# GenEst (v1.0.0) Software User Guide

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

Open File Report 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](http://www.usgs.gov/pubprod).

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

ISSN 2327-638X (online)

Contents

GenEst (v1.0.0) Software User Guide 1

Abstract 6

1.0 Introduction 6

1.1 Description of Product 6

1.1.1 Software Modules 7

1.1.2 Comparison to Other Methods 7

1.2 Installation Instructions 7

1.3 Getting Started 8

1.4 Layout 8

2.0 Data Input 9

2.1 Data Files 9

3.0 Analyses 11

3.1 General Inputs 11

3.1 Searcher Efficiency 12

3.1.1 Model Inputs 12

3.1.2 Outputs 13

3.1.2.1 Data 13

3.1.2.1 Model Table 13

3.1.2.1 Figure 13

3.1.2.1 Model Selection 13

3.2 Carcass Persistence 14

3.3 Mortality Esitmation 14

3.4 Detection Probability 14

5.1 Subsection 14

5.1.1 Example subsubsection 14

4.0 Worked Examples 15

4.1 Example 1 15

4.1.1 Example subsubsection 15

4.2 Example 2 15

4.2.1 Example subsubsection 15

5.0 Conclusions 15

Option: An unnumbered but linked header 16

Acknowledgments 17

About the Authors and Collaborators 17

References Cited 17

Appendix A. Input File Formats 19

A.1 Search Efficiency 19

A.2 Carcass Persistence 19

A.3 Search Schedule 19

A.4 Carcass Observations 19

Appendix B. Analytical Details 19

B.1 Search Efficiency 19

B.2 Carcass Persistence 19

B.3 Search Schedule 19

B.4 Carcass Observations 19

Figures

**Figure 1.** Welcome screen disclaimer for GenEst, as opened in Mozilla Firefox (v58.0.2). 5

**Figure 2.** Browsing to the example data sets within the GenEst R package folder on a Windows (10.0.16299) machine. 6

**Figure 3.** Visible uploaded data in the GenEst GUI. 6

**Figure 4.** Required general inputs. 7

**Figure 2.** Screen captures of the EoA control panel and the minimized R console. 11

Tables

**Table 1.** Files needed for each analysis 10

**Table 1.** Example format of carcass persistence data. 16

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

|  |  |
| --- | --- |
|  | significance level (distinguished from shape parameter by context) |
|  | confidence level |
|  | shape and scale parameters for persistence distributions |
|  | spatial coverage, fraction of carcasses falling into searched area; see *dwp* |
| *A* | matrix of simulated arrival intervals, [*X* D *nsim*] |
| *dwp* | density weighted proportion; fraction of the total carcasses that arrive in a searched area |
|  | sampling fraction; proportion of potential sampling units that were searched |
|  | overall detection probability within the searched area, within the period of study |
|  | search interval; number of days between searches of the same area/unit |
|  | assumed search interval for reporting an estimate of |
|  | factor by which searcher efficiency decreases with each successive search |
|  | mortality rate over a set time period |
|  | mortality or number of fatalities |
|  | estimated mortality or number of fatalities |
|  | matrix of simulated mortality given arrival in *A*, [*X* D *nsim*] |
| *nsim* | number of simulation runs |
|  | searcher efficiency, probability of observing a carcass that is present in the searched area at the time of the search |
|  | probability that a carcass persists until the first search after arrival |
| span | the length of the monitoring period |
|  | temporal coverage, fraction of total mortality expected to occur during the monitored period |
|  | number of carcasses observed in searches |

GenEst (v1.0.0) Software User Guide

by Juniper Simonis[[1]](#footnote-2), Daniel Dalthorp[[2]](#footnote-3), Manuela Huso2, Lisa Madsen[[3]](#footnote-4), Paul Rabie[[4]](#footnote-5), Jared Studyvin4 and Jeffrey Mintz2

## 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. In particular, its development was motivated by the need to accurately estimate the total number of bird and bat fatalities occurring at renewable-energy facilities. Simple counts of carcasses do not accurately reflect fatality and do not allow comparison among locations 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.

# Introduction

## 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 superpopulation sensu {Williams, 2011 #446} 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 trial experiments 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 (<package submission information>) 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.

### 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.

### Comparison to Other Methods

There are currently several estimators in use that were created specifically to estimate mortality at wind power facilities {Dalthorp, 2017 #958;Huso, 2012 #412;Korner-Nievergelt, 2012 #571;Shoenfeld, 2004 #194;Wolpert, 2015 #809}. Readers familiar with them might ask how they differ from GenEst (Table 1). 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 nominal coverage because fewer constraining assumptions are made. It is closest in conceptual development to Wolpert’s but extends the models to account for uncertainty in detection probabilities, arrival times and potential mortality that could produce the observed counts. GenEst makes few assumptions regarding persistence distribution, or decline in searcher efficiency with repeated searches for the same carcass. Rather, it includes modules to allow the user to compare potential models of searcher efficiency and carcass persistence using AICc and graphs. Only EoA has a model selection method built into the software, but can only compare persistence distributions. It cannot assess influence of potential covariates. None of the currently available estimators provides a module for estimating the fraction of carcasses expected to land in the searched area,

Table 1. comparison table [insert later].

## Installation Instructions

GenEst is available as an R package. Thus, you will need to download and install R (https://cran.r-project.org/) before using GenEst.

GenEst is available for download from the USGS GitHub repository. Within R, ensure that you have the devtools package installed and loaded into R by running

install.packages("devtools")

library(devtools)

Then download and install the GenEst package

devtools::install\_github("usgs/GenEst")

And load it into the R instance

library(GenEst)

GenEst will eventually be available from the R package repository (<http://cran.r-project.org>), at which point, the stable version of the software can be downloaded and installed locally by running

install.packages("GenEst")

And then loaded into the R instance via

library(GenEst)

Installing the GenEst package will also lead to all dependency R packages being installed and loaded into the R instance. In addition, GenEst comes with example data sets, vignettes, and help files, all located within the package folder and accessible via both the command-line and GUI routes.

## Getting Started

Initializing the GenEst GUI is as simple as running

runGenEst()

which will launch the application in your default web browser (fig. 1). 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.

## Layout

The GUI is laid out using tabs, as shown in the top left corner of fig. 1: Home, Data Input, Analyses, and About. Once the user acknowledges the disclaimer, they are able to navigate between the tabs. The Home tab is a welcome page and the About tab includes additional information about the GUI and package. The Data Input and Analyses tabs allow active user interaction for analyses.

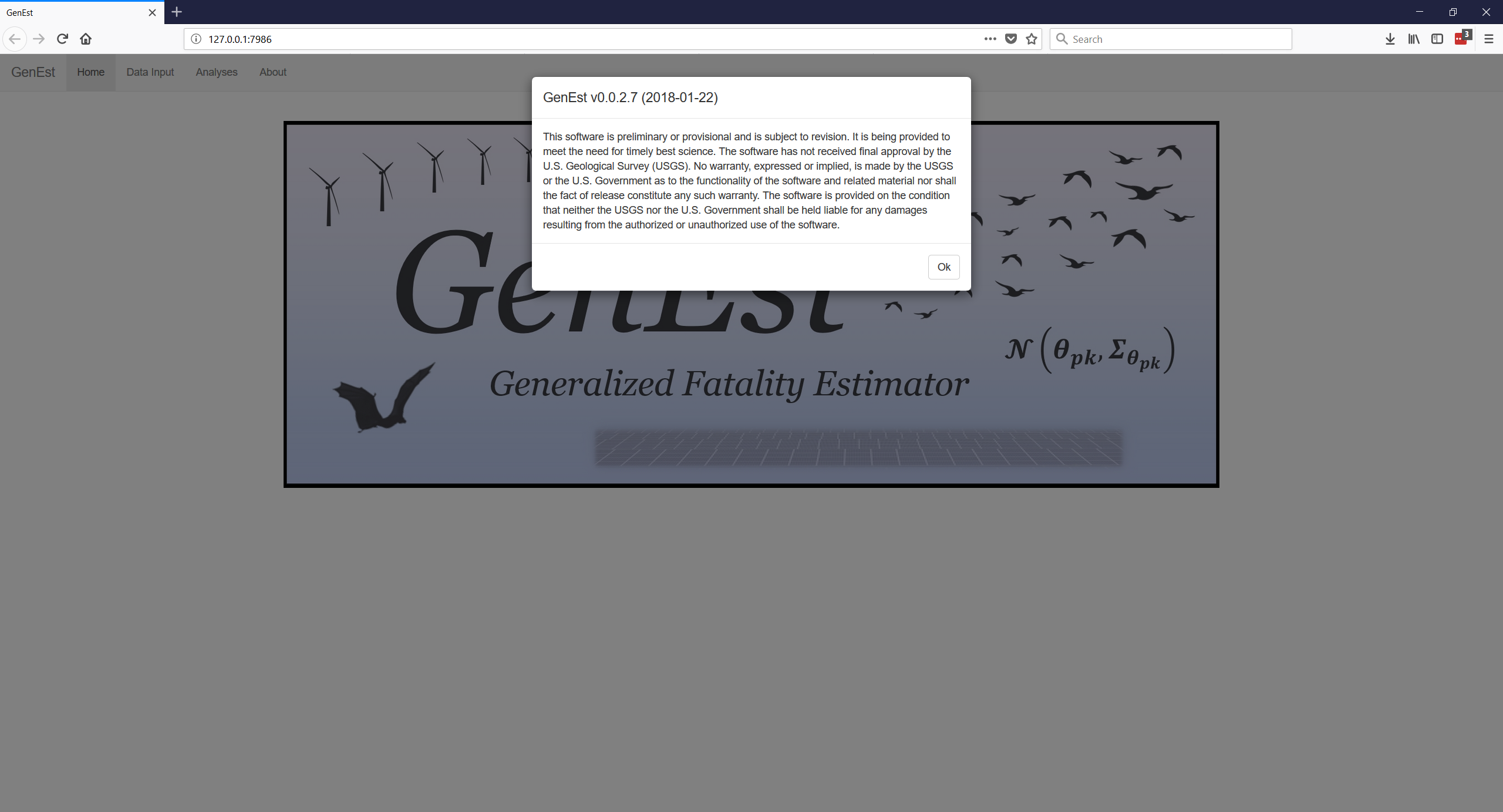


Figure 1. Welcome screen disclaimer for GenEst, as opened in Mozilla Firefox (v58.0.2).

# Data Input

## Data Files

Having acknowledged the disclaimer, the GUI is now able to accept data for use (fig. 2). 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.

Currently, the GUI can take one of each of five data files: 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 (fig. 3) as well as use throughout the GUI.

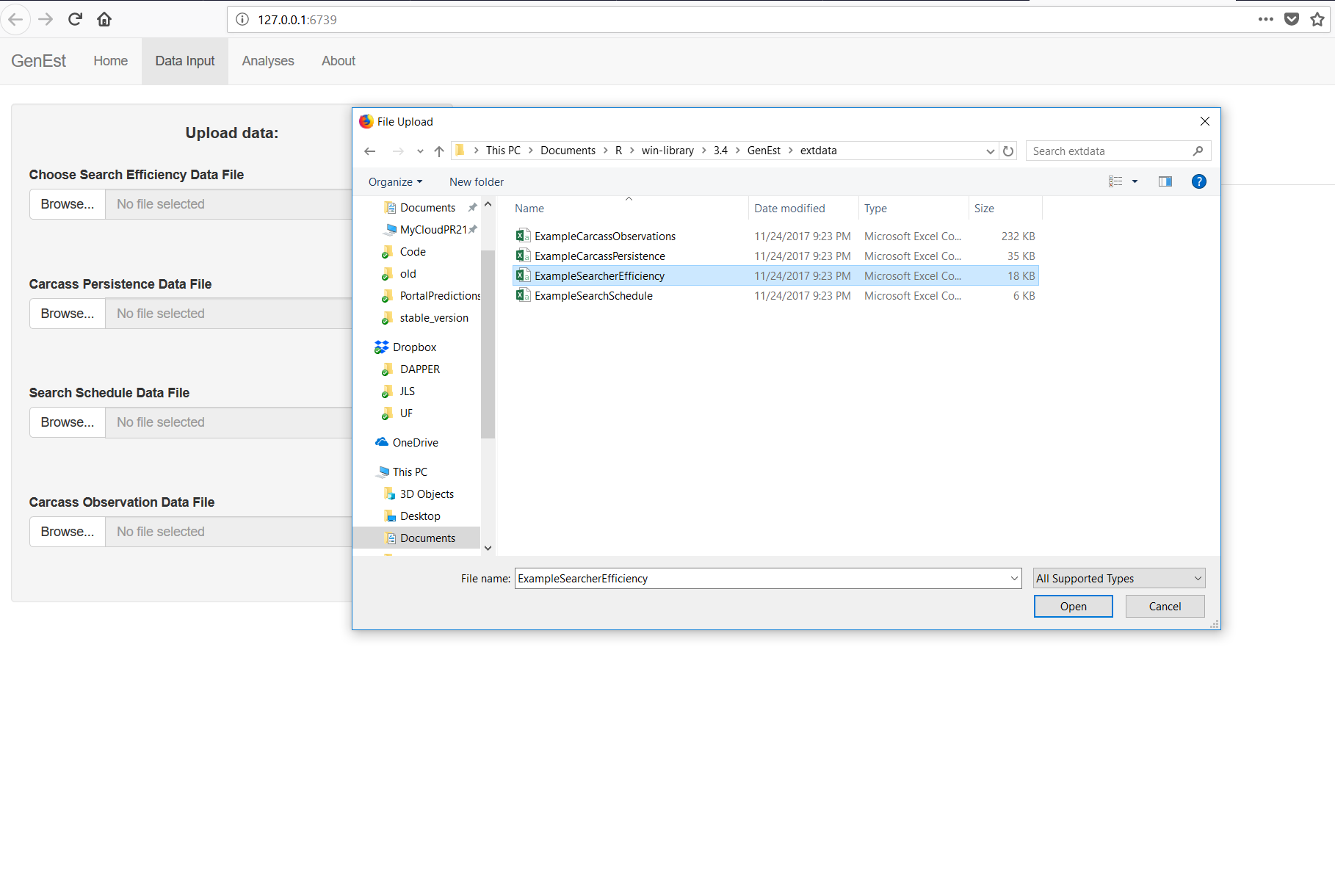


Figure 2. Browsing to the example data sets within the GenEst R package folder on a Windows (10.0.16299) machine.

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 |

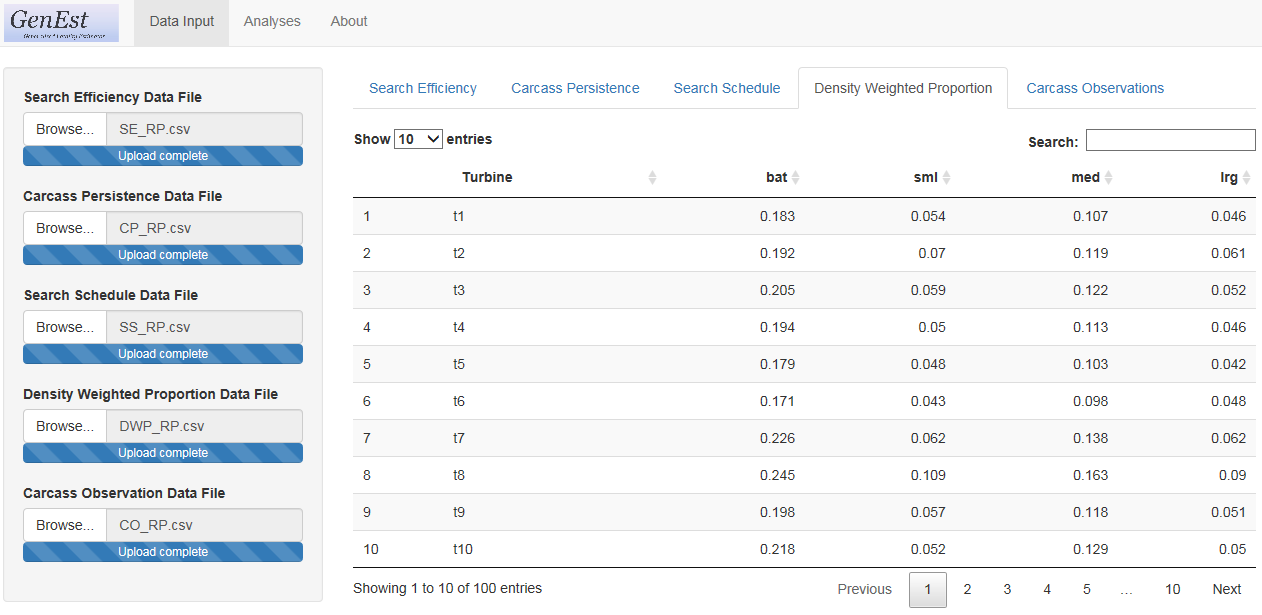


Figure 3. Visible uploaded data in the GenEst GUI.

# Analyses

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

## General Inputs

Under the “General Inputs” subtab, the user provides two values: “Number of iterations” and “Confidence Level” (fig. 4) and is asked to supply a third (optional) value. The “Number of iterations” controls the number of replicates used to estimate parameters, defaults to 1,000, and must be an integer value 1 or larger, with current support up to 10,000. 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.

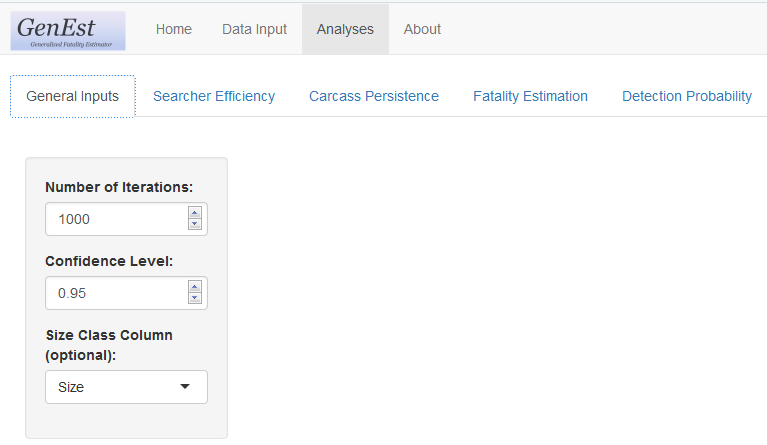


Figure 4. Required and optional general inputs.

## 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: , the probability that a carcass that is present is found during the first search after it arrived, and , the proportional reduction in the probability that a present carcass is found on each subsequent search. A value of implies that carcasses are discoverable only on the first search, and 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 discrete-valued predictors (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 (<insert\_final\_function\_names>) for the GenEst package. For full details, see appendix B.1. Users have the option to allow GenEst to estimate from the data or to fix *k* at any acceptable level (0-1). If data from only the first attempt to find a trial carcass were recorded, cannot be estimated and must be set by the user.

### Model Inputs

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

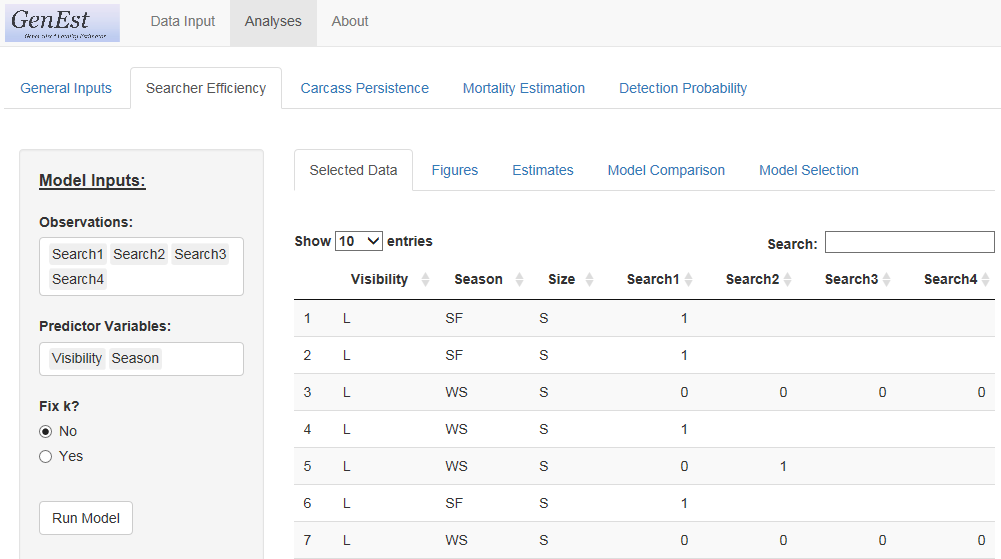


Figure 5. Searcher efficiency model inputs.

* + “Observations:”
    - The user is required to select the columns in the input data set that correspond to the search efficiency trial results for one or more searches after carcasses were placed (found = 1, missed = 0, NA = not searched of carcass no longer available).
    - The columns must be selected in order of search occasion.
    - At least one column must be selected, and currently there is no maximum number of searches allowed.
  + “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 trials
    - 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 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 between 0 and 1 (inclusive) if the searcher efficiency model is to be used for mortality estimation. If the data do allow estimation of (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.

### 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 and with the lowest AICc value (fig. 6). Several output sub-tabs for evaluating model fit are available: Selected Data, Figures, Estimates, Model Comparison, and Model Selection (fig. 6).

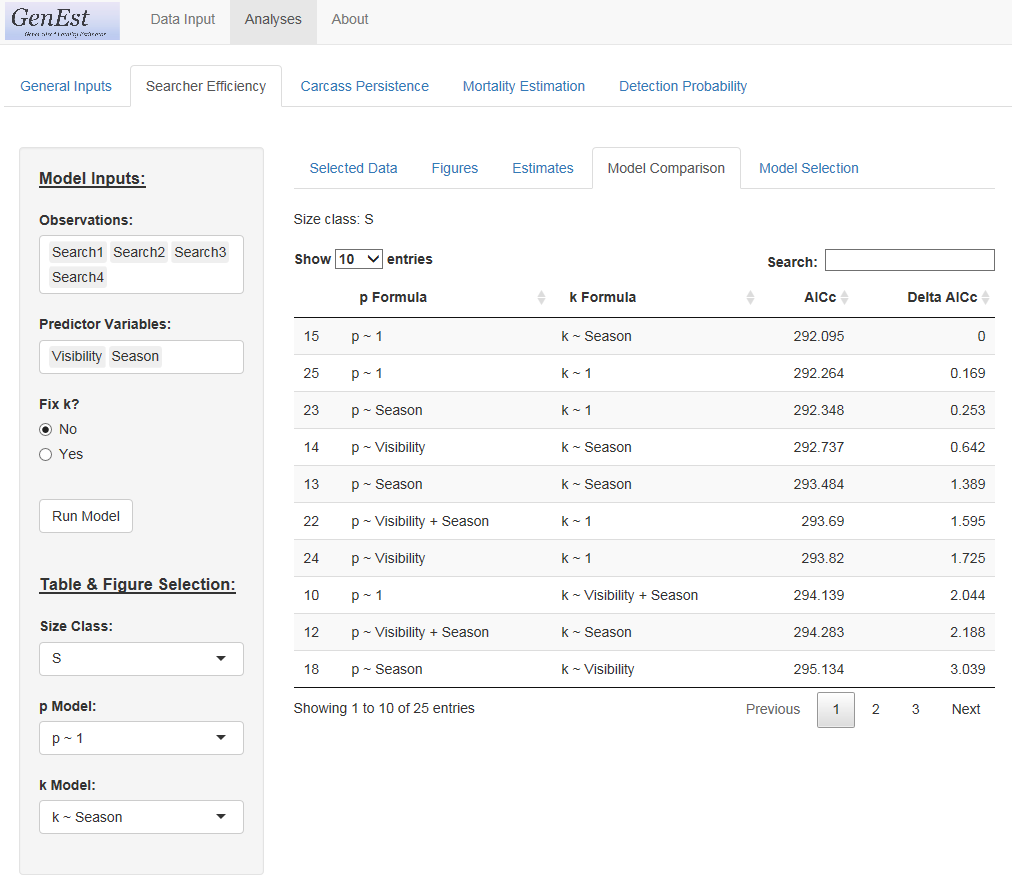


Figure 6. Options to explore results of searcher efficiency model fitting.

### 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.

### Figures

Searcher efficiency estimates from the selected model are plotted against search occasion for each possible combination of class levels of the input predictor variables (fig 7).

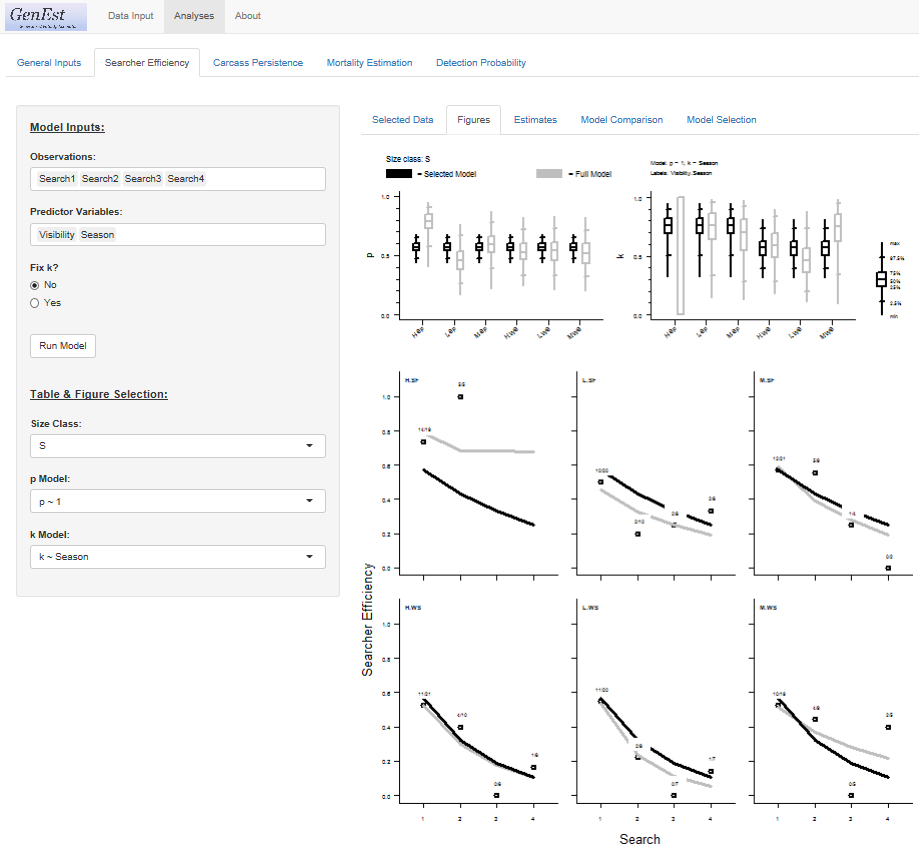


Figure 7. Example plots of estimated and observed searcher efficiency by search event. Dark lines represent selected model, gray lines represent the most complex model, i.e., separate estimates for each class combination (when possible). Open dots represent observed searcher efficiency and ratios above each open dot represent observed data used to estimate the point, i.e., found/available. Box plots at the top represent estimated and for selected model (dark lines) compared to most complex model (grey lines).

### Estimates

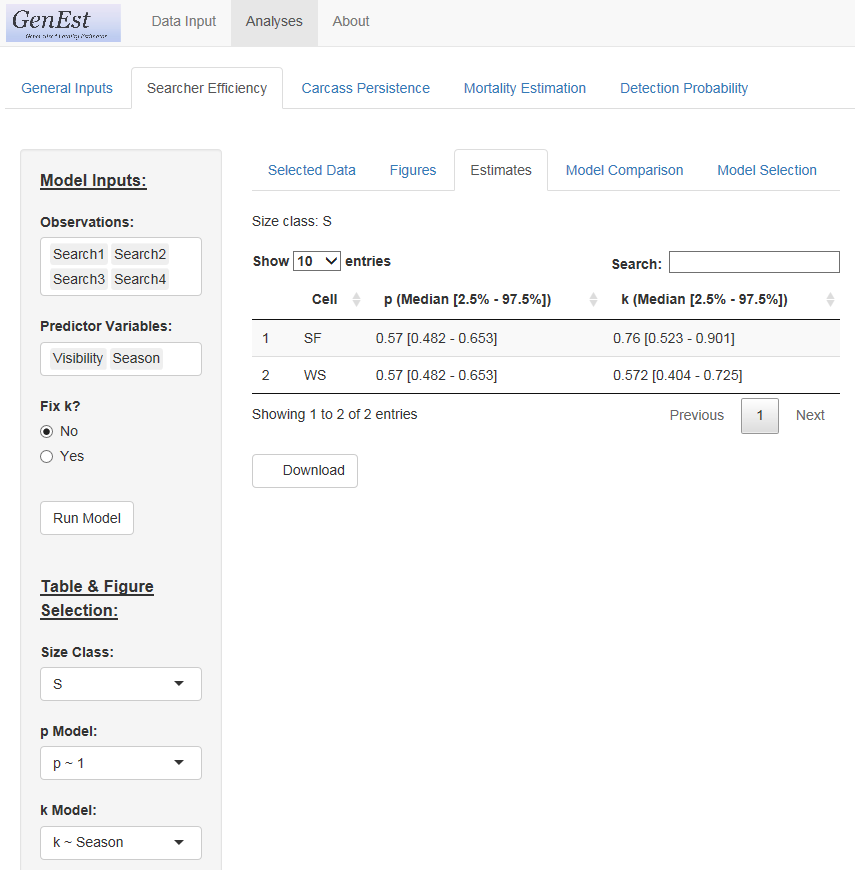
Estimates and confidence limits (α set by user in “General Inputs”) of and for each class combination of covariates for selected model are given in table format (fig 8). The table can be saved to a csv by clicking the “Download” button.

Figure 8. Estimates and confidence limits (α set by user in “General Inputs”) of and for each class combination of covariates for selected model.

### Model Comparison

The model comparison tab provides user with AICc and Delta AICc for all models fit using the covariates selected under “Model Inputs” (fig 6).

### 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 (fig 9). There is no need to “Run Model” again. The selections will be held in memory for later use in mortality estimation.

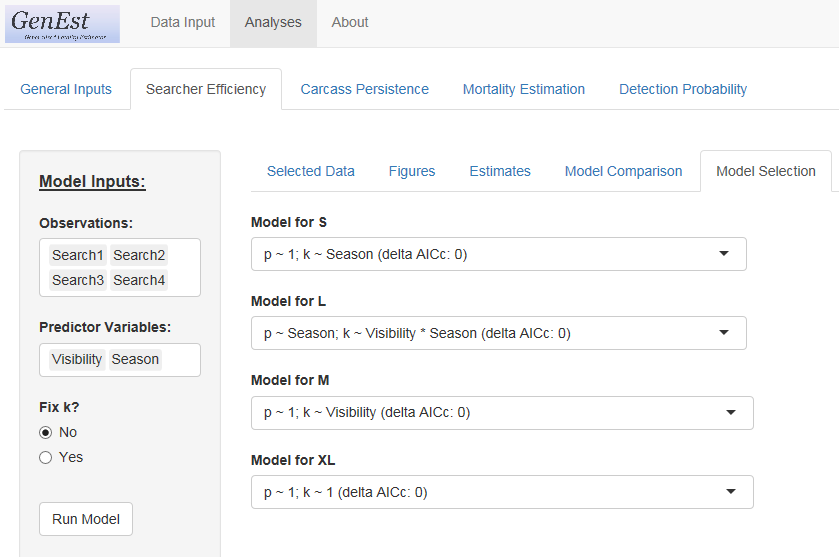


Figure 9. Model selection screen indicating which searcher efficiency model to use for each size class of trial carcasses.

## 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.

### Model Inputs

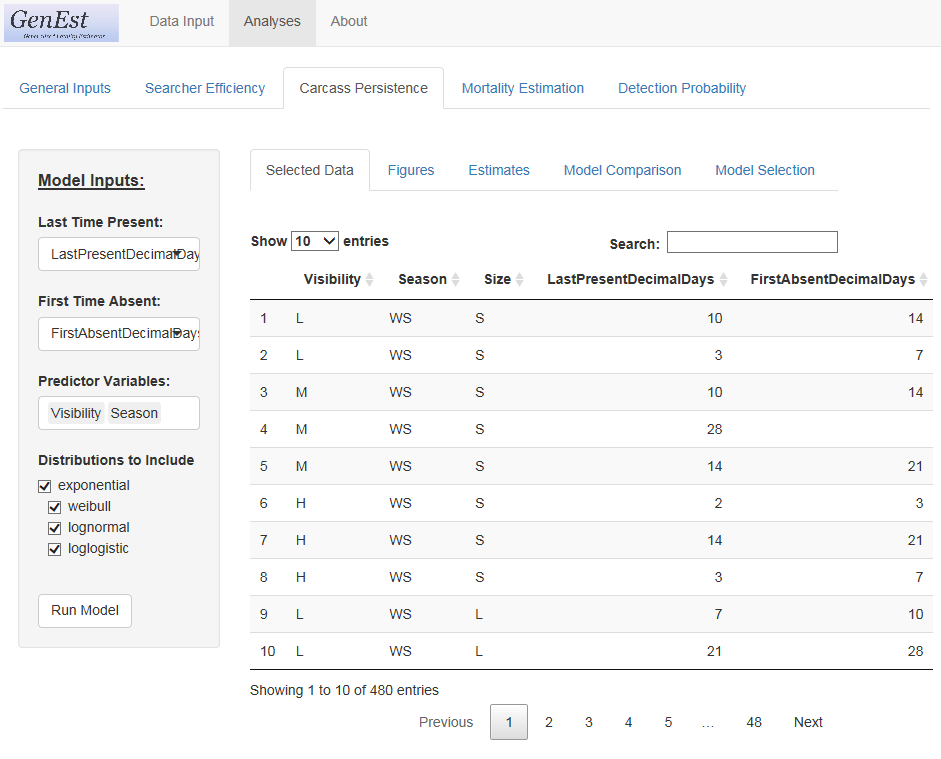
Carcass persistence model inputs are selected on the left-hand side panel of the Search Efficiency tab, Selected Data sub-tab (fig. 10). 

Figure 10. 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.

### 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 (fig. 11).

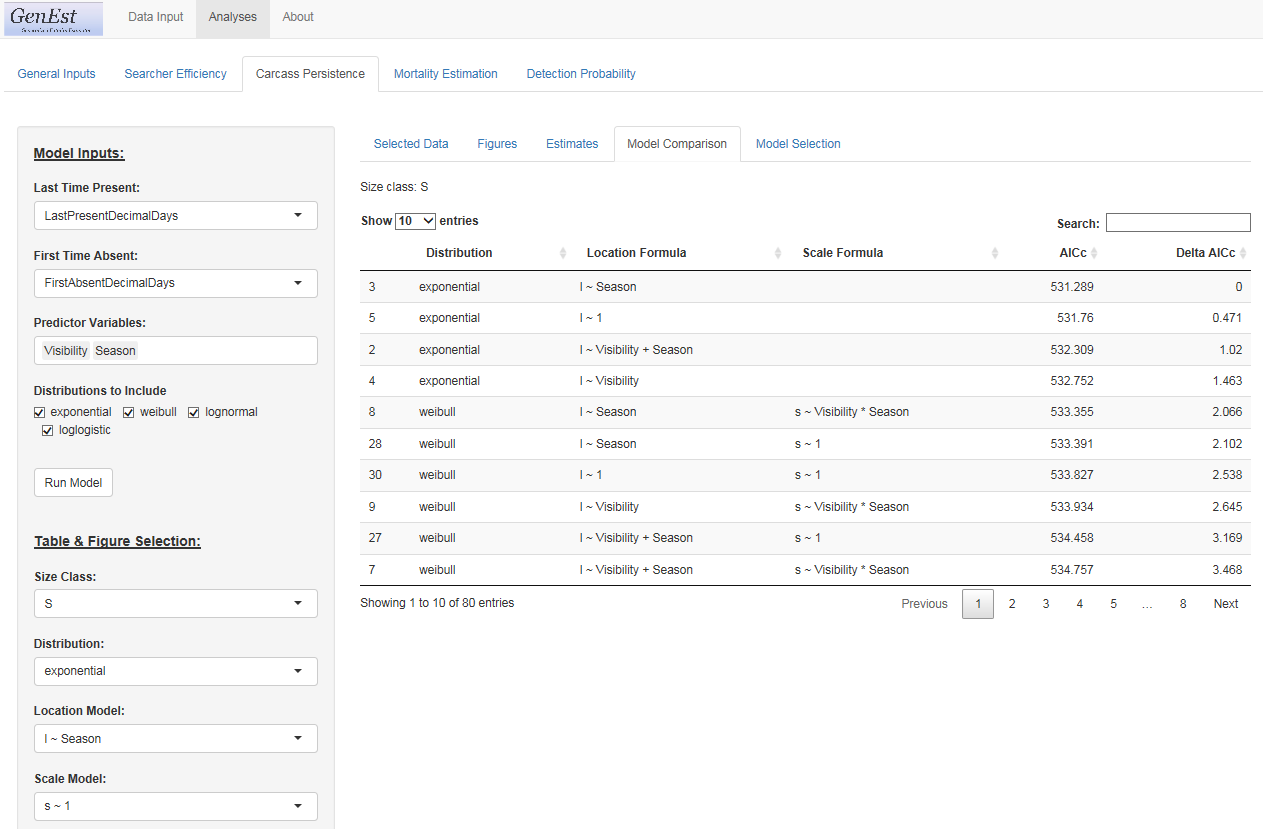


Figure 11. Options to explore results of carcass persistence model fitting.

### 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.

### Figures

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

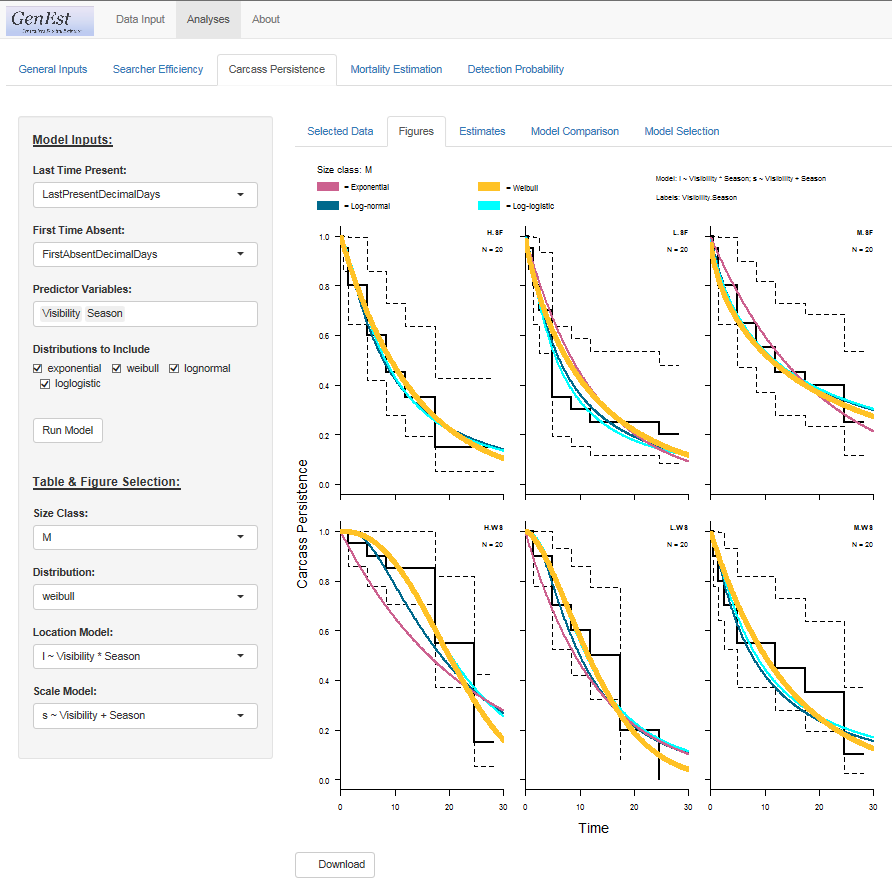


Figure 12. Example plots of estimated and observed carcass persistence by time for each combination of levels of predictor variables. Stair-step solid and dotted lines are Kaplan-Meier plots representing observed data and empirical confidence limits. All fitted persistence distributions are represented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

### 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 (fig 13). The table can be saved to a csv by clicking the “Download” button. Note: scale parameter in exponential distribution = 1, by definition.

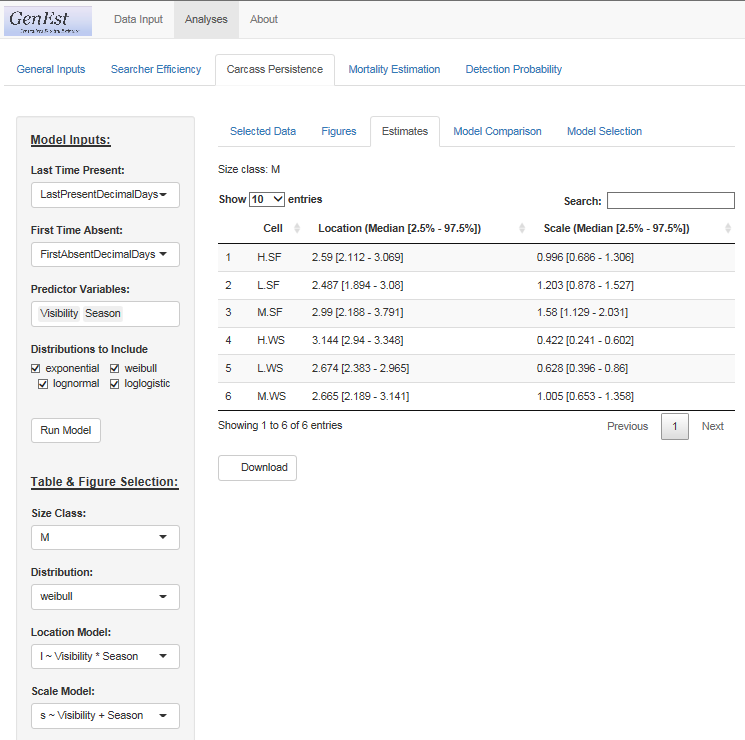


Figure 13. Estimates and confidence limits (α set by user in “General Inputs”) of location and scale parameters for each class combination of covariates for selected model.

### Model Comparison

The model comparison tab provides user with AICc and Delta AICc for all models fit using the covariates selected under “Model Inputs” (fig 11).

### 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 (fig 14). There is no need to “Run Model” again. The selections will be held in memory for later use in mortality estimation.

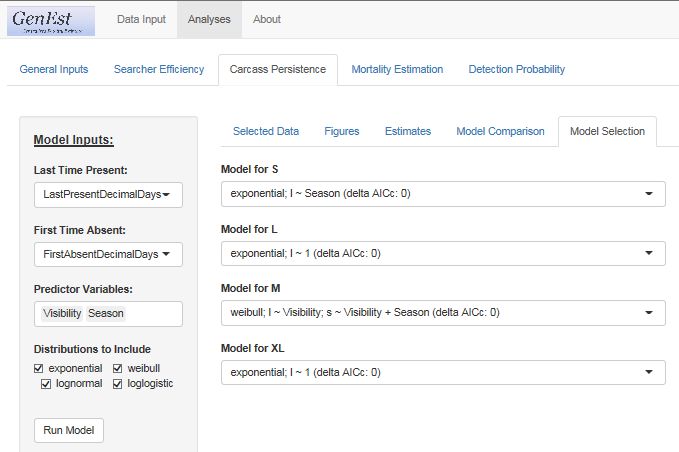


Figure 14. Model selection screen indicating which carcass persistence model to use for each size class of trial carcasses.

## 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” (fig 15**)** or “Detection Probability” (following section).

### Model Inputs

The user is asked to provide the “Fraction of Facility Surveyed” (fig 15). 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.

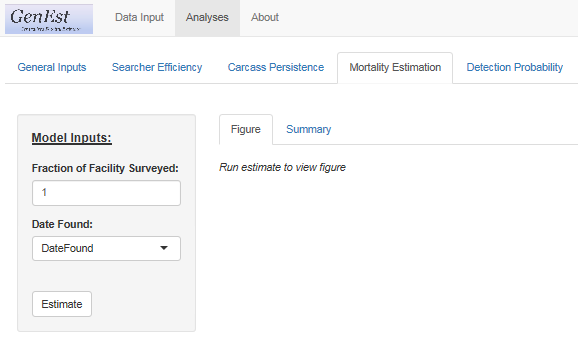


Figure 15. Model inputs screen requesting information on sampling fraction and identifying the column in the carcass observations file that represents the date on which the carcass was found by searchers.

Once the requested user input is provided, clicking “Estimate” will generate mortality estimates and confidence intervals. The program automatically generates a figure representing total mortality of all species at the site, a number that may or may not be of interest to the user (fig 16).

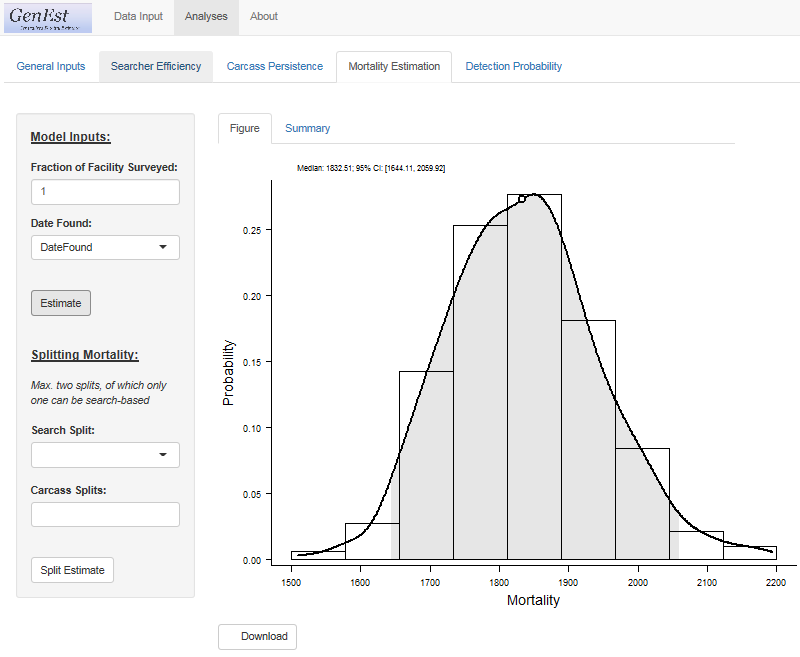


Figure 16. Estimated total mortality at the site based on user input.

### Splitting Mortality

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 search 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 (fig 17).

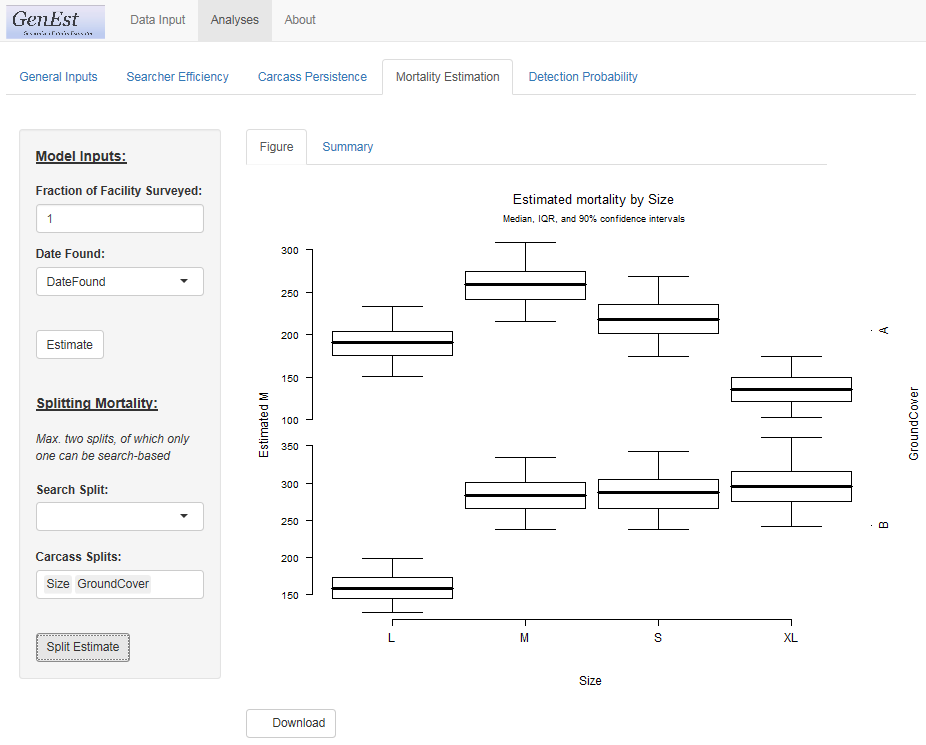


Figure 17. Estimated mortality at the site in each combination of size and ground cover.

Selecting “Summary” will open a table of estimates and selected quantiles for each split combination (fig 18). The table can be downloaded and saved as a .csv for use in reports and publications.

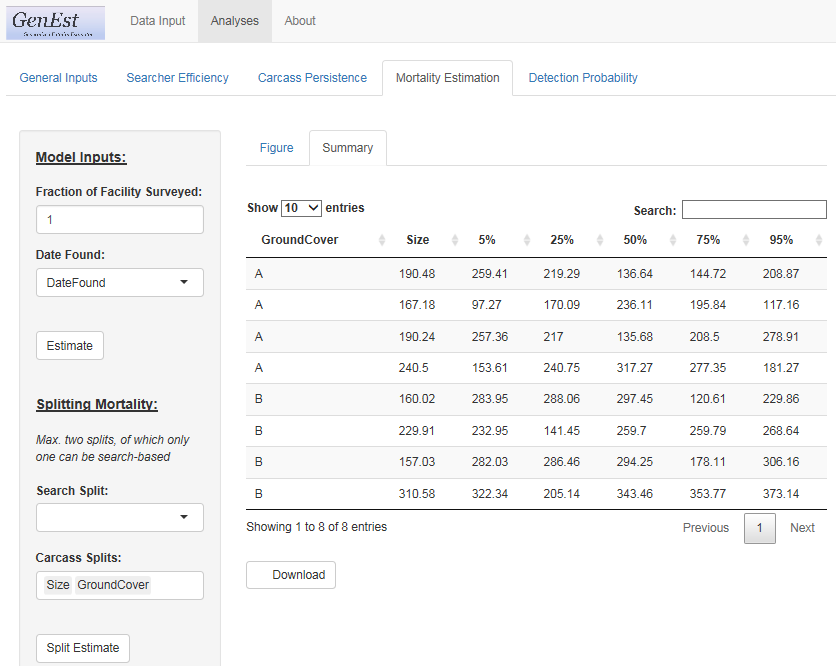


Figure 18. Estimated mortality at the site in each combination of size and ground cover. Table of quantiles used to produce associated graph.

## 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.”

### Model Input

### 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. (fig 19).

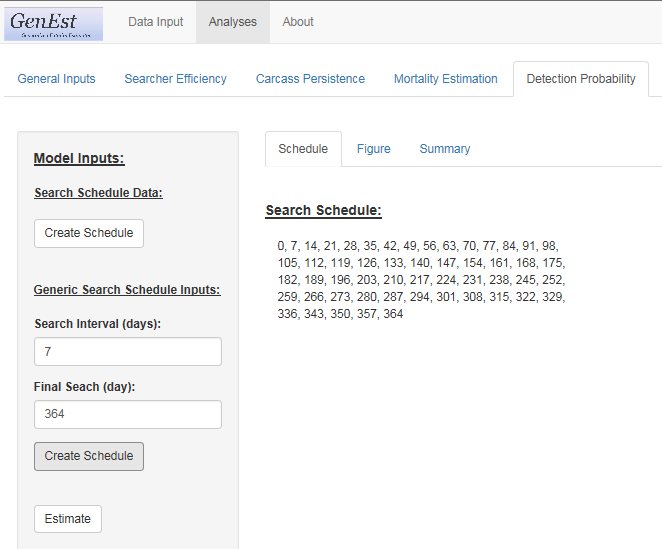


Figure 19. Model inputs screen requesting information on search schedule. Alternatively, user can input regular search interval over a set period.

### 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”

### 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 (fig 20).

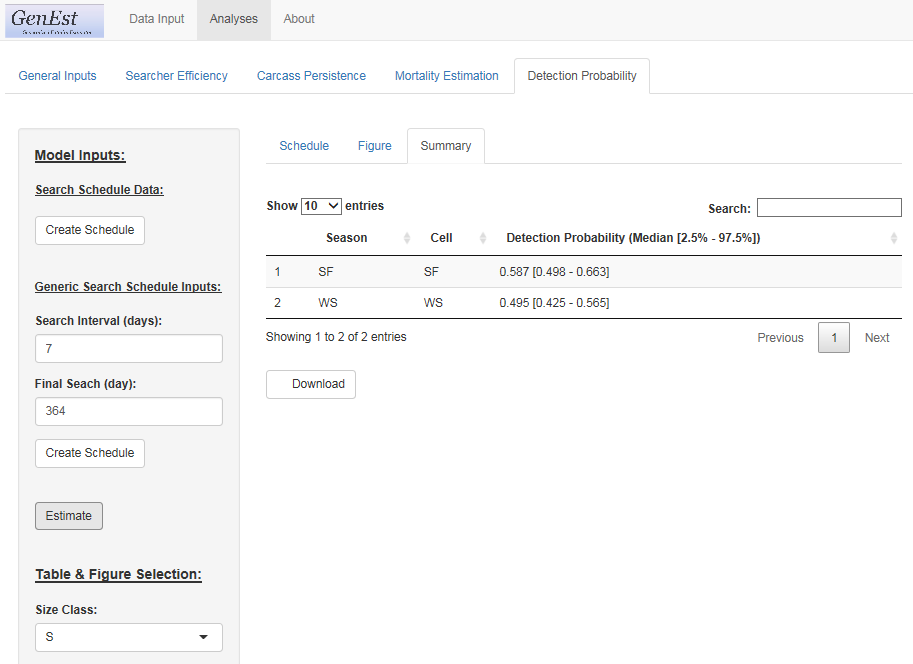


Figure 20. Summary of detection probability and confidence limits for relevant covariate classes (covariates that were included in the selected carcass persistence or searcher efficiency models) for the selected size class.

### Figure

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

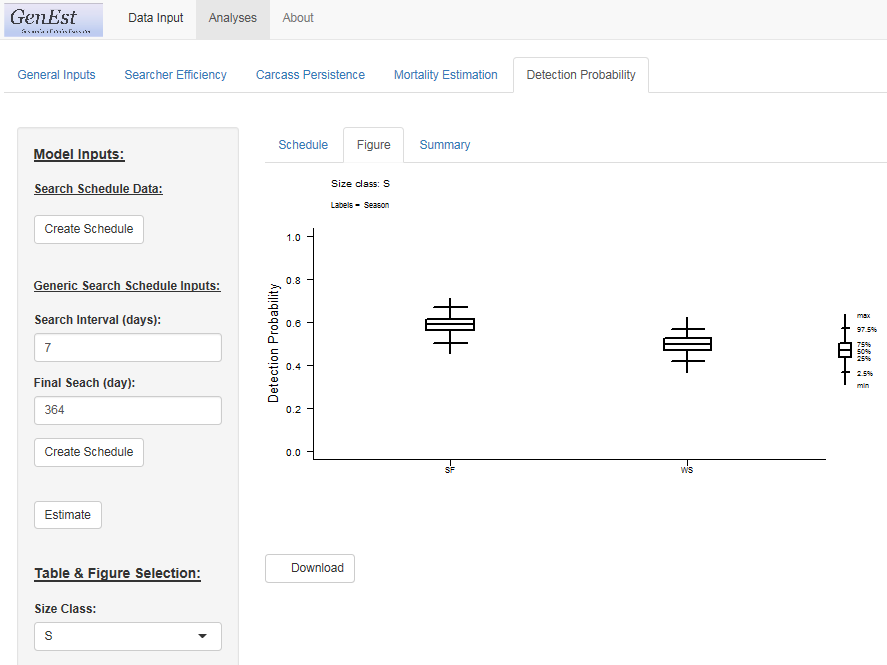


Figure 21. Box plot of estimated detection probability for relevant covariate classes (covariates that were included in the selected carcass persistence or searcher efficiency models) for the selected size class. Graphic representation of the data in the “Summary” table.

# Worked Examples

## Example 1 – Wind facility, all animals, sampled turbines

### Text.

## Example 2 – Wind facility, bats, road and pad surveys of all turbines

### Text.

## Example 3 – Solar photovoltaic facility, all birds

### Text.

## Example 4 – Solar trough facility, all birds

### Text.

## Example 5 – Solar power tower facility, all animals

### Text.

# Conclusions

* + 1. Text.

# 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 [cite estimators]. Recognizing the commonalities and differences among our approaches [cite Wolpert], 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, Lisa Madsen, Paul Rabie, Jared Studyvin and Jeffrey Mintz

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.

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.

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.

Robert Wolpert is…

Fränzi Korner-Nievergelt is an ecological statistician at the Swiss Ornithological Institute and at oikostat GmbH. She is one of the developers of the ProBat tool (www.windbat.techfak.fau.de/tools) and coauthor of several journal articles on the modelling of bird and bat mortality at wind energy facilities.

# References Cited

Chang, W., Cheng, J., Allaire, J.J., Xie, Y., and McPherson, J., 2017, shiny: web application framework for R, version 1.0.5, <https://CRAN.R-project.org/package=shiny>.

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/](%20https:/www.R-project.org/).

# 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.

## Search 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).

## 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.

## 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).

## 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.

# Appendix B. Analytical Details

This appendix provides additional information regarding the analyses underlying the GenEst software.

## Search Efficiency

* + 1. With text!

## Carcass Persistence

* + 1. With text!

## Search Schedule

* + 1. With text!

## Carcass Observations

* + 1. With text!

1. DAPPER Stats. [↑](#footnote-ref-2)
2. U.S. Geological Survey. [↑](#footnote-ref-3)
3. Oregon State University. [↑](#footnote-ref-4)
4. Western EcoSystems Technology, Inc.. [↑](#footnote-ref-5)