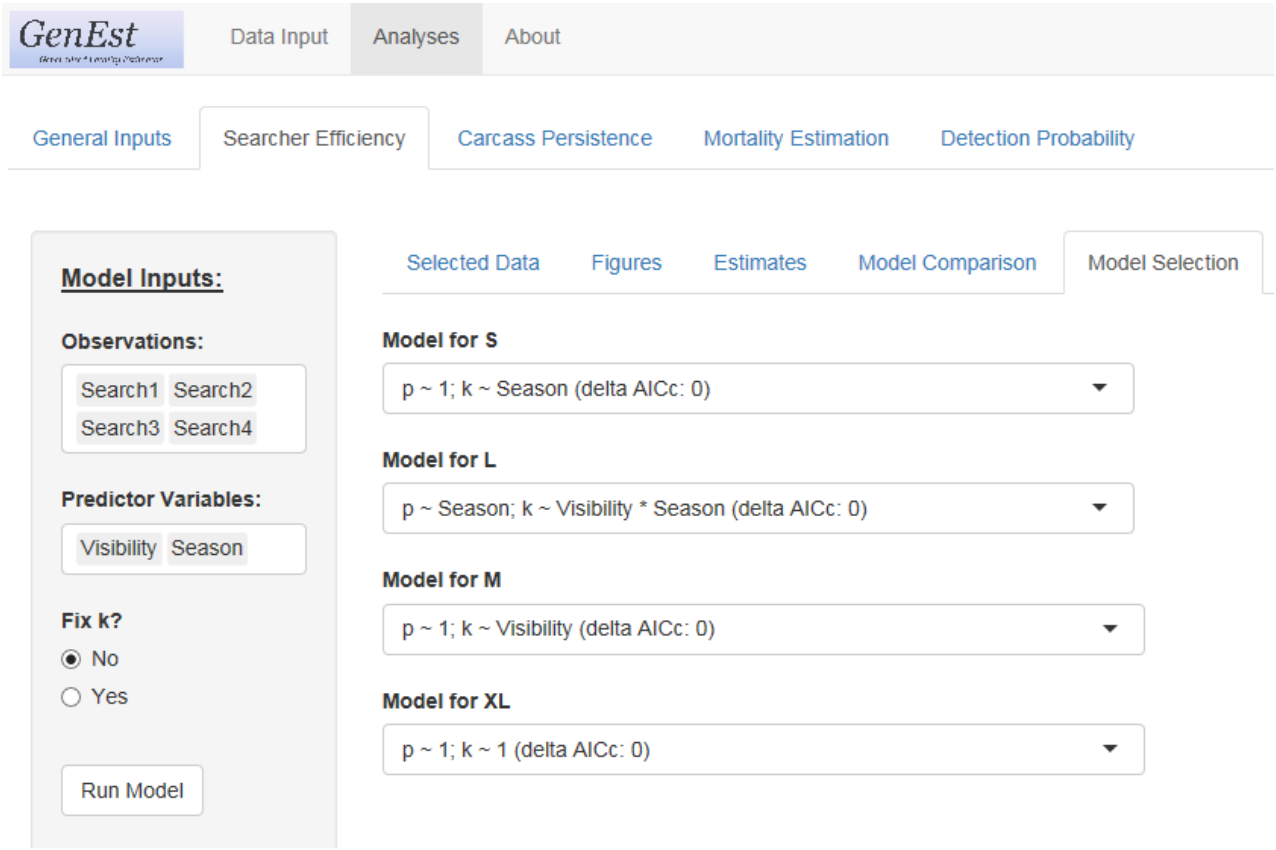

Figure 8. Model selection screen indicating which searcher efficiency model to use for each size class of trial carcasses.

.

3.3 Carcass Persistence

Carcass persistence is the probability that a carcass arriving at time 0 will continue to persist until time *t*. It is typically measured using the same trial carcasses as are used in searcher efficiency trials. Trial carcasses representing as closely as possible the species composition of turbine-caused fatalities at the facility and as closely, are placed at irregular intervals (not

necessarily immediately preceding a search) within the searched area, at locations known only to the trial conductor, not the searchers. 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 or visit and the presence or absence of the carcass. Carcass persistence often varies with characteristics of the individual carcasses or conditions of its search, e.g., size of the carcass, season, vegetation density, etc.

GenEst allows the user to fit the data to up to four persistence distributions: exponential, Weibull, lognormal and loglogistic. Carcass persistence parameters can be estimated for carcasses grouped into size classes and any combination of up to two discrete-valued predictors (e.g., 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 each class x size combination using maximum likelihood. For full details, see **appendix B.2**.

3.3.1 Model Inputs

Carcass persistence model inputs are selected on the left-hand side panel of the Search Efficiency tab, Selected Data sub-tab (**Figure 9**).

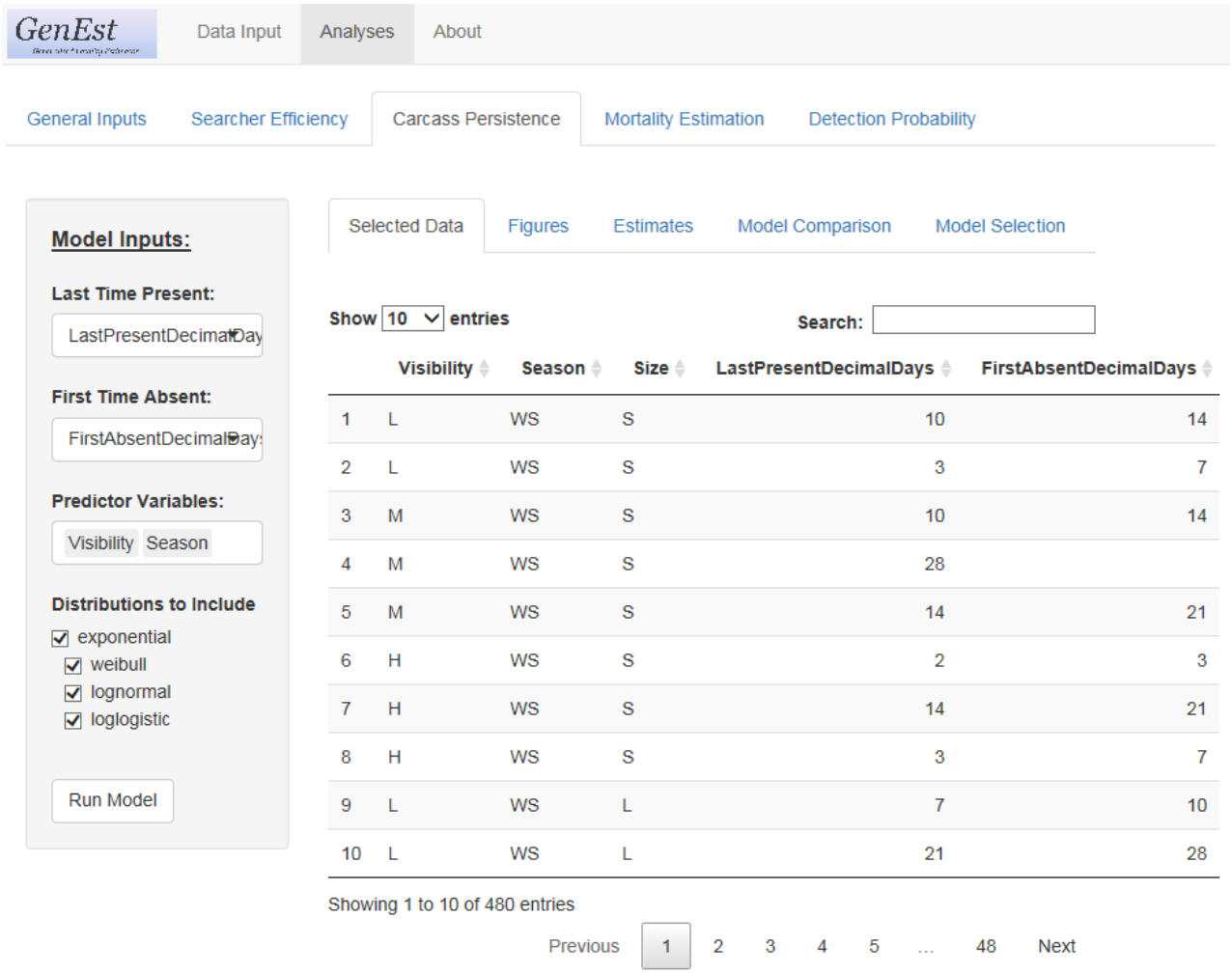

Figure 9. Carcass persistence model inputs.

- "Last Time Present:"
    - The user is required to select the column in the input data set that corresponds to the last time (in decimal days since placement) that the persistence trial carcass was known to be present.
- "First Time Absent:"

- The user is required to select the column in the input data set that corresponds to the first time (in decimal days since placement) that the persistence trial carcass was known to be absent.

- "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 & Figure Selection and Model Selection

Once all input data are selected, the user clicks on "Run Model." A section titled "Table & Figure Selection:" will appear below "Model Inputs:" on the left and the "Model Comparison" sub-tab is automatically opened, but several output sub-tabs are available: Selected Data, Figures, Estimates, Model Comparison, and Model Selection (**Figure 10**).

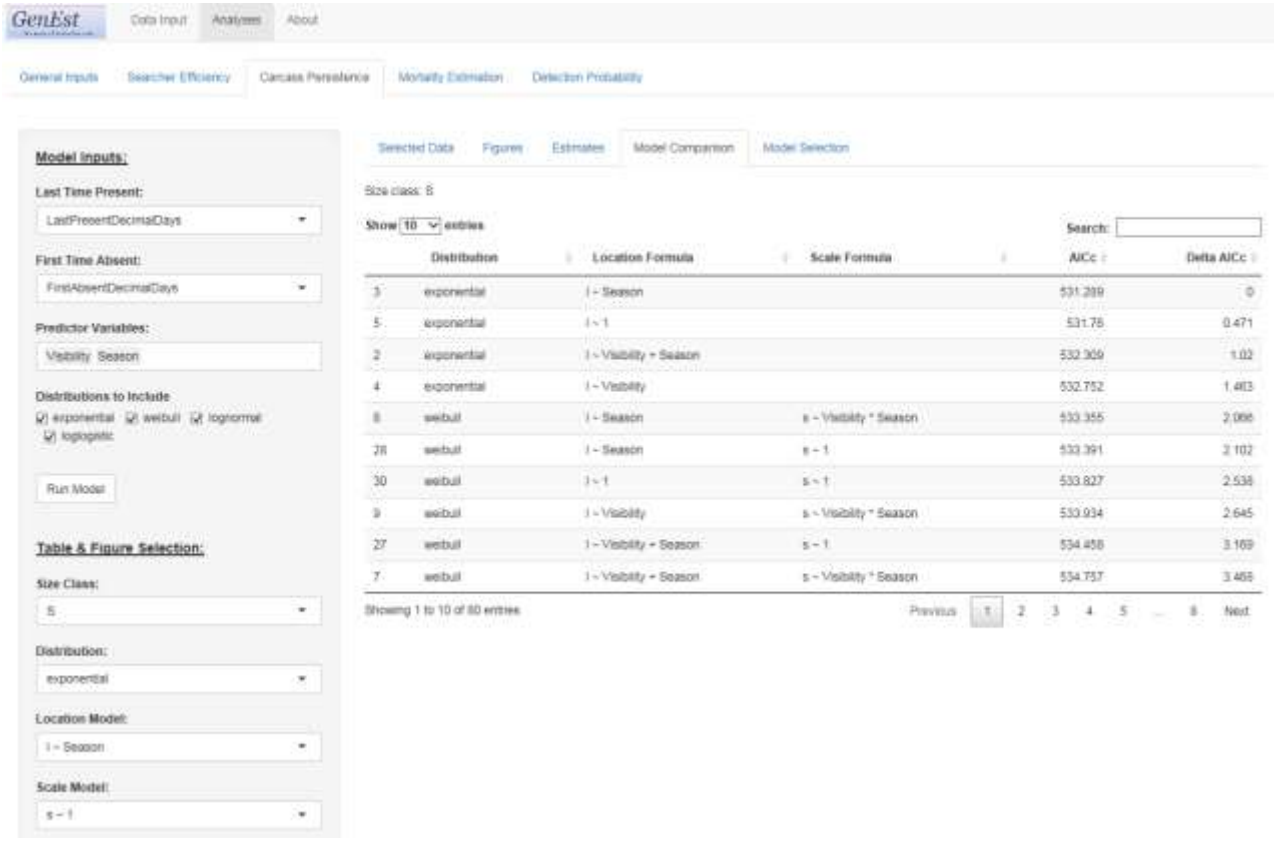

Figure 10. Options to explore results of carcass persistence model fitting.

3.3.2.1 Selected Data

Data used to fit the carcass persistence models for the selected size class (selected under Table & Figures Selection on the left screen) 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 Figures

Carcass persistence estimates from the selected model are plotted against time for each possible combination of class levels of the input predictor variables (**Figure 11**).

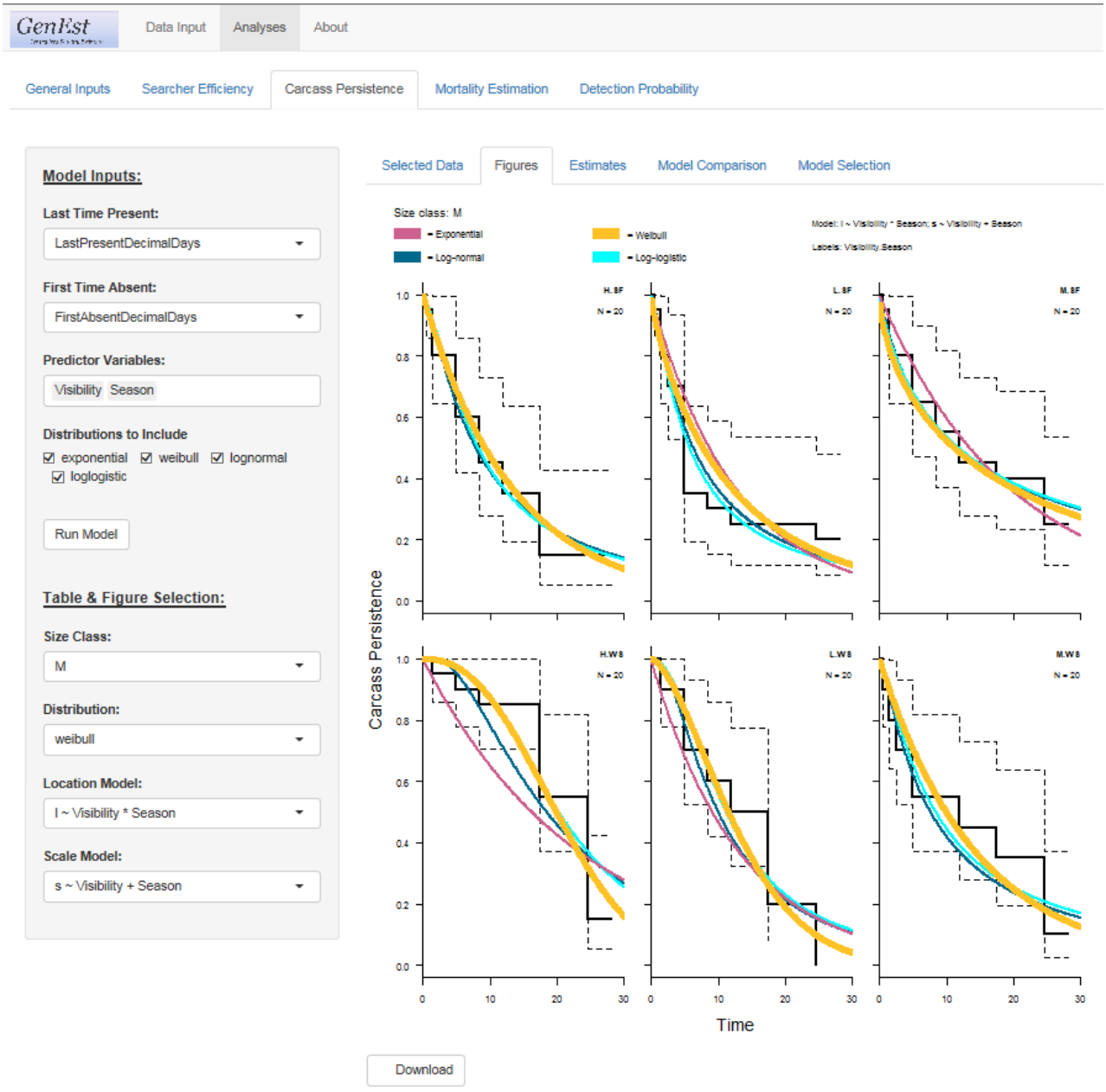

Figure 11. Example plots of estimated and observed carcass persistence by time for each combination of levels of predictor variables.Stair-step solid and dotted lines are Kaplan-Meier plots representing observed data and empirical confidence limits. All fitted persistence distributions are represented by solid curves, but the width of selected persistence distribution will be thicker than others.

3.3.2.3 Estimates

Estimates and confidence limits (α set by user in "General Inputs") of the location and scale parameters for each class combination of covariates for the selected model are given in table format (**Figure 12**). The table can be saved to a csv by clicking the "Download" button. Note: scale parameter in exponential distribution = 1, by definition.

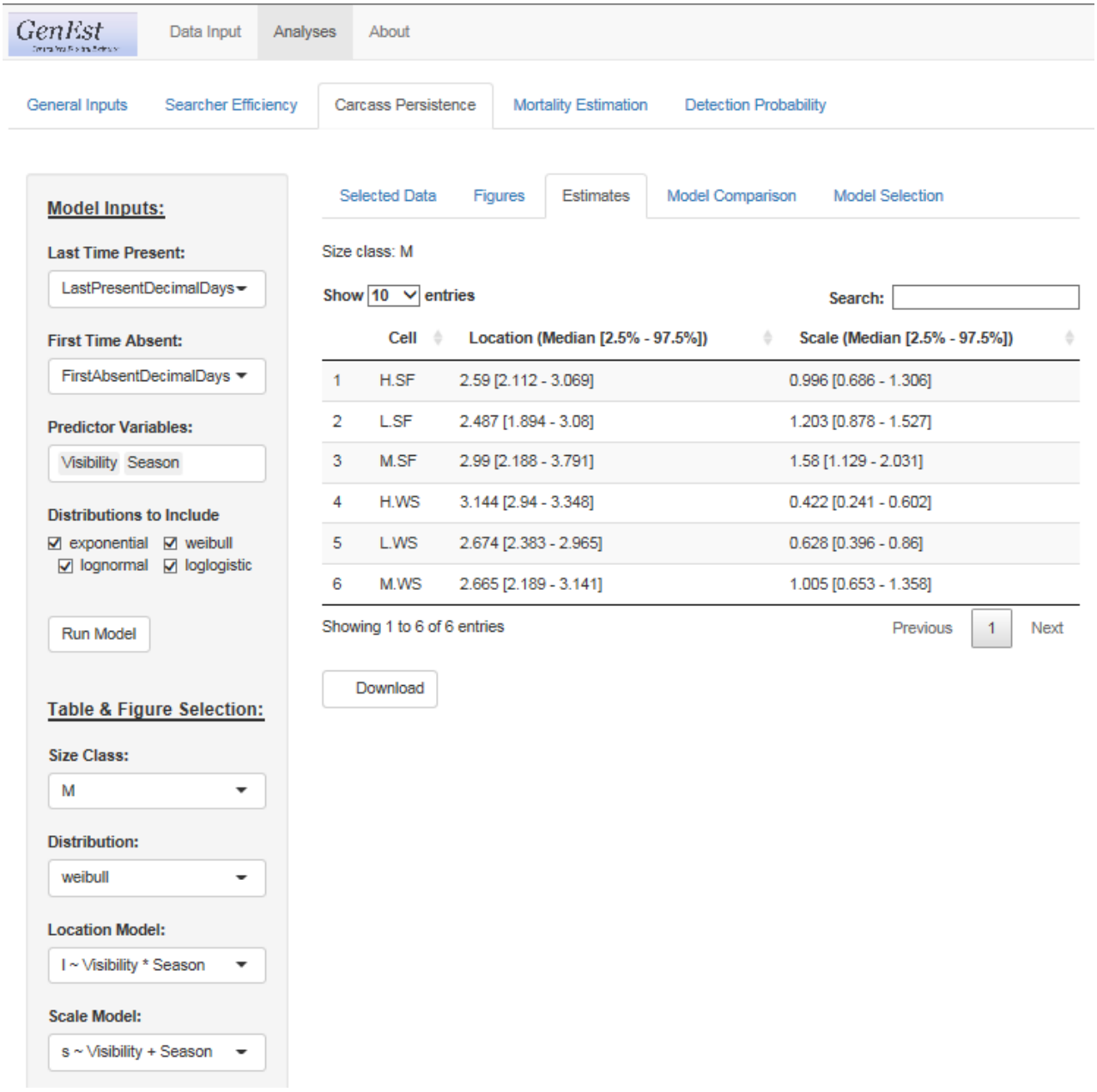

Figure 12. Estimates and confidence limits (α set by user in "General Inputs") of location and scale parameters for each class combination of covariates for selected model.

3.3.2.4 Model Comparison

The model comparison tab provides user with AICc and Delta AICc for all models fit using the covariates selected under "Model Inputs" (Figure 10).

3.3.2.5 Model Selection

The user selects the model to be used in calculating detection probability and mortality. Available models are listed in ascending order of delta AICc. The default selected models are those with delta AICc = 0, i.e., the "best" model with respect to AICc (**Figure 13**). There is no need to "Run Model" again. The selections will be held in memory for later use in mortality estimation.

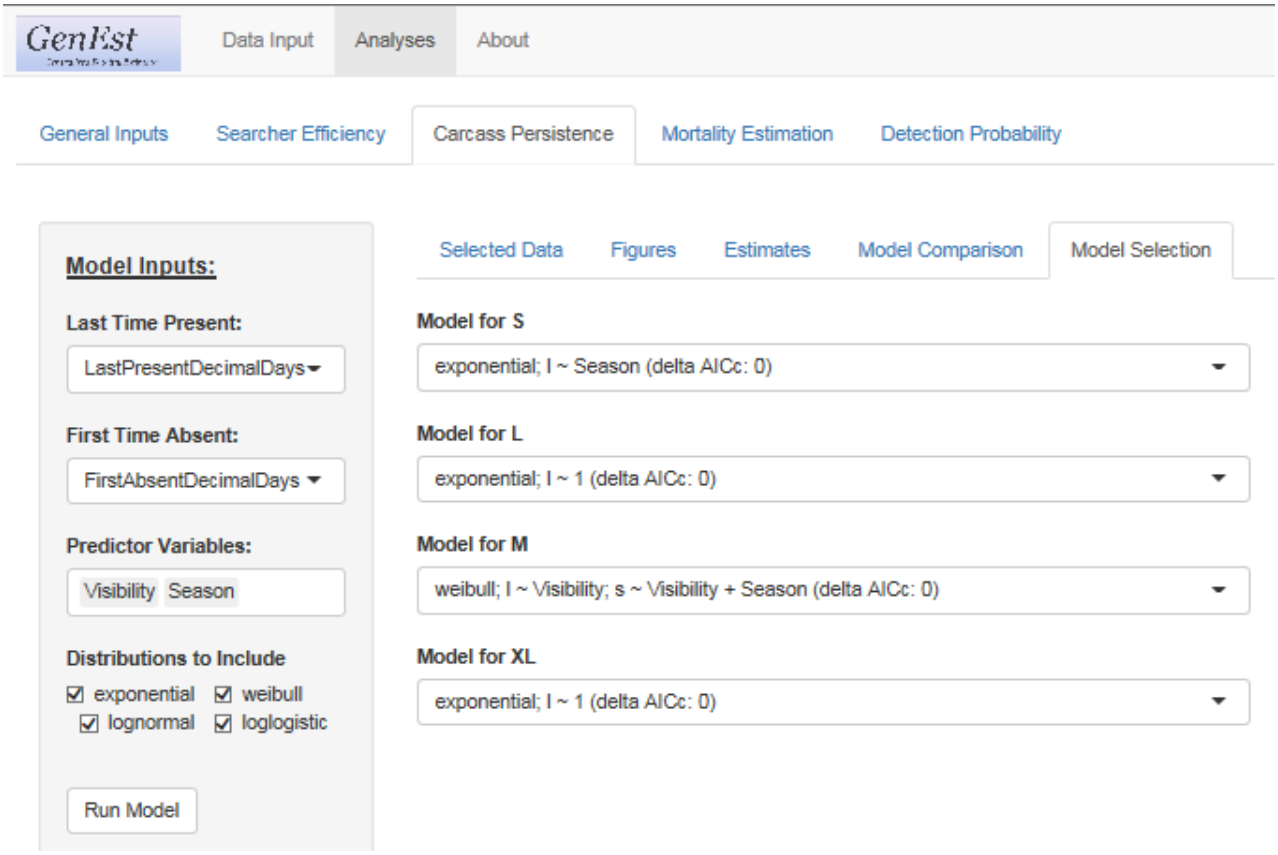

Figure 13. Model selection screen indicating which carcass persistence model to use for each size class of trial carcasses.

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" (**Figure 15**) or "Detection Probability" (following section).

3.4.1 Model Inputs

The user is asked to provide the "Fraction of Facility Surveyed" (**Figure 14**). If the facility is a wind farm or an area with clear individual sources of mortality, then a number between 0 and 1 representing the fraction of units included in the searched sample is entered. If however, the sampled areas simply represent a sampled fraction of the total area in which carcasses can arrive, the sampling fraction is entered in DWP file for every unit in the sample and the fraction of the facility surveyed is entered as 1. For example, a 100 turbine facility at which 20 turbines were selected for searching on plots cleared to 60m, the "Fraction of Facility Sampled" would be 0.2 to account for unsampled turbines and each individual turbine would have a DWP <1 to account for carcasses falling outside plots. Alternatively, a solar power tower where 100% of the area within the inner circle is searched, but one 20% of the area within each ring of reflectors surrounding a power tower in the outer circle is sampled, the "Fraction of Facility Sampled" would be 1 because all rings were sampled. 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 changes with distance [citation Solar OFR]. 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.

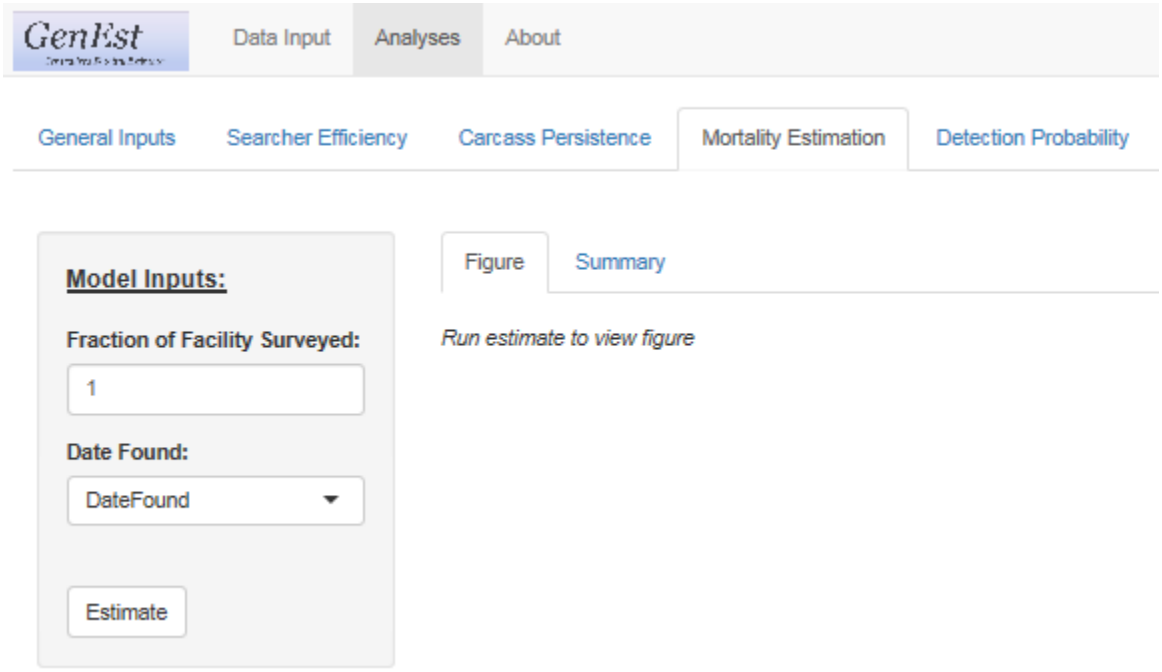

Figure 14. Model inputs screen requesting information on sampling fraction and identifying the column in the carcass observations file that represents the date on which the carcass was found by searchers.

Once the requested user input is provided, clicking "Estimate" will generate mortality estimates and confidence intervals. The 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 (**Figure 15**).

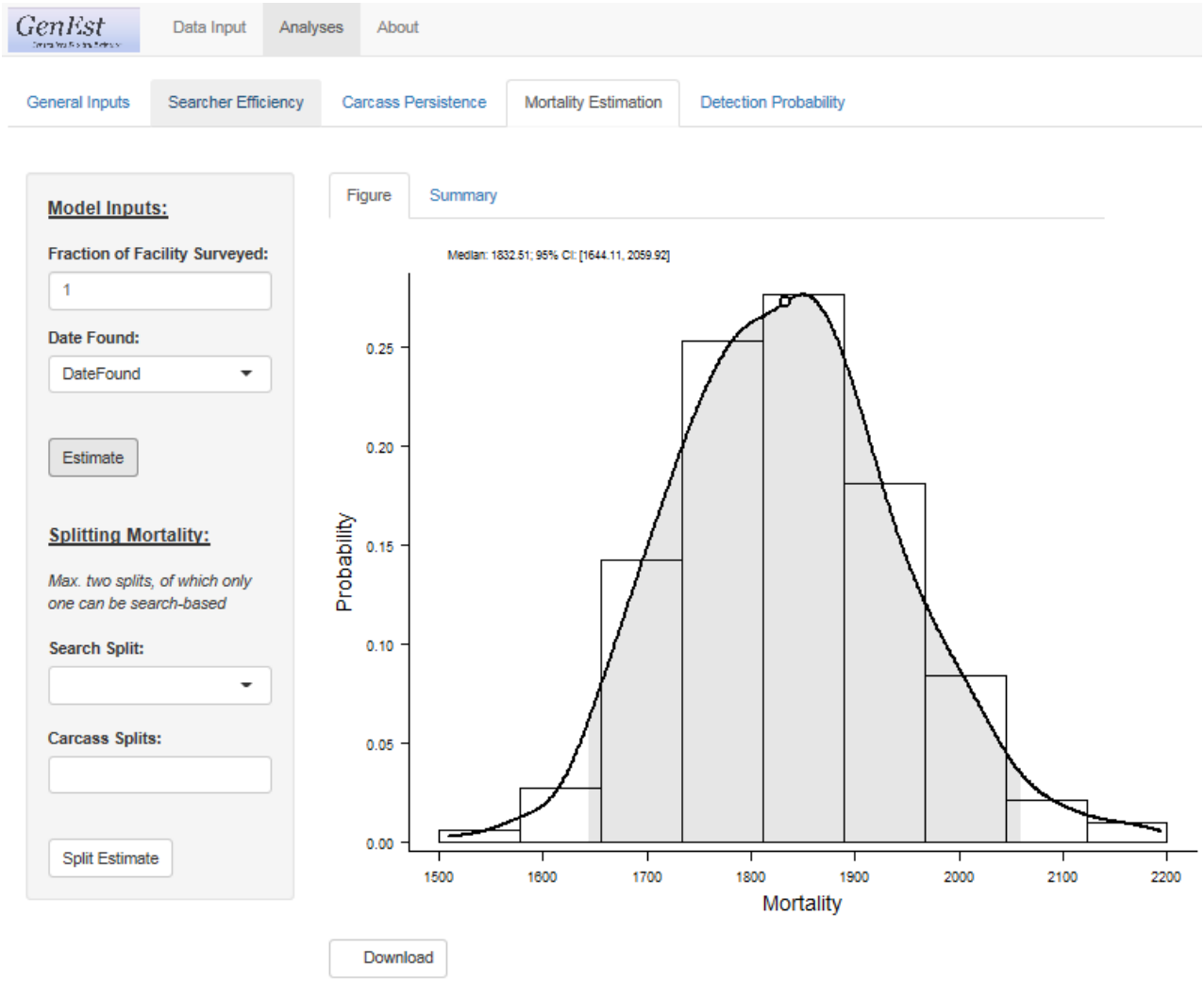

Figure 15. Estimated total mortality at the site based on user input.

3.4.1.1 Splitting Mortality Estimates

The user has the option to request estimated mortality for different subsets of the observed carcass data. These can be "Search Splits" generally referring to temporal subsets of the entire monitoring period, or "Carcass Splits" generally referring to characteristics of the carcasses, (e.g., bat vs. bird) or the conditions under which they were found(e.g., road and pad vs other). At this time, GenEst has the capacity to accommodate only two splits, only one (or none) of which can be a temporal split. Note: While GenEst can accept only two split columns, each split can have many

levels (please see examples section.) Clicking “Split Estimate” will generate a graph an a summary table of the selected split combinations (**Figure 16**).

Figure 16. Estimated mortality at the site in each combination of size and ground cover.

Selecting “Summary” will open a table of estimates and selected quantiles for each split combination (**Figure 17**). The table can be downloaded and saved as a .csv for use in reports and publications.

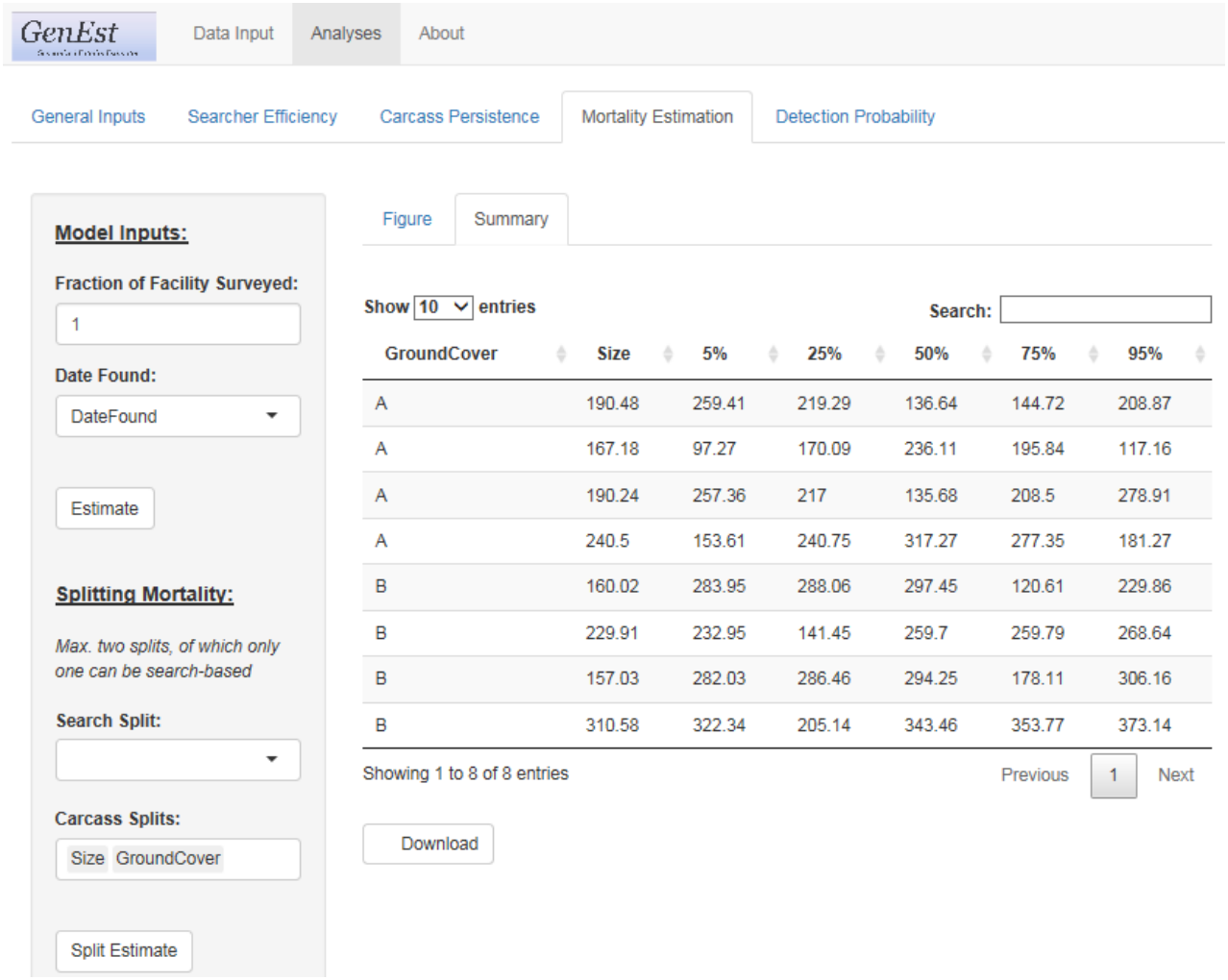

Figure 17. Estimated mortality at the site in each combination of size and ground cover. Table of quantiles used to produce associated graph.

3.5 Detection Probability

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" (prior section) or "Detection Probability."

3.5.1 Model Input

3.5.1.1 Search Schedule

Because actual search schedules may not follow intended schedules (e.g., searches are missed because of equipment maintenance, hazardous weather, unforeseen events), actual

detection probability has the potential to be unique to each carcass and possible interval in which it might have arrived. GenEst provides a generalized detection probability for relevant covariates classes based on an idealized search schedule input by the user. The user can calculate a search schedule based on the average interval in the Search Schedule data input initially. This is calculated and represented on the right when user clicks "Create Schedule." Alternatively, user can provide the "Search Schedule Data" to be used in generating generic detection probability estimates based on input values for Search Interval (7 days in the example below) and the length of time in days during which searches are conducted. (**Figure 18**).

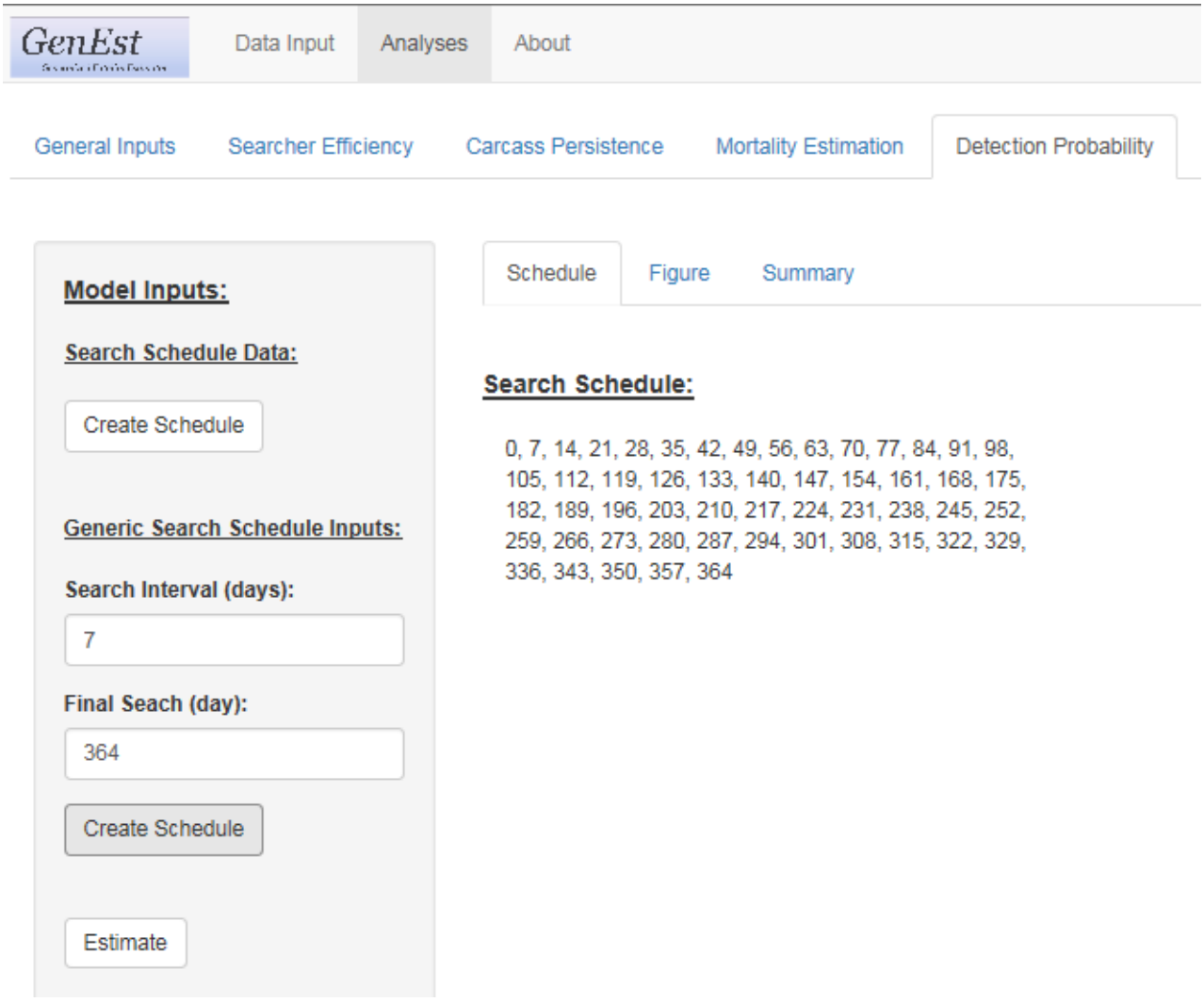

Figure 18. Model inputs screen requesting information on search schedule. Alternatively, user can input regular search interval over a set period.

3.5.2 Table and Figure Selection

When the user clicks "Estimate" The user can select the size class to be represented in the Figure and Summary table under the "Table and Figure Selection"

3.5.2.1 Summary

Selecting "Estimate" will add the "Table & Figure Selection:" box to the control panel on the left and will open a table of estimated detection probability and selected confidence interval for relevant covariates associated with the first size class. The table can be downloaded and saved as a .csv for use in reports and publications (**Figure 19**).

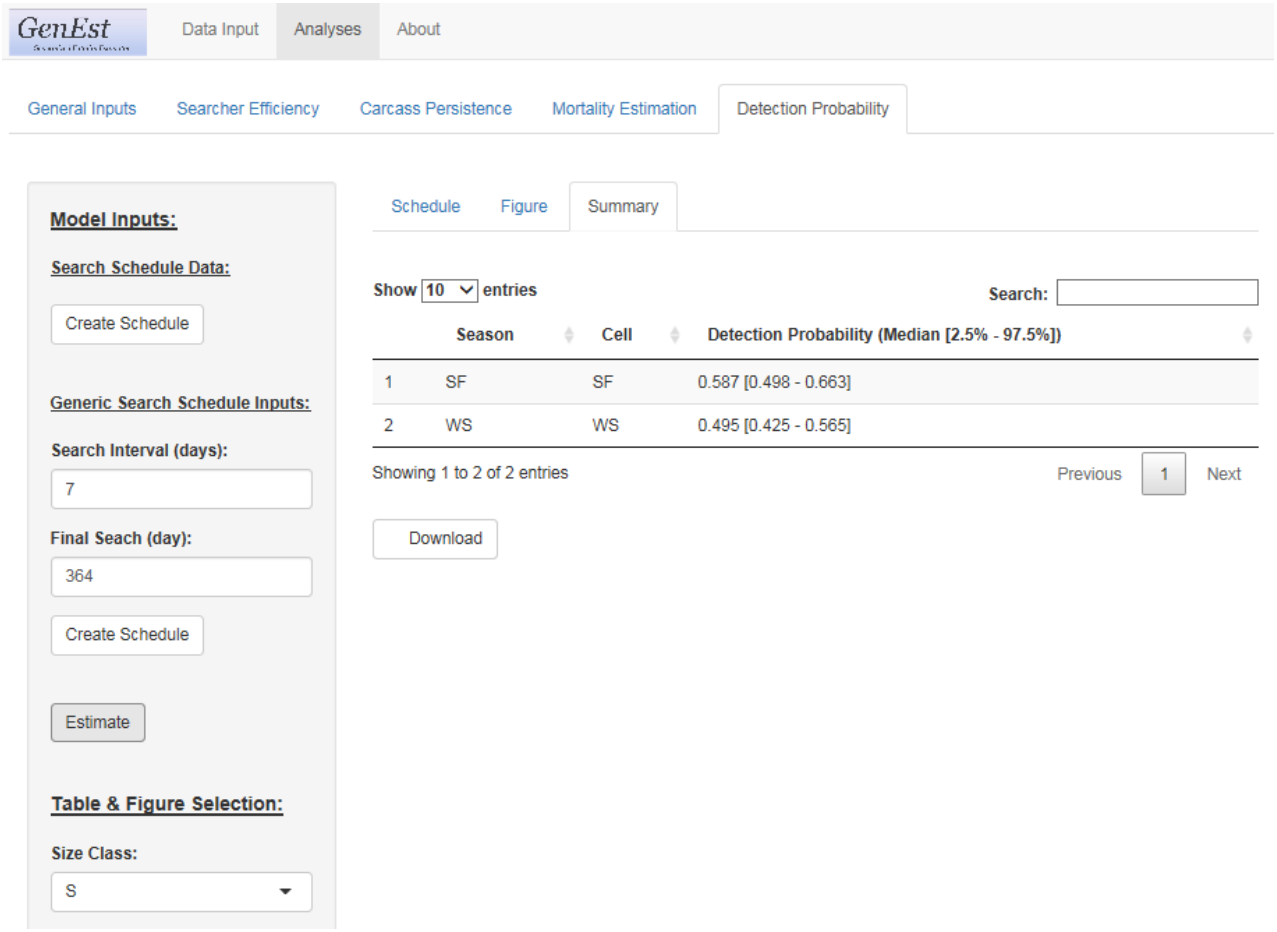

Figure 19. Summary of detection probability and confidence limits for relevant covariate classes (covariates that were included in the selected carcass persistence or searcher efficiency models) for the selected size class.

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 (**Figure 20**).

Figure 20. Box plot 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.

4.0 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 95% 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" for each file type (Searcher Efficiency, Carcass Persistence, Search Schedule, Density Weighted Proportion, and Carcass Observation) and locate the appropriate data files for SE, CP, SS, DWP, and CO. The files used for this example are located in the `extdat\wind_RPbat` folder within the GenEst library directory. As each data set is uploaded, it is displayed in the "Data Input" window (**Figure 21**).

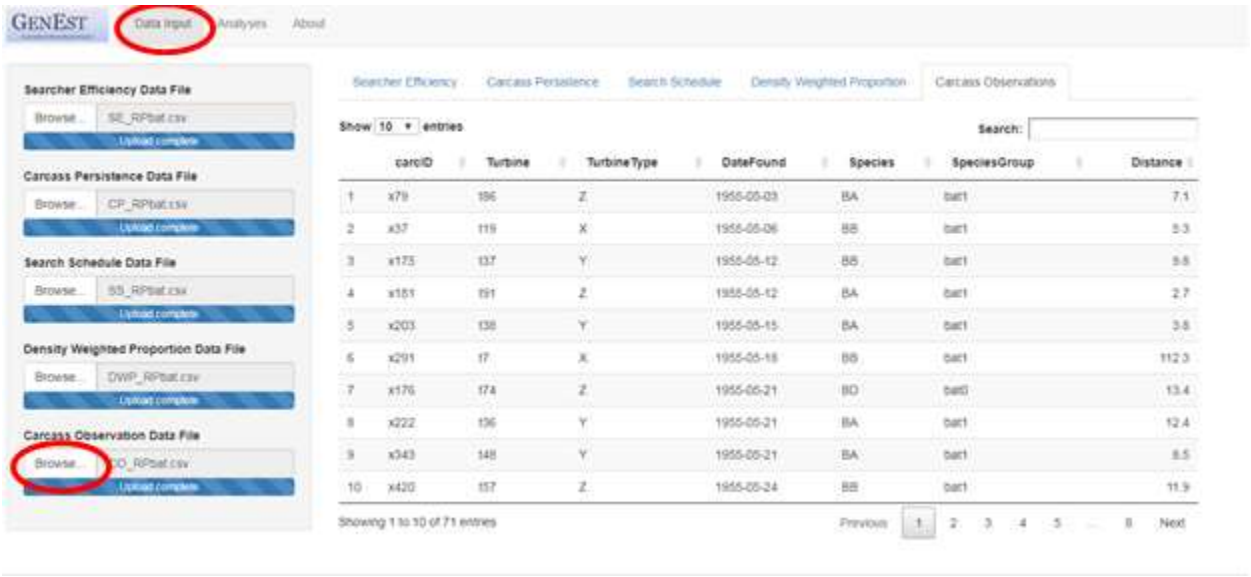

Figure 21. "Data Input" tab with "Carcass Observations" displayed after uploading.

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 (**Figure 22**).

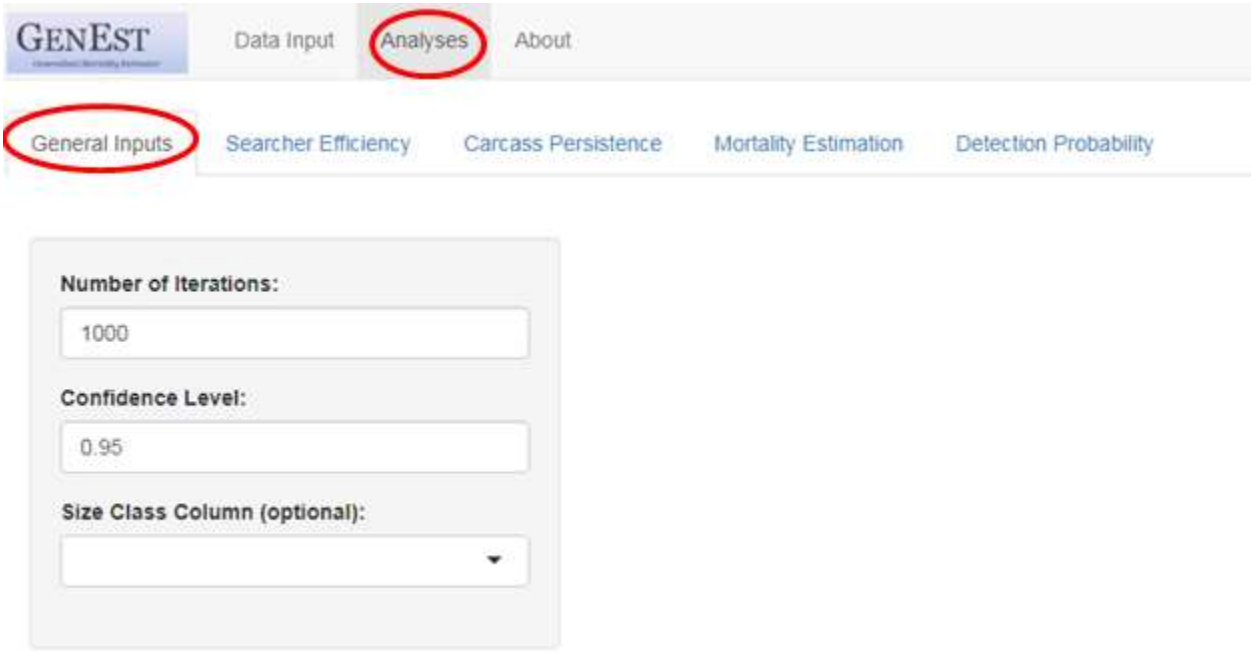

Figure 22. Starting the analysis with "General Inputs".

Confidence intervals are constructed using parametric bootstrapping, which is a simulation technique to account 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 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. Our dataset 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 (**Figure 23**).

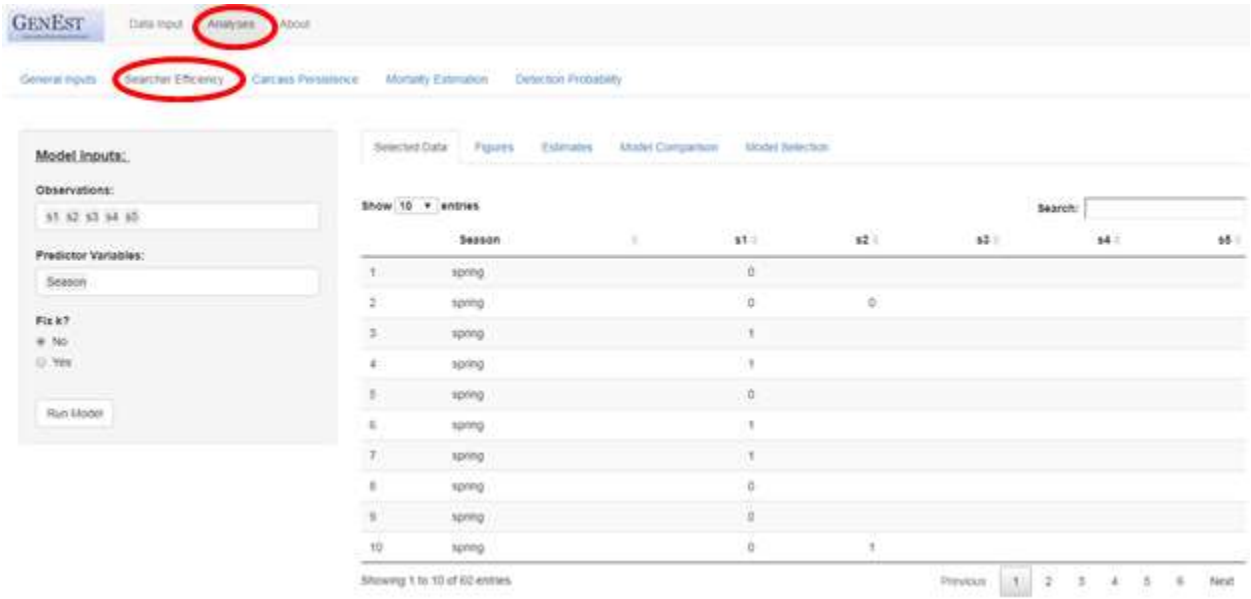

Figure 23. Model inputs for searcher efficiency analysis

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 p requires that the fates of field trial carcass data are tracked for at least one 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, user selects the names of the columns in the searcher

efficiency data file that contain the results for each search occasion (**Figure 23**). 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 for 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". Models including all combinations of covariates for both p and k are fit and displayed under the "Model Comparison" tab after user has clicked the "Run Model" button (**Figure 24**). 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 "`~ 1`"). After clicking "Run Model", uses will see the "Model Comparison" table, which shows all fitted models in order of AICc, which is a measure of quality of model fit, with better models associated with lower AICc scores. The table show 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, the models can be listed by AICc, which gives a score for the quality of the model for the given data. Complicated models that use many parameters may fit the data more closely than a simpler model but are penalized because of their complexity and relative instability. The scores have meaning only in comparison with other models'. AICc provides a rough but useful guide for model selection, but should in no way 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. When AICc differences exceed 6 or 7, preference should normally be given to the models with the lower scores. Models with differences

of less than 3 or 4 are generally considered indistinguishable by this measure. Choices among such models should be based on other criteria, such as diagnostic plots.

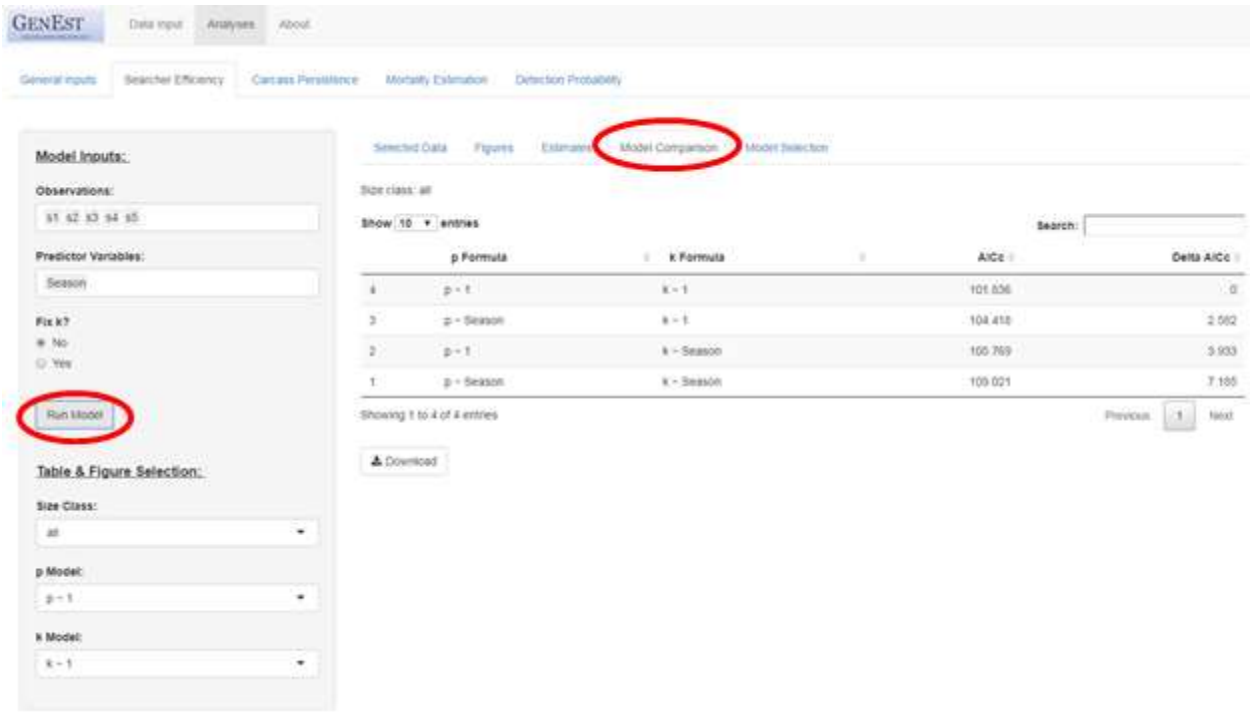

Figure 24. Searcher efficiency model comparison.

After fitting searcher efficiency models, 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 (**Figure 25**). The default is to show the model with the best AICc score in comparison with the "Reference Model", which is the most saturated (i.e., 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 tends to overfit the data, which leads to relatively large variances and introduces some instability.

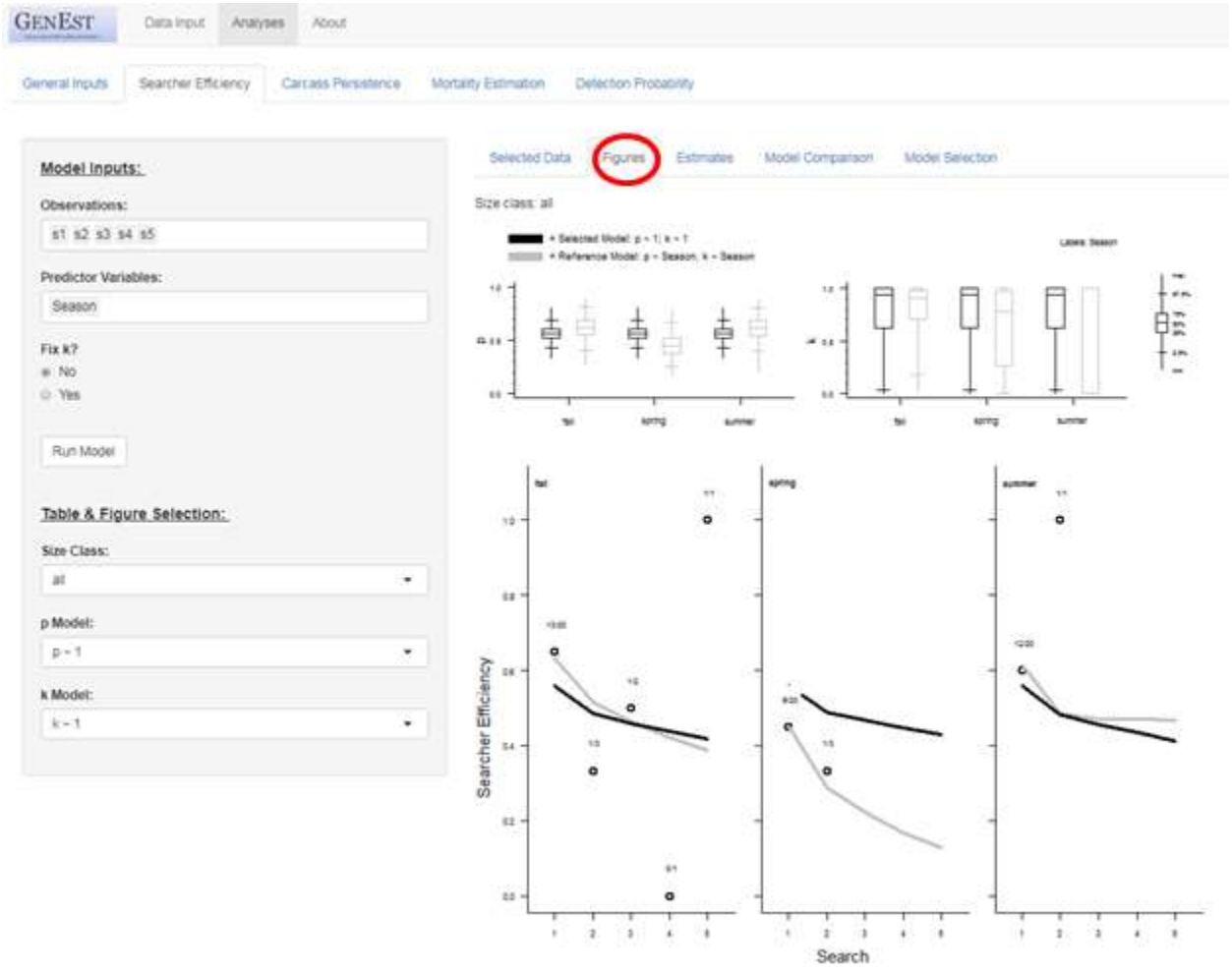

Figure 25. Searcher efficiency diagnostic plots.

The boxplots in the top section of figures for the SE diagnostic plots (**Figure 25**) 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. 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 **Figure 25**) 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. That 0-1 pattern often identifies a model that includes too many parameters. 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)

The line plots (**Figure 25**, 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 on which model to use. In this example, the model with the lowest AICc ($p \sim 1, k \sim 1$) 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, and we select that model from among the choices presented in the drop down menu under the "Model Selection" tab (**Figure 26**).

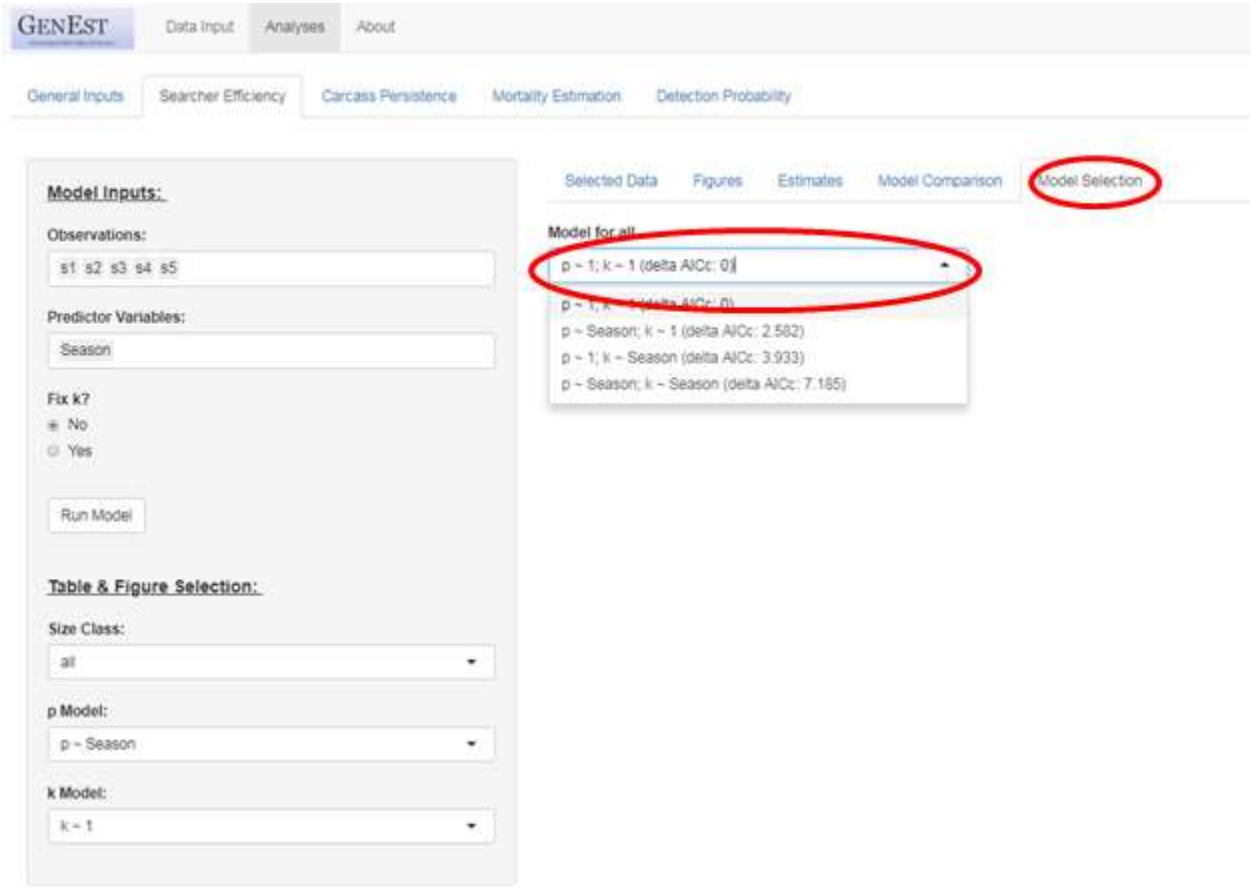

Figure 26. Searcher efficiency model selection.

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 Left and Right, indicating the left and right endpoints of the interval over which the carcass went missing. Navigate to `Analyses > Carcass Persistence` (**Figure 27**). Select "Left" for "Last time present" and "Right" 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.

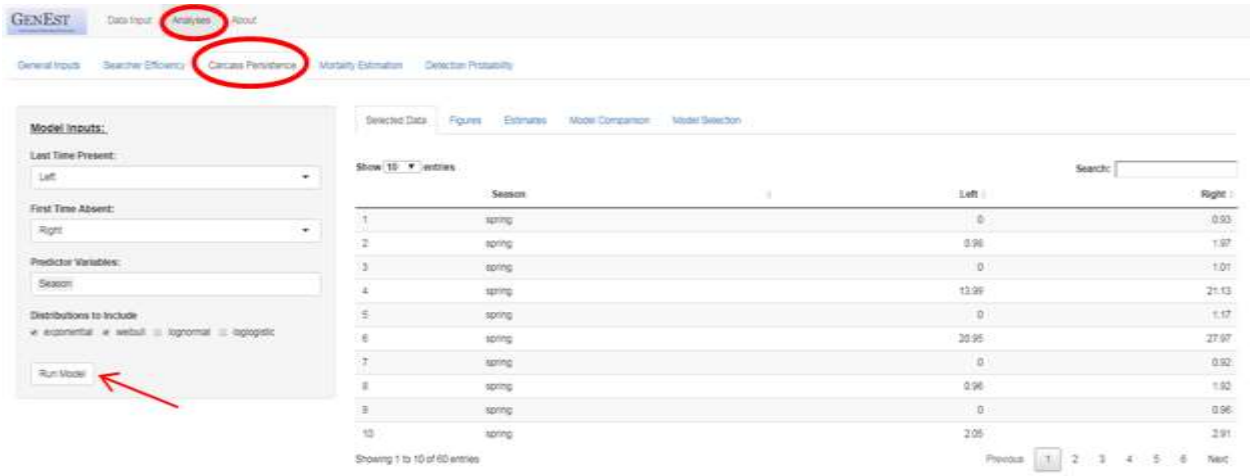

Figure 27. Carcass persistence model building.

Finally, click "Run Model" (**Figure 27**). A collection of models will be generated, ordered by ascending AICc, and displayed under the "Model Comparison" tab (**Figure 28**). Models with lower AICc scores are considered to have higher quality fits, although differences of less than about 5 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 ($l \sim 1; s \sim 1$). No exponential models are among the top ten best fits, which is not a surprise because the exponential model has only one parameter and lacks the flexibility that the 2-parameter models provide. Other models can be viewed by clicking the "Next" button (**Figure 28**).

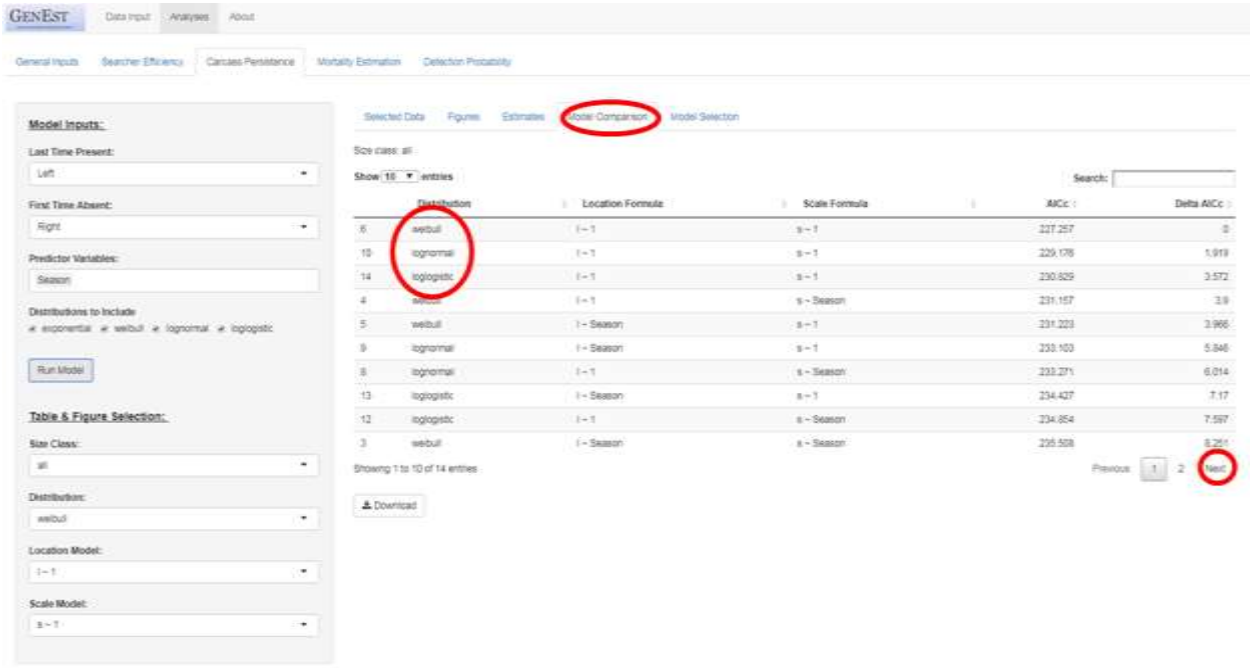

Figure 28. Carcass persistence model comparison tab.

Users may click the "Figures" tab to view fits for selected carcass persistence models (**Figure 29**). The figures show fitted line plots superimposed onto Kaplan-Meier curves that show the empirical data along with a 95% confidence interval. Users should 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 (red) is distinct from the others.

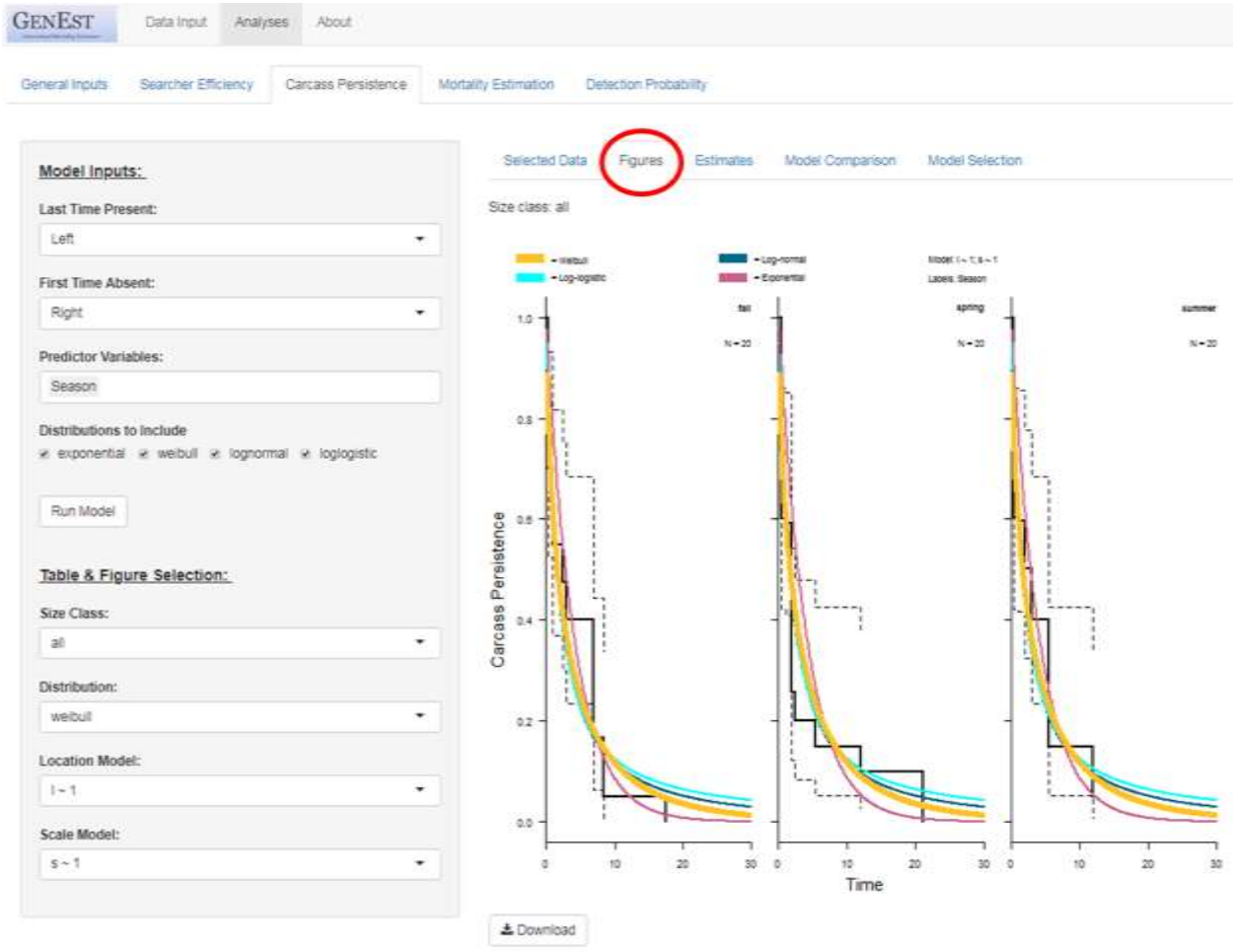

Figure 29. Fitted carcass persistence models.

To see further details about the parameter estimates, navigate to Analyses > Carcass Persistence > Estimates tab. Finally, user must choose a model from the drop down menu on the "Model Selection" tab (**Figure 30**). For this example, we will choose the model with the lowest AICc, which happens to be Weibull with location ~ 1 and scale ~ 1.

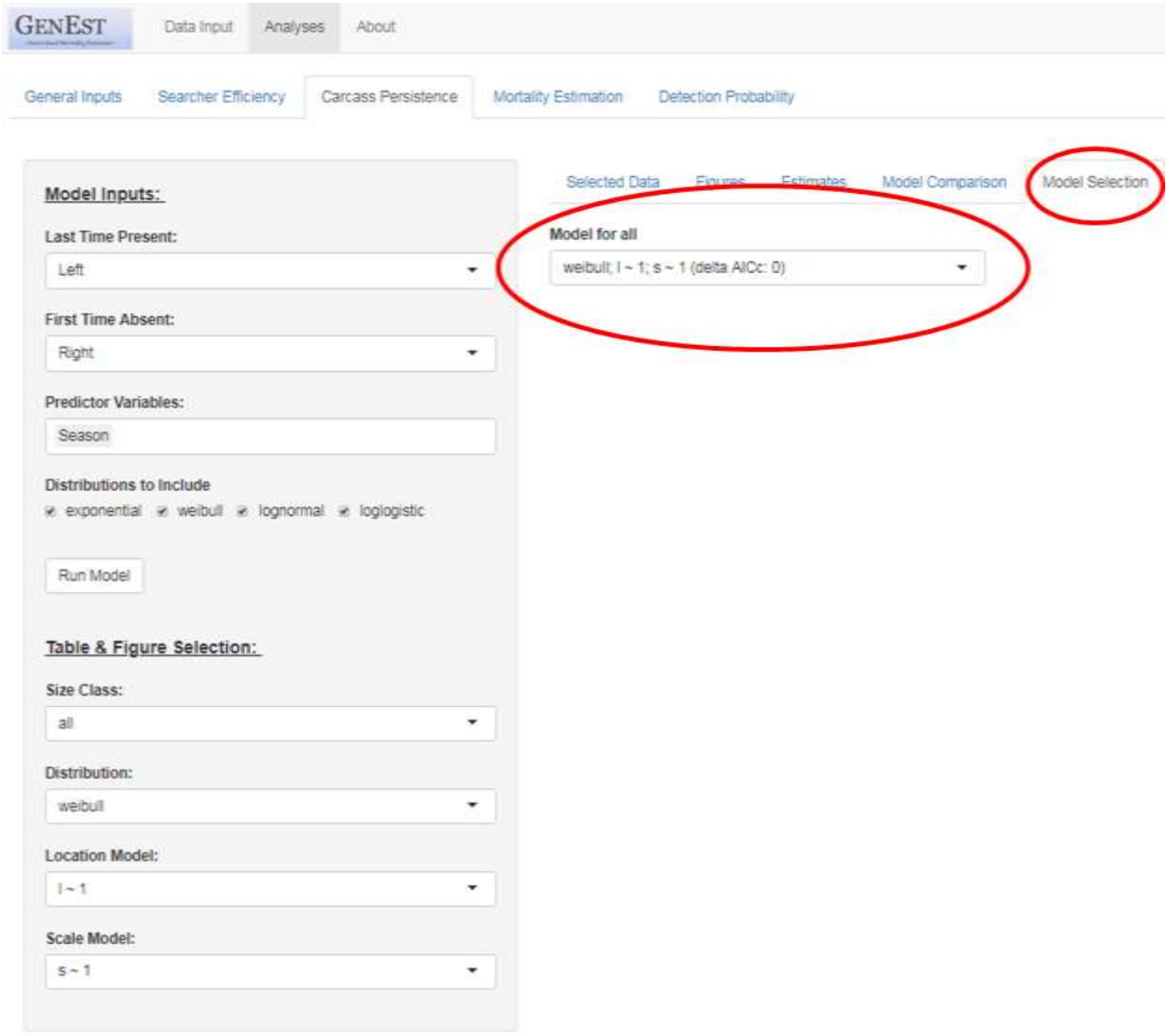

Figure 30. Carcass persistence model selection.

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 (**Figure 31**). In this example, all turbines at the site are surveyed 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 (**Figure 31**).

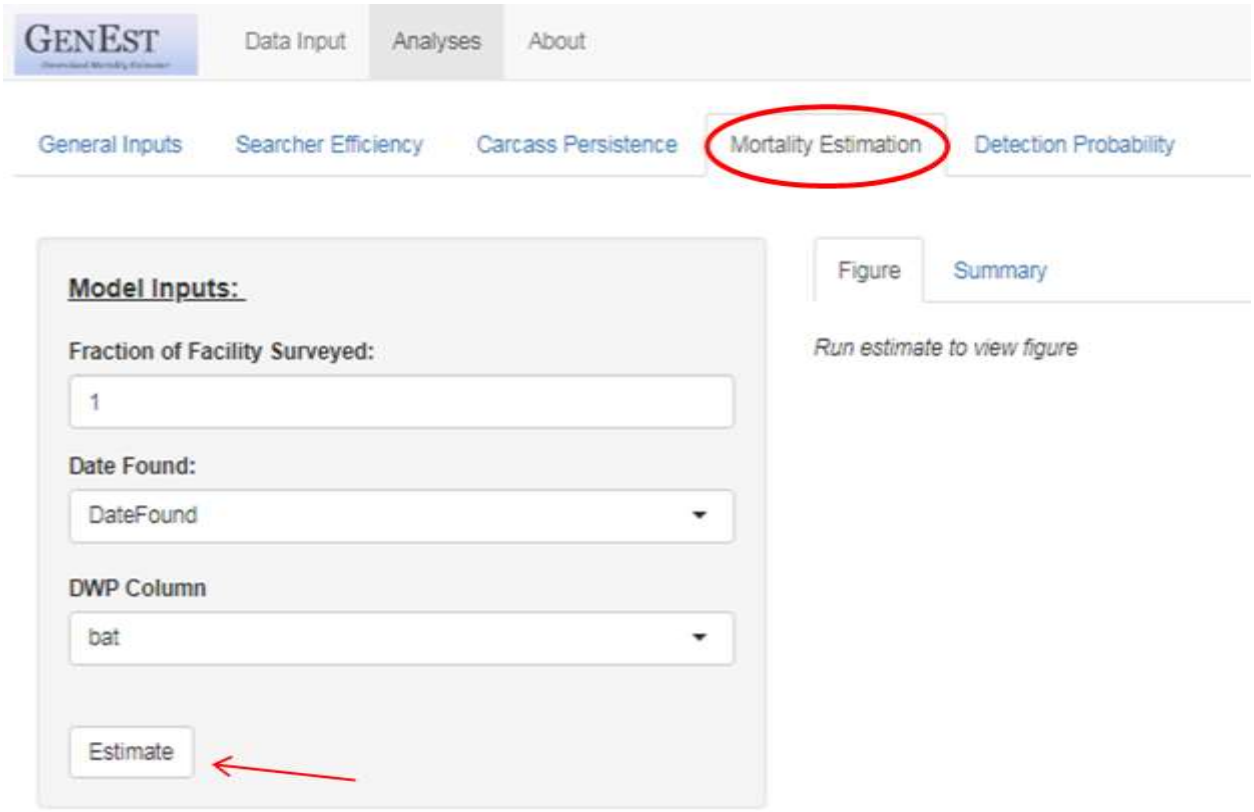

Figure 31. Mortality estimation tab.

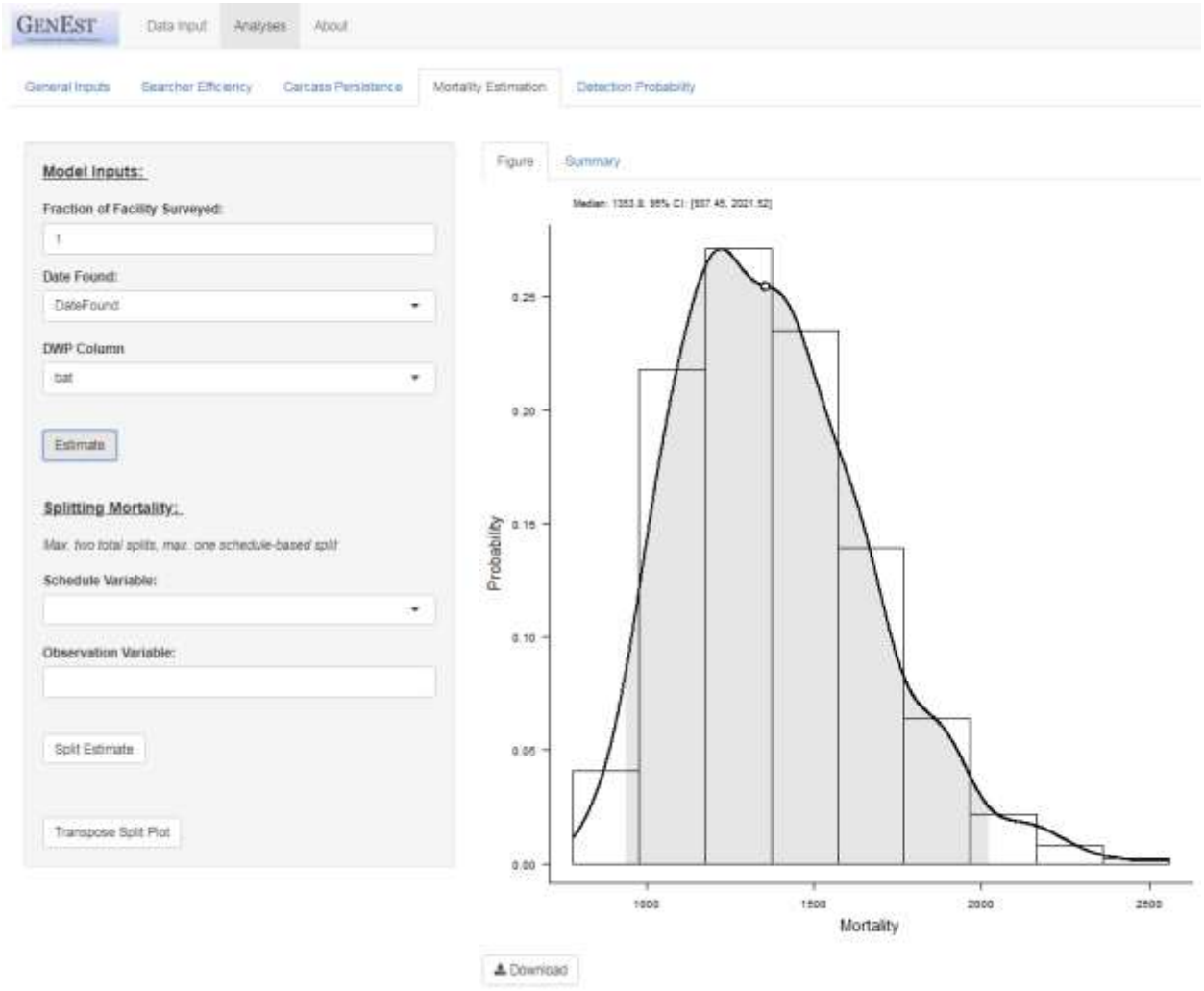

Figure 32. Total mortality and options for mortality splits.

The figure for total mortality represents the estimated mortality as a histogram of probable values, with the central confidence interval highlighted with gray shading (**Figure 32**). The median and confidence interval (95% 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. Frequently with skewed distributions, the mean is a misleading estimator because it can depend strongly on the degree of skewness. That is why the median is reported rather than the mean. Other statistics of the estimated mortality can be viewed under the "Summary" tab.

To obtain estimates for levels of temporal covariates (such as season) that can be found in the search schedule file, we select them as "Schedule Variable" splits and click "Split Estimate" to show the estimated numbers of fatalities by season (**Figure 33**). The statistics can be viewed in tabular form under the "Summary" tab.

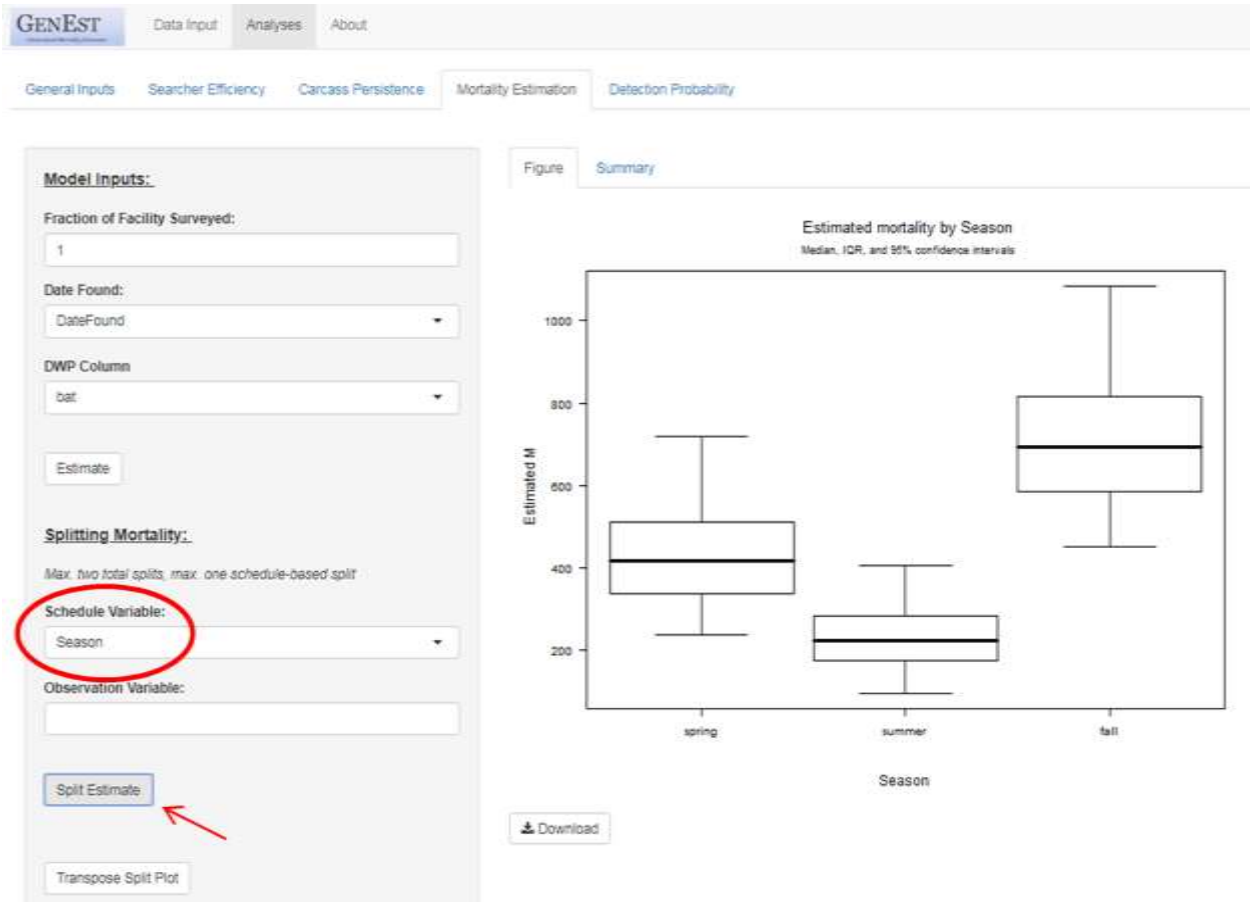

Figure 33. Mortality estimates by season.

To estimate mortality by species, first remove the season covariate by clicking in the "Schedule Variable" pressing the backspace key to delete it. Add an "Observation Variable" split for Species, and click "Split Estimate" (**Figure 34**). New boxplots and table of statistics representing mortality by species are created.

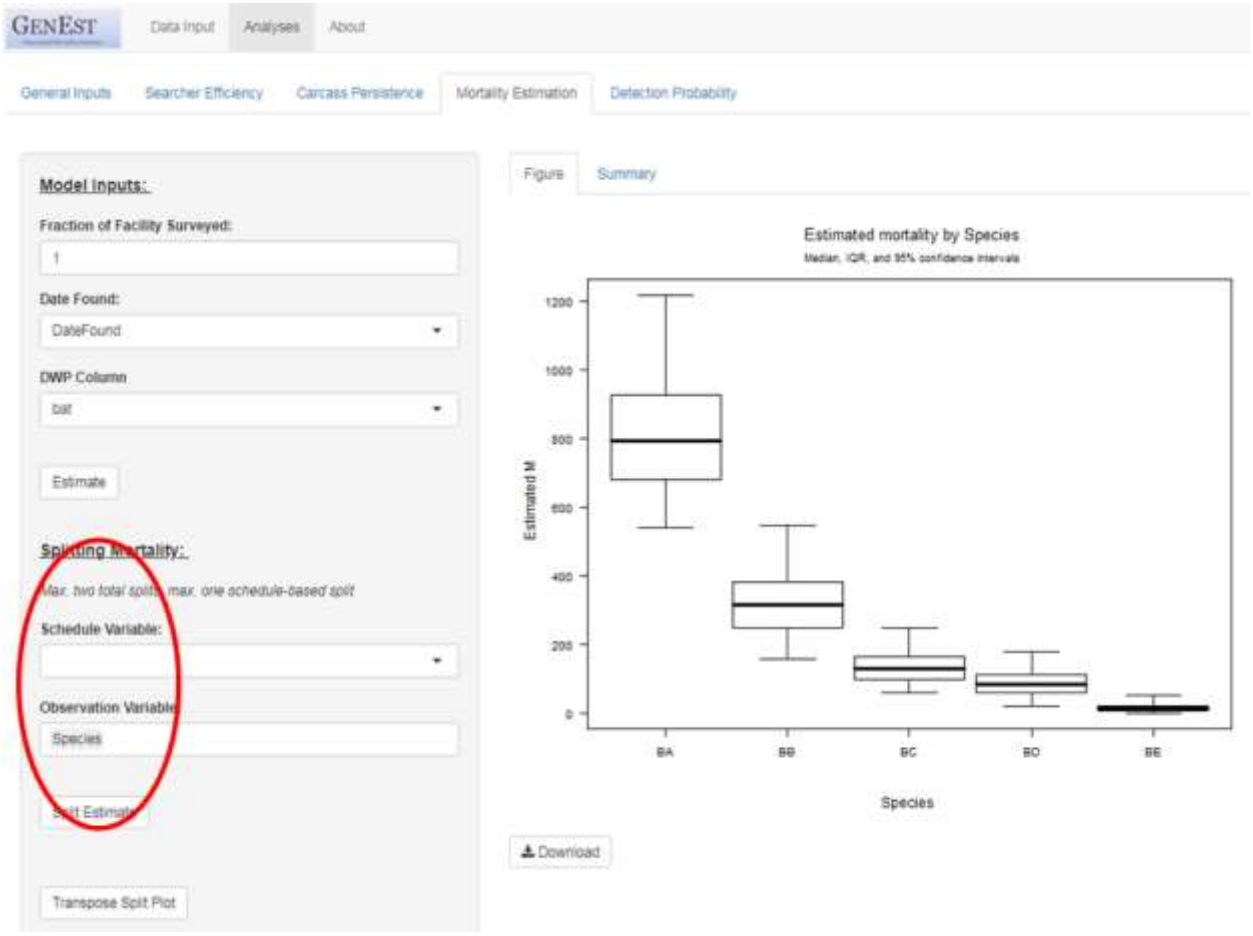

Figure 34. Mortality estimates by species.

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 a Schedule and an Observation variable. Add back the Season covariate to schedule variables, and click "Split Estimate" once more. We now end up with five panels, one for each species (**Figure 35**), 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 (**Figure 36**).

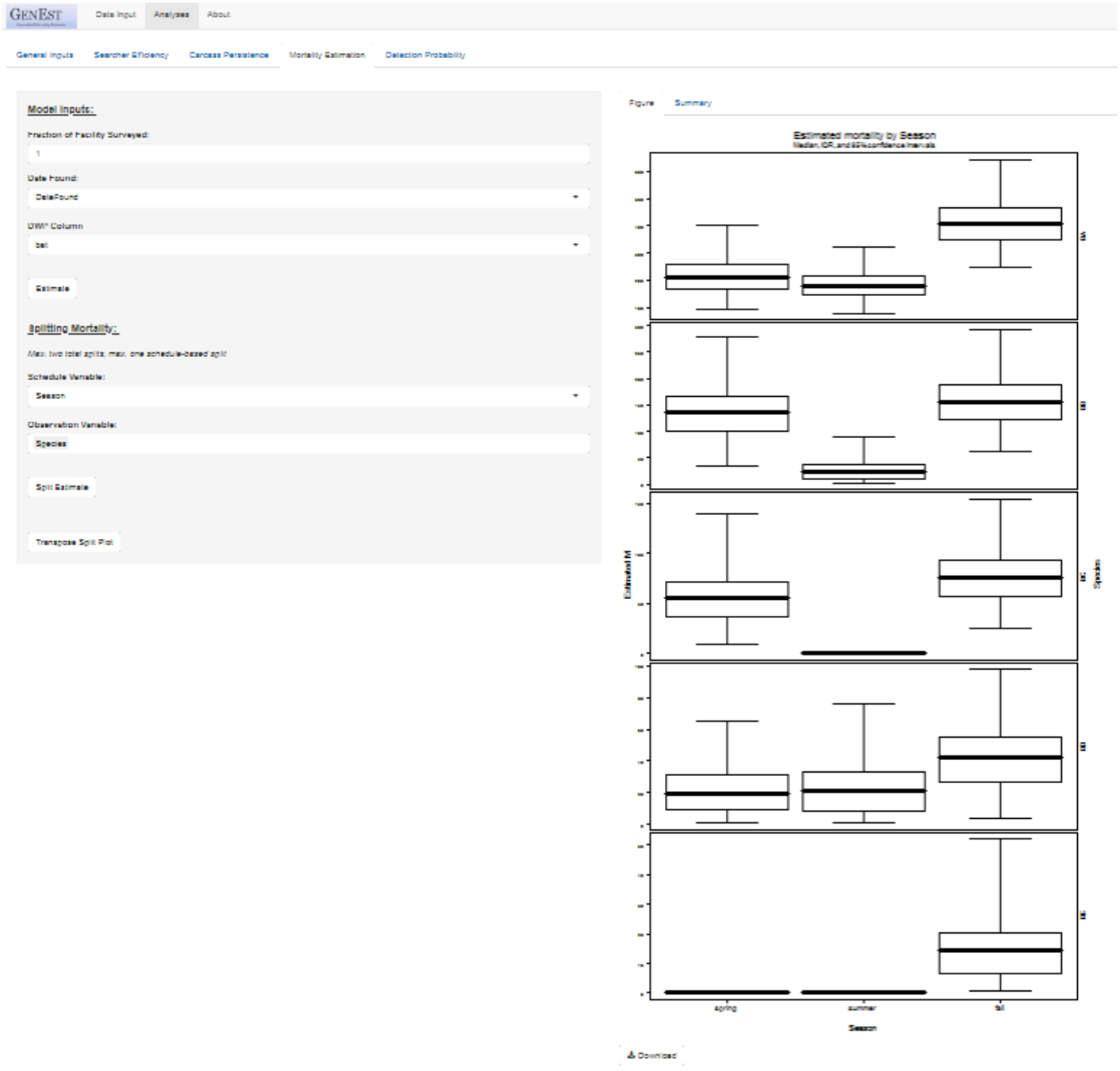

Figure 35. Seasonal mortality by species.

Figure 36. 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 (note that the monitoring period was less than one year), how bats 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 size classes of birds 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 (**Figure 37**). The files used for this example are located in the extdat\solar_pv folder within the GenEst library directory.

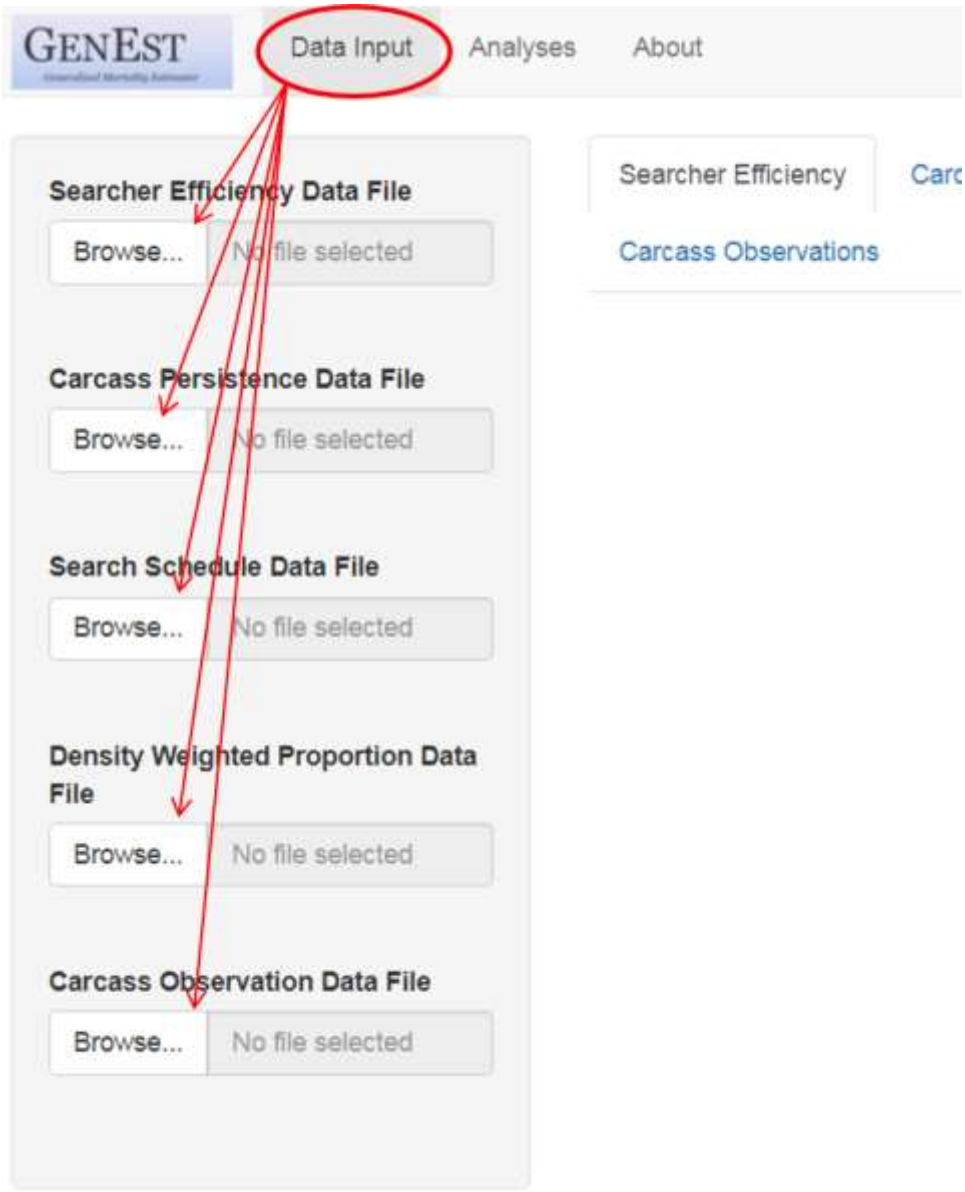

Figure 37. Data import

Figure 38. General inputs

Navigate to the Analysis > General Inputs and verify that the confidence level matches the desired 95% for our confidence interval (**Figure 38**).

Select Size from the Size Class Column drop-down. 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 calculating Searcher Efficiency and Carcass Persistence. Leaving size out would build a single SE model and a single CP covering all animals, as in the previous example.

4.2.2 Modeling Search Efficiency

Navigate to Analyses > Searcher Efficiency: Observations, and select the columns from the SE file representing the carcass searches (**Figure 39**). There were four searches, labeled Search1…Search4. In the box Predictor Variables, select Season.

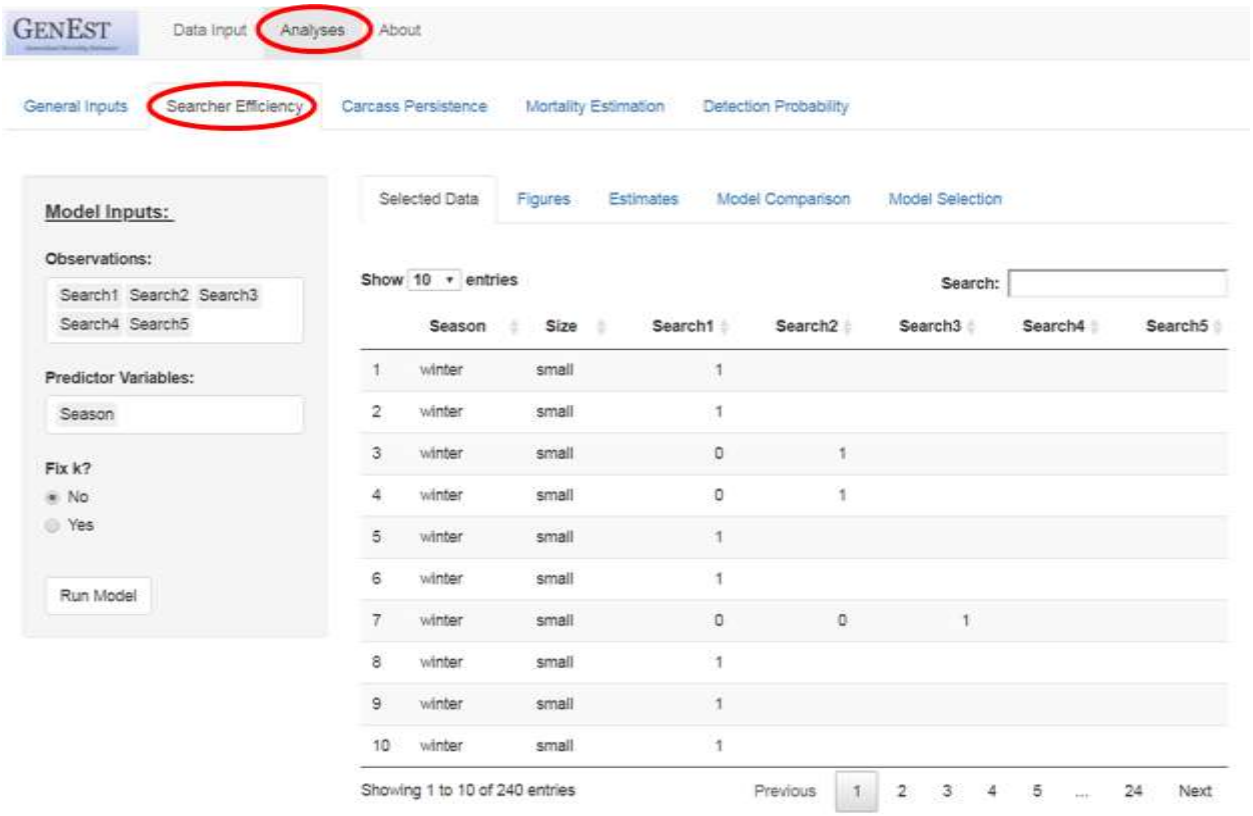

Figure 39. Data for searcher efficiency analysis.

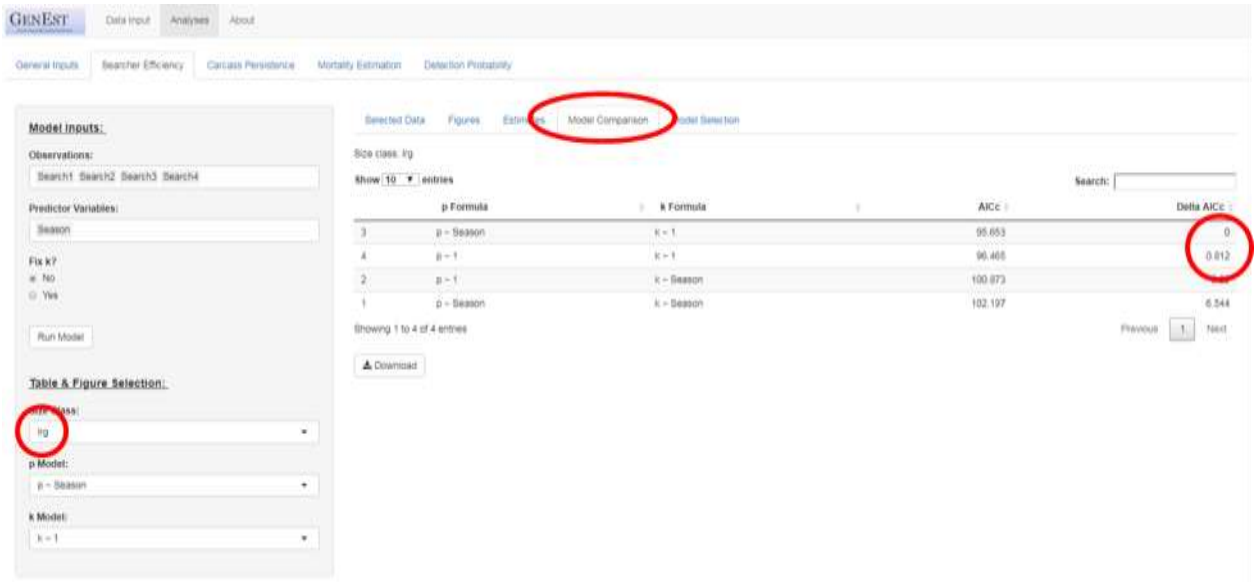

Figure 40. Searcher efficiency model comparison.

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 (**Figure 40**). As a rule of thumb, a lower AICc indicates a higher quality model, but models with AICc differences less than about 5 are generally considered indistiguishable by this measure. In this example, with the large carcasses, the $p \sim \text{Season}; k \sim 1$ and $p \sim 1; k \sim 1$ models have AICc's that differ by only 0.812, so other criteria should be used to choose between these two models.

The diagnostic plots (**Figure 41**) reveal a problem with the $p \sim \text{Season}; k \sim 1$ model. In particular, the box plots for estimated k for the selected model (**Figure 41**, 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 $p \sim 1; k \sim 1$ model does not have that problem with k and is preferable. In cases where all fitted models have unfittable k values, users may fix k at 1 or 0.

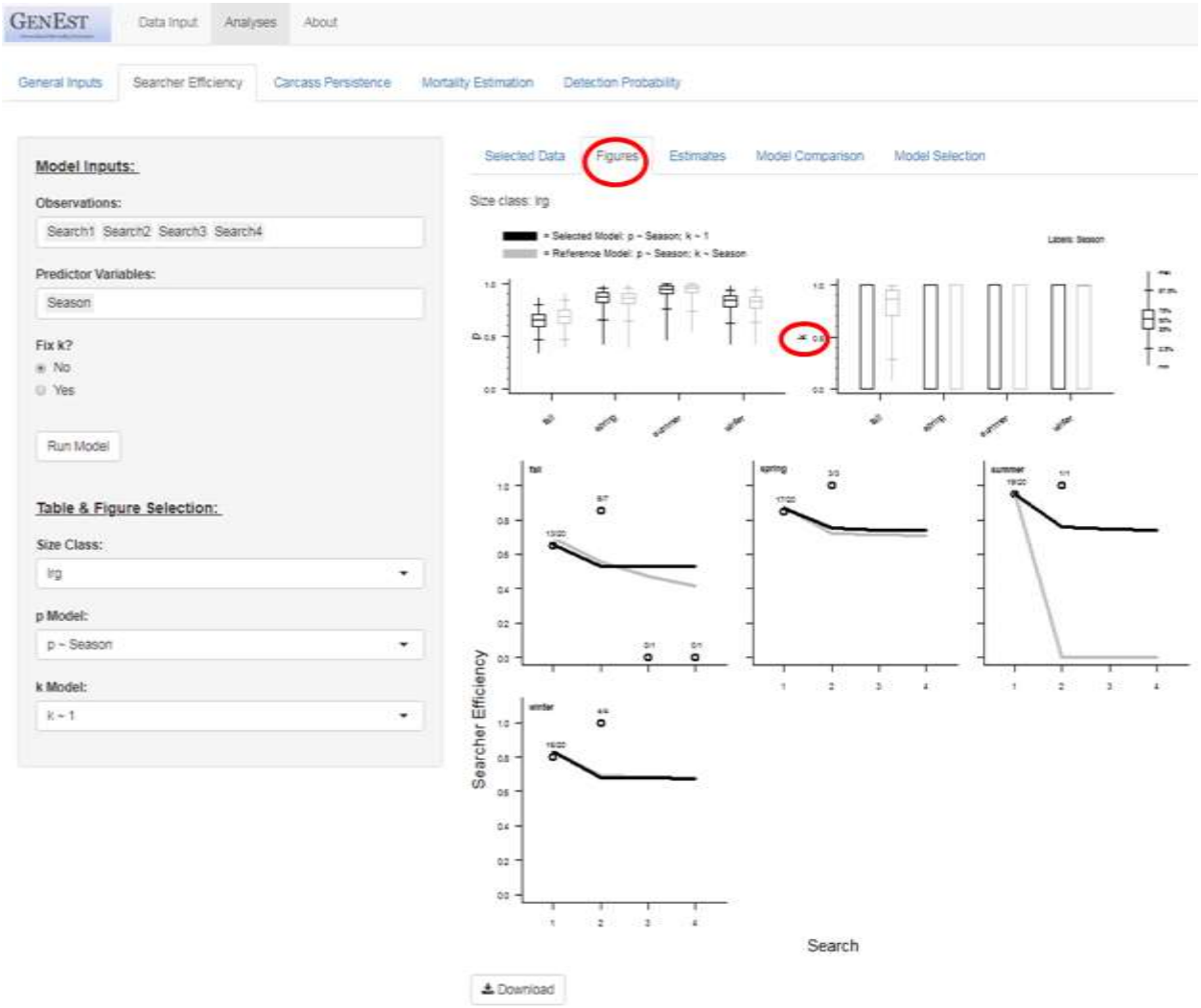

Figure 41. Diagnostic searcher efficiency plots for large carcasses at a solar PV project

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 between 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, navigate to Analyses > Searcher Efficiency >Model Selection. For this example, select the best AICc model (defaults) for small and medium birds and the $p \sim 1; k \sim 1$ model for large birds (**Figure 42**).

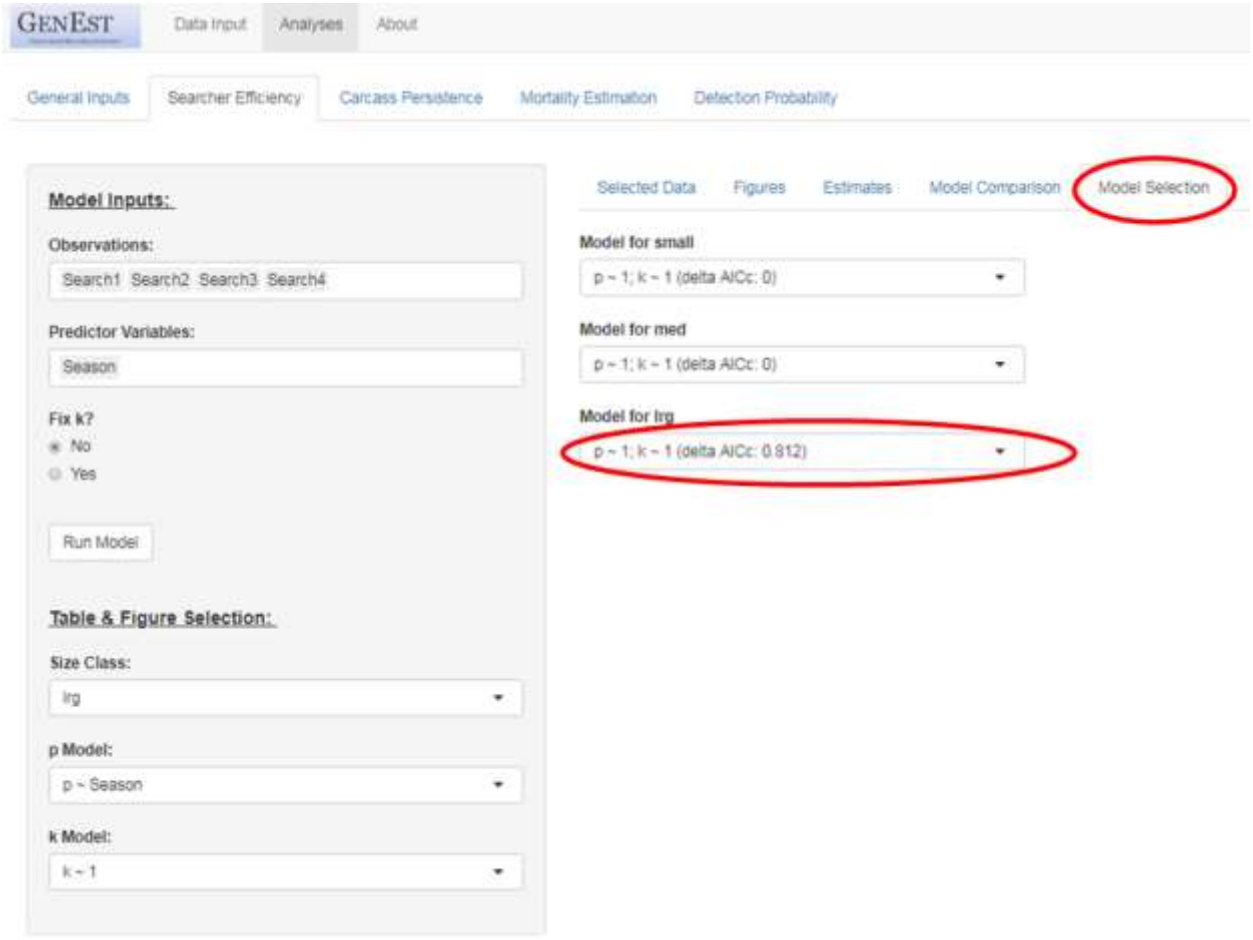

Figure 42. Selecting the SE models for each size class for the solar PV example.

This finishes the Searcher Efficiency stage.

4.2.3 Modeling Carcass Persistence

We wish to consider carcass persistence models based on Weibull or Exponential distributions, and depend in location or scale on the Season. Our CP file contains a column for the Season in which each carcass trial was performed, plus columns Left and Right, indicating the left and right endpoints of the interval over which the carcass went missing. Navigate to Analyses > Carcass Persistence. Select "Left" for "Last Time Present", "Right" 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 (**Figure 43**).

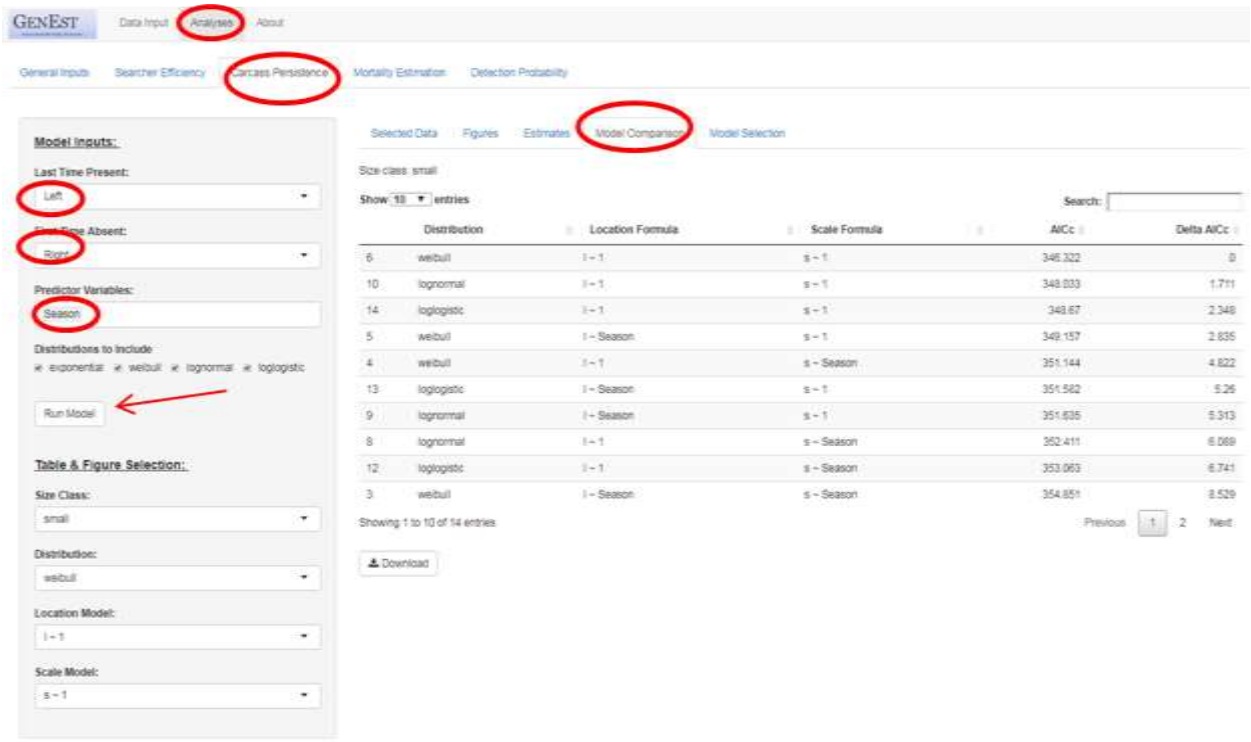

Figure 43. Fitting carcass persistence models for solar PV data set.

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 all 2-parameter models (Weibull, lognormal, loglogistic), and exponential models rank dead last by a long shot. Many of the models have AICc scores comparable to the "best", 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 fitting the best among the distributions for each particular model form. The figures for the $l \sim 1; s \sim 1$ models (**Figure 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" (~ 1) models may not be adequate.

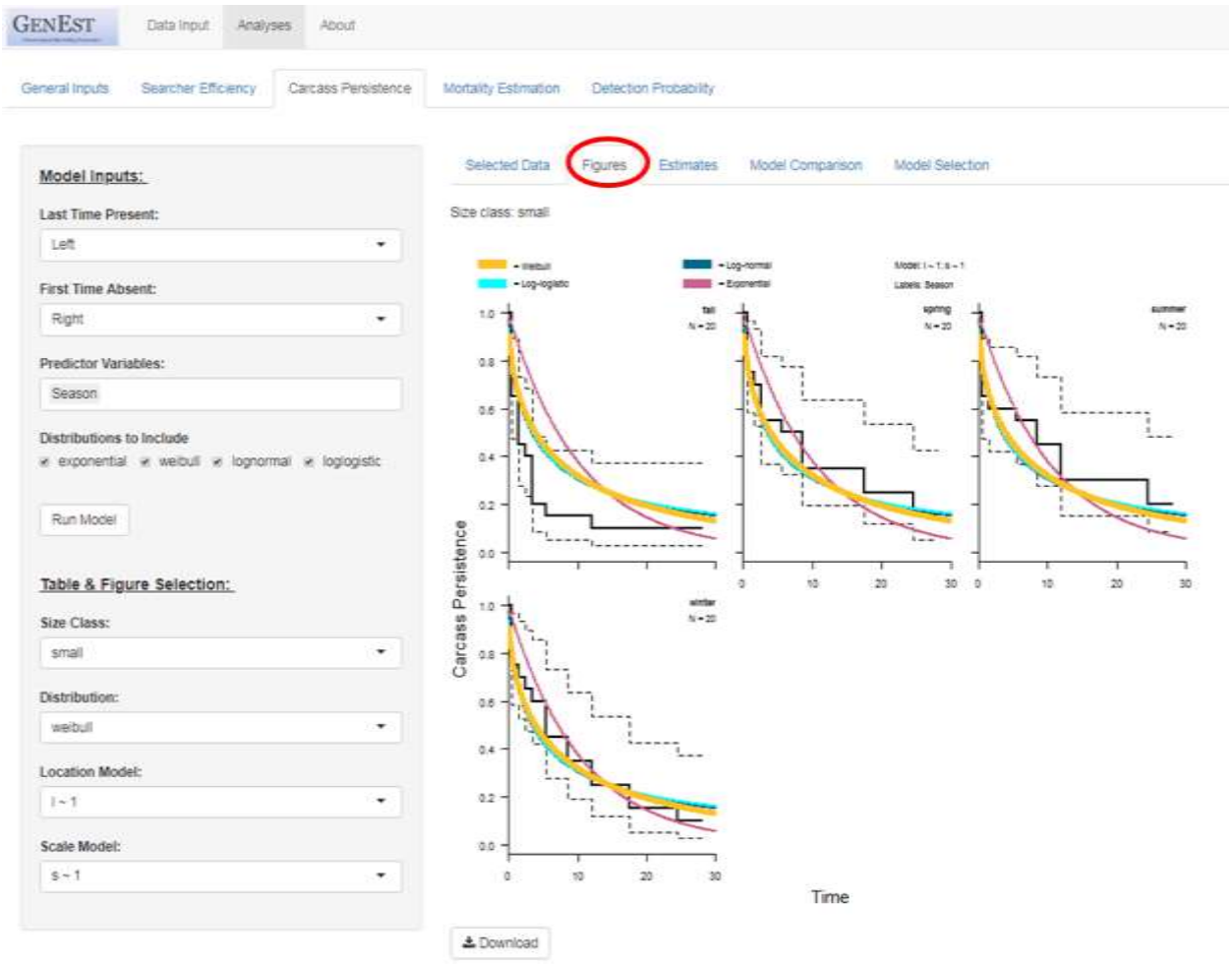

Figure 44. Fitted carcass persistence distributions for the $l \sim 1; s \sim 1$ models.

The $l \sim$ Season; $s \sim 1$ models (**Figure 45**) are comparable to the simpler $l \sim 1; s \sim 1$ models in terms of AICc, but they appear to more accurately reflect differences in CP among seasons.

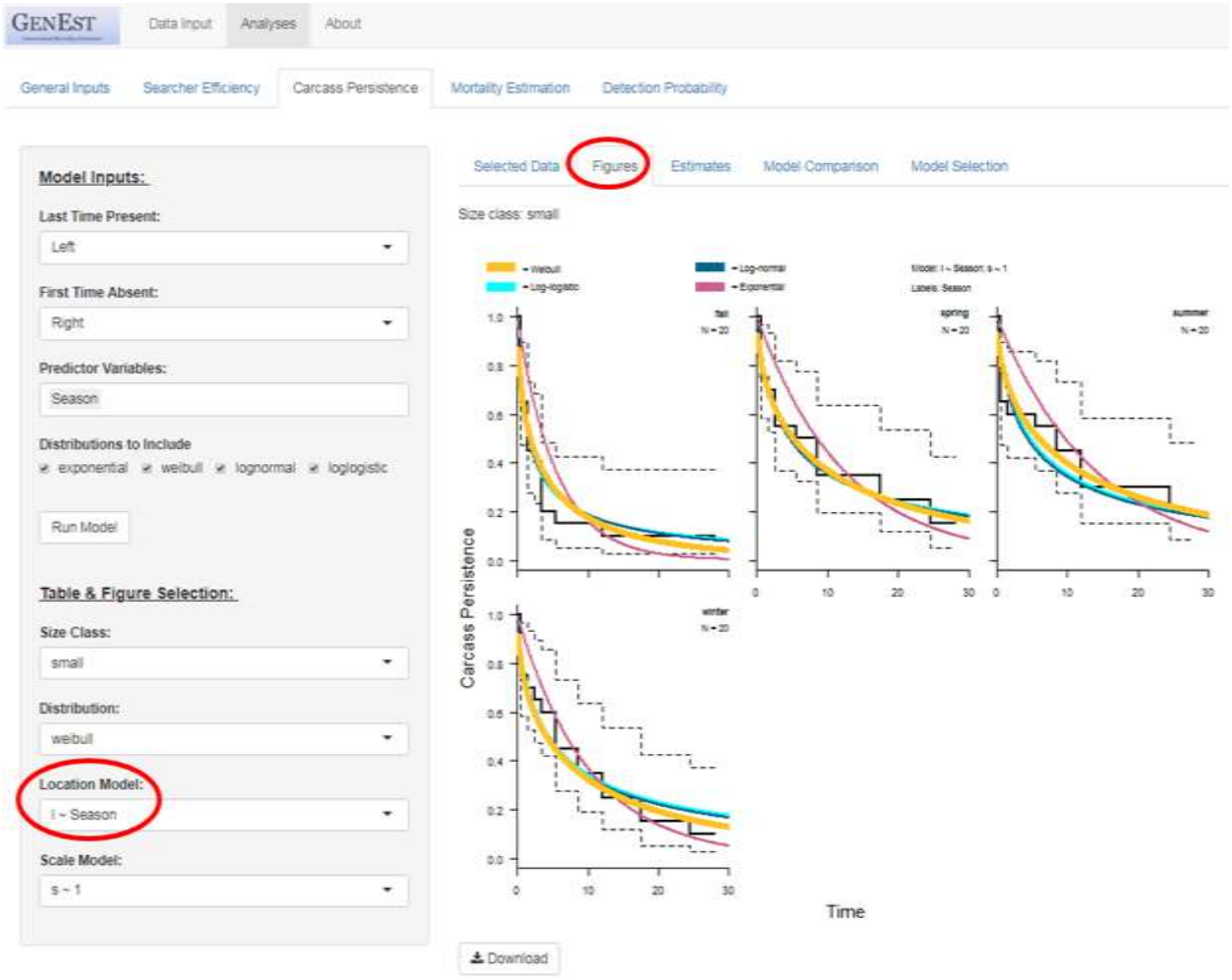

Figure 45. Fitted carcass persistence distributions for the $l \sim \text{Season}; s \sim 1$ models.

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 $l \sim \text{Season}; s \sim 1$ model (with ΔAICc = 2.835) and for medium and large carcasses, we select the model with the lowest AICc score (**Figure 46**).

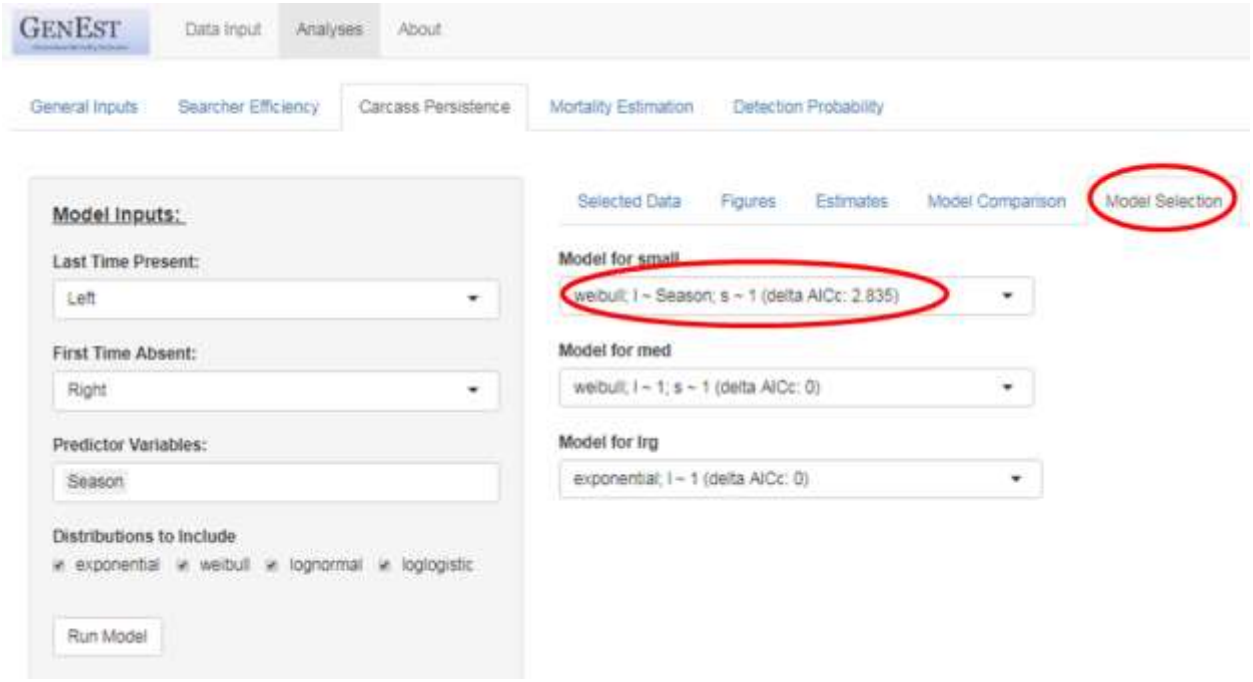

Figure 46. Selection of carcass persistence models for solar PV example

This finishes the Carcass Persistence stage.

4.2.4 Mortality Estimation

To complete the estimation of mortality, navigate to the Analyses > 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 (**Figure 47**). This is the total combined number of fatalities across all carcass size classes.

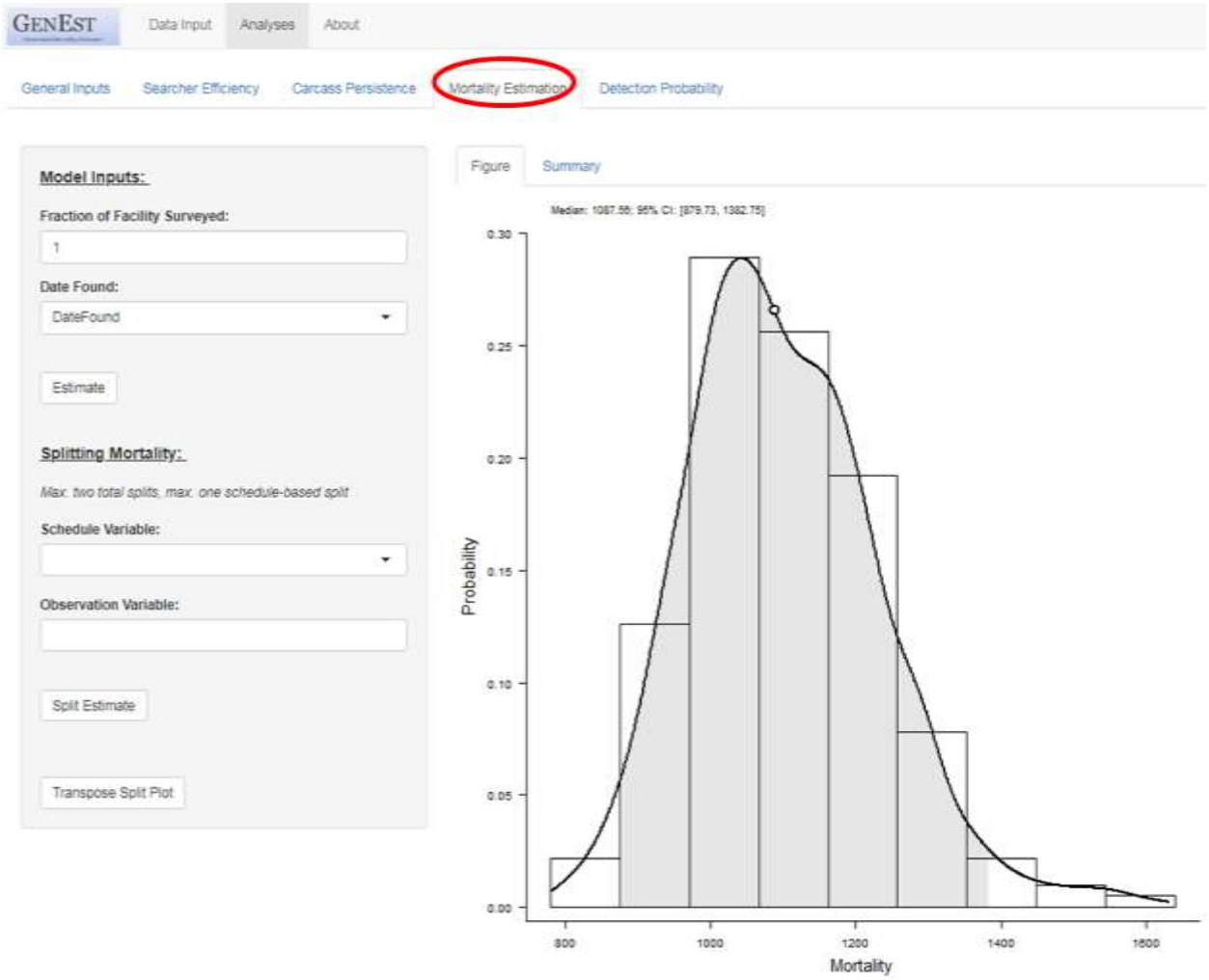

Figure 47. Total mortality in the solar PV example.

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

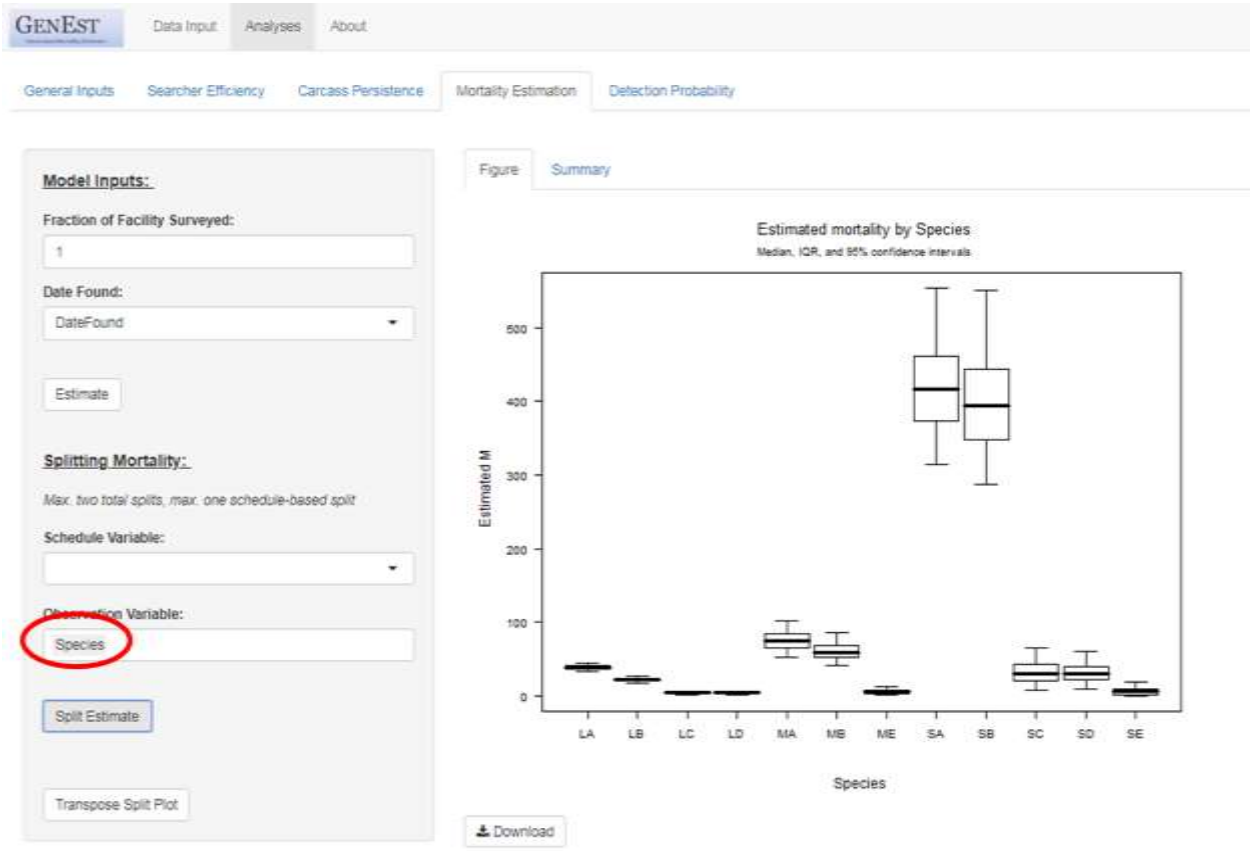

Figure 48. Mortality by species for solar PV example

To view mortality by species and season, select "Season" as the schedule variable and "Species" as the observation variable and click "Split Estimate" (**Figure 49**).

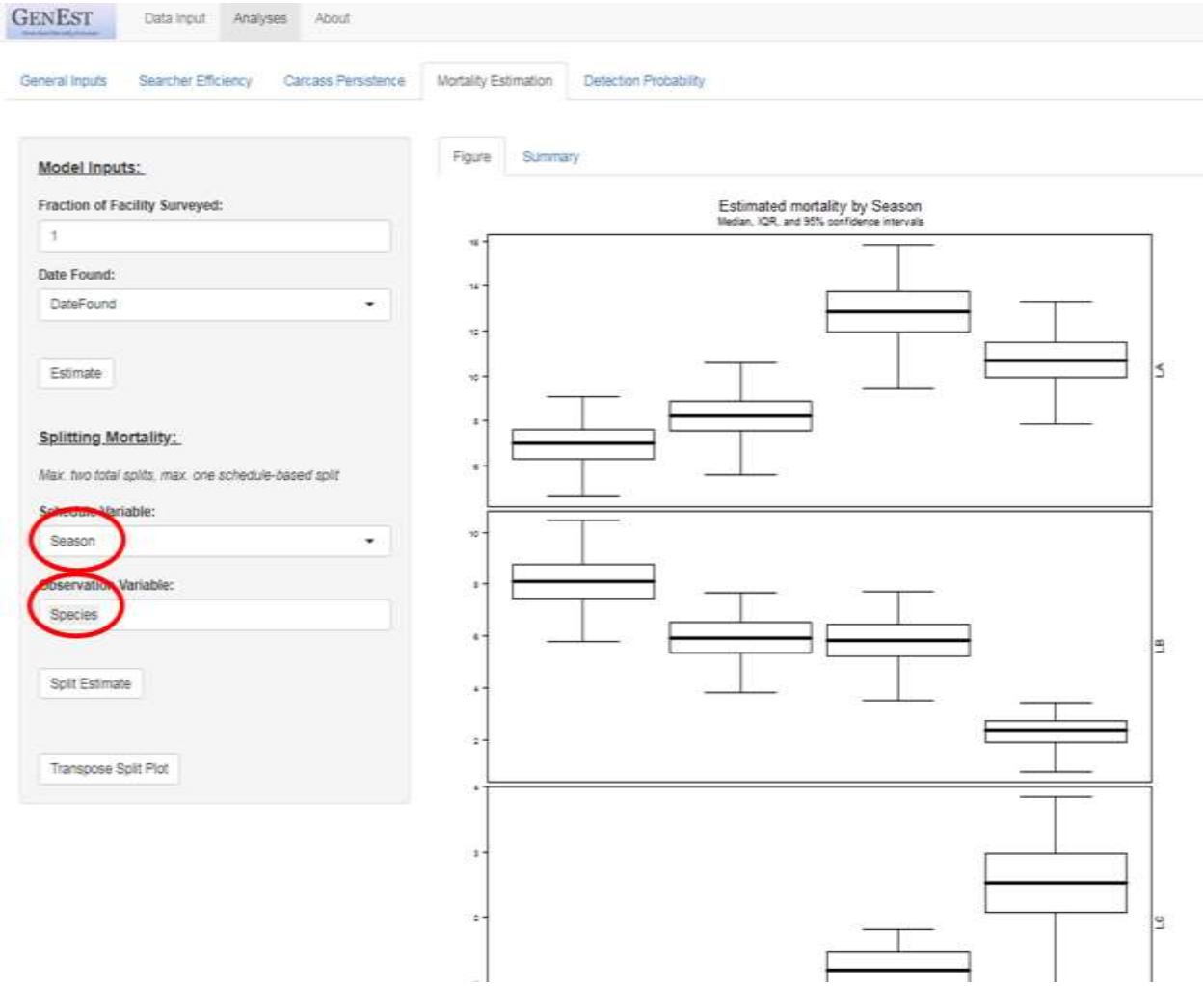

Figure 49. Mortality by species and season for the solar PV example. NOTE: The full figure is quite large and has been truncated to save space in the User Guide. It can be viewed in its entirety in the GUI.

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, navigate to Analyses > Detection Probability (**Figure 50**). 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.

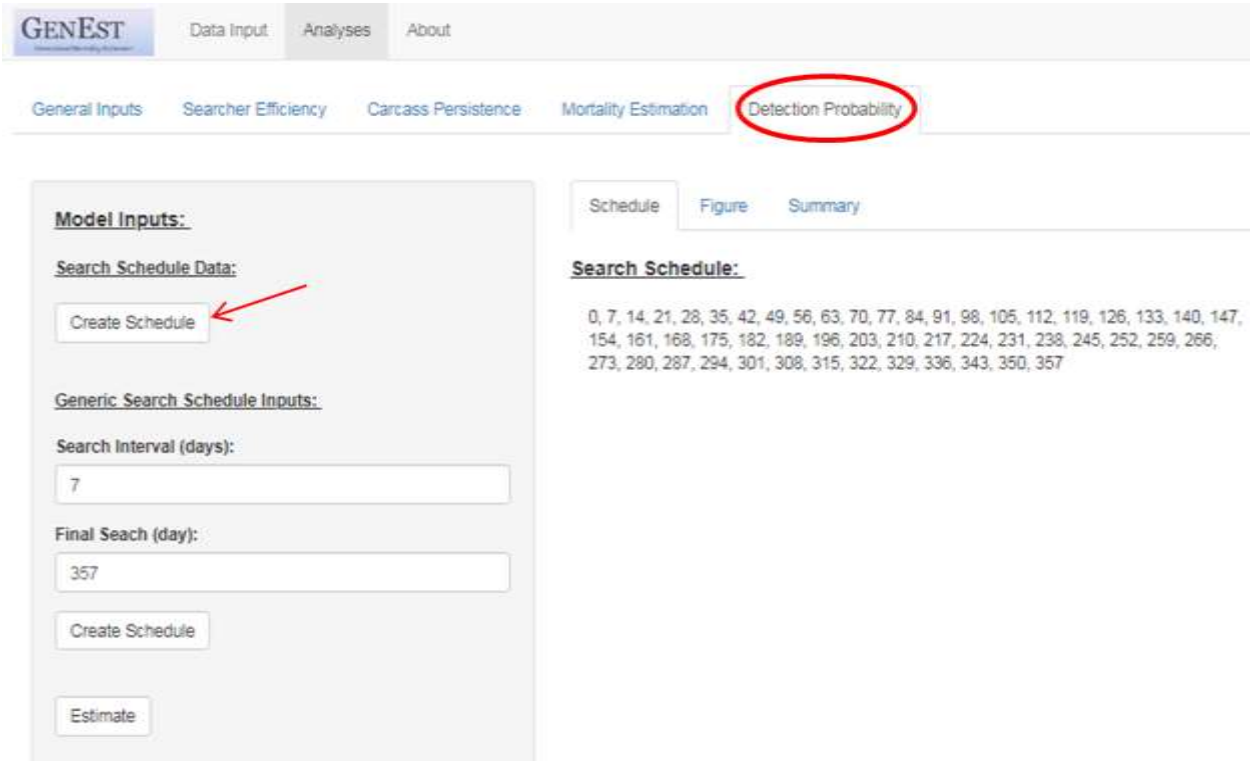

Figure 50. Average search schedule for solar PV example data.

After clicking "Estimate", a table of summary statistics is shown under the "Summary" tab (**Figure 51**). Detection probability varied from around 0.279 in the fall to 0.47 in the summer. Graphical representation of the summary statistics are given under the "Figure" tab.

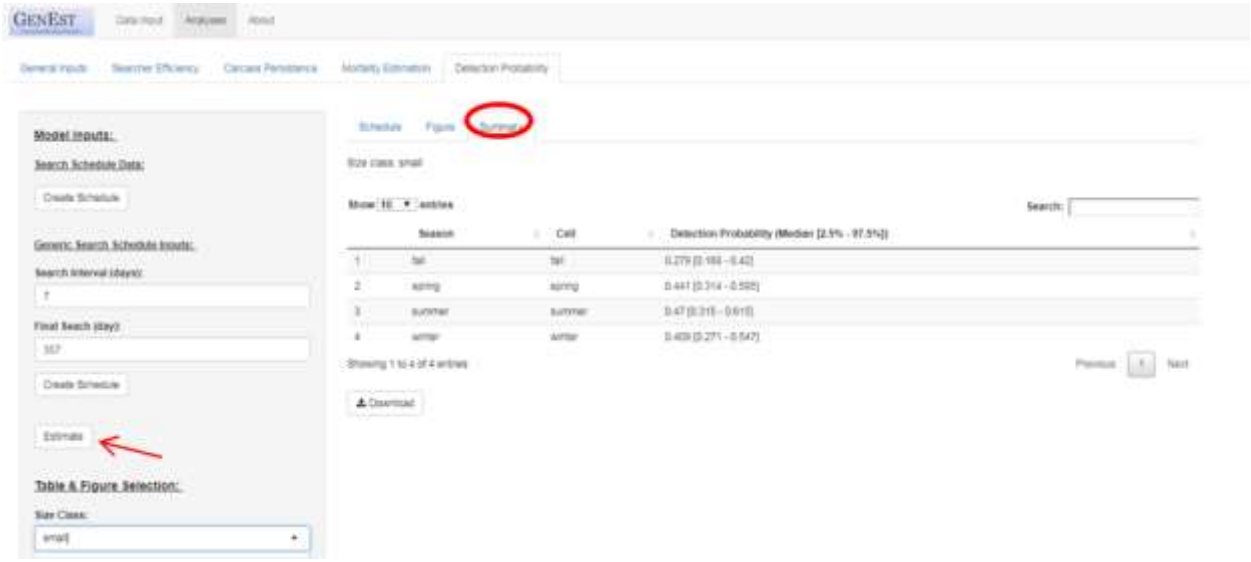

Figure 51. Detection probability by season for small birds (solar PV example).

Acknowledgments

We would like to thank the GenEst steering committee for their tireless support of this project: Cris Hein with Bat Conservation International, Taber Allison with American Wind Wildlife Institute, Jerry Roppe with Avangrid Renewables and Dave Cowan with…. 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 [stats beta reviewers] for rigorous evaluation of our statistical approach and [user beta reviewers] for critical feedback that served to make this software and its user guide more user-friendly. [We are grateful to the many USFWS managers and industry consultants who tested and provided valuable feedback on earlier version of the software.] 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, WEST Inc. LLC and Avangrid Renewables. 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). It is "generalized" statistically and also in its application – it is generalized to apply to monitoring fatalities at solar- and wind-energy facilities, but can apply to many other situations, e.g., fisheries by-catch, power lines, oil spills, etc. The general approach we've taken doesn't change with any of these. GenEst allows the user to evaluate assumptions regarding input parameters and select the approach that best reflects the situation and data.

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

Manuela Huso is a research statistician with the U.S. Geological Survey. She is coauthor of the Evidence of Absence software and the Triggering OFR as well as several journal articles on estimation of mortality at renewable-energy facilities.

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. He is author of the Evidence of Absence software and the Triggering OFR as well as several journal articles on estimation of mortality at renewable-energy facilities.

Jeff Mintz 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

Bengtsson, H. 2018. matrixStats: Functions that Apply to Rows and Columns of Matrices (and to Vectors). R package version 0.53.1. https://CRAN.R-project.org/package=matrixStats

Brown, C, 2017. sticky: Persist Attributes Across Data Operations. R package version 0.5.2. https://CRAN.R-project.org/package=sticky

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. https://CRAN.R-project.org/package=shiny

Chang, W. and Ribeiro, B. B. 2018. shinydashboard: Create Dashboards with 'Shiny'. R package version 0.7.0. https://CRAN.R-project.org/package=shinydashboard

Dalthorp, D. 2018. cbinom: Continuous Analog of a Binomial Distribution. R package version 1.3

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., Wolpert, R., Rabie, P., Stuyvin, J., Simonis, J. and Mintz, J. 2018. GenEst Statistical Models. USGS Open File Report. *In review*.

Eddelbuettel, D. and Francois, R. 2011. Rcpp: Seamless R and C++ Integration. Journal of Statistical Software, 40(8), 1-18. URL http://www.jstatsoft.org/v40/i08/.

Eddelbuettel, D.. Lucas, A., Tuszynski, J., Bengtsson, H., Urbanek, S., Frasca, M., Lewis, B., Stokely, M., Muehleisen, H, Murdoch, D., Hester, J., Wu, W., Kou, Q., Onkelinx, T., Lang, M., Simko, V., and Hornik. K. 2018. digest: Create Compact Hash Digests of R Objects. R package version 0.6.15. https://CRAN.R-project.org/package=digest

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. URL http://CRAN.R-project.org/package=mvtnorm

Grolemund, G. and Wickham, H. 2011. Dates and Times Made Easy with lubridate. Journal of Statistical Software, 40(3), 1-25. URL http://www.jstatsoft.org/v40/i03/

Hankin, R.K.S 2006. Special functions in R: introducing the gsl package. R News 6(4)

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

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., Mizrahi, D. S. 2013. Estimation of bird and bat mortality at wind-power farms with superpopulation models. Journal of Applied Ecology. doi: 10.1111/1365-2664.12100

R Core Team, 2017, R: A language and environment for statistical computing, v3.4.3: R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/.

RStudio and Inc. 2017. htmltools: Tools for HTML. R package version 0.3.6. https://CRAN.R-project.org/package=htmltools

Shoenfeld, P., 2004, Suggestions regarding avian mortality estimation: Mountaineer Wind Energy Center Technical Review Committee; Report for the West Virginia Highlands Conservancy, West Virginia, 6 p.

Therneau T. 2015. A Package for Survival Analysis in S. version 2.38, URL: https://CRAN.R-project.org/package=survival.

Warnes, G.R, Bolker, B., Lumley, T. 2015. gtools: Various R Programming Tools. R package version 3.5.0. https://CRAN.R-project.org/package=gtools

Hadley Wickham, Jim Hester and Winston Chang (2018). devtools: Tools to Make Developing R Packages Easier. R package version 1.13.5. https://CRAN.R-project.org/package=devtools

Wickham, H. 2011. testthat: Get Started with Testing. The R Journal, vol. 3, no. 1, pp. 5-10.

Kathryn A. 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 92(4): 821-828. https://doi.org/10.1890/10-0137.1

Wolpert, R. 2015. ACME: A Partially Periodic Estimator of Avian & Chiropteran Mortality at Wind Turbines. arXiv:1507.00749v1

Xie, Y. 2018. DT: A Wrapper of the JavaScript Library 'DataTables'. R package version 0.4. https://CRAN.R-project.org/package=DT

A Appendix A. Input File Formats

This appendix details the formats of the input files for the GenEst. In general, naming of columns and factor levels should be consistent across files. We will illustrate with an example dataset, available in GenEst\inst\extdata\wind_RPbat.

A.1 Searcher Efficiency

Results of the SE field trials are stored in the `SE_RPbat.csv` file.

```
## pkID Season s1 s2 s3 s4 s5
##  pk1 spring  1 NA NA NA NA
##  pk2 spring  0 NA NA NA NA
##  pk3 spring  1 NA NA NA NA
```

Required columns:

1) Columns representing the ith attempted search for each carcass. Column names should start with `s` or `S` and end with a digit, such as `s1` or `Search1`. Values must be 1 when a carcass was detected, and NA once a carcass is no longer present. Entries such as 0 for not found, or other character strings such as `NS` indicating Not Searched are also acceptable.

In this example, searcher efficiency trials were conducted on roads and pads, with a total of 60 fresh carcasses placed in the field over the course of the entire monitoring period, evenly divided among seasons (spring, summer, fall). Carcasses that were later discovered by search teams during the course of normal carcass surveys were removed from the field. Carcasses were left in the field for up to 5 searches after carcass placement.

Columns `s1, s2,` … show the fate of carcass `pkID` on the 1st, 2nd, … searches after the carcass was placed. A 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 (whether it was discovered in a previous search or removed by scavengers).

A.2 Carcass Persistence

Results of the CP field trials are stored in the `CP_RPbat.csv` file.

```
## cpID Season  Left Right
##  cp1 spring 10.03 14.09
##  cp2 spring  3.98  7.05
##  cp3 spring 20.95 28.05
```

Required columns:

1) A column containing Left: a numeric value from 0 to the value of `Right`.
2) A column containing Right: a numeric value from the value of `Left` to `Inf`.

In our example, carcass persistence trials were conducted on roads and pads. A total of 60 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 `Left` and the right endpoint as `Right`. For carcasses not scavenged by the end of the study, `Left` is the time of the last check and `Right` is `Inf`. For carcasses whose scavenging time is known exactly (e.g., the scavenging was recorded by camera), `Left = Right`.

A `Season` column was used to note the season at the time the carcass was placed in the field.

A.3 Search Schedule

The example search schedule is stored in the `SS_RPbat.csv` file. The output here has been truncated at 10 columns, whereas there are 100 turbine columns total.

```
## 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. Format should be "%Y-%m-%d CDT", or "%Y/%m/%d CDT" (time zone optional).
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 (e.g., season).

A.4 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 search unit names column: This column must have the same name as the column holding search units in DWP. Additionally, levels of search units used in this column need to be valid column names within the search schedule.
2) A column for date found: Column for date on which a carcass was observed. Format should be "%Y-%m-%d CDT", or "%Y/%m/%d CDT" (time zone optional).

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.