Elephant dung density in Sumatra

A. Background

In 2001, the elephant population of Bukit Barisan Selatan National Park (BBSNP) in Sumatra, Indonesia, was estimated from dung surveys. If each elephant produces *p* dung piles per day and dung piles remain visible for *t* days, then dung density will be:

$$S = D \times p \times t$$

where *D* is the density of elephants. If we can estimate *p* and *t*, we can calculate *D* from *S*.

The field research comprised three components (Hedges et al, 2005):

- o defecation rate estimation was based on observations of 12 captive elephants ranging freely in the nearby Way Kambas National Park in 2001-2002: 18.15 dung piles per 24 h with SE 2.53.

- o the time dung remains visible was calculated from monitoring 1302 dung piles in BBSNP for 18 months prior to the survey of dung density: 305.36 days with SE 7.33.

- o dung pile density was estimated by line transect surveys, and we'll see how this was done.

We will start with a look at the line transect data in Excel. Then we'll import the data into DISTANCE and run a first analysis with the default settings in DISTANCE.

Then we will see how to stratify the analysis, to take account of large differences in density in different parts of the park.

We will work on getting a good model for the detection function, trying different mathematical models for the curve and truncating the data set.

Finally, we'll put in the values for dung production and dung disappearance time and get DISTANCE to calculate results for elephant density and population abundance.

B. A first look at the dung density data

Open the file "BBSNP_dung_distances.csv" in MS Excel® (or other spreadsheet program) and check what it contains.

The park was divided into 2 types of zones (2 strata) on the basis of 'recce' walks carried out at the beginning of the project: in the "Low" zones less than 20 dung piles were spotted per km walked, more than 20 in the "High" zones. Most of the transects are in the "High" zone, even though it makes up only about 30% of the total area: we'll see later why this was done.

After the columns for Stratum label and Stratum area, come columns for the Transect label and Transect length. Notice that the transects vary in length; this is taken into account in the analysis.

There is a row in the spreadsheet for each dung pile recorded, with its distance from the centre of the transect line and its decay stage. In our analysis we'll use dung piles of all decay stages, so we don't need this information, but it would be possible to use, say, only dung in stages A through C2 in the analysis.

There is also a row in the spreadsheet for every transect where *no* dung piles were recorded (scroll down to the foot of the file) and the Perpendicular distance and Dung-pile decay stage columns are left blank. It's very important to include these rows, otherwise the total length of transect surveyed will be too low and the estimated encounter rate artificially high.

The data have been sorted so that all the records for each transect are together, and all the transects for each stratum are together. (If necessary, use Data > Sort… and sort by Stratum label first, then by Transect label.) Sorting this way is essential before importing the data into DISTANCE.

Close the file "BBSNP_dung_distances.csv" without saving changes (Excel will try to persuade you to save it in .xls format).

C. Importing the data into DISTANCE

Download and install DISTANCE as described on the wcsmalaysia.org web site (http://www.wcsmalaysia.org/stats/Software_summary.htm#DISTANCE).

Start DISTANCE, go to 'File > New project…' and select the folder where you want to save your project – DISTANCE will start off with the Samples folder, but that probably isn't the best place to save your own projects. Name the project something like "BBSNP_elephants".

DISTANCE saves all your work as you go along, you do not need to save frequently. (On the other hand, if you screw up and do **not** want to save your work, don't just close DISTANCE but select 'File > Revert to Backup Copy' and DISTANCE will use the backup copy created when you started the session.) If you check the target folder now, you will see three new items, files named 'BBSNP_elephants.dst' and 'BBSNP_elephants.ldb' and a folder named 'BBSNP_elephants.dat'.

The Data Import Wizard will start immediately.

Step 1 : Select 'Analyze a survey that has been completed' and click 'Next'

Step 2 : This is an information page; read through it then click on 'Next'

Step 3 : The nest survey was a Line transect survey with a Single observer, where Perpendicular distances were measured to Single objects. Click 'Next'.

Step 4 : The distances from the transect are in Centimetres, the transect length is in Metres and we want the results for dung and elephant density per Square kilometre. Click 'Next'.

In the next step, DISTANCE asks about multipliers. The density of elephants is a multiple of the density of dung piles, and we can get DISTANCE to do the calculations for us. Or we can just get the density of nests from DISTANCE and do the multiplication by hand. However, in the latter case, it is difficult to estimate confidence intervals properly, so we'll let DISTANCE do that for us.

Step 5 : Check the box next to Indirect survey of: and select Dung from the drop-down box. Click 'Next'.

DISTANCE will create fields for Dung production rate and Dung disappearance time and we will need to fill these in later.

Step 6 : Select Proceed to Data Import Wizard and click on 'Finish'.

DISTANCE will create the necessary file structure, and then the Data Import Wizard with start.

Step 1 : This is an information page; check through it then click 'Next'.

Step 2 : In the 'File containing data to import' box, select Files of type: "All files (*.*)" and browse to the file 'BBSNP_elephants.csv' and click 'OK'.

DISTANCE uses a hierarchical structure with

→ a Global layer = study area or topic (in our case BBSNP_Elephants), containing…
    → Regions (our "Low" and "High" zones), containing…
        → Transects, containing…
            → Observations.

Step 3 : The 'Lowest data layer' in our data file is Observation and the 'Highest data layer' is Region. We want to Add all new records under the first record in the parent data layer and to Create one new record for each line of the import file. Click 'Next'.

A pop-up box may appear with a message about assigning fields to columns. By default, DISTANCE expects numbers in the input file to be separated by Tabs. Our file uses commas instead, so DISTANCE thinks it has just one column. Just click on 'OK'.

Step 4 : In our file, the Delimiter is Comma, and we Do not import the first row, as that is just the column titles. (All our values are whole numbers, so the Decimal symbol setting is unimportant.) The correct grid size should now be displayed: 1327 rows x 6 columns. Click 'Next'.

Step 5 : In our case, the Columns are in the same order as they will appear in the data sheet in DISTANCE, so we can use the Shortcut. Scroll to the right and check that all the columns are indeed correctly aligned. You will see that the 'Dung-pile decay stage' column is marked '(Ignore)': that's fine, as we will not need this for our analysis. Click 'Next'.

Step 6 : Select Overwrite existing data and click 'Finish'!

DISTANCE will take a little while to read your data file and get the data into its own database format, then the Data browser will appear (see screen shot below).
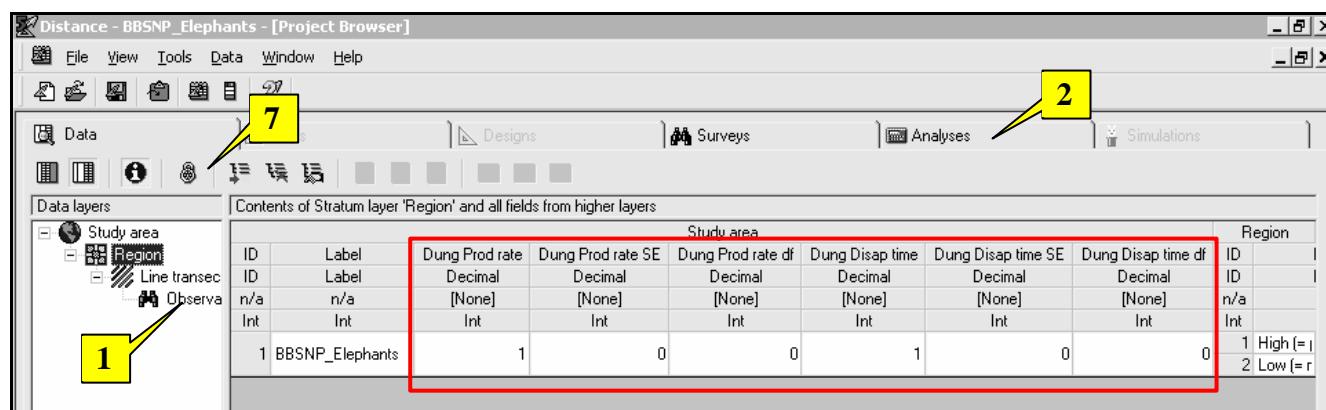
You will see that DISTANCE has created fields for the multipliers 'Dung Prod rate' and 'Dung Disap time' and inserted some default values. If we leave these as they are, DISTANCE will calculate the density and abundance of dung piles: we'll do that for a first run through and put in the proper values later.

Click on "Observations" **1** in the left-hand window to display all the data and scroll right and down to check the contents.

In the lower right corner, you will notice the transects with no observations of dung. The last transect should be no. 58 and the last dung pile no. 1313.

A little higher, you can see that transects 1 – 40 are in the "High" region; 70% of the transects are in the region which we think has a high density of elephants, even though it is less than 30% of the area of the park. This means that we must calculate separate estimates for the two regions and be careful when combining them, ie. we must stratify the data for analysis.

D. A first run through an analysis

We'll first run through a complete analysis just with the default settings, to get an overview of how it works.

Click on the "Analyses" tab **2** to open the Analysis Browser window. Drag the divider between the windows **3** to the right to see the full names. DISTANCE has already created a "New Analysis" with the default settings, and we can run that by clicking on the 'runner' icon **4** .
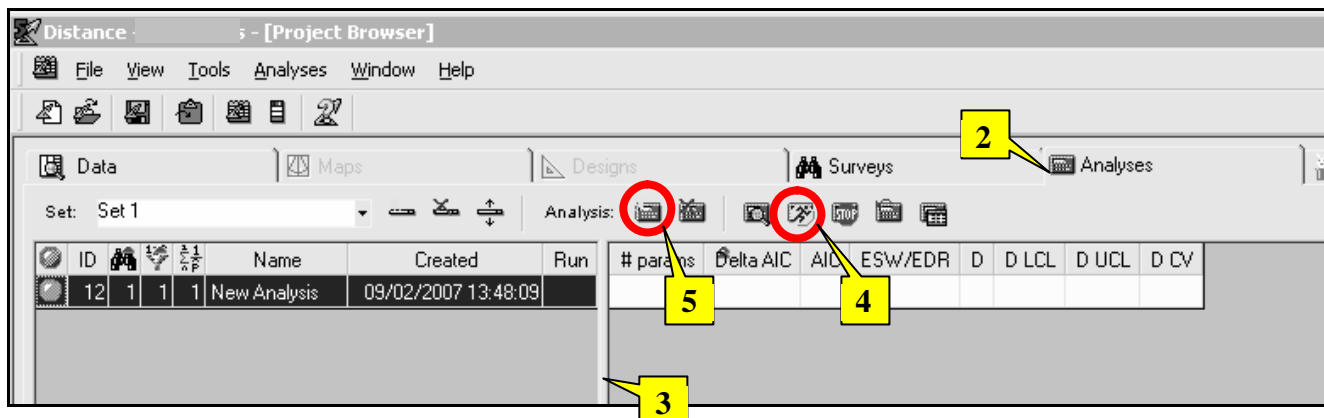
The analysis may take a minute or so to run, during which time your screen may go black – don't panic! When it's finished, the grey bullet on the far left will turn to an orange color and some numbers will appear in the columns in the right-hand panel of the browser:

# params	2	These are used to compare analyses, so we'll discuss them later, when we have more than one analysis.
DeltaAIC	0.00	
AIC	17473.48	
ESW/EDR	432.63	Effective Strip Width (or Effective Detection Radius for point transects)
D	2060.939	Estimate of the density of dung piles
D LCL	1471.018	Lower and upper 95% confidence limits of D
D UCL	2887.434	
D CV	0.170	Coefficient of Variation for D

That all looks reasonable so far. Let's have a look at the details of the analysis.

Double-click the orange bullet on the left to go to the analysis window.

The analysis window opens with the 'Log' tab, which is orange, like the bullet you just clicked. That's because the analysis produced warnings, but in this case it's only a warning that we have too many points for all to be included on the q-q plot (more on the q-q plot later). Nothing here to worry about.

Now select the 'Results' tab (bottom right). Click the "Next" and "Back" buttons at the top right of the window to move between the various results pages.

The first page has general information about the default analysis. Some of this is self-explanatory, some items we'll come back to later. The bottom half of the page has a Glossary of terms and symbols which you may find useful.

On the second page, check the four numbers top left: effort = total length of the transects (73630m); # samples = number of transects (58); width = in this case, the maximum perpendicular distance observed (1500cm); # observations = number of dung piles observed (1313).

Flip through the pages until you find a graph that looks like the one below:
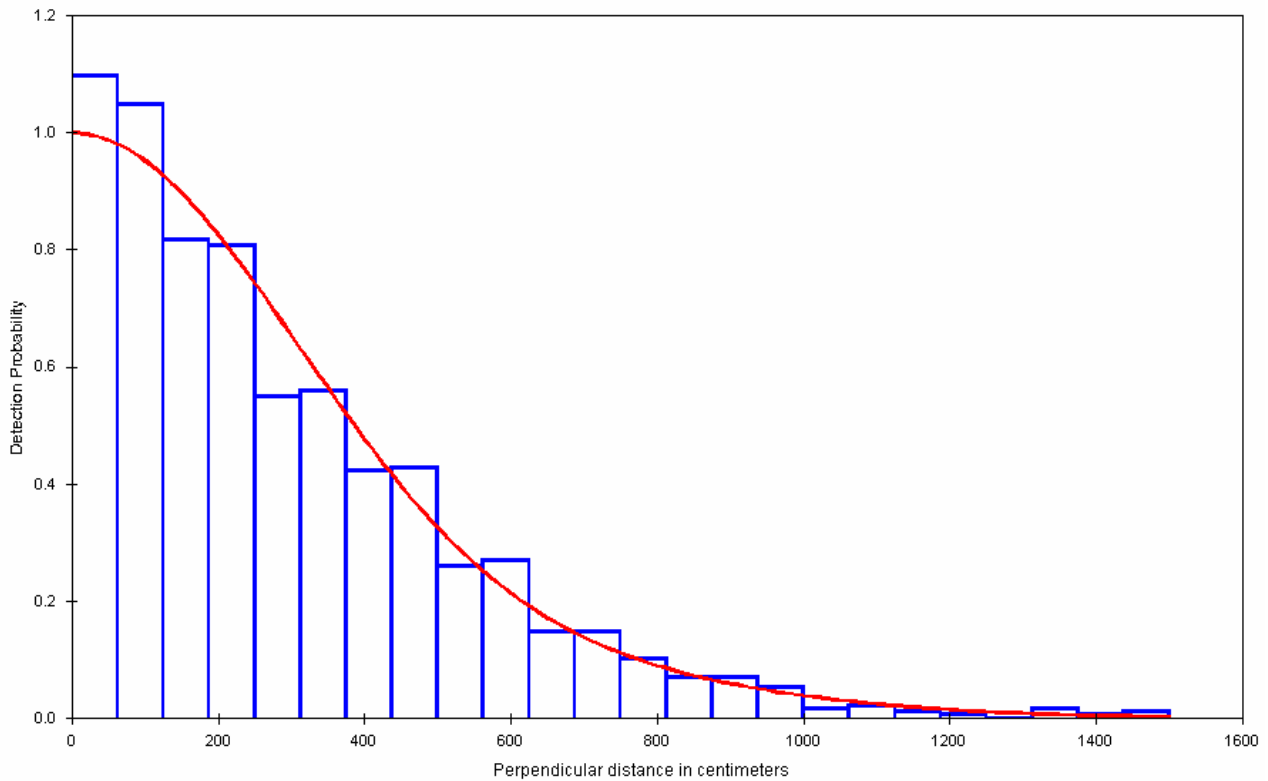
The histogram (blue) shows the distribution of the observations. It shows that the frequency of observations is highest close to the transect, with fewer and fewer observations as the distance gets larger, up to 1000cm (10m). Between 10m and 15m, the pattern is not so clear.

The curve (red) is the fitted detection function, in this case a 'normal' or Gaussian curve – or the right-hand half of a normal curve, to be exact – plus a cosine curve, adjusted to fit the observed frequencies (the histogram) as closely as possible. If you go back to the first page, you will see under the heading "Estimators" the entries "Key: Half-normal" and "Adjustments - Function: Cosines"; these are the default for a new project.

DISTANCE plots two graphs like this with different widths for the histogram bars, and these sometimes show up problems.

The critical part of the curve is the left-hand section, as our estimates of density depend on the point where the curve cuts the vertical axis. In this area, the histogram bars go way above the curve, and the fit does not look good.

Now let's look at the various measures of goodness-of-fit that DISTANCE supplies. You'll see all these if you flick back and forth among the pages of the Results tab:

o chi-squared tests: each of the histograms is followed by a chi-squared table, with actual and expected values for each of the bars of the histogram, with the chi-squared statistic[1] calculated from the difference. For both histograms the probability of getting a greater chi-squared value if the fitted curve is the correct model is very low, ie. our data are a long way from the modeled values. Not a good fit. But note that some of the numbers in the Expected Values column are pretty low: we should not be doing a chi-squared test with expected values < 4, and we should pool the values beyond a distance of 10m into 2 or 3 groups for this calculation. In fact we'll adopt a different solution to this problem, as we'll discuss below.

o qq-plot (below): this compares the actual observations (red dots) with the expected values predicted by the detection function (blue line). If the fit was perfect, all the red dots would lie

---

[1] The chi-squared statistic is named from the Greek letter chi (χ). For each bar of the histogram, calculate

$$(\text{Observed} - \text{Expected})^2 / \text{Expected}$$

then add up the results for all the bars.

exactly on the blue line. In our case, a large number of dots between 0.1 and 0.6 are well away from the line.

Empirical distribution function
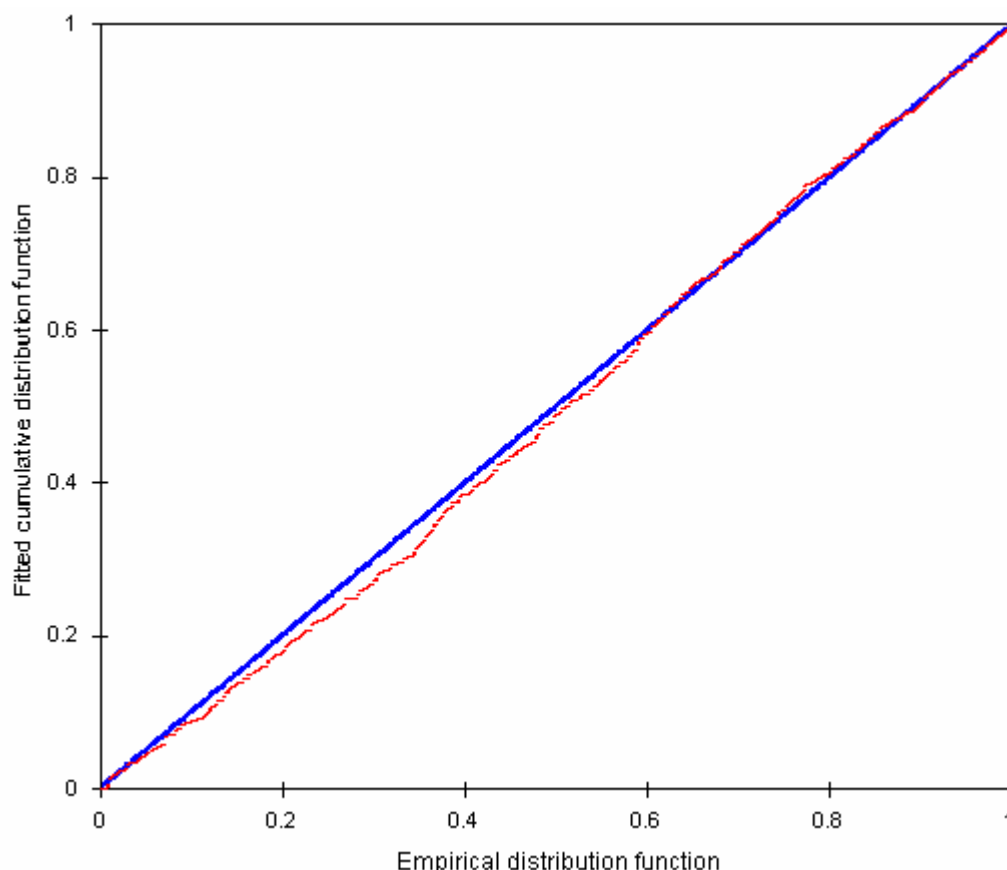
o the Kolmogorov-Smirnov and Cramér-von Mises tests are measures of how far the red dots are from the blue line, K-S using the biggest difference and C-vM taking a weighted average. Both then see how likely such a distance is if in fact the model is correct. A high likelihood (close to 1) means a good fit. The cosine-weighted C-vM test gives greater weight to the fit nearest to the transect centre line, which for us is the most important part of the curve (our density estimate depends on where the curve cuts the distance=0 axis). In this case, all three indicate a poor fit.

The estimates of densities (and total number of dung piles) are calculated from two elements: the encounter rate (n/L) and the effective strip width (ESW).

Go on to the "Estimation Summary – Encounter rates" page. This gives the encounter rate (n/L), which is the number of dung piles per kilometer of transect, plus its confidence interval.

n/L : number of objects (dung piles) observed per unit length (the encounter rate), 1313/73630 = 0.017832 dung piles per metre, or 17.832 per km.

Now go on to the "Estimation Summary – Detection probability" page. This begins with some information on the fit of the Half-normal/Cosine model used to compare models: number of parameters, log-likelihood (LnL), AIC, and two alternatives to AIC, which we'll discuss when we've run a few models to compare. You should recognize $f(0)$ and p from the theory of distance measurements. The figure we want here is:

ESW : the effective strip width; we'd see the same number of dung piles if we'd spotted *all* the dung piles within 432.63 cm of the line, and none beyond.

The last page is "Estimation Summary – Density & Abundance",

D : the density of dung piles, 2060.9 per sq km.

N : the number of dung piles in the whole park, based on the density × the areas we put in for the Regions, 5,587,200.

DISTANCE also estimates variances: you will see that the %CV for the ESW is quite low, and most of the variance in the density (and number of dung piles) is due to variance in the encounter rate. Remember that we're lumping together data from the "High" and "Low" regions, which have different encounter rates – the transects with no dung were all in the "Low" region.

The analysis we've just done ignores the division of the park into two regions and the concentration of survey effort in the "High" region. This is **wrong**.

E. Stratification

The park was divided into 2 regions on the basis of recce surveys done before the line transect surveys.

Ideally we should do separate analyses for the data for each region, but we only have 30 observations from the "Low" region, not enough to calculate a detection curve properly. The forest type is similar in both regions of BBSNP, so there is no reason to expect detection probabilities to differ. So our stratification will only affect the encounter rate.

Click on the 'Inputs' tab of the Analysis Details window for our first analysis.
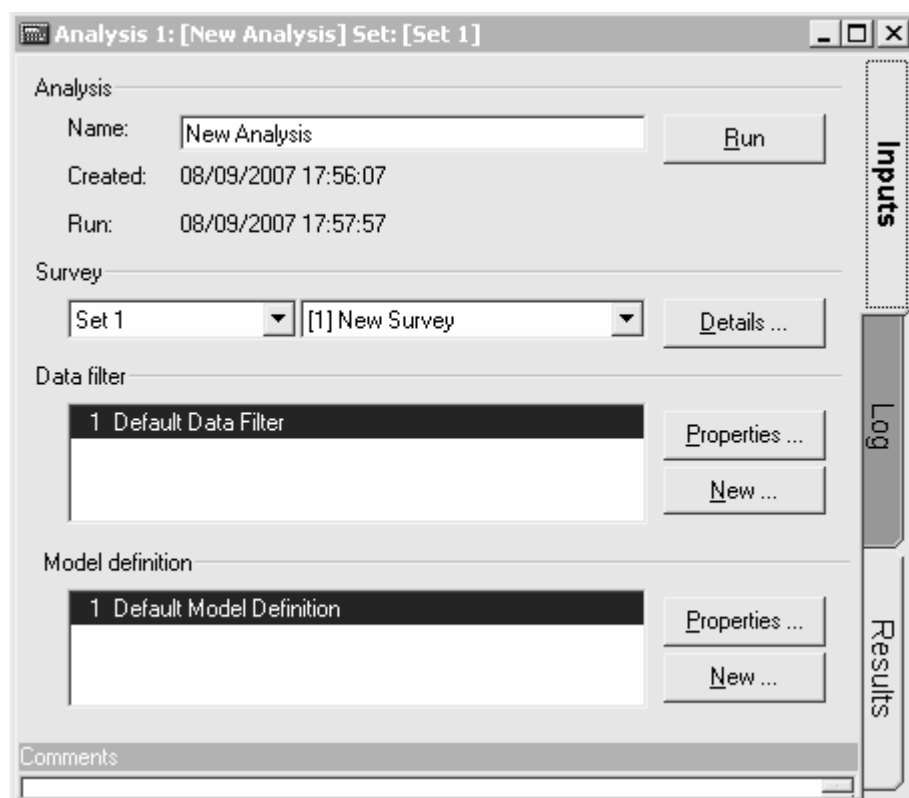
At the top is the name of the analysis, followed by the name of the survey: we only have one survey, so we can ignore this. The two important components are the Data filter and the Model definition, and we are using the default versions of each.

In the Model definition section, click on Properties. In the Model Definition Properties box which opens, go to the 'Estimate' tab (see screen shot below).

Click on the Use layer type: button and make sure that Stratum is selected as the Layer type. Then, under Quantities to estimate… check the boxes as shown: the Encounter rate should be calculated separately for each stratum, but the detection function should be the same for all strata, i.e., it should be global. And we want DISTANCE to produce density estimates for each stratum as well as a global estimate.

Now go to the 'Detection function' tab. Click on the 'Diagnostics' button and change Maximum num points in qq plot to something > 1313, say 1500. Then click on the 'Model' button and you'll see that the Detection function model is Half-normal + Cosine.

Change the Name: of the model (at the foot of the box) to something like "Strat Hnorm+cos". Click 'OK'.

A window will pop up saying that if you change the model definition you will have to run the analysis again. That's not a problem – the unstratified analysis is wrong anyway – so click 'Yes'.

Back in the Analysis Details window, change the name to 'Strat Hnorm+cos', then press 'Run'.

Now there should be no warning about too many points for the qq plot, and the green Results tab will open.

If you now go through the results pages, you will find that the detection function results and the ESW have not changed, as these are based on the global data set. However, you now have separate estimates of encounter rates for the two strata: 0.025 vs 0.0013 dung piles per metre, or 25 vs 1.3 per km. The density is about 20 times higher in the "High" region than in the "Low" region, 2940 vs 149 dung piles per sq km. And the pooled estimate of density on the last page of the results – 945 dung piles per sq km – is closer to the "Low" estimate as this region covers most of the park. As a result, it is lower than our previous unstratified estimate of 2060.

When you've finished, close the Analysis Details window and return to the Analysis tab of the Project browser.

We're now on the right track, but we still have the problem of a poorly fitting detection function.

F. Truncating the data set

If you look back at the histogram-plus-fitted-curve shown above, you'll see that the general downward trend from left to right stops around 10m (1000cm), and the last few observations don't seem to show any clear pattern. These were the same observations that caused a problem with the chi-squared test. Fitting the curve to these few outlying points can distort the results which we are really interested in, which is the area close to the line. The solution is to be brutal and discard these distant observations. In fact, looking at the chi-square table, you'll see there are only a few observations beyond 10m, so we'll truncate at 10m.

We do this with a new analysis using a different Data filter.

In the Analyses tab, click on the 'New analysis' button **5**. A new analysis appears in the main browser window: double-click on the grey bullet next to it to open the 'Inputs' tab of the Analysis Details window.

In the 'Data filter' section of the window, click on 'New…" and a new Data Filter Properties Box opens. Select the Truncation tab (see screen shot below).
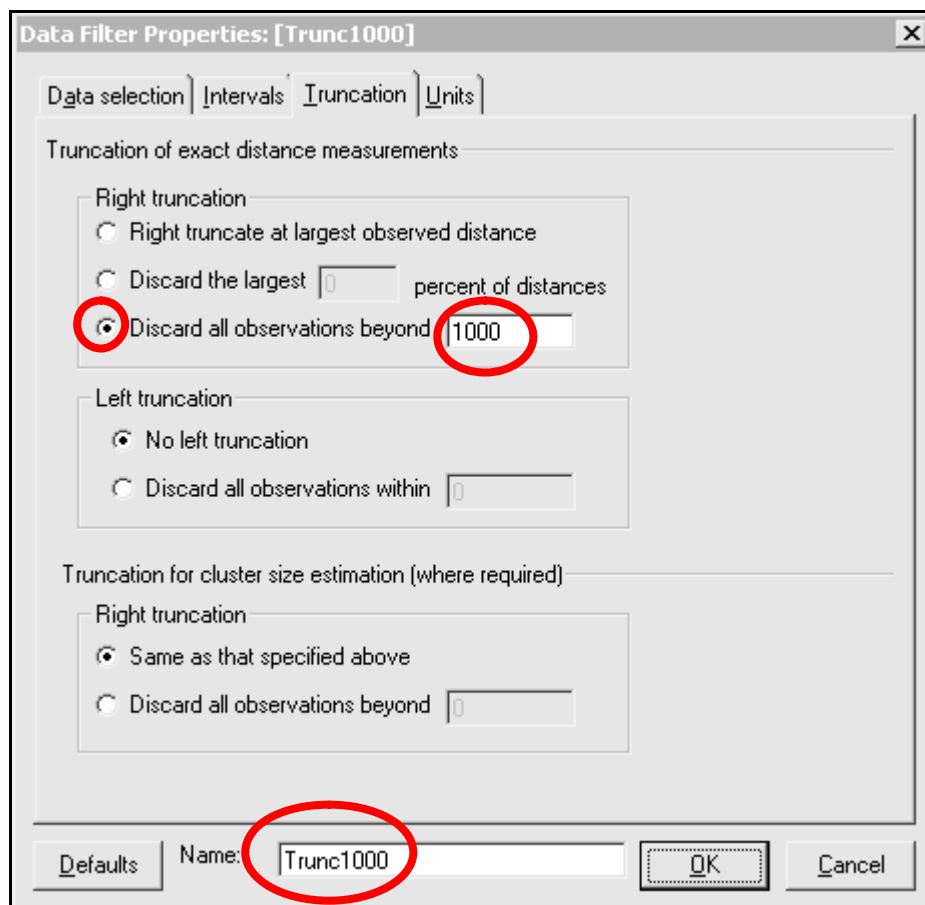
Select the <u>Discard all observations beyond</u> button and enter '1000' in the field. Change the name of the filter to something short but informative, such as 'Trunc1000', then click OK.

Back at the 'Inputs' tab of the Analysis Details window, change the name of the analysis to something like "Trunc1000 Strat Hnorm+cos" and press 'Run'.

There should be no warnings and the Results tab should open. Go to the second page and check the number of observations: 1297 instead of 1313, so we have discarded 16 data points; the width is now 1000cm. The plots of histograms and curves look somewhat better, and the K-S, C-vM and chi-square tests are better too, but still not brilliant. The ESW is 23cm less, and the density estimate has gone up a bit, from 945 to 982 dung piles per square km.

DISTANCE can fit a number of different curves to the data, and the half-normal + cosine model may not be the best. Let's go on to compare different models.

G. Constructing and comparing models

DISTANCE offers a range of models with different mathematical equations for the detection function – the red line in the histogram diagrams in DISTANCE. Look at the graphs in the Appendix to get an idea of what these are like. Each model is based on one of three key functions (uniform, hazard rate or half-normal), each of which has a fairly flat area near the transect line and then drops off[2]. To obtain a better fit, DISTANCE tries adjusting the curve by adding in other terms, either cosines or simple

---

[2] The software can also fit a fourth key function, the negative exponential, but this does not have a flatish section at the left end. That means it gives very imprecise estimates *even if* it is a good fit to the data, and it is not recommended.

polynomial terms (of the form $y = a + bx + cx^2 + dx^3 + \ldots$) or Hermite polynomials, the last being similar to a polynomial but more suitable for adjusting the half-normal curve. The number of extra terms to use is decided automatically; in the example we just used, adding one term improved the fit, but a second one did not, so just one was used. (Look at the second page in the Results tab and you'll see that DISTANCE tried Model 1 with no adjustments and Model 2 with 1 adjustment and found that Model 2 was better. Then it tried Model 3 with 2 adjustments, but Model 2 was better.)

Four combinations are recommended:

- o uniform + cosine (also known as the Fourier series)
- o hazard rate + polynomial
- o half-normal + cosine (the default option we used above)
- o half-normal + Hermite polynomial

We'll begin by setting up the 3 additional models, then set up the analyses later.

Go to the Analysis components window: select View > Analysis Components from the drop-down menus or click on the button on the toolbar just below the menus.

The Analysis Components window has two pages, one for Data Filters, one for Model Definitions.

Select Analysis Components > Model Definitions from the drop-down menus.

The 'Strat Hnorm+cos' model should be highlighted.

Click 3 times on the yellow star (third icon on the lowest toolbar) to create 3 new model definitions.

Highlight the second Model Definition ('Strat Hnorm+cos 1') and click on the pointing-finger button (5th icon on the toolbar) to open the properties box. This is a copy of the 'Strat Hnorm+cos' model, so stratification is already set up. Go to the Detection Function tab and click on 'Models'. (See the screen shot below.) Click on 'Half-normal' and change this to a <u>Uniform</u> key function, leave the series expansion as <u>Cosine</u>, and change the name at the bottom to something like "Strat Uni+cos".
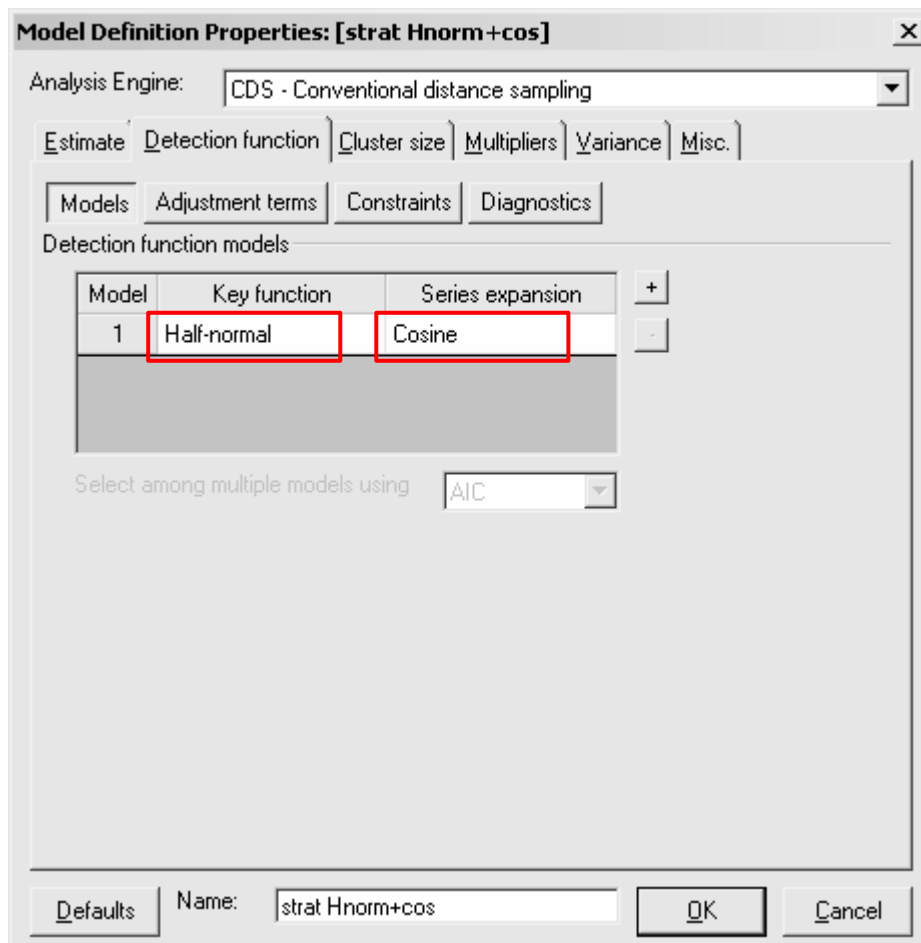
For the third Model Definition, leave the key function as <u>Half-normal</u> and change the series expansion to <u>Hermite polynomial</u>; call this "Strat Hnorm+Herm".

For the fourth Model Definition, change the key function to <u>Hazard-rate</u> and change the series expansion to <u>Simple polynomial</u>; call this "Strat Haz+poly".

Now go back to the Project Browser and the Analyses tab. Highlight the "Trunc1000 Strat Hnorm+cos" analysis and click 3 times on the New Analysis button to create 3 more analyses.
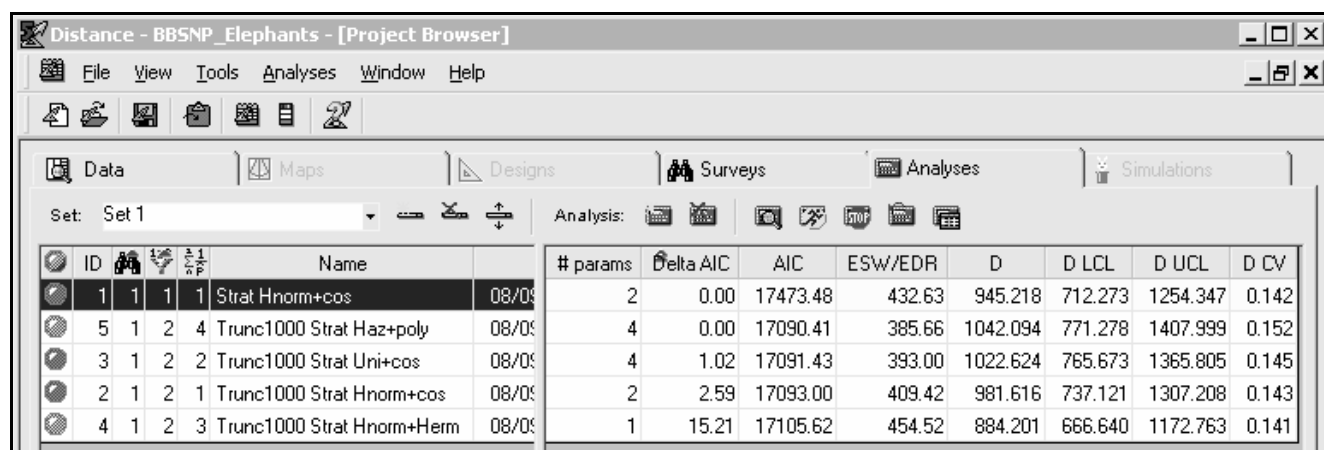
Double-click on the grey bullet next to the first new analysis. On the Inputs tab, highlight the "Strat Uni+cos" model. Change the name of the model to "Trunc 1000 Strat Uni+cos" and click 'Run'.

Similarly, change the other two analyses to use the "Strat Hnorm+Herm" and "Strat Haz+poly" and run them too.

Let's begin with the results in the Analyses tab of the browser (see screen shot below).

The four "Trunc1000" analyses should be in order, with the one with the lowest AIC at the top (if it isn't, click on the heading of the AIC or deltaAIC column). The first analysis with the untruncated data is at the top, also with a Delta AIC value of 0. We cannot use AIC to compare this analysis with the other four, as it uses a different data set. You can now delete it if you wish (use the "Delete selected analysis" button on the toolbar).

The "Trunc1000 Strat Haz+poly" model is the best, with the lowest AIC of 17090.41, but the "Uni+cos" model is close behind. The difference in the density estimates for the two best models is small, about 2% of the value.


If you don't get the same results as this, check the analysis setup. This is a bit complicated and it's easy to get the different filters, key functions and adjustments mixed up. To check, look at the second page of the results, headed "DetectionFct/Global/Model Fitting". Near the top, the Width should be 1000.000 and the # observations 1297 if the data are truncated at 1000cm. Further down is the heading "Model 2" and the next two lines tell you the key being used (Uniform, Half-normal or Hazard Rate) and the type of adjustment (Cosine, Hermite polynomial or Simple polynomial). Check that these correspond to the name of the analysis. Look at the last-but-one page: a model with 2 strata will have 2 density estimates here.

Note that the bullet next to the "Trunc1000 Strat Haz+poly" model is orange, not green, denoting that it ran with warnings. Let's look at the details.

Double-click on the bullet next to the "Trunc1000 Strat Haz+poly" model.

The Log tab opens, with three warnings at the bottom of the screen.

Two warnings are about 'monotonicity': DISTANCE assumes that detection probability declines as distance from the transect line increases. Go to the results tab and flick through the pages to the histogram and look at the right end of the curve (see figure right). The curve rises at the end, implying that detection probability is higher at 1000 cm than around 930 cm, so we
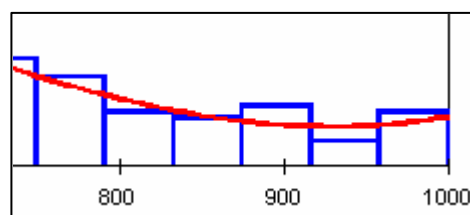
have a problem! Under the magnifying glass, the histogram bars between 800 and 1000 show the same sort of pattern that prompted us to truncate the data at 1000 cm; maybe we should truncate at 800 cm.

The third warning is about a 'lower bound'. Think of a normal (bell-shaped) curve: it is described by two parameters, location (mean) and spread (variance or s.d.). Location can take any value (no bounds), but spread cannot be negative (lower bound = 0). Spread *can* be 0, but only if all the values are identical; that rarely happens in biological problems, so it usually means that something has gone wrong! Now go to the second page of results (Detection Fct/Global/Model Fitting) and scroll down to Model 4. The 'lower bound' warning appears here, and you'll see that the second value in the 'Final parameter values:' line is 1.0000000. In DISTANCE, the 2nd parameter for the hazard rate curve cannot be less than 1, and a value of exactly 1 indicates a problem fitting the curve to the data. However, this is Model 4, which is *not* the one selected: go to the next page (Detection Fct/Global/Parameter Estimates) and you'll see that the estimate for the second parameter is 1.057. The boundary warning doesn't apply to the curve which DISTANCE uses to estimate ESW and dung-pile density.

Look at the details of the next-best model, "Trunc1000 Strat Uni+cos". This has no warnings and the curve doesn't turn up at the end, but looks oddly flat at the end and has an funny wiggle in the middle.

These models are still a bit ugly – even if they satisfy the theoretical criteria – and the observations between 800 and 1000 are causing difficulties. Let's try truncating at 800 cm and see if this improves things.

H. Truncating again

Create a new analysis and go to the input tab. Create a new Data Filter set to <u>Discard all observations beyond</u> 800 and name it "Trunc800". (See section F for details of how to do this.) Give the analysis an appropriate name and run it.

Create three more analyses using "Trunc800", so that you have one for each of the four detection function models. Run all of them.

The data set now includes 1259 observations, so we have discarded 54 observations, 4% of the total of 1313. Buckland et al (2001) suggest that truncating about 5% of observations is often reasonable, though they caution that decisions should be on a case-by-case basis.

In the Analyses tab of the Project browser, note that the "Trunc800" and "Trunc1000" analyses are separated; they are based on different data sets and the AICs cannot be compared. Of the "Trunc800" group, the "Hnorm+cos" analysis has the lowest AIC, closely followed by "Uni+cos". Look at the details of these two and they both look like good fits to the data, with reasonable values for the goodness-of-fit test. The "Haz+poly" model has a warning about boundaries, but this wasn't generated by the model used to estimate ESW and density. "Hnorm+Herm" hasn't worked: adding one cosine adjustment to a basic half-normal curve gives the best model, but adding a Hermite polynomial term is worse than the basic curve.

The output for the "Trunc800 Strat Hnorm+cos" model gives (with a bit of rounding):

- o dung-pile density 1011 km^{-2} with a 95% confidence interval of 760 to 1350 and %CV 14
- o total dung piles in the park 2.74 million with a CI of 2.06 to 3.65 million and %CV 14.

I. Converting dung density to elephant density

We can now estimate elephant density and population using estimates of dung production rate (18.15 per elephant per day) and dung disappearance time (305.36 days):

Elephant density = 1011 / (18.15 x 305.36) = 0.182 elephants per sq km

Elephant population = 2,741,500 / (18.15 x 305.36) = 494.65, say 495 elephants

However, calculating variance and confidence intervals by hand is not so easy! To do that, it's best to incorporate the multipliers into the estimation in DISTANCE.

When we imported data into DISTANCE, we specified that we wanted to use dung production and dung disappearance times as multipliers, and DISTANCE created fields for these variables and inserted default values (see the screen shot on page 3). But we did not insert our values, and carried out the analysis with the defaults. This enabled us to calculate the density of dung piles and simplified the discussion of the analysis.

Now we'll put in the multipliers and rerun the analyses to get estimates of elephant densities and population.

Go to the Data tab in the Project browser (screen shot on page 3) and scroll to the left side of the data sheet.

The 'Lock data sheet' button – the 4th button on the lower toolbar, with a padlock icon 7 – should be depressed. Click on this to unlock the data sheet.

Double-click on the '1' in the Dung Prod rate column, type in the value 18.15 and press Enter.

Repeat for the other values:

Dung Prod rate	Dung Prod rate SE	Dung Prod rate df	Dung Disap time	Dung Disap time SE	Dung Disap time df
18.15	2.53	11	305.36	7.3	1300

Now click on the padlock icon again 7 to lock the data sheet and prevent any changes being made accidentally.

Now go back to the Analyses tab.

Nothing has changed, the values displayed are the densities for dung piles. We need to rerun all the analyses.

Highlight all the analyses in the left side of the Analyses window (use Shift-click) and then click on the 'run' button (or select Analyses > Run analysis from the pull-down menus).

The figures for # params, Delta AIC, AIC and ESW/EDR have not changed, as they relate to dung detection. But the figures for density (D, D LCL, D UCL and D CV) are now *much* lower.

Let's look at the details for the best model once again, the "Trunc800 Strat Hnorm+cos" model. The first few pages deal with the detection function, and are unchanged. The last page gives the pooled estimate for the whole park, 495 elephants in total, but quite wide 95% confidence limits: 331 to 738.

The next-to-last page gives the corresponding figures for each of the two strata (regions). Densities are indeed much higher in the "High" region (0.57 km^{-2}) than in the "Low" region (0.03 km^{-2}). Also, most of the 495 elephants (89%) are in the "High" region, which constitutes less than 30% of the area of the park.

Notice that the %CV for the "Low" region (43.85%) is more than double that in the "High" region (20.79%). The same detection function, production rate and disappearance time were used for both regions, and the difference is due to the estimates of encounter rate; go back to the 'Estimation Summary – Encounter rates' page and you'll see the difference in %CVs. This is because more survey effort was devoted to the "High" region; look at the value for k, the number of transects, in each region: 40 in the "High" region and only 18 in the "Low" region. Why was the survey designed this way? Wouldn't it be better to devote equal effort to the two regions and try to even up the %CVs?

Look again at the numbers of elephants: 440 vs 55. 20.79% of 440 is 91 elephants, while 43.85% of 55 is a lot less, only 24 elephants. So when we add these together, most of the imprecision in the total comes from the 440, not the 55. In general, when a total population estimate is the sum of a big number and a small number, you want to put most effort into getting a precise value for the big number; the small number is less important.

But if you design the survey with most effort in the region with a high density of animals, it's *essential* to include stratification in the analysis. If you don't – and we didn't when we ran through the first analysis with DISTANCE's default – you are implicitly giving most weight to the region with most

effort, and the overall result will be *wrong*. (The result of our unstratified analysis was double the stratified result.)

To sum up: we estimate the elephant population in BBSNP at 495 animals, with 95% CI of 331 to 738. The population is concentrated in the "High" region, which makes up about 30% of the park, where the population is estimated at 440 animals (95% CI 290 to 667).

These figures are close to the results published by Hedges et al (2005): 498 animals, with 95% CI of 373 to 666. The difference – in particular our somewhat wider confidence interval – are probably due to changes in the software, as Hedges et al were using a previous version of DISTANCE.

J. Getting finished

DISTANCE automatically saves each of the Filters, Model Definitions and Analyses and the results as you create them or run them, so there is no need to save your work before closing. If you do **not** want to save the work you have done during the current session, you can select 'File | Revert to Backup Copy' and DISTANCE will use the backup copy created when you started the session.

To close DISTANCE, select 'File | Exit' or click on the ☒ button in the top right corner of the main DISTANCE window. When the "Are you sure…" box appears, click on 'Yes'.

Acknowledgements

Thanks to Simon Hedges and WCS Indonesia for providing the data, and to participants in our Workshop at Krau in September 2007 for piloting the lab guide and pointing out errors.

References

Buckland, S T; D R Anderson; K P Burnham; J L Laake; D L Borchers; L Thomas 2001. *Introduction to distance sampling: estimating abundance of biological populations*. Oxford University Press, Oxford, UK.

Hedges, S; M J Tyson; A F Sitompul; M F Kinnaird; D Gunaryadi; Aslan. 2005. Distribution, status, and conservation needs of Asian elephants (*Elephas maximus*) in Lampung Province, Sumatra, Indonesia. *Biological Conservation* **124**:35-48.

Tyson, M; S Hedges; A Sitompul; Sukatmoko; D Gunaryadi. Defecation rates of Asian elephants in Sumatra, and implications for elephant surveys in Southeast Asia. *J Wildlife Management* **in review**.

	Uniform + Cos	Hazard rate + polynomial	Half-normal + cosine	Half-normal + Hermite
0				
1				△
2		△		
3		△		●
4		△		●
5				

● = Generated an error message (negative variance).

△ = Generated warning messages

Detection functions for the elephant dung data set