

SOP 12: Abundance Estimation (KM_K 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

Summary

The certified versions of the “KM_E: Cumulative observation database” and “KM_J: Detection function” deliverable for the current survey year are used to estimate species-specific, on-water density and abundance within the survey area. Using software code for the R statistical package provided in Appendix D, the user specifies data inputs and then executes code to produce estimates of encounter rates, density, and abundance for use in reports and analyses of population status and trend. Output from the R code is appended to “KM_K: Abundance estimates” deliverable.

This protocol employs the R software package for statistical computing, freely available at <http://www.r-project.org/>. We also recommend use of an external editor to manage and execute the R code provided in Appendix D. Program TINN-R is a suitable graphical user interface and editor for the R programming language. Information about this software is available at <http://www.sciviews.org/Tinn-R/> and the software is freely available at <http://sourceforge.net/projects/tinn-r/>.

We selected survey transects using a two-stage procedure (see Appendix B, section B.7). The first stage employed probabilistic methods of sample selection, and these sampling probabilities were used to appropriately weight estimates for analyses of population status (see Appendix A, section A.3). The second stage added one transect to the survey sample via non-probabilistic methods; hence, this transect is appropriate for analyses of population trend but not status. Output from this SOP separates results for analyses of population status versus population trend.

Data inputs for population analyses are the estimates produced in Program DISTANCE as described in SOP 11 (detection probability across the transect width, probability of detection near the transect center line, group size for KIMU, MAMU, and unidentified murrelets, and the transect strip width) and the corrected survey transect data from “KM_E: Cumulative observation database” deliverable (groups of KIMU, MAMU, and unidentified murrelets encountered on the water on survey transects, transect length, selection probability for transects, latitude and longitude of mid-points of transects). Estimation methods (described in Appendix A) account for detection probability, incomplete species identification, uneven allocation of sampling intensity, the spatially-balanced sampling design, and

clumped distributions of murrelet populations. Output from this SOP includes estimates and associated measures of uncertainty for: encounter rates (groups/km) for KIMU, MAMU, and unidentified murrelets, species-specific adjusted encounter rates (accounting for unidentified murrelets), and species-specific density (individuals/km²) and abundance (across the survey area) accounting for unidentified murrelets.

Creation of Input Files

The Data Manager creates two comma delimited text files for import into R. The file “Encounter_Rates_YYYY.csv” (where YYYY is the survey year) is created from KM_E: Cumulative observation database deliverable and will summarize encounter rates of murrelet groups on each transect as well as transect attributes for the current survey year (Table SOP 12.1). The file “DISTANCE_Output_YYYY.csv” (where YYYY is the survey year) is created from KM_J: Detection Function deliverable and summarizes estimates of detection probability and group size for murrelets from Program DISTANCE for the current survey year (Table SOP 12.2).

Detailed Steps

A. Data Manager Tasks

1. Produce a text file summarizing encounter rates and transect attributes for the current survey year.
 - a. Be certain the KM_E and KM_J deliverables are certified for the year of interest – the file is created from their content.
 - b. Connect to the data management web site (currently [\\165.83.57.239](http://165.83.57.239)) and log in using your Windows userid and password – userid must be of the form “username@NPS”.
 - c. Under Kittlitz's Murrelets, click on the KM_K task called “Create input table SOP 12.1”.
 - d. Specify the year of interest from the dropdown.
 - e. The right truncation distance reflected in KM_J is shown as the default to use. An analyst may choose to enter another value over this default.
 - f. Press the button to preview a sample result on screen if desired.
 - i. The file should contain one row for each transect surveyed (transects with no encounters must be included an encounter rate of 0, with nine fields as described in Table SOP 12.1.
 - ii. Each scheduled transect that was not surveyed at all should appear with null (not zero) for transect distance and the four encounter rates.
 - g. Press the button to save complete results to
“[\\inpglbafs03\data\SEAN_data\Work_zone\KM\KM_K\YYYY\Encounter_rates_YYYY.csv](file://\\inpglbafs03\data\SEAN_data\Work_zone\KM\KM_K\YYYY\Encounter_rates_YYYY.csv)” where YYYY is the four-digit survey year.
2. Produce a text file summarizing estimates of detection probabilities and group sizes.
 - a. Be certain the KM_J deliverable is certified for the year of interest and the “Update database with certified KM_J” data management task has successfully completed – the file is created from this content.

- b. Connect to the data management web site (currently [\\165.83.57.239](http://165.83.57.239)) and log in using your Windows userid and password – userid must be of the form “username@NPS”.
- c. Under Kittlitz's Murrelets, click on the KM_K task called “Create input table SOP 12.2”.
- d. Specify the year of interest from the dropdown.
- e. The timestamp that the year's content was put in the database appears. Be certain it is in the appropriate range.
- f. Press the button to preview a sample result on screen if desired.
 - i. The file must contain one row with 25 fields as described in Table SOP 12.2.
 - ii. Estimates of detection probability near the transect center line (P_c) are taken from results of field experiments (see Appendix A, section A.2). These must have been already stored in the one-row database table named `tbl_detection_centerline`.
 - iii. Transect width (w ; aka ‘the right-truncation distance’) is selected during preliminary analyses (see SOP 11).
 - iv. All other fields are estimates from Program DISTANCE (see SOP 11).
 - v. Because not all transects are appropriate for population status estimates (see Appendix B, section B.7), separate attributes hold estimates of group size for population trend and status.
- g. Press the button to save complete results to
“[\\inpglbafs03\data\SEAN_data\Work_zone\KM\KM_K\yyyy\DISTANCE_Output_yyyy.csv](http://inpglbafs03\data\SEAN_data\Work_zone\KM\KM_K\yyyy\DISTANCE_Output_yyyy.csv)” (where yyyy is the survey year).

Executing R Code

The Project Leader opens the R code from Appendix D in Program TINN-R or another text editor. In the user-defined input section, the Project Leader species the input files and other parameters for analyses. Execution of the code in R produces estimates of abundance and related parameters, which are displayed via the R console and exported to a database file.

Detailed Steps

A. Project Leader Tasks

1. Install required R package: MASS (if not previously installed)
 - a. Start TINN-R and open an R GUI console using the ‘R|Start/close and connections|R GUI (start)’ file menu command.
 - b. In the R console window, use the ‘Packages|Set CRAN mirror’ file menu command to select from a list of available mirrors.
 - c. Use the ‘Packages|Install package(s)’ menu command to open a window with available packages. Navigate to and select the “MASS” package to install it.
2. Open the R code for abundance analyses

- a. Start TINN-R and open an R GUI (graphical user interface) console using the 'R|Start/close and connections|R GUI (start)' file menu command.
 - b. In the TINN-R window, use the "File|Open" menu command to select the R file containing code from Appendix D
(["/inpglbafs03/data/SEAN_Data/Work_Zone/KM/KM_K/APPENDIX_D.r"](#))
 - c. The code should be displayed as a tabbed window in the text editor.
3. Specify user-defined inputs
 - a. Navigate to the top of the file. The first section headed "BEGIN USER INPUT" should be displayed, showing attributes of the analyses defined by the user (Figure SOP 12.1).
 - b. Save your working file under a new name using the 'File|Save as" menu command.
 - c. Specify the full path of the working folder containing input files created by the Data Manager (see section1, above) and where output will be placed using the 'path' input. Note: R requires forward slashes "/" in the file path, rather than the back slashes "\" used in the Windows operating systems.
 - d. Provide names of the input files containing output from Program DISTANCE and the encounter rate data for the current survey year.
 - e. Specify the 4-digit survey year.
 - f. The area in km² for projection of abundance estimates can be specified by the user at the 'A' input, but typically projection over the area of the survey area (1,170 km²) will be desired.
 - g. Specify the desired % for confidence interval coverage in percent at the 'CI' input. The default is 95% CI coverage.
 - h. Other user-defined inputs specify characteristics of the sample of survey transects and are unlikely to change unless the survey design is altered.
 - i. The 'panels' input specifies the number of survey transect panels (i.e., 1 permanent + 3 alternating panels).
 - ii. The 'plim' input specifies the range of acceptable transect ID numbers.
 - iii. The 'coordlim' input specifies the range of acceptable latitude and longitude coordinate values in decimal degrees for transect mid-points.
 - iv. The 'xstatus' input specifies the ID numbers of transects selected using non-probabilistic methods and hence to be excluded from population status analyses.
 - i. Save the file using the 'File|Save" menu command to preserve user-defined inputs.
4. Execute analyses

- a. Execute analyses using the “R|Send|File” menu command or by clicking the equivalent button on the R toolbar (Figure SOP 12.2) to send the code to the R console.
- b. If successful, output from analyses should be displayed in 2 tables in the R console window (Figure SOP 12.3): one for population trend analyses and one for population status analyses.
- c. Output will also be written to an ACSII CSV text file “R_Output_YYYY.csv”, where YYYY is the survey year. The file will be formatted as shown in Table SOP 12.3 and output to the folder specified above in step 3.c.
- d. Estimates should be carefully checked to make sure that output is consistent with data inputs and prior results and that all required fields are non-null and that field values are within permissible ranges (Table SOP 12.3).
- e. If unsuccessful, the R console window likely will display an error message(s).
- f. The R code checks for deficiencies in the structure and values of the input files; error messages from these checks should indicate the nature of the deficiency.
- g. Otherwise, error messages generated by R may help identify the nature of the error preventing code execution.
- h. Make corrections to the files used in this process and re-execute the R code until the output is acceptable.

Formally validate and certify the final product

The SEAN data management plan requires all monitoring deliverables be formally validated against rules defined in the protocol. When all mandatory criteria have been met, 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_K, for year YYYY is ready; the Data Manager will know where to find the file named “R_Output_YYYY.csv”.

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 R_Output_YYYY.csv file into the staging area for validation at a subdirectory named with the submission number in [\\inpglbafs03\data\SEAN_Data\Staging\KM\KM_K](http://inpglbafs03\data/SEAN_Data/Staging/KM/KM_K). Rename the file to KM_K_YYYY.csv

3. Validate the submission according to the current criteria.
 - a. Invoke validation of the CSV file according to current criteria by using the web program at http://165.83.57.239/KM_DM_validate_KMK.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

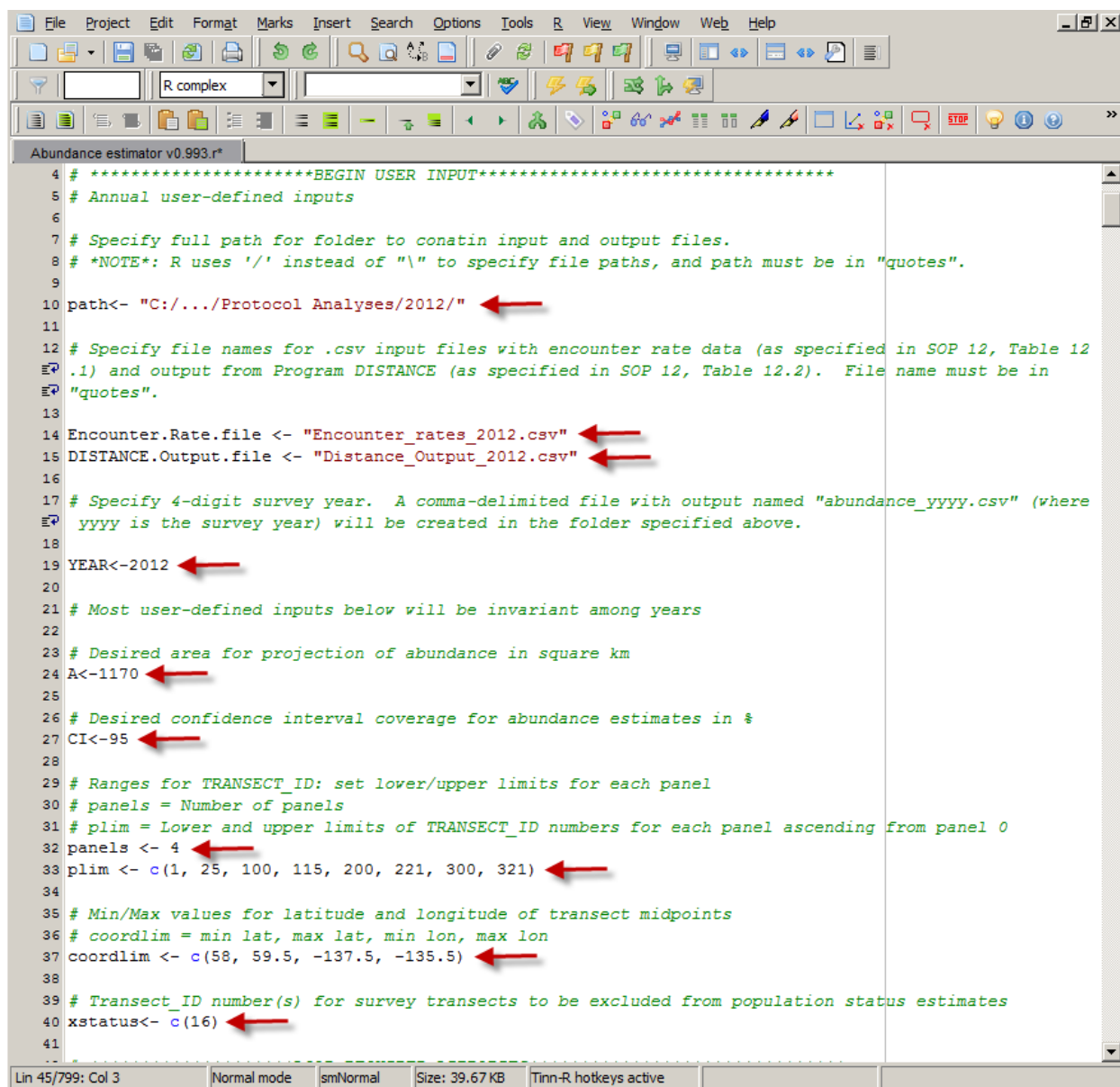
1. On receipt of a failure email:
 - a. Review the validation report and remedy the mandatory errors by revising the output file.
 - b. Initiate a new submission with the Data Manager.
2. On receipt of a success email, review any failed optional criteria:
 - a. If the criteria review indicates the validated 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 database.
 - v. Terminate the process.
- b. Use the "Update Submission Log" web tool at http://165.83.57.239/0_submission_update.aspx to mark the submission as certified.
 - i. Set status to "C".
 - ii. Set certification date to date of the certification email.
 - iii. Set certified by to the initials of the project lead.
 - iv. Mark any previously certified KM_K for that submission unit as decertified.
- c. Update the staging database to reflect the new revised abundance estimates by invoking the web program at http://165.83.57.239/KM_K_update_db_DM.aspx.
 - i. The application will revalidate the CSV file.
 - ii. If validation fails, remediate the issue.
- d. Test accessibility to updated data from Integration web site until successful.
- e. Propagate from staging to production database.
- f. Verify final deliverable is accessible from production web site.
- g. Update the deliverable tracking spreadsheet milestones with the date of completion for KM_K. The sheet is kept in development's \AuxRep\KM directory.
- h. Update the scope of the formal metadata so it includes this new date range.

SOP 12 Figures and Tables



```
Abundance estimator v0.993.r
4 # *****BEGIN USER INPUT*****
5 # Annual user-defined inputs
6
7 # Specify full path for folder to contain input and output files.
8 # *NOTE*: R uses '/' instead of '\' to specify file paths, and path must be in "quotes".
9
10 path<- "C:/.../Protocol Analyses/2012/"
11
12 # Specify file names for .csv input files with encounter rate data (as specified in SOP 12, Table 12
13 # .1) and output from Program DISTANCE (as specified in SOP 12, Table 12.2). File name must be in
14 # "quotes".
15
16 Encounter.Rate.file <- "Encounter_rates_2012.csv"
17 DISTANCE.Output.file <- "Distance_Output_2012.csv"
18
19 # Specify 4-digit survey year. A comma-delimited file with output named "abundance_yyyy.csv" (where
20 # yyyy is the survey year) will be created in the folder specified above.
21
22 YEAR<-2012
23
24 # Most user-defined inputs below will be invariant among years
25
26 # Desired area for projection of abundance in square km
27 A<-1170
28
29 # Desired confidence interval coverage for abundance estimates in %
30 CI<-95
31
32 # Ranges for TRANSECT_ID: set lower/upper limits for each panel
33 # panels = Number of panels
34 # plim = Lower and upper limits of TRANSECT_ID numbers for each panel ascending from panel 0
35
36 panels <- 4
37 plim <- c(1, 25, 100, 115, 200, 221, 300, 321)
38
39 # Min/Max values for latitude and longitude of transect midpoints
40 # coordlim = min lat, max lat, min lon, max lon
41 coordlim <- c(58, 59.5, -137.5, -135.5)
42
43 # Transect_ID number(s) for survey transects to be excluded from population status estimates
44
45 xstatus<- c(16)
```

Figure SOP 12.1. User input area in Program R code for estimation of density and abundance, displayed in the Tinn-R text editor. Red arrows indicate lines where user input is provided to the right of the '<-' operator. The folder path for data import and export is specified in line 10, and names of input files containing data for import into R are specified in lines 14 and 15. Results of analyses are exported in the file "R_Output_yyyy.csv", where yyyy is the survey year specified in line 19. Other user input typically will be invariant.

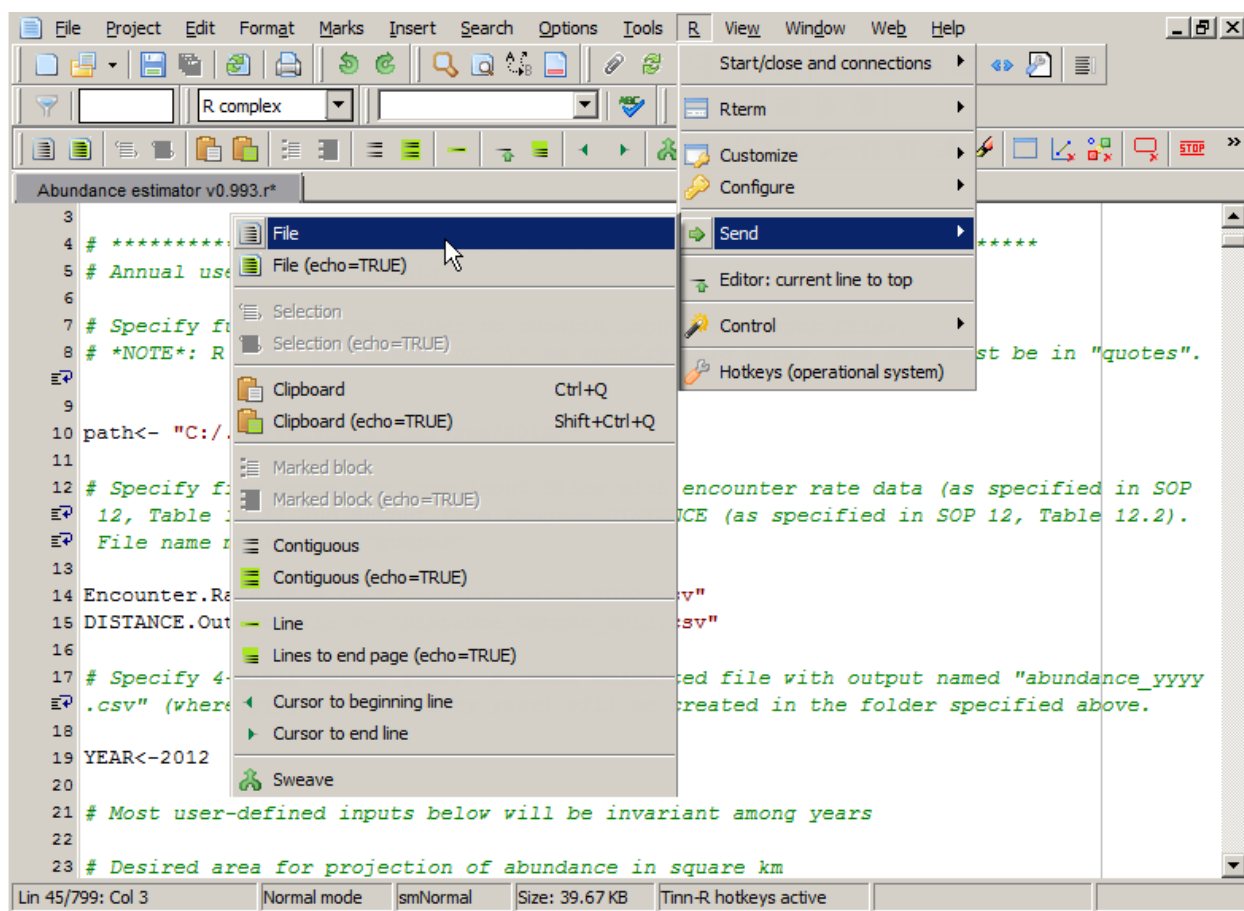


Figure SOP 12.2. After specifying user inputs and opening the R GUI console, send the code in the text editor to R using the 'R|Send|File' menu command or using the equivalent menu button indicated by the red arrow.

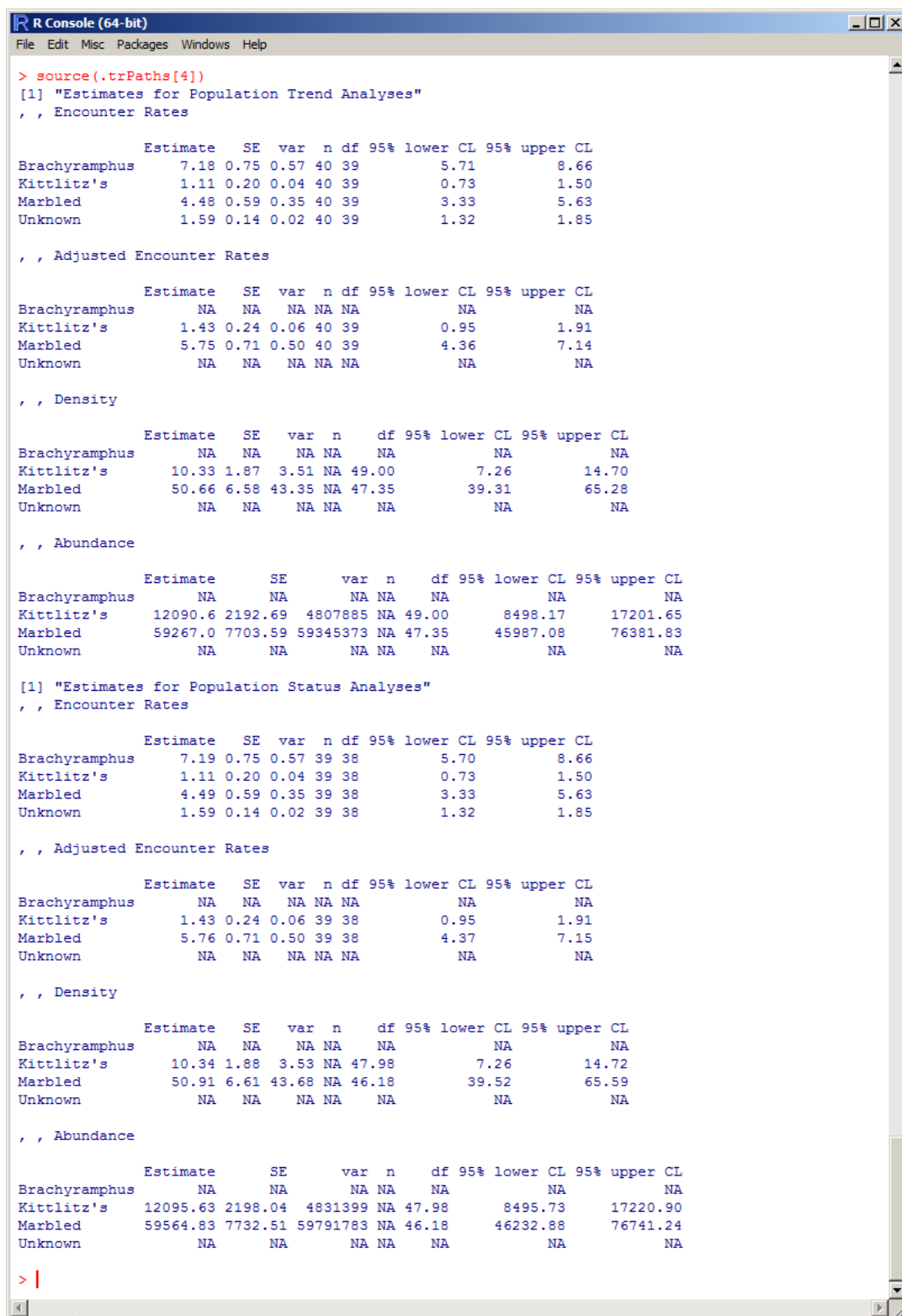


Figure SOP 12.3. Estimated population estimates for analyses of population status and trend displayed in the R GUI console. Estimates for *Brachyramphus* and unidentified murrelets are only presented for encounter rates, because other estimates are adjusted to account for unidentified murrelets.

Table SOP 12.1. Data fields for “Encounter_Rates_YYYY.csv” (where YYYY is the survey year). Encounter rates and TRANSECT_KM may vary by survey, but other fields should be invariant unless the survey design changes.

Field name	Type	Description
TRANSECT_ID	string(3) ^a	Transect ID# (3-digit numeric code)
TRANSECT_KM	numeric	Length of surveyed transect in km
ERK	numeric	Encounter rate ^b for KIMU (groups on water/transect length in km)
ERM	numeric	Encounter rate ^b for MAMU
ERU	numeric	Encounter rate ^b for UNMU (unidentified <i>Brachyramphus</i> murrelets)
ERB	numeric	Encounter rate ^b for BRMU (all <i>Brachyramphus</i> murrelets), sum of above 3
LAT_MID	numeric	Coordinate value for latitude of mid-point of transect in decimal degrees
LON_MID	numeric	Coordinate value for longitude of mid-point of transect in decimal degrees
P_INCLUD	numeric	Probability of inclusion of transect derived from sampling design

^aNumber in parenthesis indicates string length.

^bGroups encountered on water within right-truncation distance divided by transect length in km. The right-truncation distance is determined during estimation of the detection function (see SOP 11).

Table SOP 12.2. Data fields for “DISTANCE_Output_YYYY.csv” (where YYYY is the survey year). The after-hyphen suffixes denote the standard error (SE) and degrees of freedom (DF) for estimates. For group size estimates (starting with ‘S’), the second character denotes species class (‘K’ = Kittlitz’s murrelet, ‘M’ = marbled murrelet, and ‘U’ = unidentified *Brachyramphus* murrelet) and the third character indicates whether estimates are appropriate for population trend (T) or population status (S) analyses.

Field name	Type	Description
Pa	numeric	Probability of detection across transect width P_a
Pa_SE	numeric	SE for P_a
Pa_DF	numeric	Degrees of freedom (<i>df</i>) for P_a
Pc	numeric	Probability of detection near transect center line P_c
Pc_SE	numeric	SE for P_c
Pc_DF	numeric	<i>df</i> for P_c
W	integer	Transect width (meters) w
SKT	numeric	Average group size for KIMU for trend analyses S_{KT}
SKT_SE	numeric	SE of S_{KT}
SKT_DF	numeric	<i>df</i> of S_{KT}
SMT	numeric	Average group size for MAMU for trend analyses S_{MT}
SMT_SE	numeric	SE of S_{MT}
SMT_DF	numeric	<i>df</i> of S_{MT}
SUT	numeric	Average group size for UNMU (unidentified) for trend analyses S_{UT}
SUT_SE	numeric	SE of S_{UT}
SUT_DF	numeric	<i>df</i> of S_{UT}
SKS	numeric	Average group size for KIMU for status analyses S_{KS}
SKS_SE	numeric	SE of S_{KS}
SKS_DF	numeric	<i>df</i> of S_{KS}
SMS	numeric	Average group size for MAMU for status analyses S_{MS}
SMS_SE	numeric	SE of S_{MS}
SMS_DF	numeric	<i>df</i> of S_{MS}
SUS	numeric	Average group size for UNMU (unidentified) for status analyses S_{US}
SUS_SE	numeric	SE of S_{US}
SUS_DF	numeric	<i>df</i> of S_{US}

Table SOP 12.3. Data fields for the text file "R_Output_YYYY.csv" (where YYYY is the survey year).

Field name	Type	Permissible Values	Description
SURVEY_YEAR	integer	≥2010	4-digit survey year
ANALYSIS_CODE	string	TREND	Estimate for population trend analyses
		STATUS	Estimate for population status analyses
ATTRIBUTE_CODE	string	Encounter.Rate	Groups/km
		Adjusted.Encounter.Rate	Encounter rate accounting for unidentified murrelets
		Density	Density (individuals/km ²)
		Abundance	Abundance (within specified survey area)
SPECIES_CODE	string	Brachyramphus	All <i>Brachyramphus</i> murrelets
		Kittlitz's	Kittlitz's murrelet
		Marbled	marbled murrelet
		Unknown	Unidentified <i>Brachyramphus</i> murrelet
ESTIMATE	numeric	≥0	Estimate for specified attribute and species
SE	numeric	≥0	Standard error of estimate
VAR	numeric	≥0	Variance of estimate
N	numeric	≥0	Sample size ^a
DF	numeric	≥0	Degrees of freedom (<i>df</i>) or approximated <i>df</i> ^a
CL_LOWER_95_PCT ^b	numeric	real number	Lower 95% confidence limit
CL_UPPER_95_PCT ^b	numeric	≥0	Upper 95% confidence limit

^aFor estimates combining data with differing sample sizes, sample size is absent and degrees of freedom are approximated. See Appendix A, section A.3.3.

^bField name will reflect user-specified confidence interval coverage.