Supplemental Material S1. Implementing the analysis in Distance 6.2.

Our goal is to estimate the abundance of Gopher tortoises at Fort Gordon Army Installation in Georgia. Recall we have an extremely low density tortoise population in which we have systematically placed pseudo circuits of transects. Furthermore there is variability in animal density across the study area and it is challenging to obtain enough tortoise detections to reliably estimate the probability of detection. Therefore we suggest a slight modification to traditional distance analysis, based on locating usable burrows (which are more common than tortoises) and separately accounting for the proportion of burrows occupied. We will use a variance estimator for systematic designs based on overlapping post-stratification to derive a more precise estimate in contrast to the default analysis in Distance.

We will outline the steps for this procedure, including how to organize the data and run the analysis in Distance. Then we will outline how to estimate the average probability of detecting a burrow, the proportion of occupied burrows using a cluster size technique, and how to implement a post-stratification scheme to get a better estimate of the variance. Finally, we will use the mean of both encounter rate variance estimates from Distance and show how to obtain an overall density coefficient of variation through the delta method. The coefficient of variation obtained through this procedure will be used to calculate the 95% confidence intervals.

1. **Data organization**

We will organize the data so they can be conveniently imported into Distance. The study area and survey design are shown below. Sampling units consist of a pair of parallel segments 50 m apart forming a “pseudo-circuit.” Therefore each “transect” is made up of two parallel segments of 500 m long, totaling 1 km. Segments were not always 500 m long because they may have fallen at the edge of the sample region or installation boundary. The pseudo-circuits were separated by 500 m to the east-west and 300 m to the north-south, along a diagonal grid as shown below.

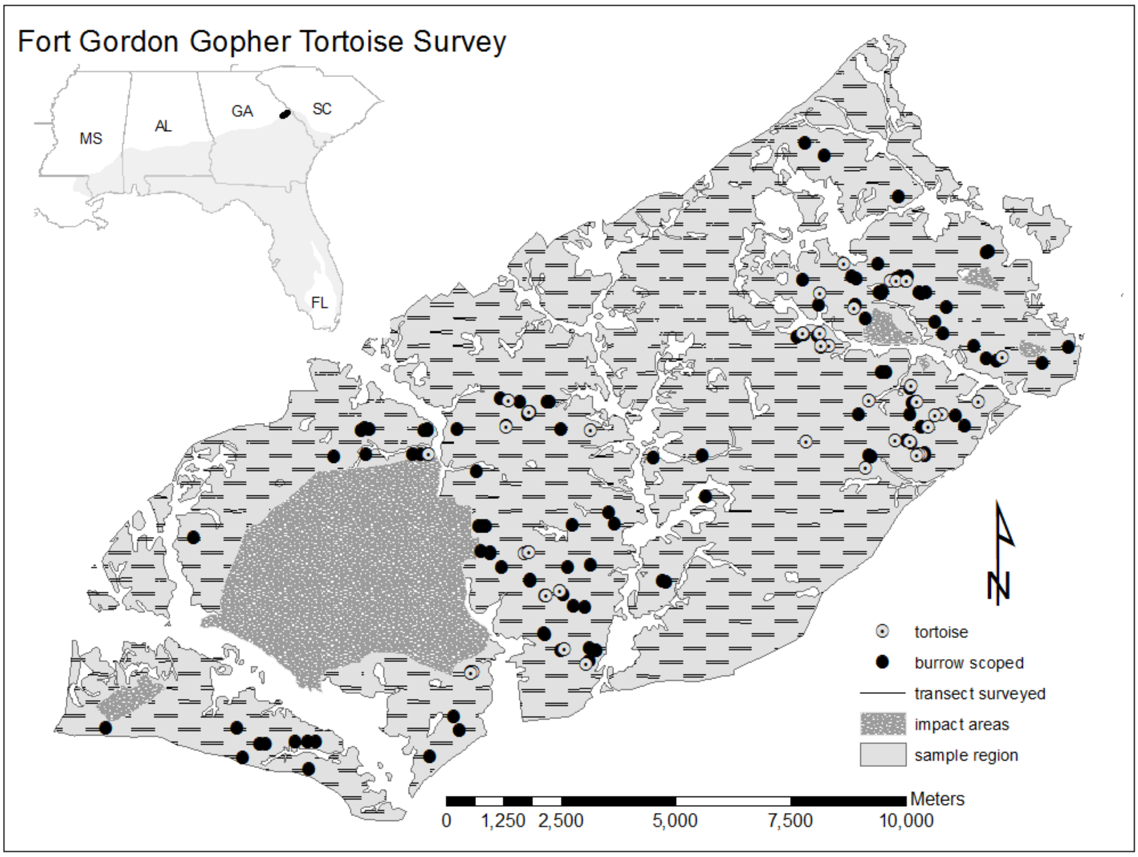


Figure 1. Gopher tortoise (*Gopherus polyphemus*) population survey using line transect distance sampling methodology during 2010 and 2011 at the Fort Gordon Army Installation, GA. Tortoise Habitat Management Unit with sample region, impact areas, 556 pseudo-circuit transects surveyed, burrows and tortoises used in analyses.

While it is possible in Distance to enter data manually, this is inefficient for all but very small datasets. Typically, the Data Import tool is used to bring the data into Distance. The data can readily be imported as a text file, where columns are separated by tabs. Below is an example of how we organized the data.

	A	B	C	D	E	F	G	H	I	J
1	transect	length	NE_dir	NW_dir	dist	tortoise	diameter	year		
2	1	445.5	194	205						
3	2	989.8	193	228						
4	3	547.1	192	240						
5	4	820.2	220	227						
6	5	915.9	221	241						
7	6	552.1	251	242						
8	7	1006.6	280	243						
9	8	550.4	307	261						
10	9	999.9	308	244						
11	10	411.4	306	282	15.5939495	0	21.8	20110331		
12	11	1007	372	259						
13	12	1006.8	373	246						
14	13	867.3	132	353	11.7388072	0	34	20110413		
15	13	867.3	132	353	12.7887269	0	39.2	20110413		
16	13	867.3	132	353	23.6464764	0	29.6	20110413		
17	14	932.1	62	386						
18	15	870.3	131	384						
19	16	998.3	108	385	6.65178086	0	32.5	20110413		
20	16	998.3	108	385	12.6084416	0	16.7	20110413		
21	16	998.3	108	385	18.0732125	-1	21.2	20110413		
22	17	984.1	42	389						
23	18	1000.9	61	390						
24	19	1000.3	109	391						
25	20	927.6	130	392	9.16433199	0	20.1	20110418		
26	21	647.7	175	383	4.76098752	0	27.6	20110418		

dataFG

Figure 2. Spreadsheet example of dataset to be imported into Distance with transect number, length, grouping distance to burrow, record of tortoise observation, diameter of burrow and date of the observation.

The first column (A) labeled "transect" contains the unique identifier for each of the pseudo-circuits. Transect length (the sum of both segments) is represented in column B. We will use a variance estimator for systematic designs based on overlapping post-stratification. This works by grouping together pairs of adjacent transects from the systematic sample. Each pair of adjacent transects is grouped into a stratum. The strata will improve variance estimation, because the systematic sample behaves more like a stratified sample than a random sample which is the default analysis in Distance. The first stratum consists of transects with ID 1 and 2, the second of transects 2 and 3, the third of transects 3 and 4 and so on to yield a total of k-1 strata and corresponding variance estimates.

Due to our systematic survey design, the closest adjacent transect to a fixed transect is the nearest diagonal – north-east or north-west located – transect. Therefore we created columns 3 and 4 from column 1. Column C "NE_dir" realigns transects in a north-east direction weaving them across the survey in a "serpentine"/zig zag fashion, whereas column D "NW_dir" orders them in a north-west direction. The perpendicular distance of the detected burrows along with the occupancy and other details are recorded in the successive columns (E, F, G, and H). We coded tortoise occupancy in "tortoise" as "1" if a tortoise was observed when the burrow was scoped, "0" if not and "-1" if occupancy was unknown.

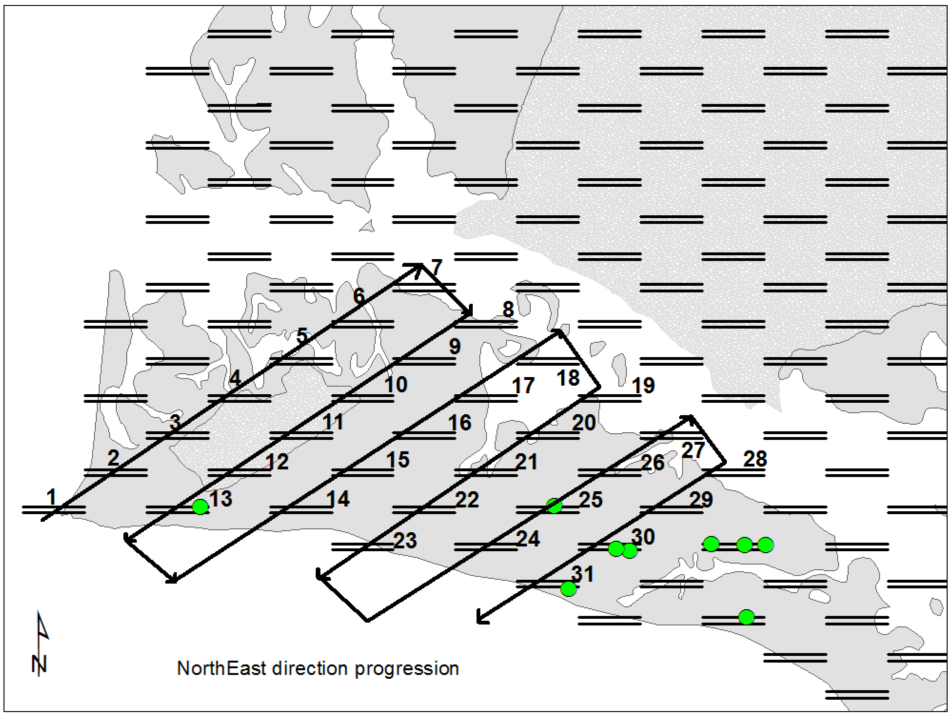


Figure 3. Close up of numbering system for pseudo-circuits with an example of the north-east direction realignments.

To analyze the variance for systematic designs using overlapping post-stratification in Distance, we needed to create as many text files as directions of the sequence of transects are defined. In this particular case, two text files, one for each order of transects depending on direction, were created. For each text file the ordered transect sequence ("NE_dir" and "NW_dir") is the transect ID, and reordered the observations (to import them to Distance for the analysis) according to the new ID transect sequence.

2. Analysis in Distance

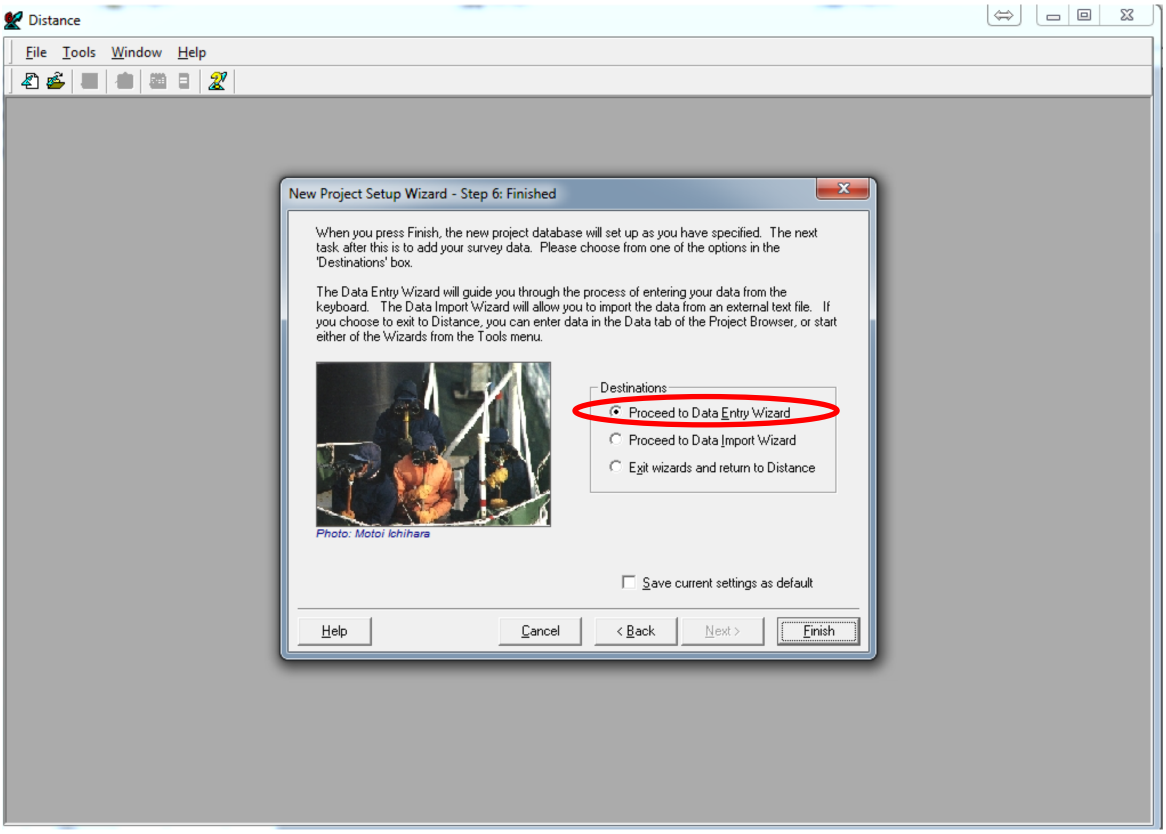
We will use a stratified estimator with overlapping strata for the systematic survey. However, we can build the stratified design following the different directions, leading to as many estimators as possible directions (two in this case). So we need to run the analysis in Distance twice, once using "NE_dir" as transect ID column and then using "NW_dir". Both analyses will differ only in the variance estimate.

- A) Using "NE_dir" as transect column

Below we outline steps for this procedure in Distance software.

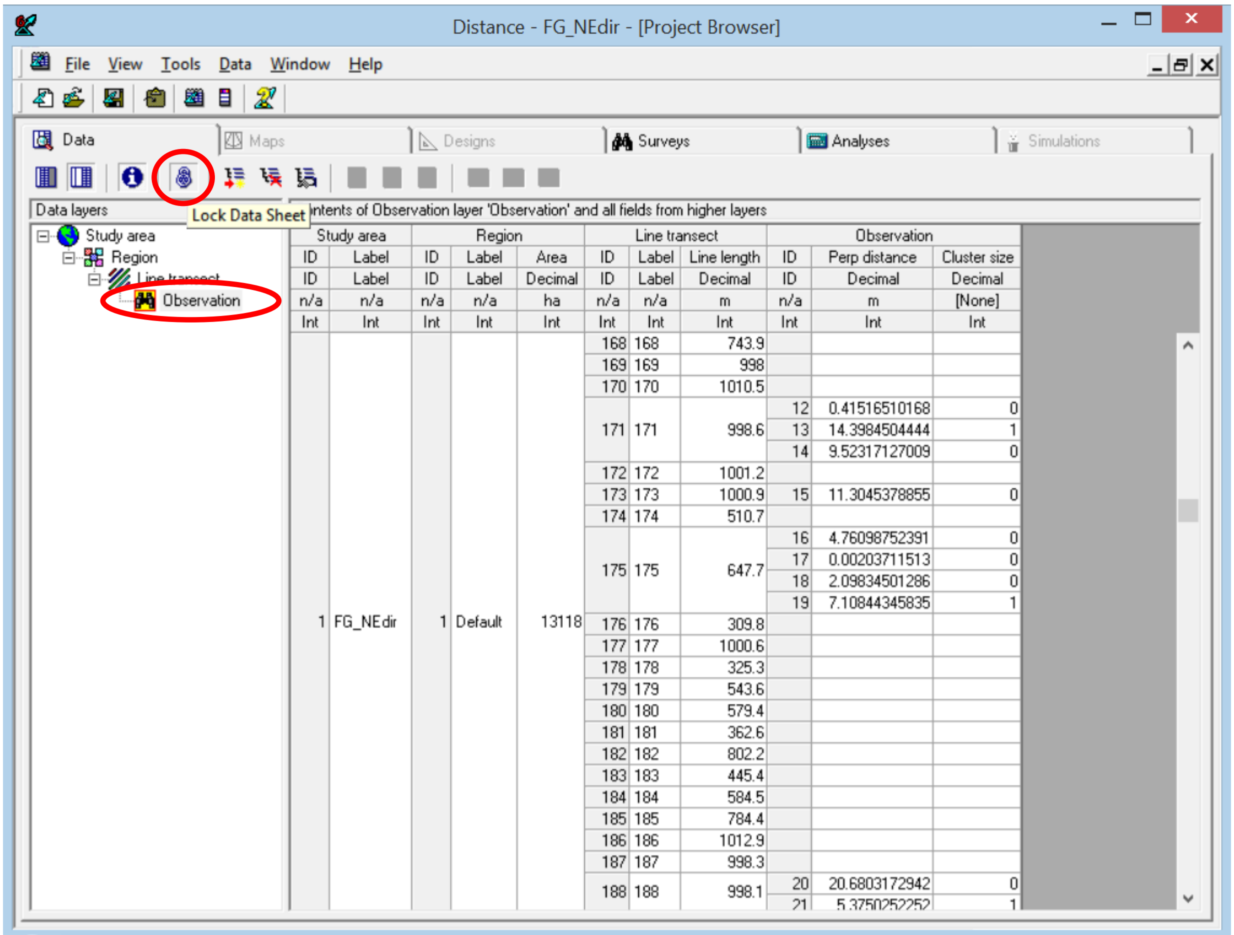
Open Distance. Click on **"Start"**. A list will be displayed. Click on **"Programs"**, then **"Distance"**. Now click on **"Distance 6.2"**. (Or double-click the **Distance 6.2** icon on the desktop.)
Open a new project (click on **File** then on **New project** ...), name it, and click on **Create.** Step through the New Project Setup Wizard. You should not need to change any of the defaults, except for the observations to **clusters of objects**, not the default single objects, and the transect units are measured in meters.

To import the data from the text file click on **Proceed to Data Import Wizard**.



2.1. Import the data to Distance

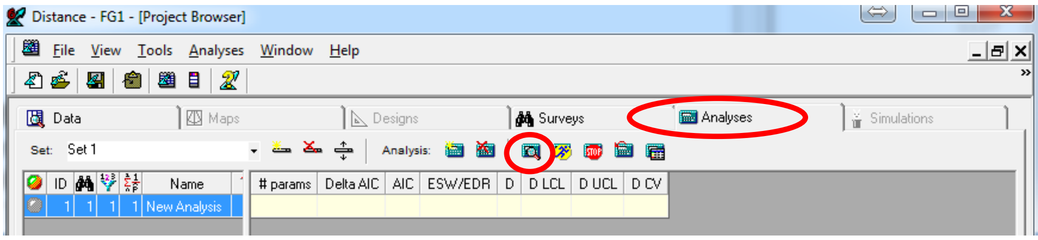
The data import wizard has six screens. The first is introductory. The second asks you for the data source - the text file to import data from. The third allows you to specify the destination of the data- which data layers to put the data into (lowest data layer: Observation; Highest data layer: Region), and how to assign rows in the text file to records in the Distance database. The fourth screen asks you to specify the delimiter used in the text file (make sure you tick the appropriate box do not import the name of the columns in the first row). The fifth asks you to match up columns in the text file with fields in the database (select "**NE_dir**" column - Line transect > Label, "length" - Line transect > Line length, "dist" – Observation > Perp distance and "tortoise" - Observation > Cluster size). The last screen allows you to check if the number of columns and rows is correct, and displays a log of any errors that occur during the import process.
Our data look like this once imported successfully.



Press the button **lock Data Sheet** to avoid making changes in the data.
Note that we sort ascending the sequence of transects "NE_dir" in transect label, to matches transect ID.

2.2.Create the first analysis

Creating a new analysis: Click on the **Analysis** tab of the **Project Browser** to show the **Analysis Browser**.

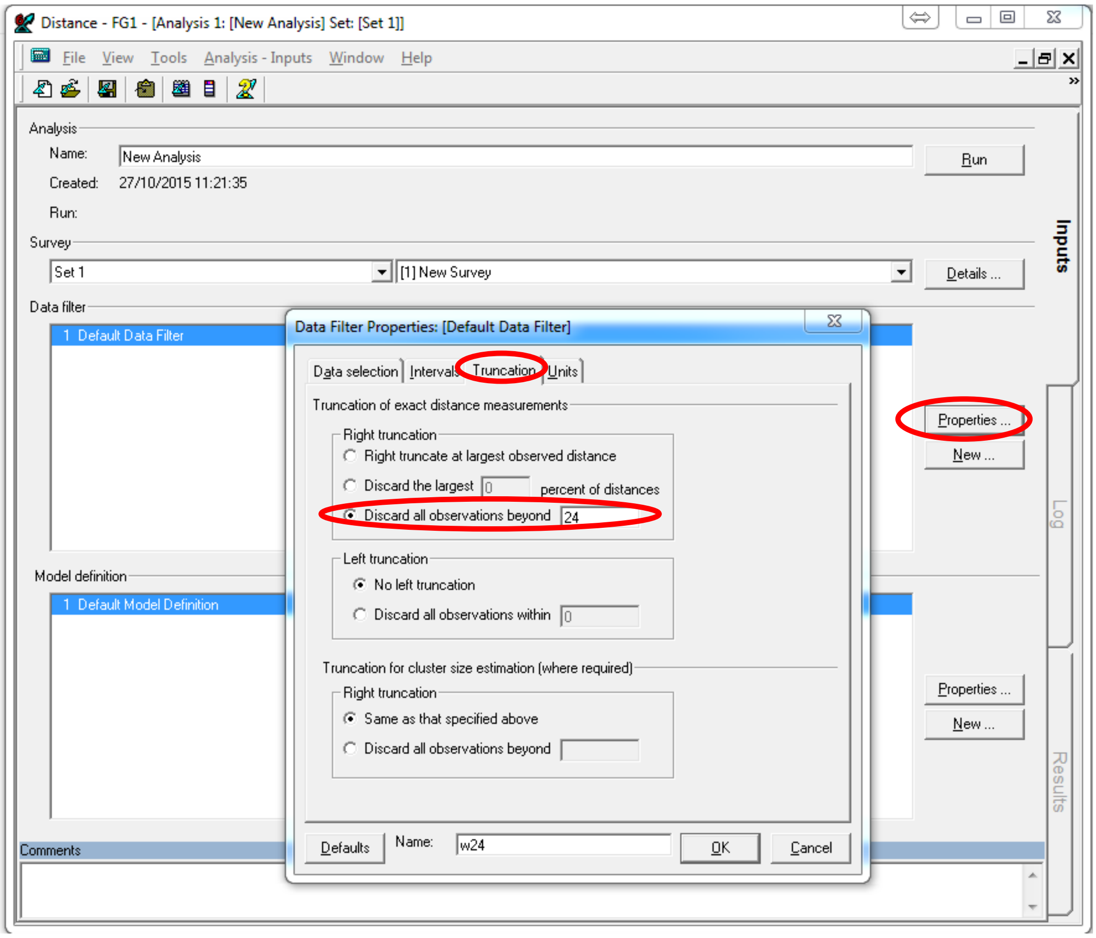


You should see one analysis listed, called "New Analysis." A grey status icon indicates that this analysis has yet to be run. Double-click on the grey status button for this analysis to open the **Analysis Inputs** tab for this analysis (you can do the same thing by clicking the 3rd button after "Analysis:" on the Analysis Browser menu bar, or by choosing **Analyses** then **Analysis Details** from the menu bar at the top).

Create the first analysis. Not need to edit the Survey in the Inputs tab.

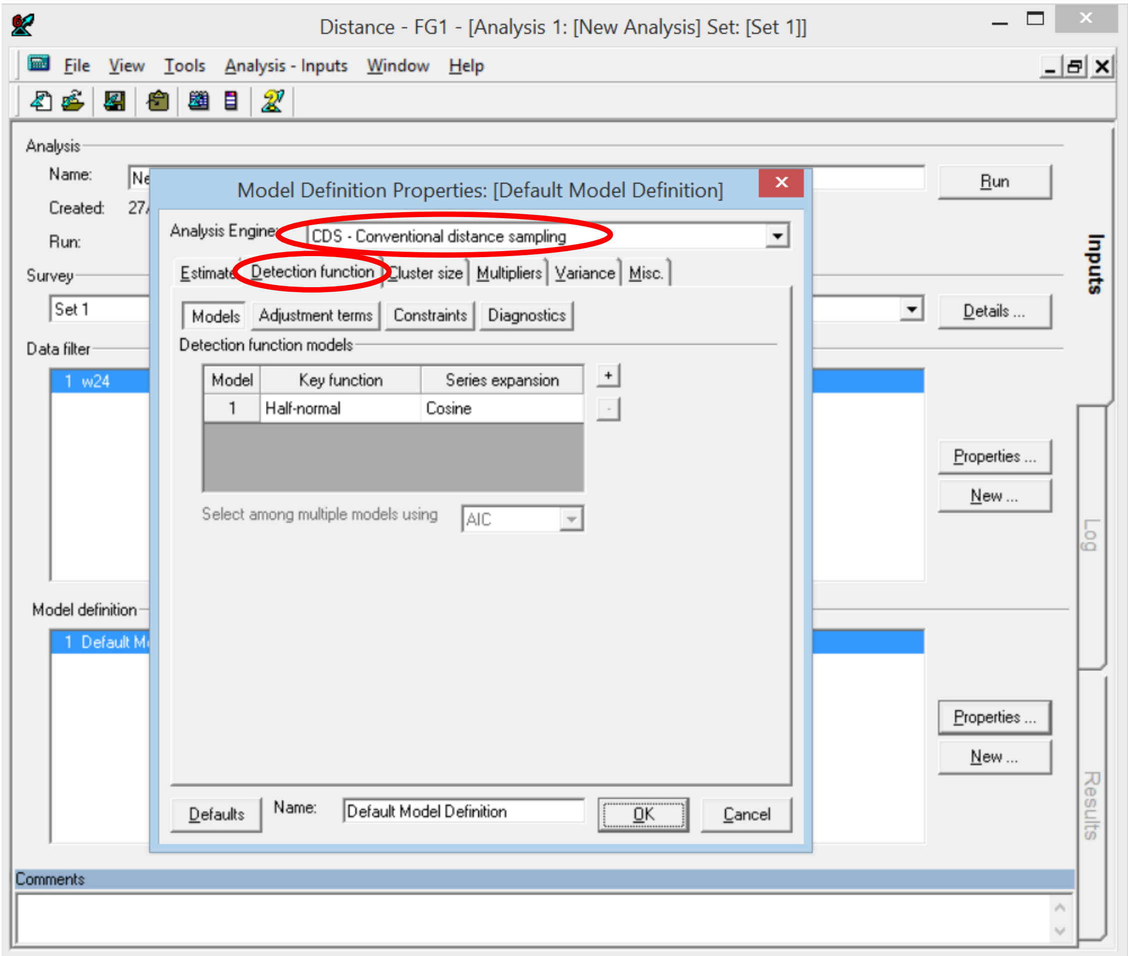
- Define the truncation distance (w = 24)

Create a filter in the **Data filter** tab to truncate the observations at 24m from the line to avoid double counting burrows due to the circuit design of the transects. Click on **Properties**, then click on the **Truncation** lab and select “Discard observation beyond” entering 24. When you have defined the filter, give it a suitable name (one that reflects the options you have set) and select **OK.**



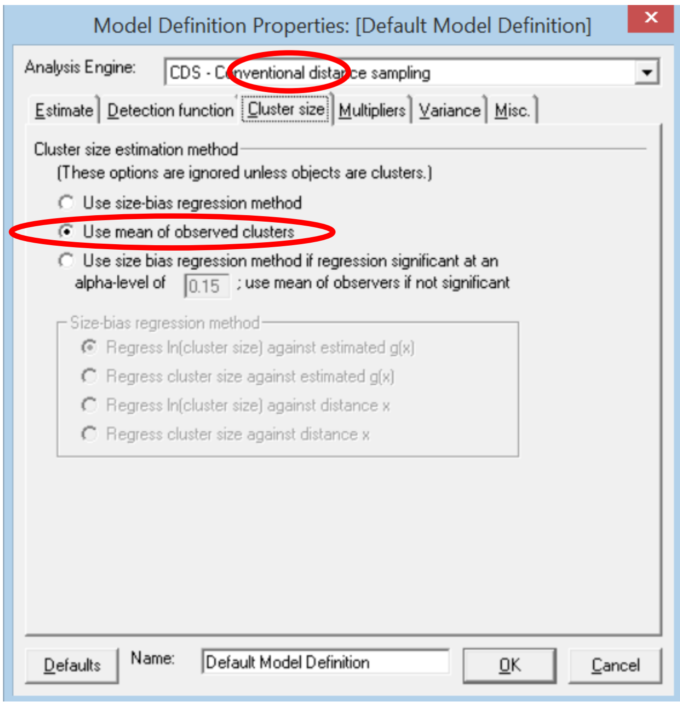
- Specify a model

Select **CDS-Conventional distance sampling** in the Analysis Engine.
In the **Detection function** > **Models** specify a key function with a series expansion, allowing selection of adjustment terms. In this case we started with a half-normal key function with a cosine series adjustment.



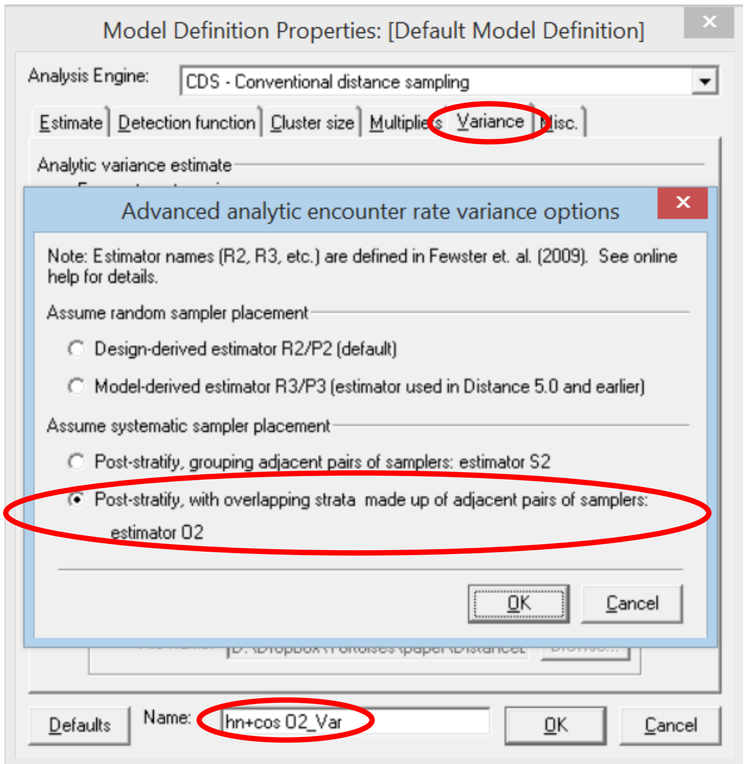
- Include burrow occupancy as cluster size

Click use mean of observed clusters in the **Cluster size** tab.



- Estimate the variance for systematic designs using overlapping post-stratification

Under Model Definition, click in the **Variance** tab, click the **Advanced** button, and select the option "**Post-stratify, with overlapping strata made up of adjacent pairs of samplers: estimator O2**". We could select the option "Post-stratify, grouping adjacent pairs of samplers: estimator S2" and estimate the encounter rate variance for systematic designs using non overlapping strata, but O2 is recommended when the design can accommodate overlapping strata in whole or part, and S2 otherwise (Fewster *et al*., 2009).

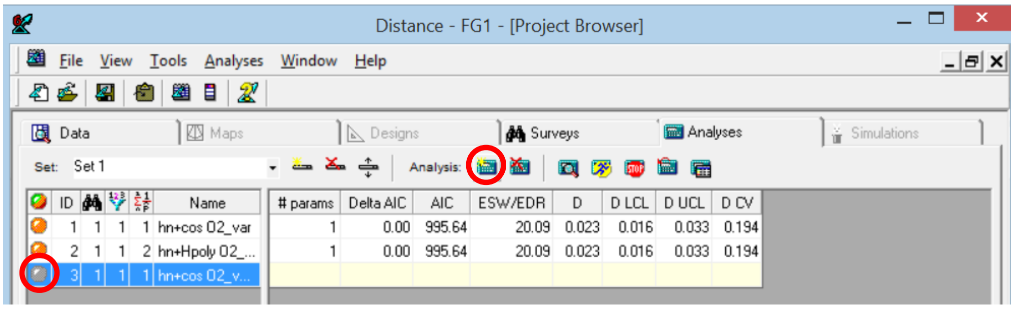


When you have defined the model, give it a suitable name and select **OK.**

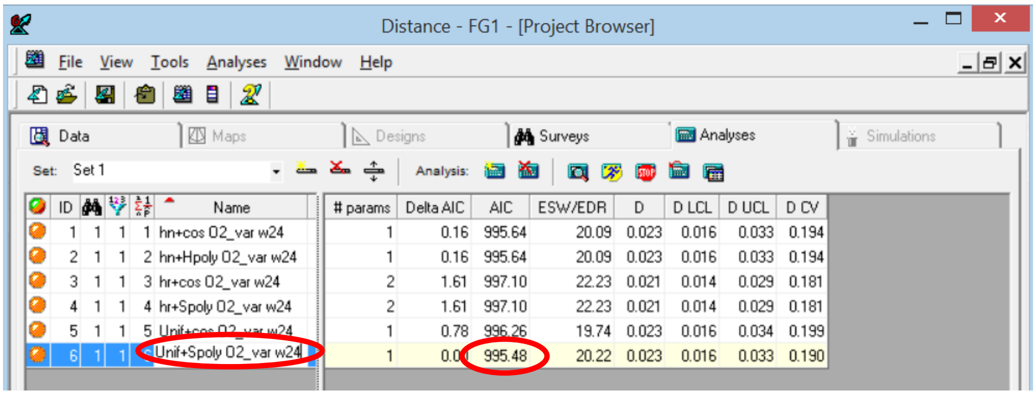
Now give your analysis a suitable name, and click the **Run** button. When the analysis finishes, it will automatically take you to the log tab if there were problems, or the results tab if the analysis ran without errors or warnings. The **Results** tab in the latter case should turn green. Click on the **Results** tab to see the results, and use the **Next** > button to move through the pages of results. (**Note:** These are the analysis details (Inputs/Log/Results) for one analysis - you can resize this window so that you can view details from multiple analyses when you have more than one analysis to compare).

2.3. Create other analysis and select the detection function for inference

Create new analysis. Return to the Analysis Browser, and click on the first button after "Analysis:" on the Analysis Browser menu bar ("New Analysis"). Double-click on the status button to go to the Analysis Details window for this new analysis.



Because the analysis has not run, you are taken to the Inputs tab. You will not need to edit the Survey or Data Filter since the truncation distance is the same, but click on **New** in the **Model definition** section. Following the steps described above, run different analysis. First specify a new model, then include burrow occupancy as cluster size and select the O2 estimator for the variance. Try different key functions with their respective series adjustments. As you create more Data Filters and Model Definitions, you may find that you want to change their order, rename or delete them. A convenient way to do this is using the **Analysis Components** window – click the 6th button from the right on the main menu bar ("View Analysis Components"). In the Analysis Components window, clicking the first button lists the Data Filters and clicking the second button lists the Model Definitions. Finally, select the model for inference considering AIC and double check GOF statistics to make sure the fit is reasonable.



It is easiest to compare results from different analyses using the Analysis Browser. You can change the default columns in the browser using the **Column Manager** (furthest button on the right of the Analysis Browser menu bar).

- B) Using "NW_dir" as transect column

Follow the same steps to carry out the analysis with the sequence of transects given by "NW_dir". Since we are using the same data, the results, except for the variance estimate will be the same.

3. Overall variance estimation

The overall density coefficient of variation obtained through the delta method, in this case, is the combination of three components 1) the variance of the encounter rate, 2) the detectability process and 3) the occupancy estimate.

$$[CV\,(D)]^2 \approx \left[CV\,\left({}^{n}/_{L}\right)\right]^2 + [CV\,(P_a)]^2 + [CV\,(P_o)]^2$$

We define the $CV\left(\frac{n}{L}\right)$ as the mean of both encounter rate variance estimates from Distance analysis.

- A) Using "NE_dir" as transect column
- B) Using "NW_dir" as transect column

First of all, we need to get the variance output for both analysis.

- A)

```
                Point         Standard     Percent Coef.        95% Percent
 Parameter     Estimate        Error      of Variation     Confidence Interval
 ---------    -----------   -----------  --------------  ----------------------
   DS        0.90622E-01   0.11623E-01       12.83       0.70520E-01  0.11646
   E(S)      0.25000       0.35714E-01       14.29        1.0000      0.33108
   D         0.22656E-01   0.43496E-02       19.20       0.15587E-01  0.32930E-01
   N          297.00        57.020           19.20        204.00       432.00
 ---------    -----------   -----------  --------------  ----------------------

Component Percentages of Var(D)
-------------------------------
Detection probability   :  13.6
Encounter rate          :  31.1
Cluster size            :  55.4


                             Estimate      %CV     df     95% Confidence Interval
                           ------------------------------------------------------
                n            157.00
                k            555.00
                L           0.42843E+06
                n/L         0.36646E-03    10.70   554.00 0.29717E-03  0.45190E-03
                Left         0.0000
                Width        24.000

                             Estimate      %CV     df     95% Confidence Interval
                           ------------------------------------------------------
Uniform/Polynomial
                m            1.0000
                LnL         -496.74
                AIC          995.48
                AICc         995.51
                BIC          998.54
                Chi-p       0.39736
                f(0)        0.49459E-01     7.07   156.00 0.43017E-01  0.56865E-01
                p           0.84246         7.07   156.00 0.73273      0.96861
                ESW          20.219         7.07   156.00  17.586       23.247

                             Estimate      %CV     df     95% Confidence Interval
                           ------------------------------------------------------
Average cluster size
                            0.25000        14.29   147.00  1.0000       0.33108

                             Estimate      %CV     df     95% Confidence Interval
                           ------------------------------------------------------
Uniform/Polynomial
                DS          0.90622E-01    12.83   681.68 0.70520E-01  0.11646
                D           0.22656E-01    19.20   420.57 0.15587E-01  0.32930E-01
                N            297.00        19.20   420.57  204.00       432.00
```

- B)

```
                Point         Standard     Percent Coef.        95% Percent
 Parameter     Estimate        Error      of Variation     Confidence Interval
 ---------    -----------   -----------  --------------  ----------------------
   DS        0.90622E-01   0.11144E-01       12.30       0.71246E-01  0.11527
   E(S)      0.25000       0.35714E-01       14.29        1.0000      0.33108
```

```
   D        0.22656E-01  0.42705E-02      18.85      0.15691E-01  0.32712E-01
   N         297.00       55.983          18.85       206.00        429.00
 ---------  -----------  -----------  --------------  ----------------------
 Component Percentages of Var(D)
 -------------------------------
 Detection probability   :  14.1
 Encounter rate          :  28.5
 Cluster size            :  57.4

                         Estimate      %CV      df      95% Confidence Interval
                        ------------------------------------------------------
                 n        157.00
                 k        555.00
                 L       0.42843E+06
                 n/L     0.36646E-03   10.06   554.00 0.30090E-03  0.44630E-03
                 Left     0.0000
                 Width    24.000

                         Estimate      %CV      df      95% Confidence Interval
                        ------------------------------------------------------
 Uniform/Polynomial
                 m        1.0000
                 LnL     -496.74
                 AIC      995.48
                 AICc     995.51
                 BIC      998.54
                 Chi-p   0.39736
                 f(0)    0.49459E-01    7.07   156.00 0.43017E-01  0.56865E-01
                 p       0.84246        7.07   156.00 0.73273      0.96861
                 ESW      20.219        7.07   156.00  17.586       23.247

                         Estimate      %CV      df      95% Confidence Interval
                        ------------------------------------------------------
 Average cluster size
                         0.25000       14.29   147.00  1.0000       0.33108

                         Estimate      %CV      df      95% Confidence Interval
                        ------------------------------------------------------
 Uniform/Polynomial
                 DS      0.90622E-01   12.30   662.31 0.71246E-01  0.11527
                 D       0.22656E-01   18.85   397.17 0.15691E-01  0.32712E-01
                 N        297.00       18.85   397.17  206.00       429.00
```

To sum up,

$CV_{NE}\widehat{\left({}^{n}/_{L}\right)} = 10.7$ and $CV_{NW}\widehat{\left({}^{n}/_{L}\right)} = 10.06$ so $CV\widehat{\left({}^{n}/_{L}\right)} = \frac{10.7+10.06}{2} = 10.38$

$C\widehat{V\,(P_a)} = 7.07$ and $C\widehat{V\,(P_o)} = 14.29$

Therefore,

$$[C\widehat{V\,(D)}]^2 \approx \left[C\widehat{V\left({}^{n}/_{L}\right)}\right]^2 + \left[C\widehat{V\,(P_a)}\right]^2 + \left[C\widehat{V\,(P_o)}\right]^2 = 10.38^2 + 7.07^2 + 14.29^2$$

$$C\widehat{V(D)} \approx 19.02$$

4. Confidence intervals

Once we have the overall density coefficient of variation, we will need to calculate the respective confidence intervals.

The 95% CI formula for the density (D) used by Distance software, assumes that D follows a log-normal distribution.

95% CI for D: $\left[\hat{D}/_{C},\ \hat{D}C\right]$ where $C = \exp\left\{1.96\sqrt{\log\left(1 + \frac{\widehat{Var}(\hat{D})}{\hat{D}^2}\right)}\right\}$

$\hat{D}$ is known but we need to know $\widehat{Var}(\hat{D})$ in order to build the 95% CI.

$$CV(\hat{D}) = \frac{SD(\hat{D})}{\hat{D}} = \frac{\sqrt{Var(\hat{D})}}{\hat{D}} \quad \text{then} \quad Var(\hat{D}) = \{\hat{D}\ CV(\hat{D})\}^2$$

Therefore,

95% CI for D: $\left[\hat{\mathrm{D}}/\mathrm{C},\ \hat{\mathrm{D}}\mathrm{C}\right]$ where $\mathrm{C} = \exp\left\{1.96\sqrt{\log\left(1 + \frac{\left\{\hat{\mathrm{D}}\ \widehat{CV(\hat{\mathrm{D}})}\right\}^2}{\hat{\mathrm{D}}^2}\right)}\right\}$