

SOP 11: Detection Function Estimation (KM_J Creation)

Version 1.0

Revision History Log:

Previous Version #	Revision Date	Author	Changes Made	Reason for Change	New Version #
	November 2012	S. Hoekman, W. Johnson, B. Moynahan, C. Sergeant	Initial version		1.0

Contents

	Page
Figures.....	SOP 11-2
Tables.....	SOP 11-3
Summary	SOP 11-4
Data Import to DISTANCE	SOP 11-5
Exploratory Analyses.....	SOP 11-9
Selecting a Detection Function	SOP 11-13
Estimating Group Size Using DISTANCE.....	SOP 11-18
Assembling the KM_J Deliverable	SOP 11-23
Formal Validation and Certification	SOP 11-24
Literature Cited	SOP 11-28
SOP 11 Figures and Tables.....	SOP 11-29
Appendix SOP 11.A: R code for plotting detection functions.....	SOP 11-53

Figures

Page

Figure SOP 11.1. Creation of a template project file in Program DISTANCE... SOP 11-29

Figure SOP 11.2. Selection of appropriate settings using the 'New Project Setup Wizard' during creation of a DISTANCE project template file. SOP 11-30

Figure SOP 11.3. The 'Import Data Wizard' of DISTANCE requires an appropriately formatted text file containing survey observation data..... SOP 11-31

Figure SOP 11.4. Creation of a 'Data Filter' in the 'Analysis Components' window of DISTANCE..... SOP 11-32

Figure SOP 11.5. The 4 tabs of the 'Data Filter Properties' window, showing appropriate settings for the "Untruncated" data filter..... SOP 11-33

Figure SOP 11.6. Creation of additional data filters 34

Figure SOP 11.7. Creating and defining the 'model definition' named "4 model AIC-selected"..... SOP 11-35

Figure SOP 11.8. Settings for the "4 models AIC-selected" model definition. .. SOP 11-36

Figure SOP 11.9. Creating and naming the "Top model export stats" model definition SOP 11-37

Figure SOP 11.10. Creating and specifying pre-defined analysis sets and analyses..... 38

Figure SOP 11.11. Creation of a new project in the 'New Project Setup Wizard' of DISTANCE using a template project file SOP 11-39

Figure SOP 11.12. Importing a formatted text file containing survey observation data using the 'Import Data Wizard'. SOP 11-40

Figure SOP 11.13. Displaying, executing, and displaying output from pre-defined analyses in the 'Exploratory' set..... SOP 11-41

Figure SOP 11.14. Setting the right-truncation distance SOP 11-42

Figure SOP 11.15. Histograms showing frequency of detections relative to perpendicular distance from the transect center line..... SOP 11-43

Figure SOP 11.16. Histograms showing frequency of detections relative to perpendicular distance from the transect center line..... SOP 11-44

Figure SOP 11.17. Binning of survey observations by perpendicular distance from the transect center line SOP 11-45

Figure SOP 11.18. Output from DISTANCE for an estimated detection function SOP 11-46

Figures (continued)

Page

Figure SOP 11.19. Selecting, executing, and viewing results from analyses for individual detection function models..... SOP 11-47

Figure SOP 11.20. Exporting the Program DISTANCE results stats file for the selected detection function model..... SOP 11-48

Figure SOP 11.21. Exporting a plot of a detection function from DISTANCE.. SOP 11-49

Figure SOP 11.22. Program R code used to produce an image of the estimated detection function and associated frequency data..... SOP 11-50

Tables

Page

Table SOP 11.1. Data columns in input file of observation data for import to Program DISTANCE..... SOP 11-51

Table SOP 11.2. Pre-defined DISTANCE analyses used in exploratory analyses, estimation of a detection function, and estimation of group size. SOP 11-51

Table SOP 11.3. Key to codes for specification of output related to estimated detection functions and group size from the DISTANCE results files..... SOP 11-52

Summary

After certification of the “KM_E: Cumulative observation database”, the current year’s survey data are used to estimate a detection function, which provides estimates of detection probability for murrelet groups relative to their perpendicular distance from the transect center line. This detection function is integral to assessing whether critical assumptions of methods were met and to population estimates. Analyses are conducted using Program DISTANCE. Exploratory analyses identify violations of assumptions and potential remedies and also identify appropriate truncation and binning strategies for data. Statistical criteria and biological considerations are used to select an appropriate structure for the detection function. Estimates of group size for each species class may also be produced in and exported from DISTANCE. The Data Manager produces necessary input files for DISTANCE analyses, imports data to DISTANCE, and imports to KM_J: Detection function the output from DISTANCE analyses. The Project Leader will conduct analyses to estimate a detection function and group sizes and export results from DISTANCE.

This protocol employs Program DISTANCE version 6.0 release 2 (Thomas et al. 2010), which is Windows-compatible software freely available at <http://www.ruwpa.st-and.ac.uk/distance/>. Because results and file structures may differ among versions, we recommend use of this version to ensure compatibility with database structure. Additional to information below and help files accessed by pressing <F1> within DISTANCE, extensive documentation is provided in an PDF file in the DISTANCE install package. Furthermore, two texts on distance sampling provide extensive theory and practical examples of estimation and use of detection functions (Buckland et al. 2001, 2004), and we will make reference to specific, short passages in external sources to augment methods and discussion in this document.

Program R and Program TINN-R are also used to create figures from DISTANCE output. Program R is a software language and statistical computing environment that is freely available at <http://www.r-project.org/>. Program TINN-R is a GUI/Editor for the R programming language. Information about this software is available at <http://www.sciviews.org/Tinn-R/>, and software is freely available at <http://sourceforge.net/projects/tinn-r/>. This protocol provides R code that can be easily executed within TINN-R to produce specified graphic output.

Outputs from this SOP are: 1) a right-truncation distance for censoring murrelet groups based on distance from the transect center line, 2) an estimated detection function and associated parameter estimates, 3) a figure of the estimated detection function, 4) an estimated detection probability for groups within the right-truncation distance, 5) estimated effective strip width, and 6) estimated group size for each species class (produced separately for analyses of population status and trend).

This SOP will be completed by executing these steps:

1. Importing data into DISTANCE
2. Exploratory analyses to prepare for subsequent analyses
3. Selection of a detection function and estimation of parameters
4. Estimation of group size
5. Exporting results from DISTANCE

Throughout the detailed steps below (numbered text), black text indicates actions essential to the SOP. However, successfully completing these tasks will require the Project Leader to apply judgment to avoid potential pitfalls. We have used blue text to indicate information guiding such decisions, including definition of key terms and concepts, advice for identifying and remedying problems, and references to supporting material.

Data Import to DISTANCE

Necessary data from the repository will be imported to a DISTANCE project via a formatted text input file. A DISTANCE project template will import appropriate settings, data structure, and preliminary analyses. The “KM_E: Cumulative observation database” repository deliverable must be certified before data may be imported.

Detailed Steps

A. Data Manager Tasks

1. Create a new directory to hold the year's work named
“[\\inpglbafs03\data\SEAN_data\Work_zone\KM\KM_J\yyyy\](#)” where yyyy is the four-digit year.
2. Create the ‘DFplot.r’ file to be used in plotting the detection function
 - a. ‘DFplot.r’ is an unformatted text file containing Program R code that can be used to create an image file containing a plot of the detection function.
 - b. Copy the DFPlot.r file into the new directory from ...\\KM\KM_J\.
 - c. Edit DFPlot.r so its filenames and paths for output table and plot reflect the survey year.
 - d. In the event a previous copy of DFPlot.r is not available, the file can be recreated using a text editor to open a new text document, copying the text from the Appendix into the file, and saving the unformatted text under the file name ‘DFplot.r’.
3. Produce a comma-delimited text file of murrelet observation data for import into DISTANCE
 - a. Connect to the data management web site (currently \\165.83.57.239) and log in using your Windows userid and password – userid must be of the form “username@NPS”.
 - b. Under Kittlitz's Murrelets, click on the KM_J task called “Create input table SOP11.1”.
 - c. Specify the year of interest from the dropdown.
 - d. View the sample result on screen if desired.
 - e. Press the button to save complete results to
“[\\inpglbafs03\data\SEAN_data\Work_zone\KM\KM_J\yyyy\input_yyyy.txt](#)” where yyyy is the four-digit survey year.
 - f. The generated file will have these characteristics:

- i. Each observation of a murrelet group with complete distance and angle data will occupy one row in the file.
 - ii. The file contains eight columns in the order specified in Table SOP 11.1. Columns contain headers for field names as specified in Table SOP 11.1.
 - iii. The angle field in column 5 is adjusted before import to DISTANCE.
 1. The source Angle field in Tbl_observation ranges from 0-360°, increasing from 0° opposite the direction of travel on the transect path.
 2. In DISTANCE, permissible angle values range from 0-180°, increasing with deviation from the direction of travel on the transect path.
 3. Angle values from Tbl_observation is adjusted as:
 - i. Range 180-360°, subtract 180°
 - ii. Range 0-179°, subtract from 180°
 - iv. The file is sorted in ascending order by the 'Transect' field.
 4. Create a new DISTANCE project for the survey year
 - a. A template file is used to define settings, data structure, and pre-defined analyses for the DISTANCE project for each survey year.
 - i. There should be both a template.dst file as well as a template.dat directory in [\\inpglbafs03\data\SEAN_data\Work_zone\KM\KM_Jtemplate\](#).
 - ii. Copy them both to the current 'yyyy' directory.
 - b. If no template project file set is available, define a template from scratch.
 - i. Create the base project files.
 1. Start Program DISTANCE and use the New Project Setup Wizard to name the file "template.dst" (Figure SOP 11.1).
 2. Proceed through successive steps of the New Project Setup Wizard, specifying settings as shown in Figure SOP 11.2. Click 'Finish' to start the Import Data Wizard.
 - ii. Populate the base project with observation data.
 1. Select a properly formatted text file (as in section 1.1 above) containing observation data Figure SOP 11.3.1.
 2. Proceed through successive steps of the Import Data Wizard, specifying the data structure as in Figure SOP 11.3.2-6.
 3. Finishing the Import Data Wizard will create the project template file and open the 'Data' tab of the 'Project Browser'. Left click on the 'Observation' data layer to confirm that the data have been properly imported.
 - iii. Specify pre-defined analysis components.
 1. Start DISTANCE and load the project template file created above.

2. Open the 'Analysis Components' window to view the Data Filters (Figure SOP 11.4). Rename the Default Data Filter to "Untruncated" and open the 'Item Properties' window for this data filter.
3. Select each of the tabs in the 'Item Properties' window, ensure settings match those shown in Figure SOP 11.5, and close the window.
4. Copy the "Untruncated" data filter to create 2 new filters named "Truncated" and "Truncated Bins A", and open the 'Item Properties' windows to apply settings to each as shown in Figure SOP 11.6.1-3.
5. Copy the "Truncated Bins A" filter create the filter "Truncated Bins A Status". Open the 'Data Selection' tab in the 'Data Filter Properties' window (Figure SOP 11.6.4), and in the data selection box use the drop down menu to select the 'Sample' layer type.
6. In the text box to the right, enter the text inside the brackets: [Label <> '016']. This creates a data filter that excludes Transect 016 when estimating group size for analyses of population status.
7. From the 'Analysis Components' window, left-click the model definitions button to list the Model Definitions (Figure SOP 11.7.1). Rename the default model definition to '4 models AIC-selected' and open the 'Model Definition Properties' window by right-clicking on the ID number and selecting 'Item Properties'.
8. Successively select each tab near the top of the 'Model Definition Properties' window and apply settings as shown in Figure SOP 11.7.2-6 and Figure SOP 11.8. Note that the 'Detection Function' tab has 5 separate lower level tabs, with settings shown in Figure SOP 11.9.
9. From the 'Model Definitions' window, create a new model definition named 'Top model export stats' as shown in Figure SOP 11.9. Open the "Model Definitions Properties" window, select the 'Misc.' tab, check the 'Create results stats file' box, and enter the output path
[\\inpglbafs03\data\SEAN_data\Work_zone\KM\KM_J\yyyy\DFstats_yyyy.txt](\\inpglbafs03\data\SEAN_data\Work_zone\KM\KM_J\yyyy\DFstats_yyyy.txt).
10. Right-click on the '4 models AIC-selected' definition and create an additional 4 model definitions with names corresponding to the model definitions in four rows in Table 11.2 after the "Top model export stats" definition. For each,

- navigate to the 'Models' lower-level tab in the 'Detection function' tab of the 'Model Definition Properties' window (see Figure SOP 11.8.1), and use the "-- button to the right of the 'Detection function models' box to reduce the number of models to 1. Then, select the appropriate key function and series expansion for each using the drop down menus.
11. Create 2 additional model definitions by copying '4 models AIC-selected' definition corresponding to the next 2 model definitions in Table 11.2 with only key functions. For each, reduce the number of models to one and select the appropriate key function in the 'Models' lower-level tab of the 'Detection function' tab in the 'Model Definition Properties' window (Figure SOP 11.8.1). In the 'Adjustment terms' lower-level tab, reduce the 'Maximum terms' to '0' using the down arrow button. (Figure SOP 11.8.2).
- iv. Specify pre-defined analysis sets and analyses.
 1. Navigate to the 'Project Browser' using 'View|Project Browser' in the menu bar.
 2. As shown in Figure SOP 11.10.1, select the 'Analyses' tab and then rename the analysis set "Exploratory".
 3. Create a total of 3 models in the 'Exploratory' set, and rename each as shown in Figure SOP 11.10.1.
 4. Open the 'Analysis Details' window for each analysis and define the 'data filter' and 'model definition' (Figure SOP 11.10.1-2) as listed in Table SOP 11.2.
 5. As shown in Figure SOP 11.10.3, use the drop down menu to select a new analysis set and rename it "Estimation". Create a total of 9 analyses in this set and name as in Figure SOP 11.10.4.
 6. Open the 'Analysis Details' window for each analysis and define the 'data filter' and 'model definitions' as listed in Table SOP 11.2.
 7. The project template file is saved automatically; save on exiting from DISTANCE is not needed.
 - c. Import survey observation data to a Program DISTANCE project file.
 - i. Start Program DISTANCE and use the New Project Setup Wizard to create a new project named "DF_YYYY.dst", where YYYY is the survey year (Figure SOP 11.11.1).
 - ii. In successive steps of the New Project Setup Wizard (Figure SOP 11.11.2-4), specify the project template file that will be used to import settings and preliminary analyses, and then proceed to the 'Data Import Wizard'.

- iii. To begin the 'Data Import Wizard', select the formatted text input file ("input_yyyy.txt" created above) for the current year's survey (Figure SOP 11.12.1).
- iv. Proceed through successive steps of the Import Data Wizard, specifying the data structure as in Figure SOP 11.12.2-5.
- v. Finishing the 'Import Data Wizard' will create the project file and open the 'Data' tab of the 'Project Browser'. Left click on the 'Observation' data layer to confirm that the data have been properly imported (Figure SOP 11.12.6).
- vi. The project file is saved automatically; save on exiting from DISTANCE is not needed.

Exploratory Analyses

The purpose of analyses in this SOP is to estimate an appropriate detection function for the current year's survey observations using DISTANCE. The detection function is a curve describing the probability of detection of murrelet groups during surveys relative to perpendicular distance from the transect center line. Several alternative 'key functions' may be used to model the detection function, and several alternative 'series expansion terms' may also be added to these key functions. The ultimate goal of analyses is to select an appropriate model for estimation of the detection function, but exploratory analyses are necessary to lay the groundwork for further analyses.

The Project Leader will create a DISTANCE project file to conduct exploratory analyses of the current year's survey data. Data is imported to DISTANCE using a project template file, which will provide appropriate settings and pre-defined analyses. These analyses will serve as a starting point for visualizing data, estimating preliminary detection functions, and assessing adherence of data to assumptions of statistical methods. Based on results, the Project Leader will define additional analyses as needed. The Project Leader will apply statistical criteria and judgment to select an appropriate right-truncation distance for censoring observations and appropriate bins for grouping observations by distance. When necessary, suitable remedies will be applied to mitigate problems arising from violations of assumptions.

Detailed Steps

A. Project Leader Tasks

1. Execute preliminary models to visualize data
 - a. In DISTANCE, open the project file for the current year's survey data, navigate to the 'Project Browser', and select the 'Exploratory' analysis set to view the pre-defined exploratory analyses (Figure SOP 11.13.1).
 - b. Pre-defined analyses: The 5 pre-defined analyses are listed in Table SOP 11.2. These analyses primarily differ in how data are truncated and binned.
 - c. Alternative models: All predefined analyses consider 4 alternative models for estimating the detection function, and these models consist of a 'key function' paired with a 'series expansion'. Alternative models are the 'half-normal' key

function paired with the 'hermite polynomial' or 'cosine' series expansion and the 'hazard' key function paired with the 'simple polynomial' or 'cosine' series expansion.

- d. Akaike's Information Criterion (AIC; Burnham and Anderson 2002): In each analysis, the model with the lowest AIC value indicates the most parsimonious model, which adequately explains the data with the minimum number of parameters. Parameter estimates and figures for each analysis are from this 'best' model only.
 - e. Select the 'Untruncated' analysis by left-clicking on it and then left-click the 'Run' button to execute this analysis (Figure SOP 11.13.1). This analysis includes all observations of murrelet groups and uses exact distance data.
 - f. Open the 'Analysis Details' window and select the 'Results' tab. Use the 'Back' and 'Next' buttons or the drop down menu to page through results to locate the 'Detection Fct/Global/Plot: Detection Probability 1' and 'Detection Fct/Global/Plot: Detection Probability 2' windows (Figure SOP 11.13.2), which show the distance data displayed in histograms with differing interval width.
 - g. The 'Log' tab: Selecting this tab will provide specifications for this analysis as well as any warnings (yellow/orange) or errors (red). Warnings are common and usually do not indicate serious problems, but errors require resolution. Descriptions of warnings and errors are found in the appendix of the DISTANCE documentation.
 - h. Detection Probability Plot: The estimated detection function from the AIC-selected model is shown in red; probability of detection of murrelet groups (y-axis) is shown relative to perpendicular distance in meters from the transect center line. The blue histogram shows the frequency of detections at different distances from the center line; the frequencies are scaled to match the scale of the detection function. The histogram serves to visualize the raw data.
2. Select a right-truncation distance
- a. Using the 'Detection Probability' plots from the "Untruncated" analysis, select the perpendicular distance from the transect center line (in 10's of meters) beyond which the probability of detection is less than ~0.15.
 - b. Right-truncation distance: This is the perpendicular distance from the transect center line beyond which all observations are censored. Observations become rare and uneven at large distances, which can create problems for estimating a detection function. Therefore, right-truncation typically is applied where data become sparse. For additional discussion, see section 4.3 of Buckland et al. (2001).
 - c. Selecting the right-truncation distance requires judgment and should be based on the overall trend of declining detection probability, not just local variation. An estimated detection function may provide a useful guide if it adequately

- fits the data. As an example, for the data presented in Figure SOP 11.13.2, observations were right-truncated at distances >230 m.
- d. The selected right-truncation distance should be applied to all future analyses by editing the 'Data Filters' pre-defined and user-defined analyses. In the 'Analysis Components' window open the 'Data Filters' window (Figure SOP 11.14.1). For the 'Truncated', 'Truncated Bins A', and 'Truncated Bins A Status' analyses, open the 'Data Filter Properties' window and select the 'Truncation' tab (Figure SOP 11.14.2). Select 'Discard all observations beyond' and then enter the selected right-truncation distance. When prompted, confirm the change by selecting 'Yes' in the pop-up window.
3. Consider binning observations based on distance
- a. Shape criterion: To meet statistical assumptions, the detection function should 1) be monotonically and smoothly declining, 2) should asymptotically approach zero at larger distances, and 3) should have a "shoulder," meaning that the detection function should be relatively flat (slope ~0) at small distances.
 - b. Purpose of binning data: Binning data (i.e., grouping observations based on intervals of perpendicular distances) may be useful in improving the robustness of estimates by improving adherence to shape criterion.
 - i. Selection of bins: Typically, data can be grouped in 7-10 bins with little loss in precision of estimates relative to analyses using exact distances. Selection of "cut-points" for bins depends on distribution problems observed in the data (below), but selection of the first cut-point >0 is critical. Selection of bins is likely to be an iterative process. Exploratory visualization of the data will identify problems, and several alternative binning strategies may be considered. Fitting detection functions to the data may also reveal poor goodness-of-fit and motivate further investigation of binning strategies.
 - ii. Heaping: "Heaping" occurs when the frequency of observations at specific distances is inflated relative to frequencies at nearby distances. Heaping often results from systematic errors by observers during data acquisition. For example, in estimation of distances and angles to murrelet groups, observers might favor "convenient" values (e.g., distances of 50 or 100 m, angles of 0° or 45° relative to the path of the boat), which can result in peaks and troughs in detection frequencies (Figure SOP 11.15.A). Placing binning cut-points at mid-points between peaks can improve adherence to the shape criterion and improve the fit of the detection function to the data (Figure SOP 11.15.B). Overuse of the 0° angle can result in a spike in observations on the transect center line (see Thomas et al. 2010 Figure 2a). In such case, the width of the first bin should be sufficient to smooth the spike by averaging out local peaks and troughs.

- iii. Movement prior to detection: Distance sampling assumes animals are detected at their initial locations. If animals near the transect center line attempt to evade by moving away from the boat, there may be a dearth of observations at short distances and a surplus of observations at intermediate distances (Figure SOP 11.16.A). If the magnitude of evasive movement is not too large, this problem may be mitigated by increasing the width of the first bin to include most groups that are moving away from the transect center line (Figure SOP 11.16.B).
 - iv. (Over-) Guarding the center line: If observers devote too much attention to searching near the center line relative to waters slightly further away, then detection probability may decline steeply at intermediate distances (see Figure 2.1 in Buckland et al. 2001) and hence violate shape criteria. Having observers distribute search effort over a wider area can resolve this problem.
 - v. Abrupt changes in frequencies of detections violate shape criteria and hamper achieving adequate goodness-of-fit for detection functions. Bins that straddle such changes will make changes more gradual.
- c. In DISTANCE, select the 'Analyses' tab in the 'Project Browser' and navigate to the 'Exploratory' analysis set (Figure SOP 11.17.1). Select the 'Truncated' analysis and execute it.
 - d. Open the 'Analysis Details' window of the 'Truncated' analysis and select the 'Results' tab. Use the 'Back' and 'Next' buttons or the drop down menu to page through results to locate the 'Plot: Detection Probability 1' and 'Plot: Detection Probability 2' windows (Figure SOP 11.17.2).
 - e. Use the histograms and estimated detection functions to identify any distribution problems. If none are present, then binning is not required and you may proceed to "Selecting a Detection Function" and conduct analyses using exact instead of binned distances values. However, binning is typically useful.
 - f. The example histogram in Figure SOP 11.17.2 shows some evidence of evasive movement and 'heaping.' Evasive movement appears slight, with a peak in observations at ~15 m. The heaping at ~120 m likely reflects high usage of the 150 m distance estimate by observers in combination with typical angular deviation of ~25° from the center line.
 - g. If binning is justified, open the 'Analysis Components' window, select the 'Data Filters' tab, and open the 'Data Filter Properties' window for the 'Truncated Bins A' filter (Figure SOP 11.17.3).
 - h. The 'Intervals' tab will show a pre-defined number of bins and cut-points (Figure SOP 11.17.4); adjust as needed.
 - i. Navigate to the 'Analyses' tab of the 'Project Browser' to select and execute the 'Truncated with Bins A' analysis. Open the 'Analysis Details' window

- and the 'Results' tab, and navigate to the 'Plot: Detection probability' window to view a histogram using the defined bins.
- j. The example histogram in Figure SOP 11.17.5 shows how binning has minimized problems seen in 11.17.2. The 0-30 m bin has 'leveled' the peak associated with evasive movement, and placing a cut-point at 120 m near the peak of the 'heaped' observations has distributed the excess observations into adjacent bins and 'smoothed out' the peak.
 - k. It may be useful to consider alternative binning strategies. If so, create new 'data filters' and 'analyses' to visualize alternatives. In the 'Data Filters' tab of the 'Analysis Components window, right-clicking on an existing data filter and selecting 'New Item' will copy existing settings to a new data filter, which can be renamed (e.g., 'Truncated Bins B', etc.) by double-clicking on the data filter name. Follow procedures in Figure SOP 11.17.3-4 to define bins. In the 'Analyses' tab of the "Project Browser," right-clicking on an existing analysis and selecting 'New Analysis' will copy settings to a new analysis, which can be renamed (e.g., 'Truncated with Bins B', etc.) by double-clicking on the analysis name. Right-click on the analysis and select 'Item Properties' to open the 'Analysis Details' window and select the appropriate data filter on the 'Inputs' tab (Figure SOP 11.17.6).
 - l. For each analysis, navigate to 'Plot: Detection Probability' in the 'Results' tab of the 'Analysis Details' window to evaluate how well shape criteria are met. Proceed upon selecting an appropriate binning strategy.

Selecting a Detection Function

After exploratory analyses to select a right-truncation distance and apply an appropriate binning strategy, the Project Leader will conduct analyses to select a model for estimation of a detection function. Program DISTANCE will use AIC to automatically select among alternative models the model that provides the most parsimonious fit to these data. However, AIC selection is blind to several critical criteria for robust estimation of a detection function, so the Project Leader will assess whether estimated detection functions achieve adequate goodness-of-fit and meet shape criteria. Problems may suggest revisions to models and/or binning strategies.

Detailed Steps

A. Project Leader Tasks

1. Execute AIC model selection
 - a. In the 'Analyses' tab of the 'Project Browser' select the 'Estimation' analyses set. Right-click on the '4model AIC selected' analysis and select 'Item Properties' to open the 'Analysis Details' window. On the 'Inputs' tab and select the data filter that was selected during the exploratory analyses (as shown in Figure SOP 11.17.6). If binning was unnecessary, this will be the 'Truncated' filter, otherwise it will be the data filter with the chosen binning strategy.

- b. Return to the 'Analyses' tab of the 'Project Browser,' execute the '4 model AIC selected' analysis, open the 'Analysis Details' window, and select the 'Results' tab (procedure shown in Figure SOP 11.17.1-2).
 - c. **AIC model selection:** AIC values are used to select the most parsimonious model structure. Lower AIC values indicate higher parsimony, and generally models within ~2 units of the top model also have substantial support. AIC selection is applied in 2 contexts during model selection. First, for each of the 4 alternative key function + series expansion combinations, AIC values are used to assess if the series expansion (which adds an additional parameter) increases parsimony relative to the key function alone. If the series expansion provides no improvement, it is dropped. Second, AIC is used to compare the 4 alternative models. In each analysis, AIC values for each model structure considered are reported in the 'Model Fitting' windows in the 'Results' tab of the 'Analysis Details' window. DISTANCE reports results only from the top model. AIC values can also be used to compare among analyses. In the 'Results' tab of the 'Project Browser,' DISTANCE reports the 'Delta AIC' (deviation of the top model in each analysis relative to the top overall model among analyses) for each analysis. *It is critical to realize that AIC values can only be compared when underlying data are identical.* Both the sample size as well as the binning strategy must be identical.
 - d. Use the drop down menu or the 'back' and 'next' buttons to navigate through the 'Model Fitting' windows in the 'Results' tab. The initial windows describe AIC selection relative to fitting the series expansion term for each of the 4 alternative models. The final window shows the AIC-selected top model among all 4 alternatives. This window also displays the '# of samples' (i.e., # of transects), the 'Width' (the right-truncation distance), and the '# of observations' (# of murrelet groups observed). Navigating to the next window ('Parameter Estimates') gives the same information along with estimates for parameters in the detection function (parameters shown in equation describing selected model) as well as probability of detection probability 'p' and the effective strip width 'ESW' (Figure SOP 11.18.1).
 - e. **Effective strip width:** This is the perpendicular distance within which as many groups as detected as are missed outside it. Equivalently, if all groups were detected within this distance, then the number of groups would be identical to the actual number of groups observed. The effective strip width provides an overall measure of detection probability relative to distance that facilitates comparisons among estimates with different detection functions and right-truncation distances.
 - f. **Detection probability:** The overall probability of detecting murrelet groups within the right-truncation distance.
2. Evaluate goodness-of-fit and detection function shape

- a. The output for assessing goodness-of-fit will differ depending on whether binned or exact perpendicular distance data were used in analyses. For binned data, the window 'Chi-sq GOF Test' in 'Results' tab of the 'Analysis Details' window will display a χ^2 goodness-of-fit test with bins matching those defined in the data filter. For analyses using exact distance data, the bins for the χ^2 test are defined by DISTANCE. For analyses using exact data, there will also be a q-q plot in the window 'Plot: qq-plot'.
- b. Goodness-of-fit and DISTANCE: Although several quantitative methods for selecting models and assessing goodness-of-fit exist in DISTANCE, they are of limited utility in guiding selection of a reliable detection function. While quantitative methods can be useful tools, application of sound principles and judgment are necessary to robust estimation of a detection function.
- c. χ^2 goodness-of-fit test: This test provides a quantitative assessment of fit of the model to the data by comparing expected versus observed frequencies of observations in bins defined by perpendicular distance from the transect center line. *P*-values <0.05 indicate statistically significant lack-of-fit. However, the χ^2 test has significant limitations for assessing fit of estimated detection functions. First, the power of the test increases with sample size. Thus, as the number of observed groups increases the magnitude of deviation between the data and the detection function required to generate a statistically-significant difference decreases. Second, results are strongly dependent on the bins used to make comparisons, and the test is sensitive to poor binning strategies. Third, the test treats all bins equally and does not recognize the critical importance of bins at small distances. Fourth, the test does not consider shape criteria required for reliable estimation of detection functions. In sum, *P*-values should not be used as a comprehensive assessment of adequacy of estimated detection functions but can indicate potential goodness-of-fit problems. For individual bins, high χ^2 -values can be useful in highlighting at what distances goodness-of-fit problems are occurring.
- d. q-q plots: Quantile-quantile plots provide a graphical measure of assessing fit of the model to the data are available for analyses using exact distance data. These plots compare the expected cumulative distribution function (blue line) from the estimated detection function to the actual distribution of the data (red dots). Data are sorted in increasing order, and the *x*-axis increases from 0 distance (left intercept) to the right-truncation distance (right intercept). Most points falling near the line indicates good fit, but systematic departures indicate lack of fit at specific distances.
- e. Assessment of the detection function should focus on whether adherence to shape criteria (defined at step A.3.a in the 'Exploratory Analyses' section above) and goodness-of-fit are adequate.

- f. Navigate to the 'Plot: Detection Probability' window the 'Results' tab to view the estimated detection function (red) and the histogram with frequencies of observed groups (blue) in defined bins (Figure SOP 11.18.2). This plot allows visual assessment of the shape of the detection function and the location and magnitude of deviations between predictions from the detection function and the histogram data.
- g. Navigate to the 'Chi-sq GOF Test' window the 'Results' tab to view results from the χ^2 goodness-of-fit test (Figure SOP 11.18.3). The test will use the user-defined distance bins if present; otherwise, DISTANCE will conduct 2 tests (displayed in separate windows) using auto-generated bins. The sum of the χ^2 values across distance bins ('cell' in the χ^2 table) gives the total χ^2 value, which is used to generate an overall P -value for the null hypothesis that the predicted and actual frequencies of observations are not different. See caveats on χ^2 testing above. The χ^2 values for each bin are useful for assessing goodness-of-fit at specific distance bins as well as comparing the magnitude of deviations among bins.
- h. Goodness-of-fit and detection distances: Distance sampling requires that all detection probabilities are scaled relative to that at the x -intercept, and it is usually assumed that detection is 100% at distances near 0. Therefore, it is critical that detection is accurately modeled near the x -intercept, and consequences of poor goodness-of-fit at small distances are more severe than at other distances. However, data at all distances are treated equally by current model-fitting and assessment techniques. Ideally, the detection function near the x -intercept should have a strong "shoulder" and close correspondence between predicted and actual observations. When data are binned, these guidelines should hold for the 2 bins at smallest distances. If goodness-of-fit is poor near the x -intercept, a detection function that is relatively too high or low can over- or under-estimate detection probability across all distances. A particularly troublesome case arises when detections are much more frequent in the 1st (smallest distance) relative to 2nd bin, as different models can have similar goodness-of-fit yet produce very different estimates (see Figure 2.5 and associated text in Buckland et al. 2001). If alternative binning strategies cannot produce a broader "shoulder" in the data distribution, then models with steep slopes at the x -intercept (i.e., that "climb the y -axis") should generally be avoided.
- i. If the AIC-selected detection function meets shape criteria and achieves adequate goodness-of-fit, then proceed to heading 3 below. Otherwise, the iterative process of selecting distance bins and fitting detection function models should continue until results are suitable.
- j. Consider altering binning strategies to produce histograms of frequencies of observations more conducive to achieving suitable detection functions. If frequencies of detections at large distances are hindering model fitting, it may

also be worthwhile to adjust the right-truncation distance to improve distribution of the data.

- k. Problems in the AIC-selected detection function may highlight inadequacies in the selected binning strategy. The functional form of detection models may not be flexible enough to achieve adequate goodness-of-fit, especially if large changes in detection frequencies occur between adjacent bins. If so, changes to the binning strategy (A.3 in 'Exploratory Analyses' section above) may mitigate problems, although binning cannot fully resolve severe problems in data distribution.
- l. If binning has been optimized for these data, consider alternatives to the AIC-selected detection function model, either with different key function + series expansion combinations or for a key function without a series expansion term. To execute analyses consisting of individual detection function models, navigate to the 'Estimation' analysis set in the 'Analyses' tab of the 'Project Browser' (Figure SOP 11.19). Select and execute desired analyses. Compare the 'Delta AIC' values in the right pane to assess support for each model. In the 'Results' tab of the 'Analysis Details' window for each analysis, navigate to the 'Plot: Detection Probability' and 'Chi-sq GOF test' windows to assess detection function shape and goodness-of-fit.
- m. Problems may also arise from the inadequacy of the functional form of the AIC-selected detection function. Because AIC-selection procedures do not consider shape criteria or the heightened importance of goodness-of-fit near distance 0, it may be the AIC-selected model is inferior to models with slightly higher AIC values (i.e., ΔAIC of <2). Thus, it may be worthwhile to assess performance of models with different key function + series expansion combinations. Inclusion of series expansion terms can sometimes lead to undesirable or implausible curvature in the detection function, so considering key functions with fewer or no series expansion adjustment terms may improve detection function shape.
- n. AIC values can only be used to compare alternative models of detection function for the same data set (i.e., identical sample and identical binning strategy). Therefore, if alternative binning strategies are considered, alternative detection functions models from different data sets can be compared using the χ^2 goodness-of-fit test and assessment of detection function shape. If alternative binning strategies are considered, it may be useful to create a new analysis set to segregate results using different data sets.
- o. The iterative process of truncating data, selecting a binning strategy, fitting alternative models should lead to a detection function meeting shape criteria and achieving adequate goodness-of-fit. AIC model selection and goodness-of-fit tests are useful tools in this process, but ultimately the Project Leader

must employ judgment to select a final detection function that is statistically reasonable and biologically plausible.

3. Applying the selected model to the 'Top Model Export Stats' analysis
 - a. The selected binning strategy and selected model for estimation of a detection function should be applied to the analysis 'Top Model Export Stats' in the 'Estimation' model set.
 - b. Select the 'Analyses' tab in the 'Project Browser' window, and right-click the 'Top Model Export Stats' analysis and open the 'Analysis Details' window (Figure SOP 11.20.1).
 - c. Select the 'Inputs' tab and then select the 'Data filter' containing the binning strategy selected for estimation of the detection function.
 - d. Confirm that the 'Top model export stats' is selected in the 'Model definition' (Figure SOP 11.20.2).
 - e. Open the 'Model properties' window and select the 'Detection function' tab and the 'Models' lower-level tab (Figure SOP 11.20.2-3).
 - f. Reduce the number of models to one and select the key function and series expansion used to estimate the detection function (Figure SOP 11.20.3).
 - g. In the 'Adjustment terms' lower-level tab (Figure SOP 11.20.4), manually input the number of adjustment parameters (i.e., terms in the series expansion). If a series expansion was included, there will normally be 1 parameter. If no series expansion was included in the detection function model, set this number to "0".
 - h. Open the 'Model definition properties' window and select the 'Misc.' tab (Figure SOP 11.20.5).
 - i. Enter a path for the output text file
"[\\inpglbafs03\data\SEAN_data\Work_zone\KM\KM_J\yyyy\DFstats_yyyy.txt](#)", where yyyy is the survey year, and confirm the change (Figure SOP 11.20.6).
 - j. Running this analysis will export the results file (section 7) and also will generate estimates of group size (section 6). Be sure to verify that results match those for the selected detection function model.

Estimating Group Size Using DISTANCE

Because encounter rates are calculated for groups of individuals, estimates of average group size are required to estimate density and abundance. However, if large groups are more likely to be detected at large distances, then the average calculated from observed groups will be positively biased.

Regression techniques can test for presence of such size bias and, if present, correct for it by predicting average group size for detection probability = 1. Because of the strong shoulder observed in the detection function for these surveys (Hoekman et al. 2011a, b), it was unlikely that a linear regression of group size on perpendicular distance to the transect center line would provide adequate fit to our data. Therefore, we took the approach of regressing $\ln(\text{group size})$ on the estimated detection probability for each group from the selected detection function (Buckland et al. 2001).

Analyses should be conducted separately for each species class (KIMU, MAMU, and unidentified *Brachyramphus* murrelets). Average group size should be preferred, except when strong evidence supports use of expected group size from regression analyses. All survey transects should be used for estimation of group size for analyses of population trend. However, one transect (016) selected using non-probabilistic methods should be excluded from estimation of group size for analyses of population status (see Appendix B, section B.7.3).

Detailed Steps

A. Project Leader Tasks

1. Generating and viewing estimates of group size for analyses of population trend
 - a. In DISTANCE, open the project file for the current survey year, navigate to the 'Project Browser', and select the 'Estimation' model set. Select the 'Top model export stats' analysis in the "Estimation" model set and run it by right-clicking and selecting 'Run analysis' option.
 - b. Right-click this analysis again, and then select the "Analysis details" option and select the 'Results' tab.
 - c. Use the drop-down menu or the 'Back/Next' keys to navigate to the pages with the 'Cluster size-Estimates' for each species class (K/M/U). On these pages, '# observations' gives the sample size, ' $\Pr(T < t) =$ ' gives the *P*-value of the regression of $\ln(\text{group size})$ on predicted detection probability, estimates for 'Expected cluster size' are for estimates of group size when detection probability = 1 from the regression analyses, and estimates for 'Mean cluster size' are simply the average group size.
 - d. Pages of the cluster size results with the 'Regression plot' title give a text based graphical presentation of the regressions. The estimated regression line is represented by asterisks "*", while open circles "o" represent one or more data points.
 - e. The 'Estimation summary – expected cluster size' window contains a summary of estimates for each species class along with degrees of freedom and 95% confidence intervals. On this page, estimates from the regression analyses are labeled "*r*" (correlation coefficient), "*r-p*" (*P*-value), and "E(S)" (expected group size).
2. Interpreting estimates of group size
 - a. For each species class, the Project Leader must select the best estimate of group size for estimation of density and abundance.
 - b. Typically, average group size should be preferred. However, predicted group size for detection probability = 1 from regression analyses may be preferred when strong evidence supports size bias in detection of groups. Strong evidence should include "statistically significant" results from regression analyses (i.e., $P < 0.05$), a negative slope parameter estimate, and satisfactory adherence to assumptions of regression analyses.

- c. For regressions with negative slope coefficients and $P < 0.05$, examine regression plots to assess whether results are robust to data outliers. In particular, one or a few large groups with low predicted detection probability may create “high leverage points” that unduly inflate to the steepness of a negative slope coefficient. If the “statistical significance” of a regression depends on these high leverage points, averages may provide more reliable estimates of group size.
- d. If the influence of data outliers on results is unclear, users confident in use of DISTANCE may find it useful to compare results with and without outliers. If regression estimates only receive strong support with outliers, then average estimated group size likely will be more reliable. To filter outliers for a given species class:
 - i. In the ‘Cluster size – Regression plot’ page in the ‘Results’ tab of the analysis under consideration, note a cut-point on the on the y-axis (“log of cluster size”) that separates outliers from other data, and transform this value to natural units by taking the inverse log.
 - ii. Select ‘View|Analysis Components’, click the ‘Data Filters’ button, select the data filter used for the analysis under consideration, right-click and select ‘New Item’ to create a new data filter. If the settings of the original filter have been successfully copied, the new filter will have the name of the original with a “1” appended. The new filter can be renamed by double-clicking on the name.
 - iii. Left-click to highlight the new filter, right-click and select ‘Item properties’ to open the ‘Data filter properties’ box. On the ‘Intervals’, ‘Truncation’, and ‘Units’ tabs, confirm that the selections are identical to the data filter used by the original analysis under consideration.
 - iv. On the ‘Data selection’ tab (shown in Figure SOP 11.6.4), left-click the “+” button to the right of the ‘Data selection’ box to add a row in the box. In the drop-down menu under ‘Layer type’, select “Observation”. Left-click the text box immediately to the right and enter “[Cluster size] < X” (without quotations), where X = the inverse log of cut-point value chosen above. Left-click ‘OK’ to save changes.
 - v. Select ‘View|Project Browser’ and navigate to the ‘Estimation’ analysis set. Select to highlight the analysis under consideration, and right-click and select ‘New analysis’. If analysis settings have been successfully copied, the new analysis will have the name of the original with “1” appended. The new filter can be renamed by double-clicking on the name.
 - vi. Select and right-click this new analysis and choose ‘Analysis details’ to open the ‘Analysis details’ window. On the ‘Inputs’ tab, confirm that the ‘Model definition’ is the one for the model under

- consideration, and then double left click on the newly-created data filter to select it. Click the 'Run' button to execute the new analysis.
- vii. Select the 'Results' tab and navigate to the group size regression results for the species class in question. Confirm that the outliers have been filtered from analyses. Compare the slope coefficients and *P*-values between this analysis and the original analysis.
 - viii. Comparison of analyses with and without outliers is useful to determine effects of outliers on results, but should not be used for estimation of density and abundance. If support for regression analyses depends on few outliers, estimates of average group size may be preferred over estimates from regression analyses.
 - e. Output exported from DISTANCE will contain estimates and associated parameters for both average group size and predicted group size from regression analyses. The Project Leader must use statistical criteria above and judgment to select the most appropriate estimates of group size for each species class.
3. Specifying selected estimates of group size
- a. The selected estimates of group size for each species class must be manually specified in the "DFstats_YYYY.txt" file (created at A.3.i in the 'Selecting a Detection Function' section above), and the edited file is saved in the comma delimited file format (.csv). A text editor or spreadsheet software can be used for editing; here we provide directions using Microsoft Excel 2010.
 - b. In Excel, select the 'File' tab, and click 'Open'. In the drop down menu to the right of the 'File name' box, select 'All files'. Navigate to the folder containing "DFstats_YYYY.txt", select it, and click 'Open'.
 - c. In the 'Text Import Wizard – Step 1', select 'Delimited' original data type and click 'Next'.
 - d. In the 'Text Import Wizard – Step 2', select the 'Space' delimited file type, select 'Treat consecutive delimiters as one', and click 'Next'.
 - e. In the 'Text Import Wizard – Step 3', select the column data format 'General' and click 'Finish'. The text file should be displayed in a spreadsheet. Data in each row are coded as specified in Table SOP 11.3, with columns numbered from left to right, starting with the first column containing values. In Excel, sequential columns will be labeled "A, B, C . . ." rather than "1, 2, 3 . . .". Excel may create an additional column at the left containing no values; hence column 1 in Table SOP 11.3 may correspond to column "B" in Excel.
 - f. The 1st column from the left (labeled 'B' in Excel) should be the first column with values, which specify the species class for group size estimates on that row: 1 = Kittlitz's murrelet, 2 = marbled murrelet, and 3 = unidentified murrelet.

- g. The 4th and 5th columns ('E' and 'F' in Excel) specify rows with group size estimates: values of 3 and 1 correspond to estimates of average group size, 3 and 4 to regression estimates.
 - h. For each species class, place a "Y" in the 2nd column ('C' in Excel) on the row with the selected estimate of group size and an "N" in the row of the other estimate of group size. Other rows of this column should contain the value '0'; leave these unchanged.
 - i. Save the edited file in the comma delimited file format (.csv). In Excel, select the 'File' tab, click on 'Save As', and in the drop down menu next to "Save as type:" select 'CSV (comma delimited)'. Enter a path for the output file "[\\inpglbafs03\data\SEAN_data\Work_zone\KM\KM_J\yyyy\DFstats_yyyy.csv](#)", where yyyy is the survey year, and click 'Save'.
- 4. Generating and viewing estimates of group size for analyses of population status
 - a. In the 'Analysis Components' window, open the 'Data filters' window, select and right-click on the 'Truncated Bins A Status' filter to open the 'Data filter properties' window (Figure SOP 11.17.3). Enter the binning strategy used to estimate the detection function (Figure SOP 11.17.4). This will be the binning strategy from the data filter selected at step A.3.c in the 'Selecting a Detection Function' section above.
 - b. As in Figure SOP 11.20.1, select the 'Analyses' tab in the 'Project Browser' window, and select and right-click the 'Top Model Export Stats Status' analysis and open the 'Analysis Details' window.
 - c. On the "Inputs" tab, confirm that the 'Truncated Bins A Status' model filter and the 'Top model export stats status' model definition are selected (Figure SOP 11.20.2).
 - d. Open the 'Model properties' window as in Figure SOP 11.20.2, and apply the model selected for estimation of the detection function (Figure SOP 11.20.3-4) as detailed in steps under the A.3 heading in the 'Selecting a Detection Function' section above.
 - e. Open the 'Model definition properties' window and select the 'Misc.' tab (Figure SOP 11.20.5).
 - f. Enter a path for the output text file [\\inpglbafs03\data\SEAN_data\Work_zone\KM\KM_J\yyyy\DFstats_status_yyy.txt](#), where yyyy is the survey year, and confirm the change (Figure SOP 11.20.6).
 - g. In the 'Project Browser,' select the 'Top model export stats status' analysis in the "Estimation" model set and run it by right-clicking and selecting 'Run analysis' option. This generates estimates of group size and the associated output file for analyses of population status.
 - h. Using the 'Top Model Export Stats Status' analysis, repeat steps in A.1 above to assess whether estimates of average group size or estimates of expected group size from regression analyses should be used in analyses of population

status. Typically, group size results will be very similar to those for analyses of population trend.

- i. Repeat steps in section A.3, except substituting “DFstats_status_YYYY.txt” at step b and “DFstats_status_YYYY.csv” at step i.

Assembling the KM_J Deliverable

After selecting a detection function and estimates of group size for use in analyses of population status and trend for the current year's survey data, the Project Leader will 1) create a figure of the estimated detection function and 2) assemble this figure and the.csv files containing output from Program DISTANCE in an archive file. The Data Manager will import these into deliverable KM_J: Detection function.

Detailed Steps

A. Project Leader Tasks

1. Export the detection function from Program DISTANCE and create a figure in the .png image file format.
 - a. In DISTANCE, open the project file for the current survey year, navigate to the 'Project Browser', and select the 'Estimation' model set. After executing the 'Top model export stats' analysis, select and right-click on this analysis and open the results window by choosing 'Analysis Details'.
 - b. Select the 'Results' tab, navigate to the 'Plot: Detection Probability' window using the drop down menu, and then select 'Analysis – Results|Copy Plot to Clipboard' from the menu (Figure SOP 11.21).
 - c. Using a text editor such as Notepad, open a text file and paste the clipboard material into this file.
 - d. Save the file as unformatted text using the path
“\\inpglbafs03\data\SEAN_data\Work_zone\KM\KM_J\YYYY\DFplot_YYYY.txt”, where YYYY is the survey year.
 - e. Open Program TINN-R and use the 'File|Open' menu command to open the file 'DFplot.r' in the editor (Figure SOP 11.22).
 - f. Open an R GUI window using the 'R|Start/close and connections|Rgui (start)' menu command (Figure SOP 11.22).
 - g. In the user input section at the top of the 'DFplot.r' file, edit 'path' and 'outpath' input lines to specify the full path to the input file 'DFpoly_YYYY.txt' and the full path for the output file 'DFplot_YYYY.png', where YYYY is the survey year. Input lines also allow specification of text on the x- and y-axes.
 - h. Other features of the plot (e.g., font size, line width, image type for output, etc.) can be modified below the user-input area using the R computing language.
 - i. After making adjustments, executing the R code using the 'R|Send|File (echo=TRUE)' menu command or corresponding menu button will create an

image file containing a plot of the detection function along with a histogram showing the frequency of observations of murrelet groups in each distance category (Figure SOP 11.22). The plot will also be displayed in the R graphical output window.

2. Use WinZip or similar tool to assemble the KM_J deliverable.
 - a. Include the two .csv files containing output from Program DISTANCE created above (“DFstats_YYYY.csv” and “DFstats_status_YYYY.csv”) and one graphic of the detection function created above.
 - b. Name the final product
[\\inpglbafs03\data\SEAN_data\Work_zone\KM\KM_J\YYYY\KM_J_YYYY.zip](\\inpglbafs03\data\SEAN_data\Work_zone\KM\KM_J\YYYY\KM_J_YYYY.zip)

Formal Validation and Certification

The SEAN data management plan requires all monitoring deliverables be formally validated against rules defined in the protocol. When they have met all mandatory criteria, the Project Leader may declare them certified in writing. Once certified, the Data Manager must install them in repositories and attach them to electronic dissemination services.

Detailed Steps

A. Project Leader Tasks

1. Request validation by sending a notification email to the Data Manager, specifying in the message body that deliverable KM_J, for year YYYY is ready; the Data Manager will know where to find the file named “KM_J_YYYY.ZIP”.

B. Data Manager Tasks

1. On receipt of the submission, assign the next formal Submission Number to this file, as found in the master Submission_Log table.
 - a. Use the “Update Submission Log” web tool at http://165.83.57.239/0_submission_update.aspx.
 - b. Complete Submission_Log details up through the Submission_Date column.
2. Copy the file into the staging area for validation at a subdirectory named with the submission number in [\\inpglbafs03\data\SEAN_Data\Staging\KM\KM_J](\\inpglbafs03\data\SEAN_Data\Staging\KM\KM_J).
3. Validate the submission according to current criteria.
 - a. Invoke validation of the ZIP file according to current criteria by using the web program at http://165.83.57.239/KM_DM_validate_KMJ.aspx.
 - b. Validation summary results automatically are recorded in the Submission_Log.
4. If submission fails any mandatory criteria, reply with a “failure email” that includes:
 - a. The submission number assigned
 - b. The deliverable ID
 - c. The protocol ID
 - d. Documentation listing all the criteria failed
5. If submission passes mandatory criteria, reply with a “success email” that includes:
 - a. The submission number assigned

- b. The deliverable ID
- c. The protocol ID
- d. Documentation listing any specific optional criteria failed
- e. Request to certify deliverable as complete

C. Project Leader Tasks

1. On receipt of a failure email:
 - a. Review the validation report and remedy the mandatory errors by revising the analysis.
 - b. Initiate a new submission with the Data Manager.
2. On receipt of a success email, review any failed optional criteria:
 - a. If the criteria review indicates the validated data are acceptable:
 - i. Reply with a "certification email" stating the deliverable is certified and may be disseminated.
 - b. If the validated data are unacceptable:
 - i. Reply with a "withdrawal email", stating the deliverable is withdrawn.
 - ii. Take remedial action to obtain and resubmit a corrected deliverable, if possible.

D. Data Manager Tasks

1. On receipt of a withdrawal email:
 - a. Mark the withdrawal in the Submission_Log's Status column using the web tool.
 - b. Terminate the process.
2. On receipt of a certification email:
 - a. Verify no sensitive information is in the deliverable. Products containing sensitive information cannot be disseminated. If sensitive:
 - i. Mark the sensitive designation in the log's submission status column using the web tool.
 - ii. Notify the submitter of the situation.
 - iii. Complete the deliverable tracking spreadsheet cells referring to dissemination with the word SENSITIVE.
 - iv. Do not attempt to install the sensitive item in the auxiliary repository.
 - b. Copy the submitted file to the integration environment at
AUXREP\KM\KM_J\
 - c. Create an appropriate web link accessible from the KM main page.
 - d. Test accessibility until successful, then propagate from integration to production environment.
 - e. Verify deliverable is accessible from production web site.
 - f. Mark the certification in the Status column in Submission_Log using the web tool
 - g. Update the deliverable tracking spreadsheet milestones with the date of completion for KM_J.

- h. Update the scope of the formal metadata so it includes this new date range.
- i. Update the database to reflect the new or revised status and trends values by invoking the web program at http://165.83.57.239/KM_DM_update_KME_with_KMJ.aspx which accomplishes:
 - i. Importing the estimated detection function and estimates of group size for population trend analyses.
 - 1. The DFstats file exported from DISTANCE to “[\\inpglbafs03\data\SEAN_data\Work_zone\KM\KM_J\yyyy\DFstats_yyyy.txt](#)”, where yyyy is the survey year, is a space-delimited text file with 10 columns and a variable number of rows.
 - 2. The content in each row is coded by values in rows 1,4, and 5 (Table SOP 11.3).
 - 3. The right-truncation distance should be parsed from the text file.
 - 4. The following estimated parameters along with estimated coefficients of variation (CV) and degrees of freedom (df) should be parsed from the text file:
 - i. Detection probability within right-truncation distance
 - ii. Effective strip width
 - iii. Parameter estimates for detection function
 - 5. For each of the above estimated parameters, the standard error should be calculated as the product of the parameter estimate and its CV.
 - 6. The rows for the key function and series expansion provide integer values that correspond to the combination of key function + series expansion used in the estimated detection function (Table SOP 11.3). A series expansion is not always present.
 - 7. The number of estimated parameters m for the detection function varies and will be the sum of the # of key function parameters and the # of series expansion parameters. The coded values in column 5 for these estimates will range from 101 to 101 + (m -1). The key function parameter estimates are provided before those for the series expansion.
 - 8. All estimated parameters from the detection function should be parsed from the text file. Parameter estimates for the key function should be labeled in sequence prefixed by an ‘A’. Parameter estimates from the series expansion should be labeled in sequence prefixed by a ‘B’.

9. Estimates of average group size for each species class are for analyses of population trend.
- ii. Importing the estimates of group size for population status.
 1. Use the DFstats_status file, a space-delimited text file with 10 columns and a variable number of rows.
 2. The content is coded identically to the 'DFstats' text file (Table SOP 11.3).
 3. Only the estimates related to average group size for each species class should be parsed from this file and imported to the KM_E database.

Literature Cited

- Buckland, S. T., D. R. Anderson, K. P. Burnham, J. L. Laake, D. L. Borchers, and L. Thomas. 2001. Introduction to Distance Sampling: Estimating Abundance of Biological Populations. Oxford, New York, New York, USA.
- Buckland, S. T., D. R. Anderson, K. P. Burnham, J. L. Laake, D. L. Borchers, and L. Thomas. 2004. Advanced Distance Sampling: Estimating Abundance of Biological Populations. Oxford, New York, New York, USA.
- Burnham, K. P., and D. R. Anderson. 2002. Model Selection and Multi-model Inference, second edition. Springer, New York, New York, USA.
- Hoekman, S. T., B. J. Moynahan, and M. S. Lindberg. 2011a. Monitoring Kittlitz's and marbled murrelets in Glacier Bay National Park: 2009 annual report. Natural Resource Technical Report NPS/SEAN/NRTR—2011/440. National Park Service, Fort Collins, Colorado, USA. Retrieved October 17, 2013, from http://science.nature.nps.gov/im/units/sean/AuxRep/KM/KM_F/KM_2009_Annual_report.pdf
- Hoekman, S. T., B. J. Moynahan, and M. S. Lindberg. 2011b. Monitoring Kittlitz's and marbled murrelets in Glacier Bay National Park: 2010 annual report. Natural Resource Technical Report NPS/SEAN/NRTR—2011/441. National Park Service, Fort Collins, Colorado, USA. Retrieved October 17, 2013, from http://science.nature.nps.gov/im/units/sean/AuxRep/KM/KM_F/KM_2010_Annual_report.pdf
- Thomas, L., S. T. Buckland, E. A. Rexstad, J. L. Laake, S. Strindberg, S. L. Hedley, J. R. B. Bishop, T. A. Marques, and K. P. Burnham. 2010. Distance software: Design and analysis of distance sampling surveys for estimating population size. *Journal of Applied Ecology* 47:5–14.

SOP 11 Figures and Tables

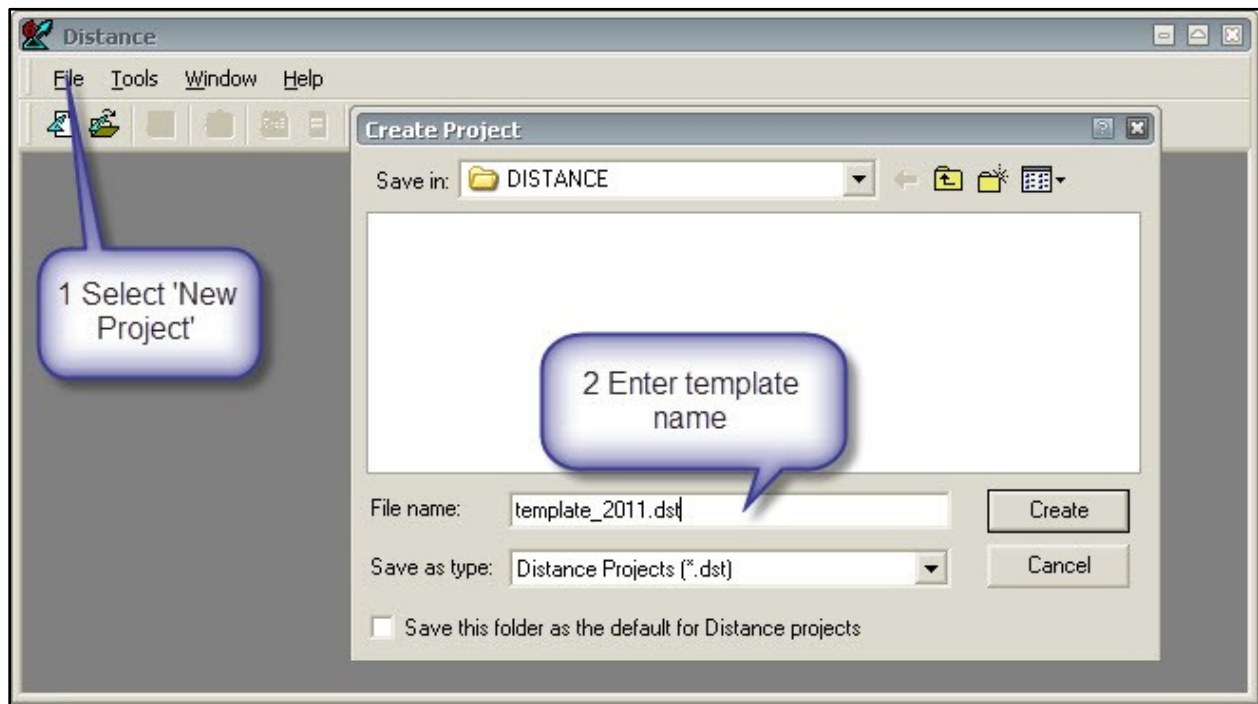


Figure SOP 11.1. Creation of a template project file in Program DISTANCE.

New Project Setup Wizard - Step 1: Type of Project 1

The new project 'template_2011' has been created.

Distance can now help you to set up the project ready for use. Select from the options below, and click on the 'Next' button to step through each of the screens that follow.

If you are not sure what to enter at any point, click on the 'Help' button or press F1 to find out more about the options.

I want to:

- ☒ Analyze a survey that has been completed

New Project Setup Wizard - Step 3: Survey Methods 2

In this screen, you tell Distance about your survey methods. Click 'Help' to find out more about each option.

Type of survey:

- ☒ Line transect
- ☐ Point transect
- ☐ Cue count

Observer configuration:

- ☒ Single observer
- ☐ Double observer

Distance measurements:

- ☐ Perpendicular distance
- ☒ Radial distance and angle

Observations:

- ☒ Single objects
- ☐ Clusters of objects

Sampling fraction: This option has been moved to the Multipliers page.

New Project Setup Wizard - Step 4: Measurement Units 3

Please specify the measurement units for your data.

If you want to analyze the data using different units, you can do so after completing this wizard (in the Units tab of the Data Filter). Click 'Help' for more information.

Units of original measurements:

- Distance:
- Transect:
- Area:

New Project Setup Wizard - Step 5: Multipliers 4

Multipliers are constants that are used to scale the final density estimate. Click 'Help' to find out more about multipliers.

Tick the boxes beside the multipliers you wish to add. Distance will then create fields in your dataset for you to enter the multiplier value and, if appropriate, its standard error (SE) and degrees of freedom (DF). You enter these values later, when you are entering the rest of your data.

Add multipliers for:	Fields added	Create fields for SE and DF?
☐ Surveys where sampling fraction is not 1	Sampling fraction	N
☐ Surveys where g(0) is less than 1	g0	Y
☐ Cue count surveys	Cue rate	Y
☐ Indirect surveys of	Production rate	Y
	Disappearance time	Y
☐ Other	Generic multiplier	Y

New Project Setup Wizard - Step 6: Finished 5

When you press Finish, the new project database will set up as you have specified. The next task after this is to add your survey data. Please choose from one of the options in the 'Destinations' box.

The Data Entry Wizard will guide you through the process of entering your data from the keyboard. The Data Import Wizard will allow you to import the data from an external text file. If you choose to exit to Distance, you can enter data in the Data tab of the Project Browser, or start either of the Wizards from the Tools menu.

Destinations:

- ☐ Proceed to Data Entry Wizard
- ☒ Proceed to Data Import Wizard
- ☐ Exit wizards and return to Distance

☐ Save current settings as default

Buttons: Help, Cancel, < Back, Next >, Finish

Figure SOP 11.2. Selection of appropriate settings using the 'New Project Setup Wizard' during creation of a DISTANCE project template file. Gold boxes indicate the sequence of steps; red arrows highlight user input.

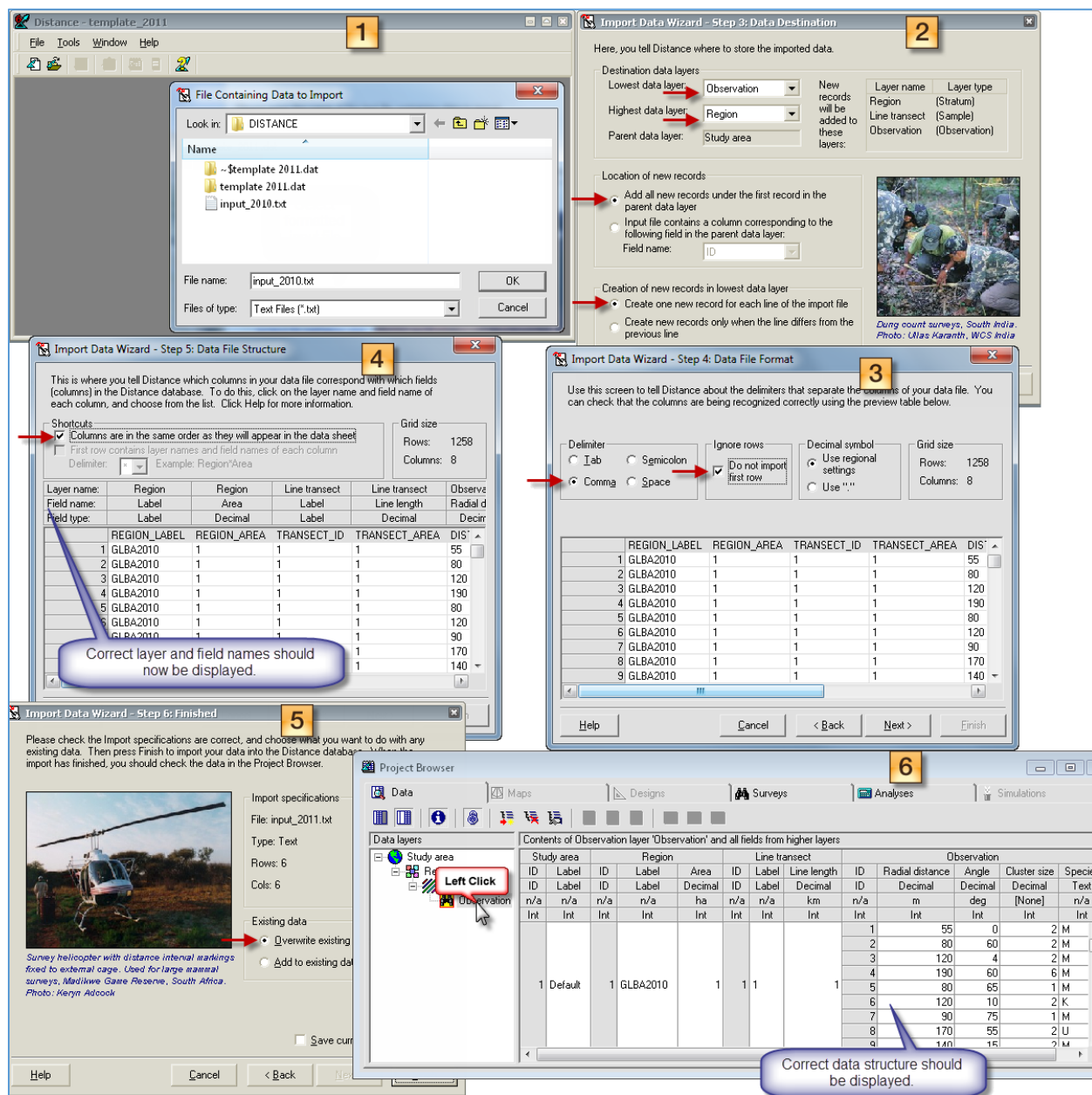


Figure SOP 11.3. The 'Import Data Wizard' of DISTANCE requires an appropriately formatted text file containing survey observation data. Successive steps of the Wizard allow specification of data structure. Upon completion of the Wizard, clicking on the 'Observation' data layer in the 'Data' tab of the 'Project Browser' window displays the imported data. Gold boxes indicate the sequence of steps; red arrows highlight user input.

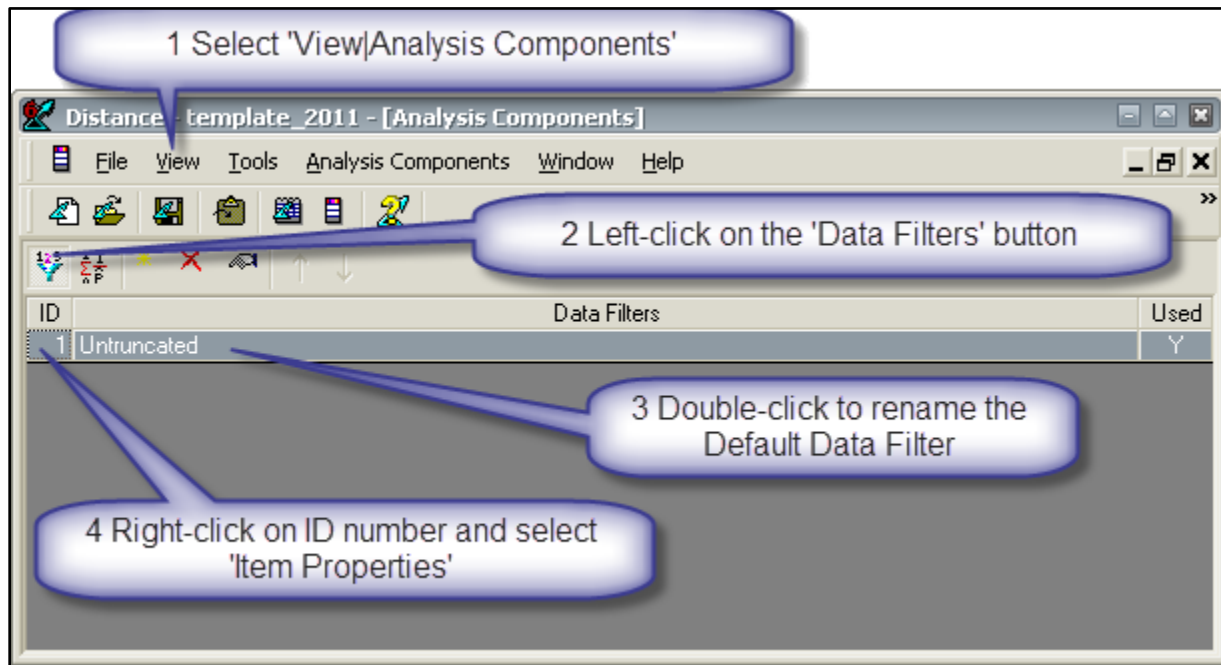


Figure SOP 11.4. Creation of a 'Data Filter' in the 'Analysis Components' window of DISTANCE. Selecting 'Item Properties' opens the 'Data Filter Properties' window shown in Figure SOP 11.5.

The figure displays four screenshots of the 'Data Filter Properties: [Untruncated]' window, showing the settings for the 'Untruncated' data filter across its four tabs: Data selection, Distance intervals, Truncation, and Units of analysis.

- Data selection tab:** The 'Layer type' and 'Selection criteria' fields are empty. A callout bubble points to the 'Selection criteria' field with the text 'Leave blank'. The 'Selection criteria' section provides examples: 'Label = 'Region 1'', 'Species = 'W/WA' AND Year = 2006', and '[Cluster size] >= 3 AND Beaufort IN (0, 1, 2, 3)'. The 'Name' field is set to 'Untruncated'.
- Distance intervals tab:** The 'Transform distance data into intervals for analysis' checkbox is unchecked. A callout bubble points to this checkbox with the text 'Leave unchecked'. The 'Number of intervals' is set to 1. The 'Interval output' is set to 'Manual'. The 'Cutpoints' table shows values 0 and 1 for the first two intervals.
- Truncation tab:** The 'Right truncation' section has 'Right truncate at largest observed distance' selected. The 'Left truncation' section has 'No left truncation' selected. The 'Truncation for cluster size estimation (where required)' section has 'Same as that specified above' selected. Red arrows point to these three selected options.
- Units of analysis tab:** The 'Distance' is set to '[Same as distance data]'. The 'Angle' is set to 'Degrees'. The 'Length' is set to '[Same as length data]'. The 'Area' is set to '[Same as area data]'. A callout bubble points to the bottom of the tab with the text 'Leave all at default values'.

Figure SOP 11.5. The 4 tabs of the 'Data Filter Properties' window, showing appropriate settings for the "Untruncated" data filter. Red arrows highlight used input. Left-clicking 'OK' returns to the 'Analysis Components' window.

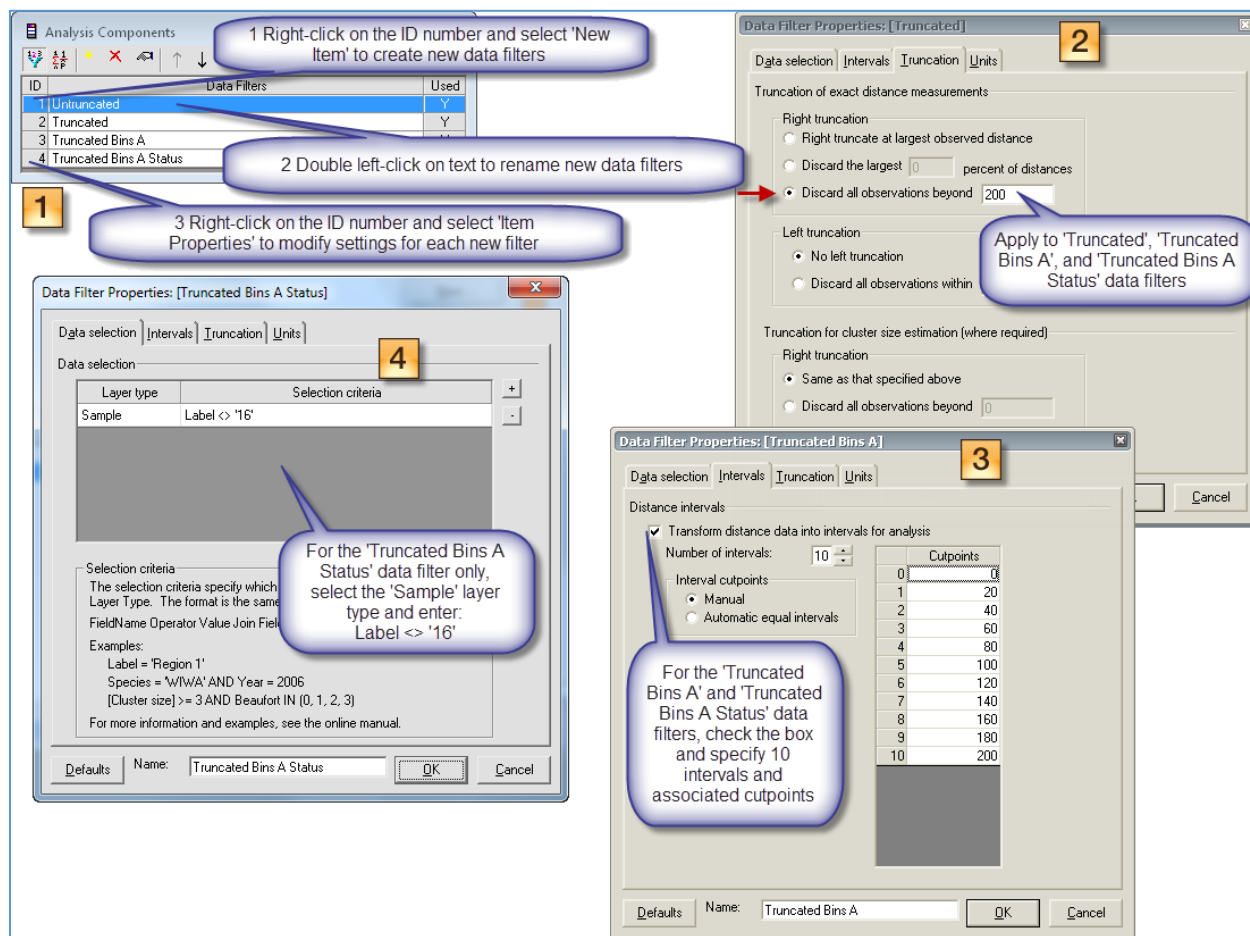


Figure SOP 11.6. Creation of additional data filters in the 'Analysis Components' window of DISTANCE. Step 2 shows settings changes applied to both the "Truncated" and "Truncated Bins A" data filters on the 'Truncation' tab of the 'Data Filters Properties' window. Step 3 shows settings changes applied only to the "Truncated Bins A" data filter in the 'Intervals' tab of the 'Data Filter Properties' window.

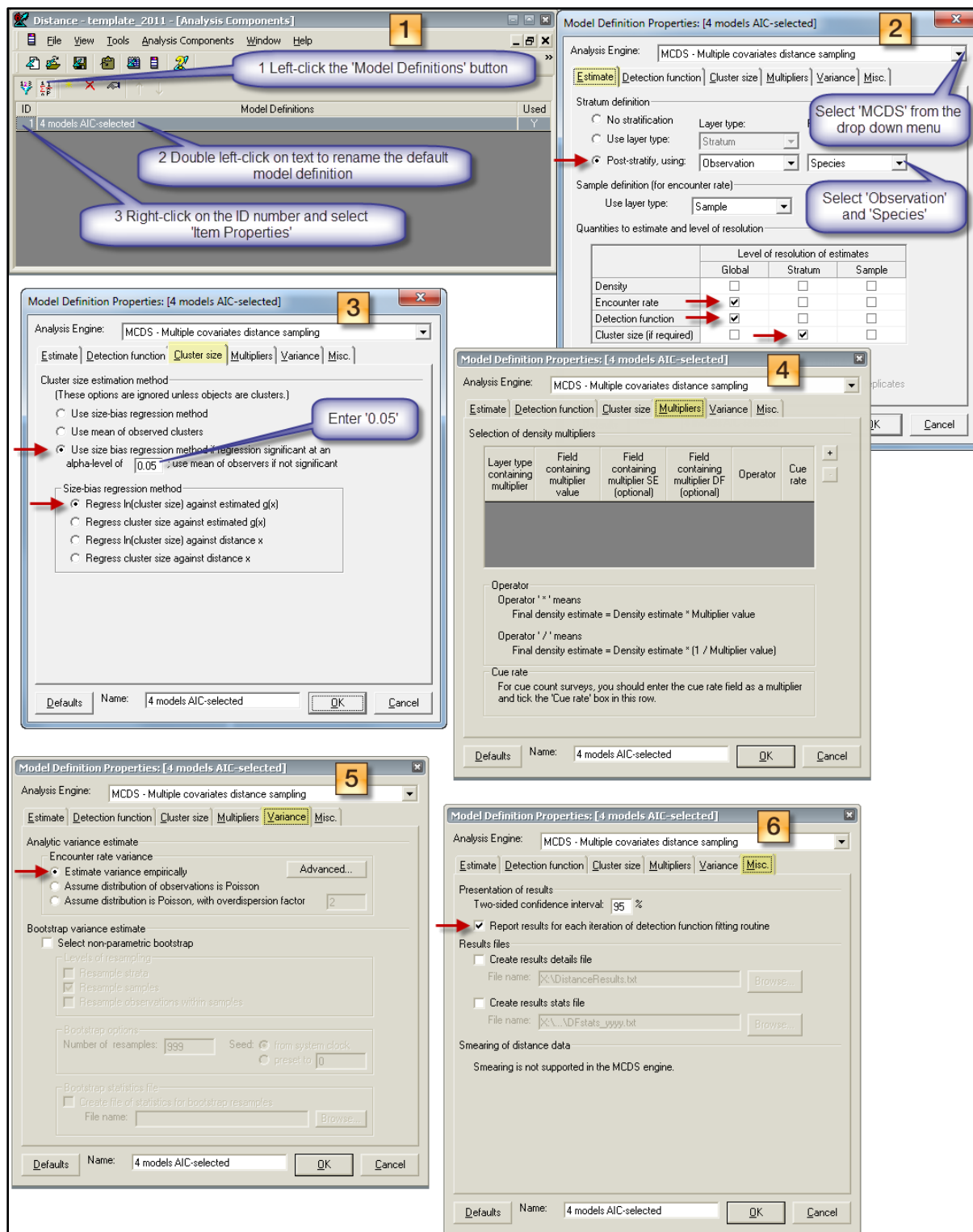


Figure SOP 11.7. Creating and defining the 'model definition' named "4 model AIC-selected". In step 1, the 'model definition' is named in the 'Analysis Components' window. Selecting 'Item Properties' opens the 'Model Definition Properties' window. Steps 2-6 show settings in all tabs except 'Detection function', which is shown in Figure 11.8. Red arrows highlight user input.

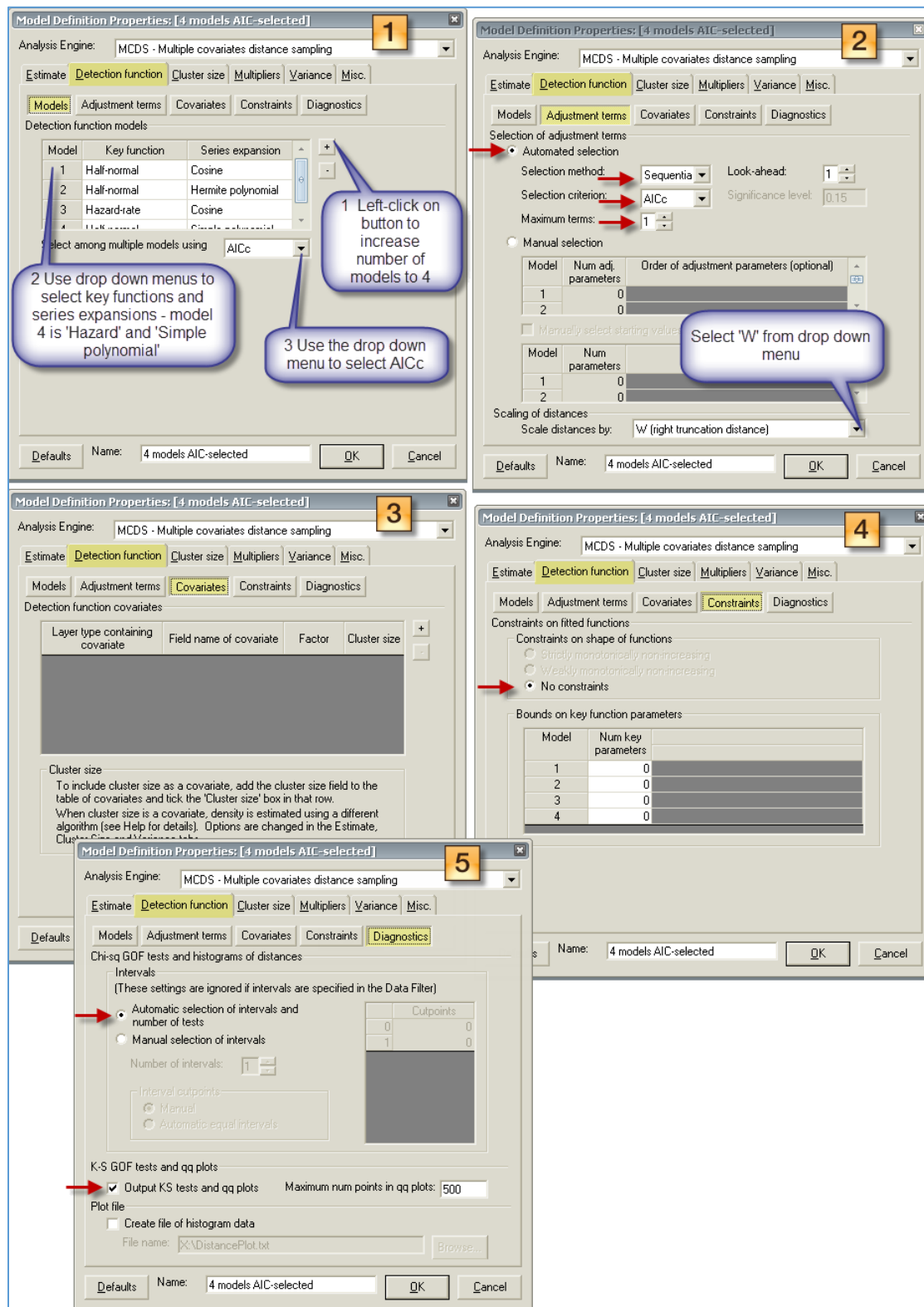


Figure SOP 11.8. Settings for the “4 models AIC-selected” model definition within the ‘Detection function’ tab of the “Model Definitions Properties” window. This tab has 5 separate lower-level tabs, each shown in a separate window. Red arrows highlight user input.

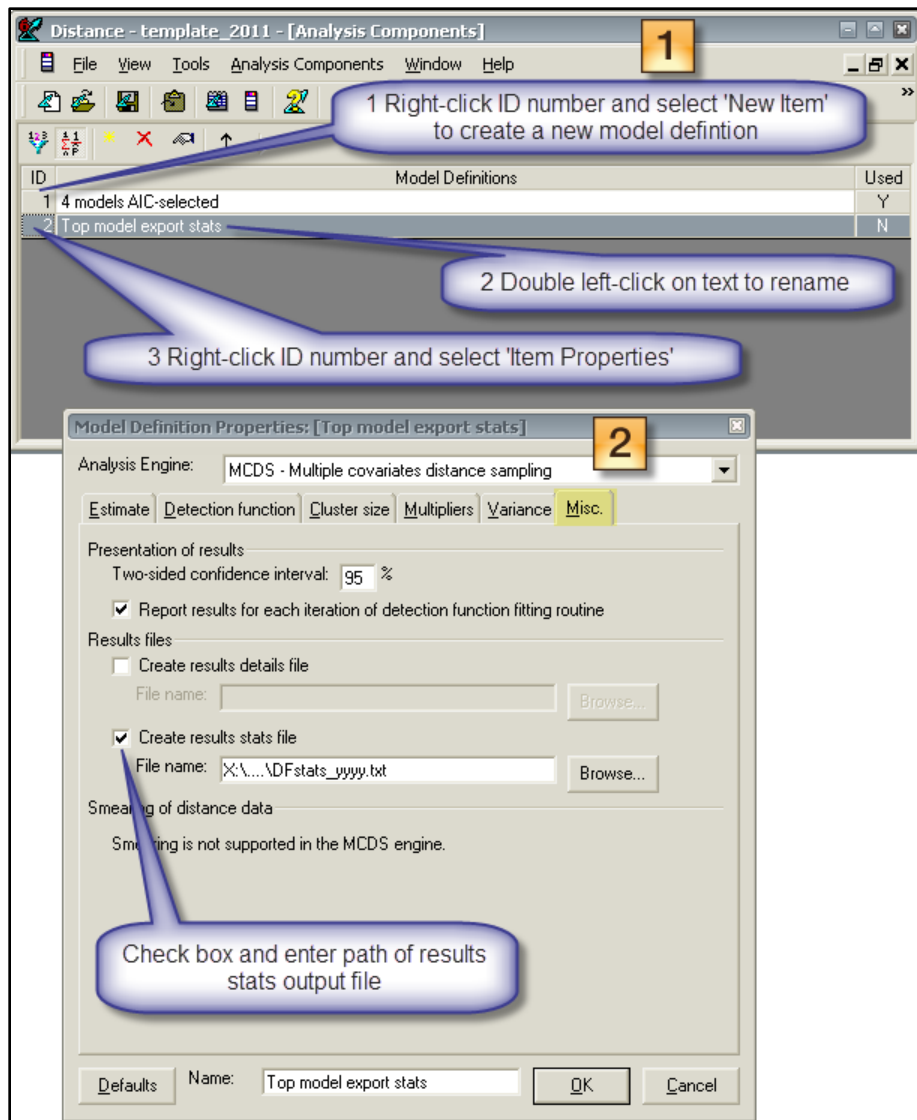


Figure SOP 11.9. Creating and naming the “Top model export stats” model definition in the ‘Analysis Components’ window. Settings for this model definition match those for the “4 models AIC-selected” model definition, except for the creation of a results stats output file. This file’s output path is specified in the ‘Misc.’ tab of the ‘Model Definition Properties’ window.

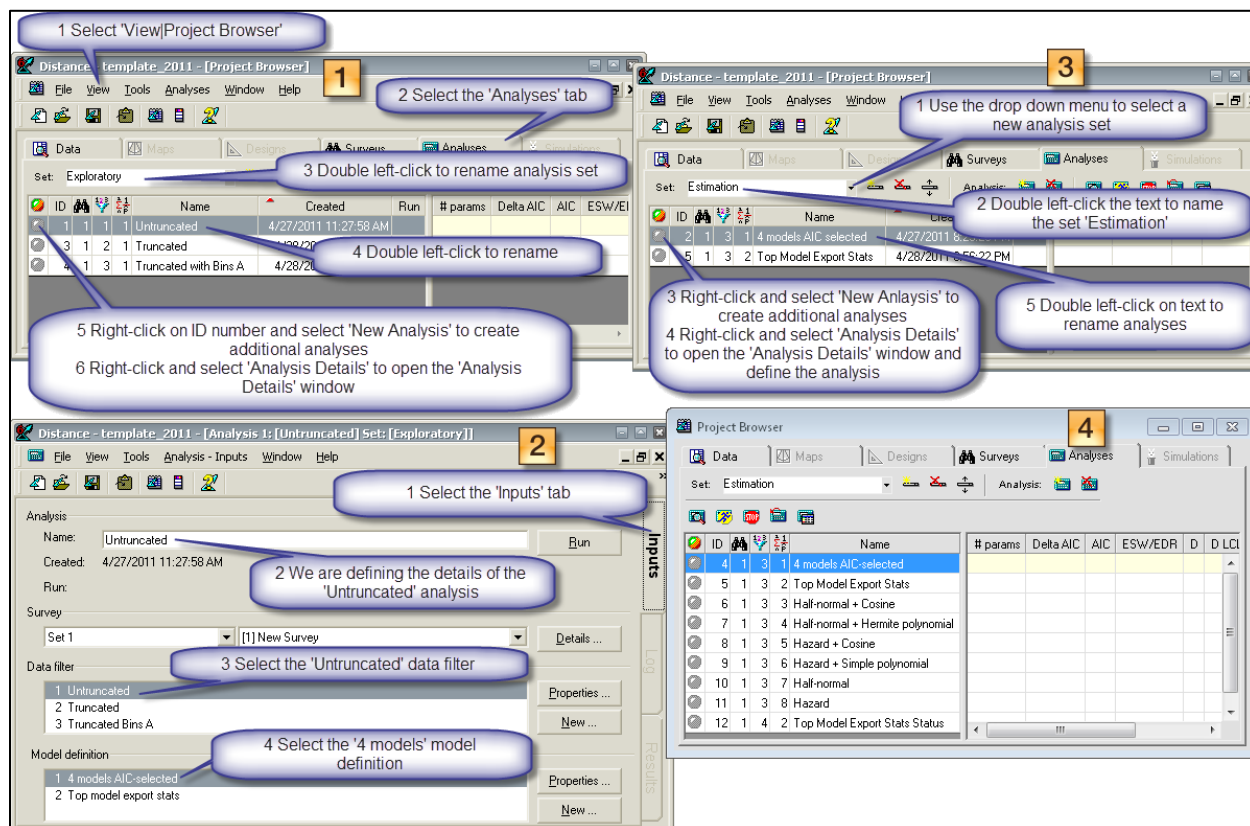


Figure SOP 11.10. Creating and specifying pre-defined analysis sets and analyses in the 'Analysis Components' window. The analysis sets and the names, data filters, and model definitions of analyses are listed in Table SOP 11.2.

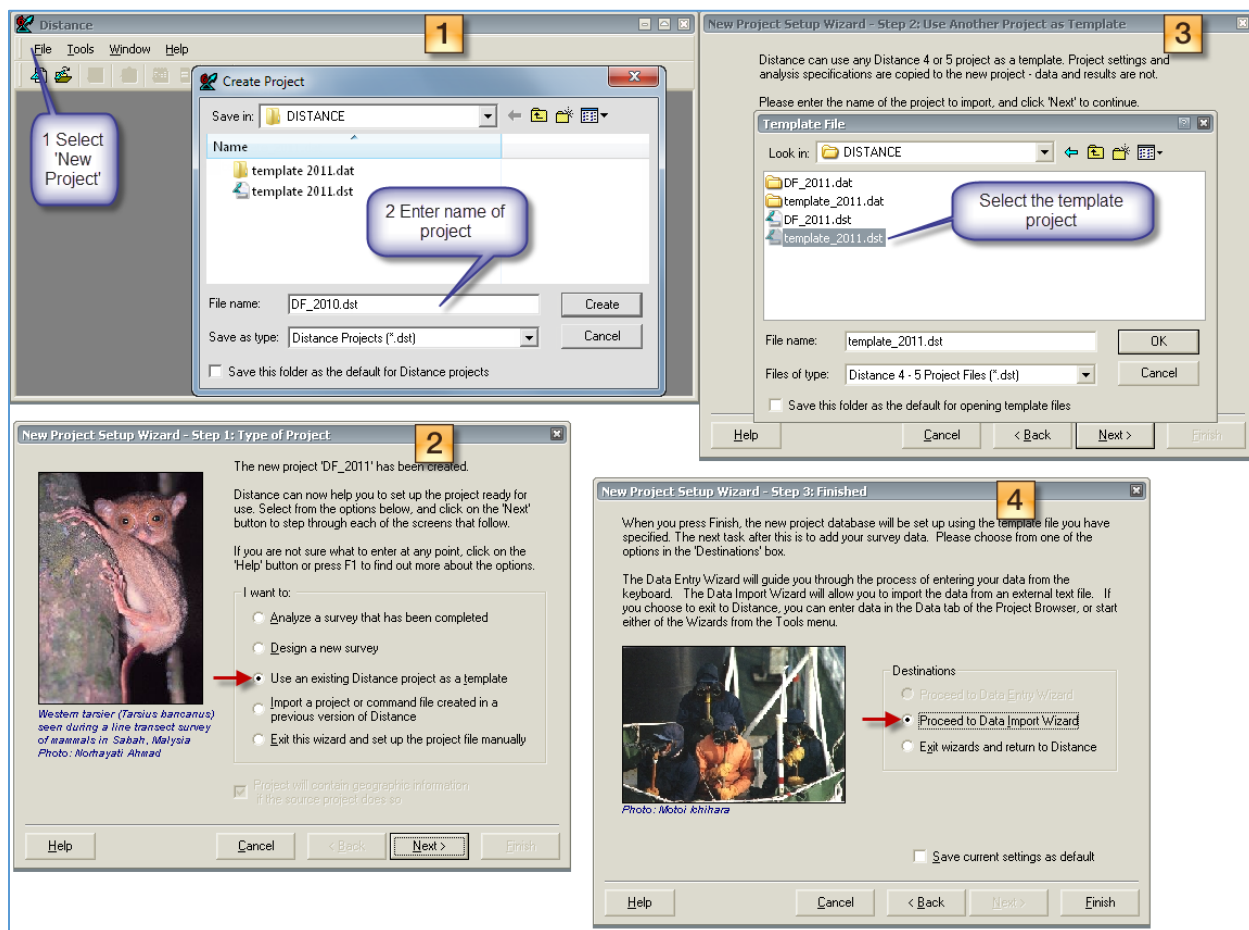


Figure SOP 11.11. Creation of a new project in the 'New Project Setup Wizard' of DISTANCE using a template project file to import pre-defined settings and analyses.

Glacier Bay National Park and Preserve Kittlitz's murrelet monitoring protocol version KM-2012.1
SOP 11 copied verbatim from original Natural Resource Report NPS/SEAN/NRR—2013/735

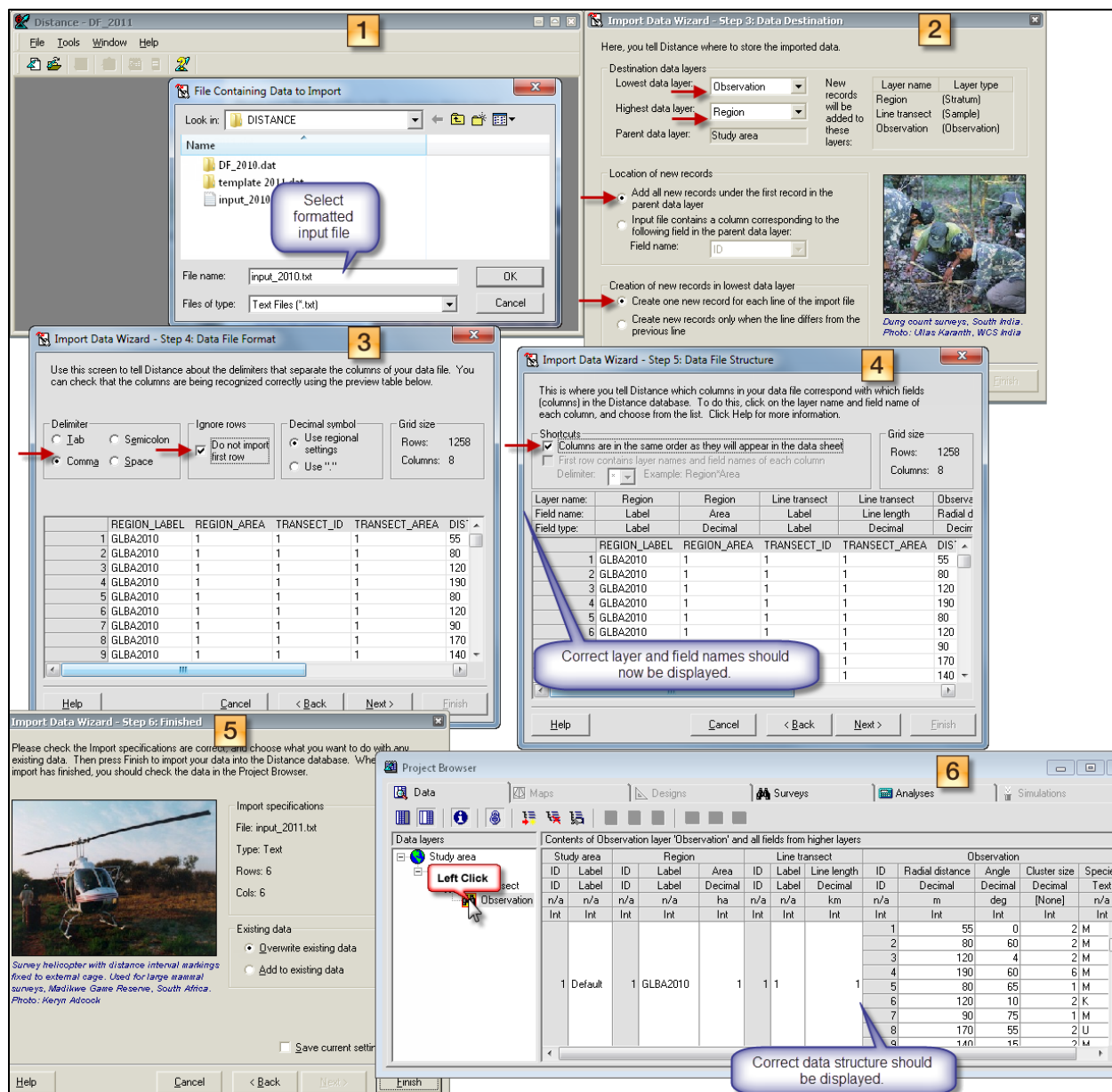


Figure SOP 11.12. Importing a formatted text file containing survey observation data using the 'Import Data Wizard'. Gold boxes indicate the sequence of steps; red arrows highlight user input. Upon completion, selecting the 'Observation' data layer in the 'Data' tab of the 'Project Browser' should display the imported data with correct layer names, field names, and data structure.

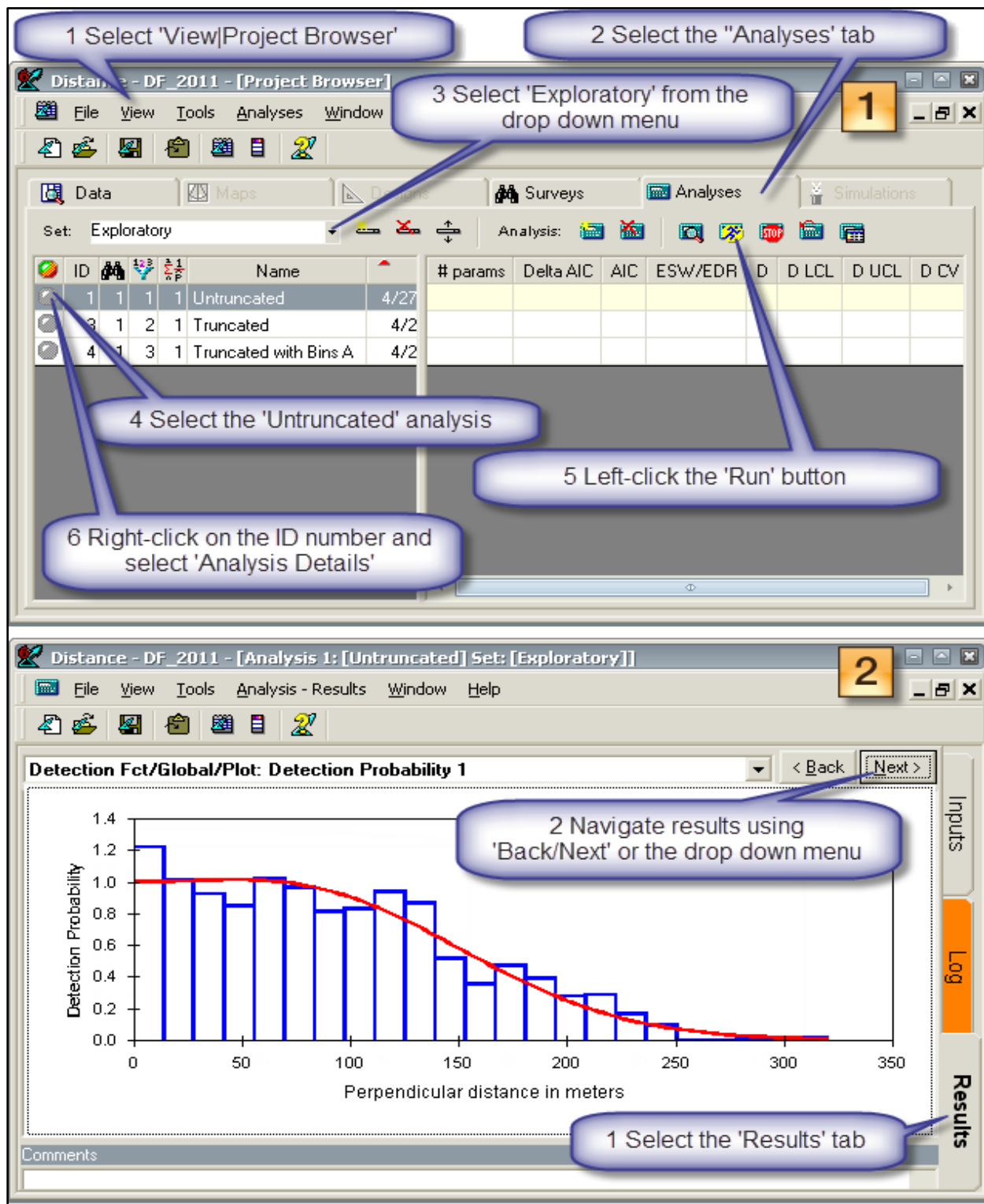


Figure SOP 11.13. Displaying, executing, and displaying output from pre-defined analyses in the 'Exploratory' set.

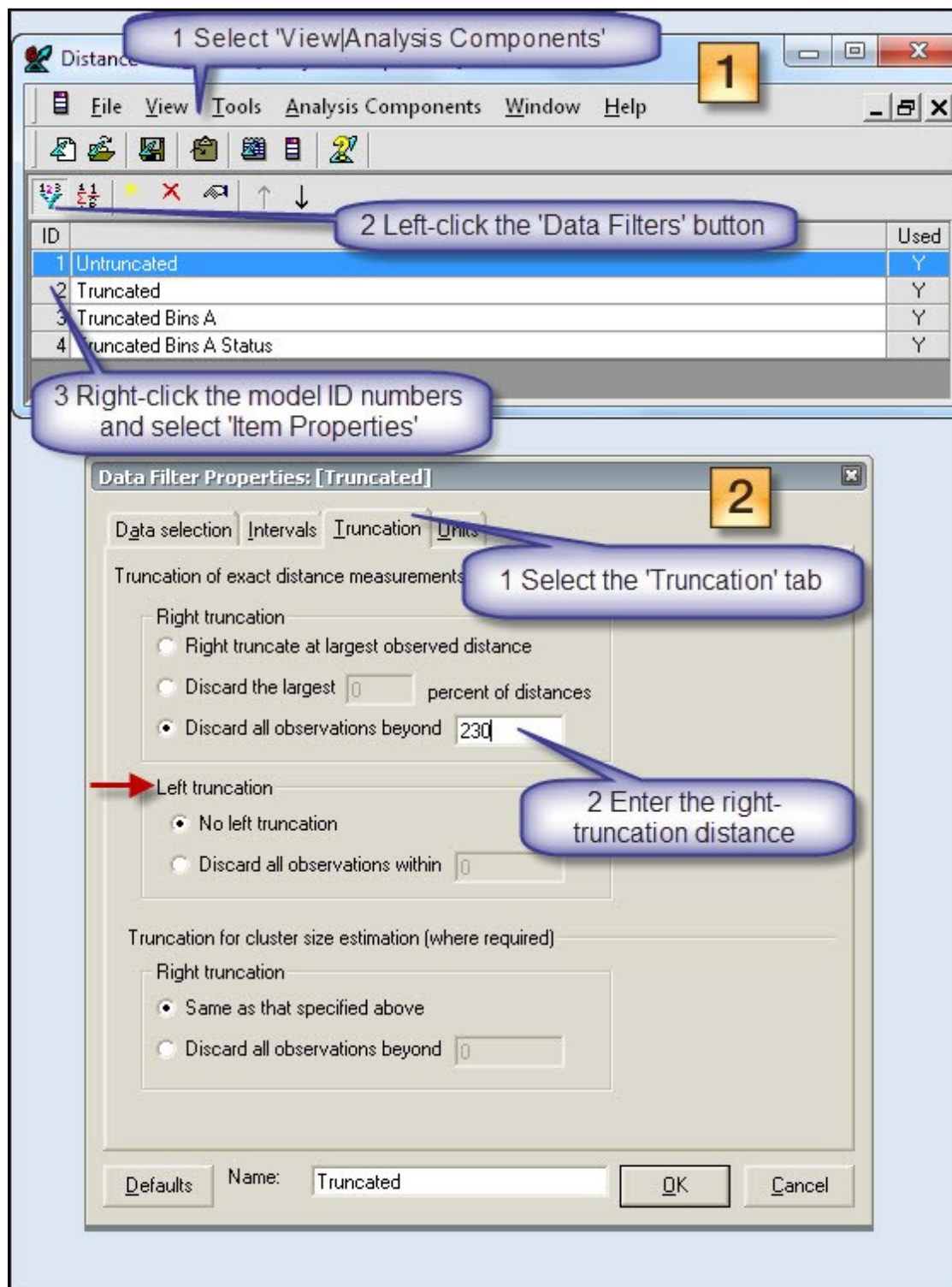


Figure SOP 11.14. Setting the right-truncation distance in the 'Data Filter properties' window. This setting should be applied to all 'Truncated' data filters.

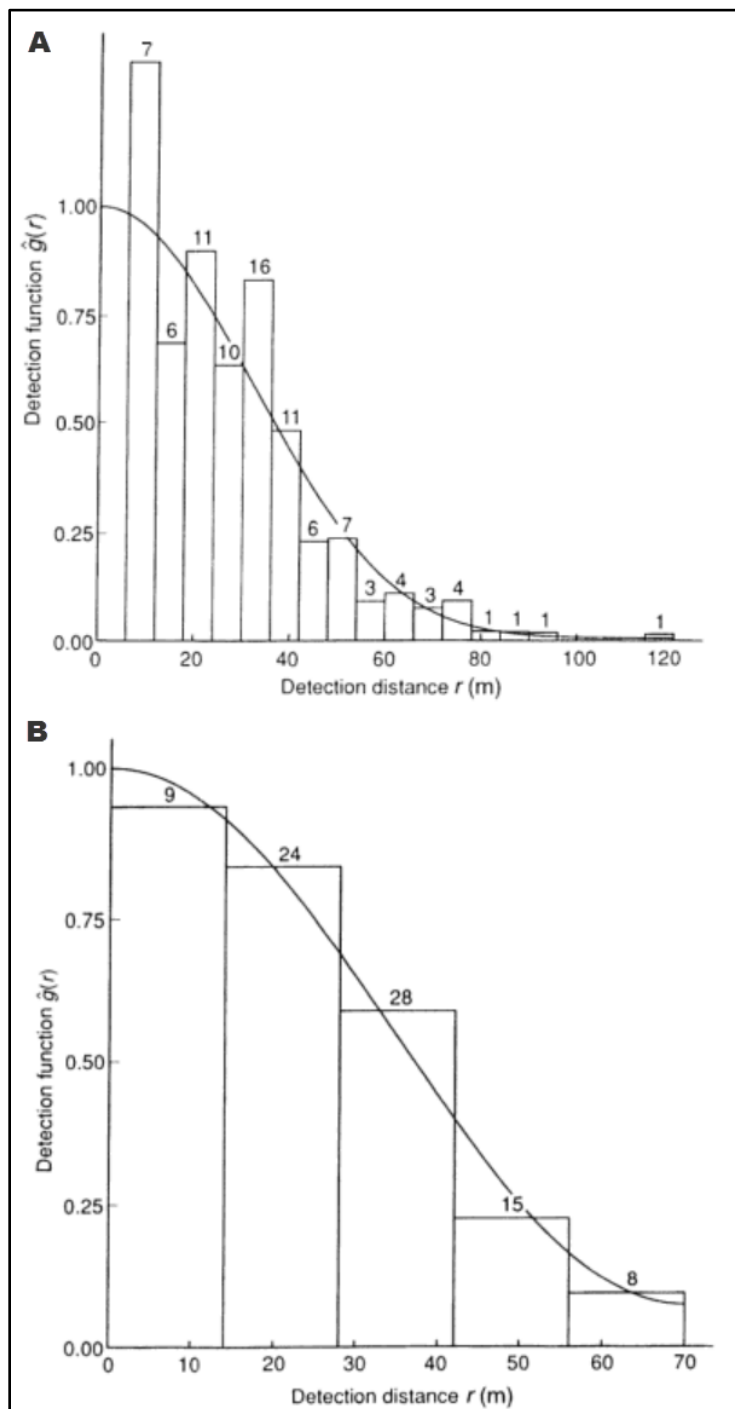


Figure SOP 11.15. Histograms (bars) showing frequency of detections (y-axis) relative to perpendicular distance from the transect center line (x-axis). Frequencies are scaled relative to estimated detection functions (lines); number of detections shown above bars. Untruncated data in relatively narrow bins (A) shows evidence of 'heaping' of observations at ~10, 20, and 30 m. Right-truncating and binning data by placing cut-points at mid-points between peaks (B) mitigates deleterious effects of heaping and helps meet shape criteria (Buckland et al. 2001).

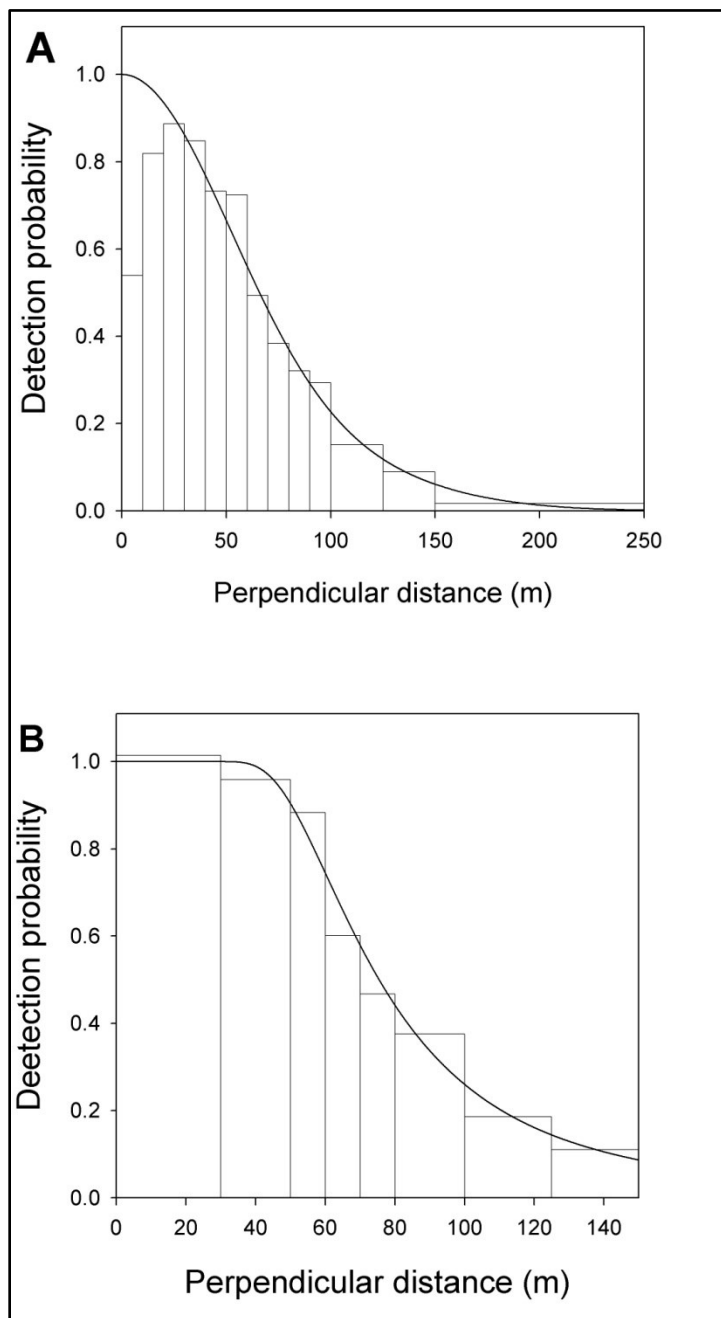


Figure SOP 11.16. Histograms (bars) showing frequency of detections (*y*-axis) relative to perpendicular distance from the transect center line (*x*-axis). Frequencies are scaled relative to estimated detection functions (lines). Untruncated data in relatively narrow bins (A) shows evidence of evasive movement (increasing frequencies from 0 to ~25 m). Right-truncating and binning data by creating a wide bin that ‘captures’ evading groups in the lowest distance bin improves the distribution by creating a stronger “shoulder” (slope ~0 at small distances) in the data distribution (Buckland et al. 2001).

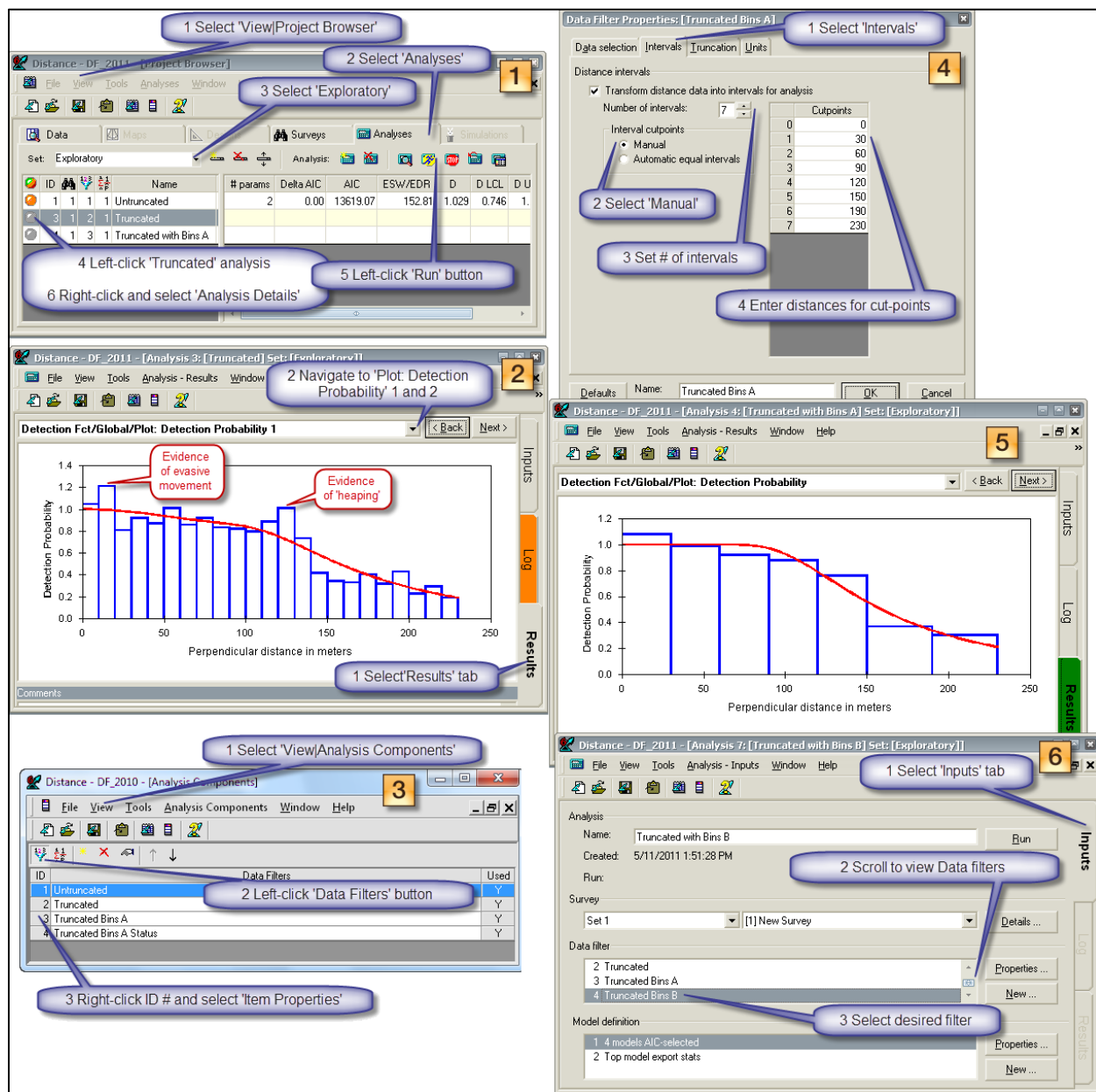


Figure SOP 11.17. Binning of survey observations by perpendicular distance from the transect center line. Distribution problems identified in exploratory analyses (1-2) may motivate binning data by distance for individual 'Data Filters' in the 'Analysis Components' window (3-4) to improve distributions (5). Alternative binning strategies can be compared by creating additional 'Data Filters' (6).

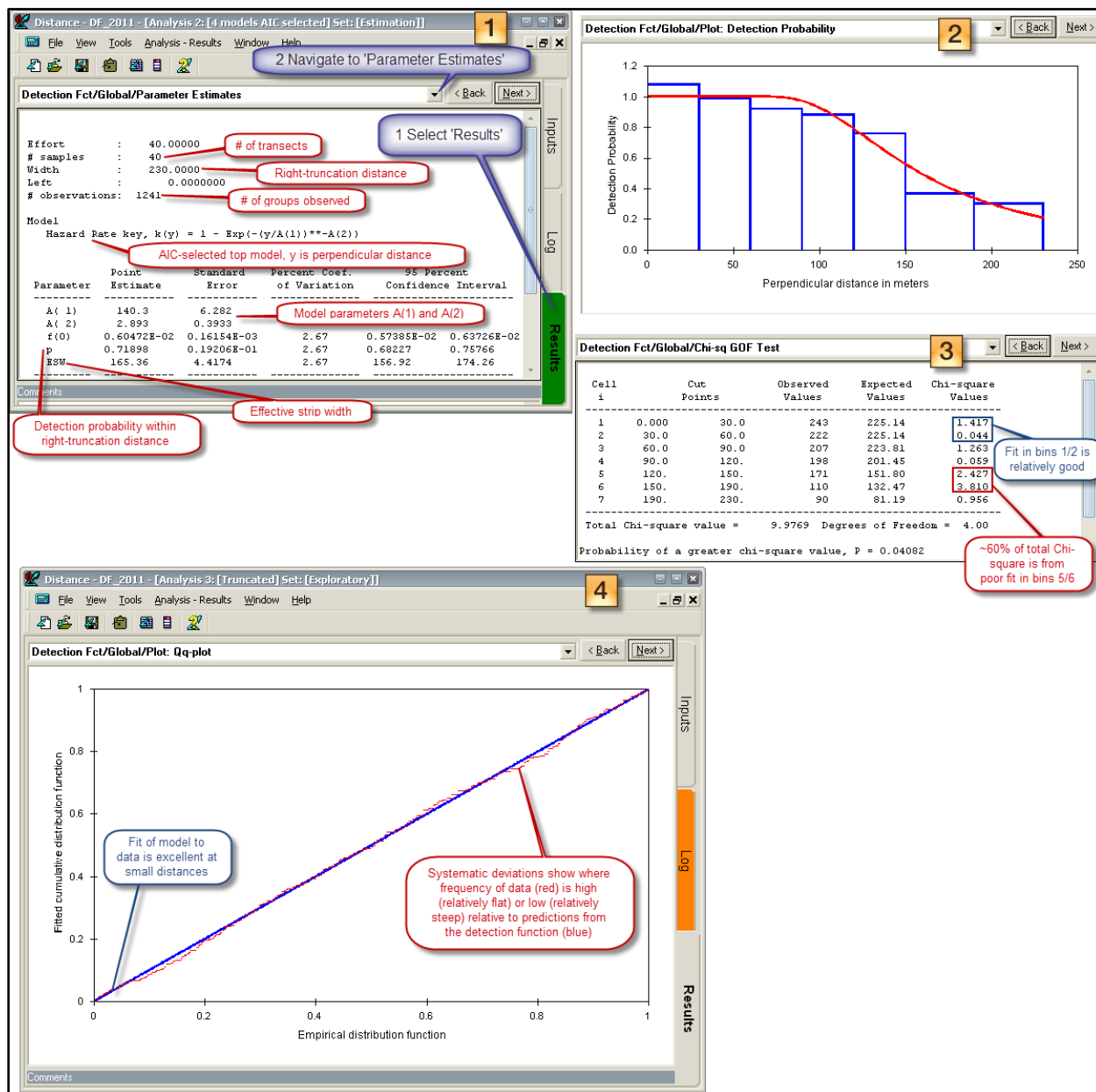


Figure SOP 11.18. Output from DISTANCE for an estimated detection function. The 'Parameter Estimates' window (1) shows samples, the AIC-selected top model, and quantitative output from this model. The 'Plot: Detection Probability' windows displays frequency data (blue) and the estimated detection function (2), and the 'Chi-sq GOF Test' window (3) contains the associated χ^2 goodness-of-fit test. For un-binned data, the q-q plot (4) displays of goodness-of-fit.

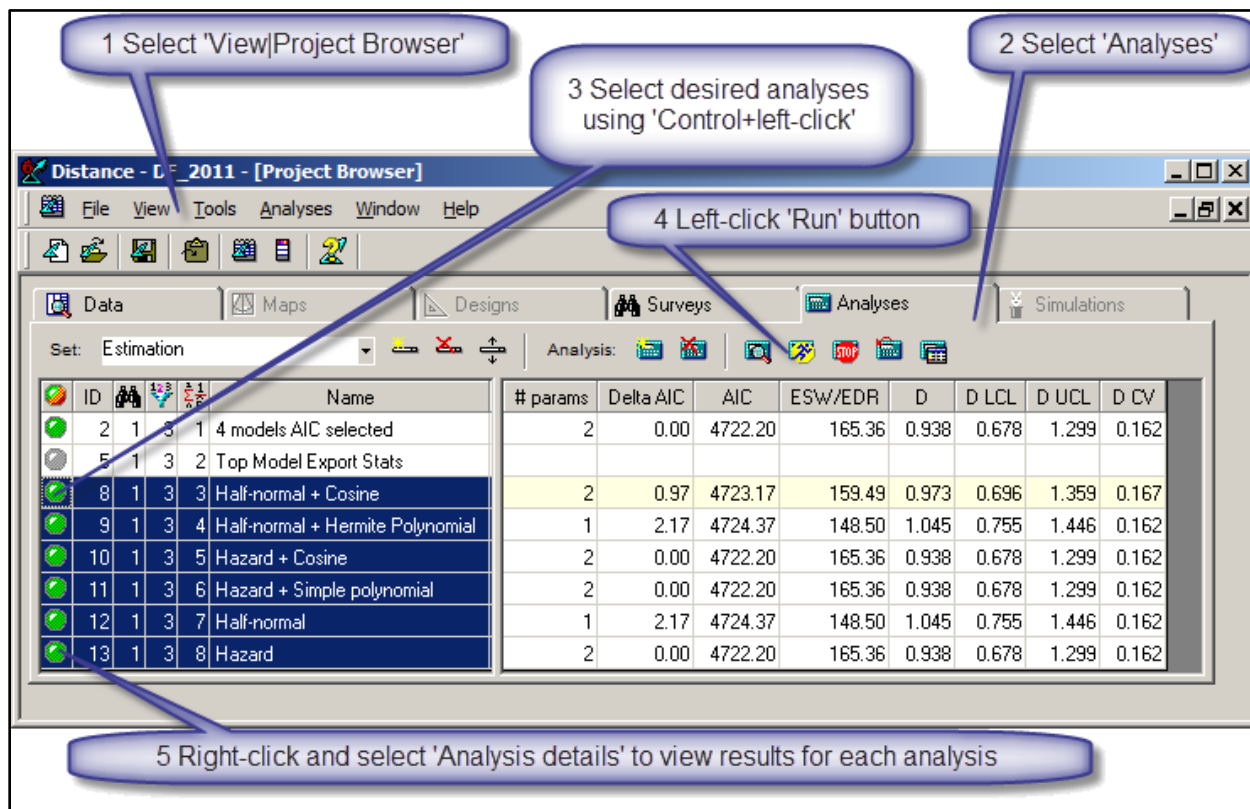


Figure SOP 11.19. Selecting, executing, and viewing results from analyses for individual detection function models. The 'Delta AIC' column gives differences in AIC values relative to the top model in the set (i.e., identical data and binning strategy).

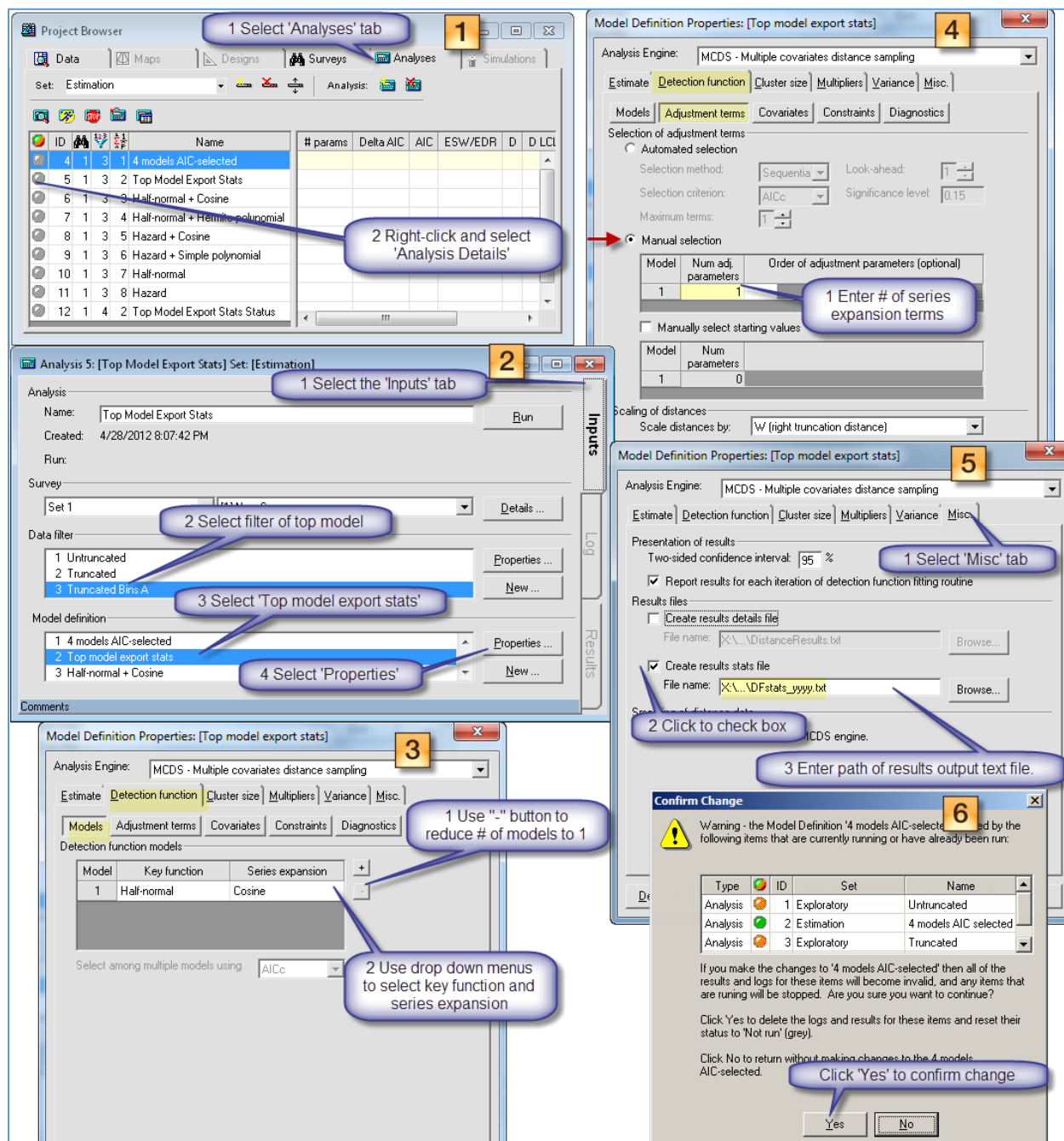


Figure SOP 11.20. Exporting the Program DISTANCE results stats file for the selected detection function model. After adding the path of the stats file to the 'Model Definition' for the analysis used to estimate the selected detection function model, execute the analysis to generate the file.

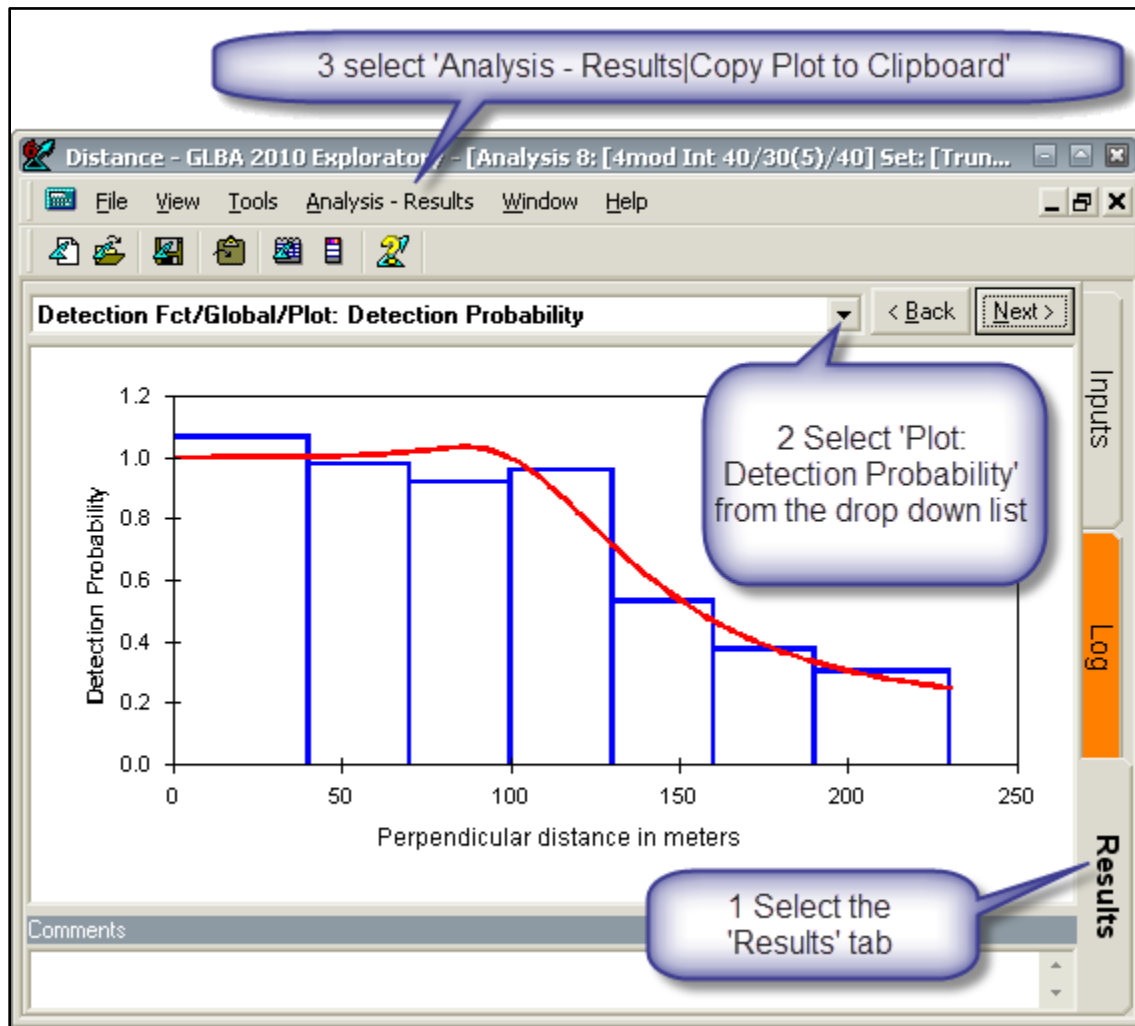


Figure SOP 11.21. Exporting a plot of a detection function from DISTANCE. The plot data are copied to the clipboard and then pasted into a text file.

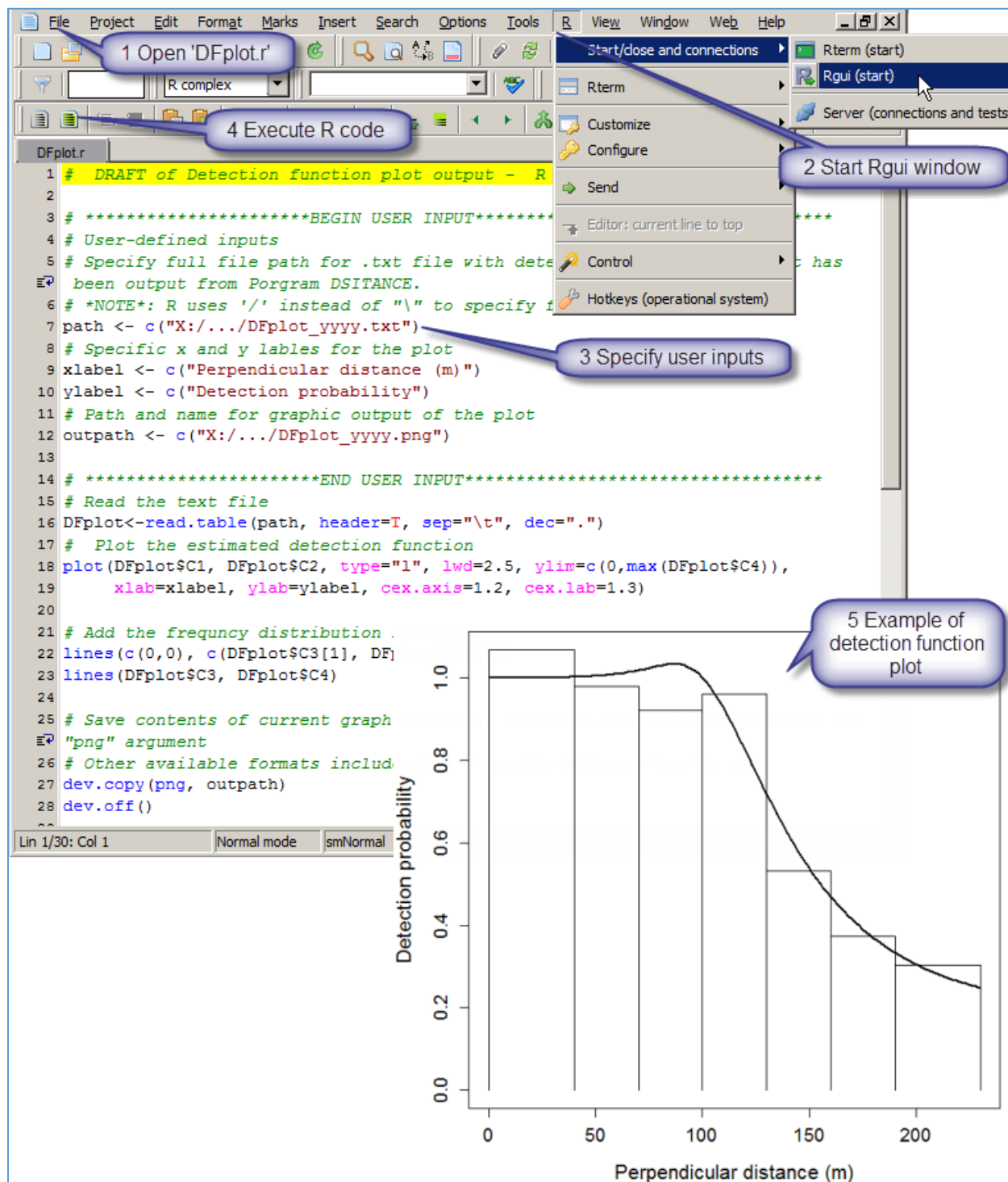


Figure SOP 11.22. Program R code used to produce an image of the estimated detection function and associated frequency data. Red text indicates user-specified inputs.

Table SOP 11.1. Data columns in input file of observation data for import to Program DISTANCE. Layer and field names in the first two columns specify hierarchical data naming structure applied within DISTANCE.

Program DISTANCE			Input file		
Column	Layer name	Field name	Data type	Field name	Repository field source or specified value
1	Region	Label	String	REGION_LABEL	GLBAyyyy ^a
2	Region	Area	Integer	REGION_AREA	1
3	Line transect	Label	Integer	TRANSECT_ID	TRANSECT_ID (Tbl_observation)
4	Line transect	Area	Integer	TRANSECT_AREA	1
5	Observation	Radial distance	Integer	DISTANCE	DISTANCE (Tbl_observation)
6	Observation	Angle	Integer	ANGLE	ANGLE (Tbl_observation) ^b
7	Observation	Size	Integer	GROUP_SIZE	GROUP_SIZE
8	Observation	Species	String	SPECIES	SPECIES (Tbl_observation) ^c

^aSurvey year is specified by yyyy.

^bAngle values recalculated as described at steps in A.3.f in the 'Data Import to DISTANCE' section.

^c"K" = KIMU, "M" = MAMU, "U" = UNMU.

Table SOP 11.2. Pre-defined DISTANCE analyses used in exploratory analyses, estimation of a detection function, and estimation of group size.

Analysis set	Analysis name	Right-truncation	bins	Data filter	Model definition
Exploratory	Untruncated	N	N	Untruncated	4 models AIC-selected
Exploratory	Truncated	Y	N	Truncated	4 models AIC-selected
Exploratory	Truncated with bins A	Y	Y	Truncated bins A	4 models AIC-selected
Estimation	4 models AIC selection	Y	Y	Truncated bins A	4 models AIC-selected
Estimation	Top Model Export Stats	Y	Y	Truncated bins A	Top model export stats
Estimation	Half-normal+Cosine	Y	Y	Truncated bins A	Half-normal+Cosine
Estimation	Half-normal+Hermite polynomial	Y	Y	Truncated bins A	Half-normal+Hermite polynomial
Estimation	Hazard+Cosine	Y	Y	Truncated bins A	Hazard+Cosine
Estimation	Hazard+Simple polynomial	Y	Y	Truncated bins A	Hazard+Simple polynomial
Estimation	Half-normal	Y	Y	Truncated bins A	Half-normal
Estimation	Hazard	Y	Y	Truncated bins A	Hazard
Estimation	Top Model Export Stats Status	Y	Y	Truncated bins A Status	Top model export stats

Table SOP 11.3. Key to codes for specification of output related to estimated detection functions and group size from the DISTANCE results files (e.g., "DFstats_YYYY.txt" and "DFstats_YYYY.csv") . Output content in each row is coded by values in the 4th and 5th columns.

Output	ID code values by column				Values from file			
	Col 1	Col 2	Col 4	Col 5	Col 6 ^a	Col 7 ^b	Col 10 ^c	SE ^b
Right-truncation distance	0	0	1	6	value			
Detection probability within right-truncation distance	0	0	2	5	estimate	CV	df	col6-col7
Effective strip width	0	0	2	6	estimate	CV	df	col6-col7
Key function	0	0	2	13	integer ^d			
Series expansion	0	0	2	14	integer ^e			
# key function parameters	0	0	2	15	value			
# series expansion parameters	0	0	2	16	value			
Parameter estimates for detection function	0	0	2	101...	estimate	CV	df	col6-col7
Parameter estimate for average group size	integer ^f	Y/N ^g	3	1	estimate	CV	df	col6-col7
P-value for group size regression	integer ^f	0	3	3	value			
Parameter estimate for expected group size from regression	integer ^f	Y/N ^g	3	4	estimate	CV	df	col6-col7

^aAll are floating point numeric, except those specified as integer.

^bFloating point numeric

^cInteger

^d1 = uniform, 2 = half-normal, 3 = negative exponential, 4 = hazard rate

^e0 = no series expansion 1 = simple polynomial, 2 = Hermite polynomial, 3 = cosine

^f1 = KIMU, 2 = MAMU, 3 = UNMU

^gY = estimate selected for density/abundance analyses, N = estimate not selected for density/abundance analyses. Column 2 is manually edited by the Project Leader (see step A.3 in the 'Estimating Group Size Using DISTANCE' section for additional detail).

Appendix SOP 11.A: R code for plotting detection functions

```
# Detection function plot output - R statistical computing environment code

# *****BEGIN USER INPUT*****
# User-defined inputs
# Specify full file path for .txt file with detection function data that has been
output from Program DISTANCE.
# *NOTE*: R uses '/' instead of "\" to specify file paths.
path <- c("\\inpglbafs03\\data\\SEAN_data\\Work_zone\\KM\\KM_J\\yyyy\\DFplot_yyyy.txt")
# Specific x and y labels for the plot
xlabel <- c("Perpendicular distance (m)")
ylabel <- c("Detection probability")
# Path and name for graphic output of the plot
outpath <-
c("\\inpglbafs03\\data\\SEAN_data\\Work_zone\\KM\\KM_J\\yyyy\\DFplot_yyyy.png")

# *****END USER INPUT*****
# Read the text file
DFplot<-read.table(path, header=T, sep="\t", dec=".")
# Plot the estimated detection function
plot(DFplot$C1, DFplot$C2, type="l", lwd=2.5, ylim=c(0,max(DFplot$C4)),
      xlab=xlabel, ylab=ylabel, cex.axis=1.2, cex.lab=1.3)

# Add the frequency distribution histogram
lines(c(0,0), c(DFplot$C3[1], DFplot$C4[1]))
lines(DFplot$C3, DFplot$C4)

# Save contents of current graph window as image in .png file format using the
"png" argument
# Other available formats include jpeg ("jpeg"), bitmap ("bmp"), etc.
dev.copy(png, outpath)
dev.off()
```