

# Introduction to Distance Sampling

## Automated Survey Design Exercises

### 1. Point transect survey of North-eastern Mexico

#### Reviewing the data

Extract and open the project **MexicoUnPrj** from the archive **MexicoUnPrj.zip**. This project contains data from 4 states in North-eastern Mexico. Let's begin by reviewing the data. On the top menu-bar select **File, Project Properties**. The **General** tab gives you information about the location of the project file and its associated data folder (MexicoUnPrj.dat). The **Geographic** tab gives you information about the default geo-coordinate system of the geographic data, and the default map projection. The geo-coordinate system is used to locate the geographic data (which is stored in decimal degrees of latitude and longitude) on the earth's surface. The projection is used to convert these data from the curved surface of the earth into a flat plane that can be used for displaying maps and designing surveys. The resulting projection has linear units, such as metres or kilometres. If you are planning a survey that will take place over a small geographic area, and you are inputting your data by hand, then you don't need to worry about geo-coordinate systems or projections and can set both these options to [None]. In this example, however, the survey area is quite large and the projection chosen will make some difference to the results. Click **Cancel** to close the Project Properties window without saving any changes you have made.

Click on the **Data** tab of the **Project Browser** to view the **Data Explorer**. In the left-hand pane, under Data Layers, you can see that there are four layers in the project: "Mex", "MexStrat", "Grid1" and "Grid2". You can tell the layer types by looking at the icons beside the names: Mex is a Global layer, MexStrat is a Stratum layer and Grid1 and Grid2 are Coverage layers. When you open a new project, the Global layer is selected by default, so the layer Mex is now selected. Click on the **Data Layer Properties** button on this tab (7th button from left) to find out more about this layer. The **Layer Properties** window opens, and under the **Geographic data** tab, you can see that the geographic data is stored in a shapefile called Mex.shp in the data folder, and that the shapes in this layer are Polygons (i.e. solid shapes). Click **Cancel** to return to the Data Explorer.

In the right-hand pane of the Data Explorer, you can see its fields: ID, Label, Area and Shape. There is one record, with ID = 1. The Shape field holds the geographic information for that record. Because this layer holds polygons, the shape record has the word "Polygon" in it. Double click on this word to open the **Shape Properties** window. This is where you edit the geographic information inside Distance (an alternative is to edit the shapefile Mex.shp from outside of distance, using a GIS package such as ArcGIS). Click **Cancel** to return to the Data Explorer.

The coverage layers Grid1 and Grid2 contain a grid of points that will be used for determining probability of coverage for our survey designs. If you click on "Grid 1" in the left-hand pane of the Data Explorer, its records open in the right-hand pane, and you can see that it has 177 records. A better way to look at the grid points is to view them in a map. Click on the **Maps** tab in the **Project Browser**. Click on the **New Map** button (3rd button along) to create a new map. Double-click on the words "New Map" to edit the name of the map, and call it "Grid 1". To view the map, click the **View Map** button (5th button along), or double click on the map's ID. A **Map Window** opens.

The map starts life blank. You add layers to the map by clicking the **Add Layer to Map** button (7th button along). Click this button and select "Mex" from the list of layers. Then click the button **Add Layer** button again and select "Grid 1" from the list. Now you can see the grid points.

You could also add the points from Grid 2 to the same map. If you do this, you will see that the grid points for Grid 2 are much closer together than those for Grid 1. (Grid 2 was generated with a spacing of 20 km, while for Grid 1 the spacing was 50km.) The points for Grid 2 obscure those from Grid 1 – you can change the order of the map layers by clicking on a the legend "Grid 2" in the left-hand pane of the map and holding the mouse button down while you drag it down to below "Grid 1". Click the [X] button in the top right corner of the Map Window to close the map. (Say "Yes" if it asks you to save changes.)

In the **Data Explorer**, click on the MexStrat layer to see those data. You can see that there are 4 strata. If you want to see where they are, you could create a new map in the **Maps** tab and add the MexStrat layer to the map. If there are layers on the new map that you don't want, you can remove them with the **Delete Selected Layer** button in the **Map** window.

When you've finished exploring the data, move on to create a new design.

### Creating a new design

Click on the **Designs** tab of the **Project Browser**. To create a new design, click the **New Design** button (1st one after the word "Design:"). A new record appears in the left-hand pane, called "New Design". Double click on the name, and edit it to call the new design "150 random points". If you need more room, click and drag to the right the vertical splitter that divides the Designs window into two. Click the **Show Details** button (3rd one after the word "Design") to open the **Design Details** window. Look under "Type of design" to see the sampler and design class; the default sampler is "Point" and the default design class is "Simple Random Sampling". Click the **Properties** button to set the properties for this design. The **Design Properties** window opens. The options you see on the design properties tabs depend on the type of design. In this example, choose the following options:

- Under Stratum layer, choose the stratum layer "MexStrat".
- Under Design coordinate system, make sure the box "Same coordinate system as stratum" is unchecked. The projection should say "Plate Carree" and the units "Metre".
- In the **Effort Allocation** tab, under Edge Sampling select the "Plus" option. Uncheck the box "Same effort for all strata". A list of the four strata in the MexStrat layer appears. Under "Allocation by stratum", click the "Percentage from" radio button, and enter "150" as the number of points. In this example, we will put most of our effort into the two Baja strata (perhaps because this is where we think most of the animals of interest live). Under "Effort %" enter 10 for Sinaloa, 10 for Sonora, 40 for Baja Sur (south) and 40 for Baja Norte (north).
- In the **Sampler** tab, select Kilometre for the point sampler radius units. Let's imagine we're surveying for a very vocal species and that our truncation distance will be 5km, so we enter 5 under radius (for this example we'll assume same sampler properties for all strata).
- Lastly, in the **Coverage Probability** tab, click on "Estimate by simulation" and enter 100 as the number of simulations. This is far too few for an accurate simulation, but will do for the purposes of demonstration. Under grid layer, choose "Grid 2", which is the one with the grid points closer together.

Now click OK to close the **Design Properties** window. The properties window closes and we are back with the design details.

### Automated generation of new surveys

Click the **Run** button on the **Design Details** window. A window pops up offering you two choices: (1) Calculate coverage probability statistics, and (2) Generate a new Survey. Choose the second option, and give the new Survey a useful name like "150 points survey" and the new layer a name like "150 points". Then click OK. A **Survey Details** window opens, and the status bar at the top of the Distance window says "Running Survey". At this point you have to be patient while the survey runs. Distance is creating a set of randomly located survey points, based on the design. When it is finished, the **Survey Details Results** tab opens, and you can review some statistics about the new survey. Click the "Next >" button to see a map of the points – you should be able to see that there are more in the

Western strata (Baja) than the eastern. Click "Next>" again to see a list of the points, with latitude and longitude for each. (You could, for example, use this to make a survey plan to take into the field. To copy this text to another file, press the "Copy current window" button, 4<sup>th</sup> from the left on the top toolbar. Then open, say, a Word document and click Paste to copy it there. You can also copy the map of points by displaying the map and pressing the copy button, or choosing the menu item Survey – Results | Copy Map to Clipboard)

Click on [X] to close the **Survey Details** window, and click on the **Surveys** tab of the project browser. You can see that your new survey has been added there. If you select it and click the "Show Details" button (3rd from left after the word "Survey") you get back to the **Survey Details** window **Results** tab. Click on the **Inputs** tab and then **Properties ...** button. Under **Data Layers**, you can see that the new Sample data layer "150 points" has been entered as the lowest sample layer. Close the **Survey Properties** and **Survey Details** windows, and click on the **Data** tab of the **Project Browser**. You can see that the new sample data layer "150 points" has been added below the "MexStrat" data layer.

### Design statistics

Go back to the **Design Details** window for your design, and click **Run** again. This time, choose the top option (Calculate probability of coverage statistics) and click OK. You have to wait while Distance generates multiple simulated surveys and uses these to work out the probability that each grid point will be covered by the survey. When it has finished, you can see the results in the **Results** tab, and a map of coverage probability by pressing the "Next >" button. In theory, this design should produce an even probability of coverage within stratum. However, you can see that there is considerable variation. Why is this? What would happen if you repeated the run with more simulation runs (say 500, or 1000)? (Before you spend a lot of time running simulations with this project, read the next section.)

### Working with projected raw data

There is a technical problem with the way the geographic data are stored in MexicoUnPrj. Each time you view a map or run a survey design, the data have to be projected from the latitude and longitude format in which they are stored using the projection you have specified (Plate Caree in this case). This takes some computer time, so if you're doing lots of survey design work there's a trick to make things more efficient. The trick involves projecting the raw data files.

We used ArcGIS to project the shapefiles in MexicoUnPrj using the Plate Caree projection, and stored this new data in the project MexicoPrj. So rather than being stored in latitude and longitude, the data in MexicoPrj is stored in meters. Run a second instance of Distance, and then extract and open the project **MexicoPrj**. Look under the **Project Properties**, and you will see that the GeoCoordinate system and Projection are both set to [None], and that the units are meters. So, we've projected the raw data, and so long as we project all the data layers the same way we don't need to tell Distance anything about the coordinate systems used.

As a check that the data really are projected, go the Data Explorer and double-click on the global layer's Polygon record. The first value is something like x= -12594701 y=3230255 – this gives the number of meters of that point on the polygon from some reference point on the earth. If you do the same thing in the MexicoUnPrj project, you'll see that the first value is something like x=-113 y=29, which is the latitude and longitude of that point.

If you're going to do lots of experimenting with the Mexico data this evening, or at home, it's better to use the MexicoPrj project, as you'll find the probability of coverage simulations run quite a bit faster. Meanwhile, move on to the next exercise.

## **2. Entering geographic data into Distance, and generating Coverage grids**

The purpose of this exercise is to show you how to enter geographic data by hand into Distance, and how to generate Coverage grid layers.

### **Create a new project and enter data**

On the top menu-bar select **File, New Project** (or click the toolbar button). In the **Create New Project** dialog box give it the File Name "Trapezium" and then click on **Create**. The new project setup wizard starts up. Under "I want to", select the option to "design a new survey", and click **Next**. Then click **Finish**.

The **Project Browser** will open up, showing the **Data** tab. Click on the menu **File | Project properties**, and look under the **Geographic** tab to confirm that there is no geographic coordinate system for this project (i.e. non-earth referenced), and that the default units are metres. Click **OK** to close the **Project Properties** window.

In the **Data** tab of the **Project Browser**, you can see that Distance has created a global data layer called Study Area, with default fields ID, Label and Shape. Double click on the word "Polygon" to open the **Shape Properties** window to edit the new survey region. Click on the **Insert Point** button 4 times and fill in the following (X,Y) coordinates: (0,0), (0,100), (120,20) and (120,0). Click **OK** to return to the Data Explorer.

### **Generate a coverage grid layer**

To generate a coverage grid layer click on the **Create New Data Layer** button (5<sup>th</sup> from left) in the **Data** tab of the **Project Browser**. Enter "TrapGrid" as your Layer Name and set the Parent Layer to "Study Area" and the Layer Type to "Coverage". You should now be able to click on the **Properties...** button. In the **Grid Properties** that pops up set the "Distance between grid points" to 2.5 and the "Units of distance" to "Metre". Once you press **OK** you should proceed to add the grid points to the layer. This may take a few moments.

Create a new map on the **Map** tab of the **Project Browser** and add your new global and coverage layers to take a look at them.

### **Creating a new design**

Click on the **Designs** tab of the **Project Browser** and then the **New Design** button. Rename your design "equal angle zigzag" and then click the **Show Details** button to open the **Design Details** window. Select the "Line" sampler and set the design class to "Equal Angle Zigzag". Click the **Properties** button to set the following properties for this design:

- As the Trapezium survey region is non-earth referenced you don't need to make any changes on the **General Properties** tab.
- In the **Effort Allocation** tab, under "Effort determined by" select the Sampler angle option. In the "Allocation by stratum" section set the Line length units to be Metres. Make sure the "Update effort in real time" check box is ticked. As there is only one survey stratum it does not matter whether the "Same effort for all strata" check box is ticked or not. The "Absolute values" radio button is the only one available when effort is determined by sampler angle. Enter 75 in the "Angle" (measured in degrees) column of the table. The "Length" column should now read 463.644. The accuracy of this approximation of zigzag length depends on the shape of the survey region, but should be relatively accurate for convex survey regions.
- In the **Sampler** tab, set the width to 1 meter.
- Lastly, in the **Coverage Probability** tab, click on "Estimate by simulation" and enter 100 as the number of simulations. Under grid layer, choose previously created "TrapGrid".

Click OK to close the **Design Properties** window and return to the design details.

## Design statistics

Run your design to work out the coverage probabilities - this design will take a while to run! In the second page of the **Design Details Results** tab that opens when its finished, take a look at the coverage probabilities map. Note how uneven these probabilities are and how they increase as the trapezium height decreases for the equal angle zigzag design.

## Additional investigations

If you are particularly interested in zigzag surveying, you might want to come back to this exercise after completing exercise 3, and compare the coverage probabilities of the three different types of zigzag designs. You can do this when you get home. For now, skip ahead to exercise 3.

For work on your own:

Create two new designs - one for the equal spaced zigzag and one for the adjusted angle zigzag. To facilitate comparisons, you want to set properties for both that are somewhat equivalent to those for the equal angle design. You can see the mean trackline length for the equal angle design in its Results tab (about 460 metres). You can then set the effort allocation for the two new designs to be the same as this value. Make sure that the Coverage probability tab shows "Estimate by simulation" and that you have an appropriate Grid Field Name.

Try creating a few surveys for each design, so you can see how they differ. Then run the coverage probability simulations. How do the coverage probabilities for the 3 designs differ? You may need more simulations to see a strong difference between the equal spacing and adjusted angle design.

## 3. *Systematic parallel line aerial survey of marine mammals in St Andrews bay*

### Reviewing the data

Open the project archived in **StAndrews.zip**. This project contains the survey region for an aerial survey of porpoise, common dolphins and seals in and around St Andrews bay. (For locals: the nearer St Andrews bay region has been extended in an easterly direction out past Bell Rock, as there are some pockets of deeper water out there that are of interest with regard to the distribution of cetaceans. The survey region has a chunk missing due to a no-fly zone around Buddo Ness, just below Carnoustie). To take a look at the survey region create a new map in the **Maps** tab and add the layer **StAndrews** to the map.



The small survey plane permits a total flight time of approximately 250 km (excluding the flight time to and from the landing strip in Fife Ness, just down the coast). A systematic line sampling design is going to be used. The design should maximise the amount of on effort surveying time whilst ensuring that the constraint of 250 km total effort (including off effort transit time between sampler lines) is achieved. The aim of this exercise, therefore, is to decide on a systematic line spacing that gives about 200 km on-effort trackline with the total trackline length constrained to 250 km. To do this create a number of systematic line sampling designs with different line spacings, generate the design statistics for these designs and then the statistics for the total trackline length to the on-effort trackline length for different designs.

Before proceeding to the design stage you need to generate a coverage grid layer, as this will be needed to generate design statistics.

## Generate a coverage grid layer

To generate a coverage grid layer click on the **Data** tab of the **Project Browser** and then the **Create New Data Layer** button (5<sup>th</sup> from left). Enter "Grid5" as your Layer Name and set the Parent Layer to "StAndrews" and the Layer Type to "Coverage". You should now be able to click on the **Properties...** button. In the **Grid Properties** that pops up set the "Distance between grid points" to 5 and the "Units of distance" to "Kilometre". (This is too far apart for estimating probability of coverage, but we know coverage is even for this design, so choosing a wide spacing makes the simulations run faster.) Once you press **OK** you should proceed to add the grid points to the layer. This may take a few moments.

Now create and generate a couple of designs with a spacing of your choice (some suggested spacings include 4.5, 5, 5.5 & 6 km)

## Creating a new design

Click on the **Designs** tab of the **Project Browser** and then the **New Design** button. Rename your "New Design" something like "systematic line test" and then click the **Show Details** button to open the **Design Details** window. Select the "Line" sampler and set the design class to "Systematic Random Sampling". Click the **Properties** button to set the following properties for this design:

- On the **General Properties** tab under Stratum layer, the StAndrews stratum layer should be selected. Under Design coordinate system, the design coordinate system should be "Non-earth referenced". (The data have already been projected from the OSGB 1936 geo-coordinate system using the transverse mercator projection – the same trick we used for the MexicoPrj project.)
- In the **Effort Allocation** tab, under Edge Sampling select the "Minus" option. In the "Allocation by stratum" section set the Line length units to be Kilometres. Make sure the "Update effort in real time" check box is ticked. As there is only one survey stratum it does not matter whether the "Same effort for all strata" check box is ticked or not. Click the "Systematic line spacing" radio button and enter the line spacing in the "Spacing" column of the table. When you enter a 5 km spacing for instance the "Length" column should then read 226.203 and the "Samplers" column 8. The accuracy of this approximation of on-effort line length and total number of line samplers depends on the shape of the survey region, but should at least give you some indication of what to expect.
- In the **Sampler** tab, select Kilometre for the line sampler width units. Set the truncation width to 2 km.
- Lastly, in the **Coverage Probability** tab, click on "Estimate by simulation" and enter 100 as the number of simulations. This is too few to give accurate coverage probabilities, but sufficient for the on-effort and total trackline length statistics. Under grid layer, choose previously created "Grid 5". Make sure the Grid field name is the same as your design name.

Click OK to close the **Design Properties** window and return to the design details.

## Design statistics

For each design run Distance generates multiple simulated surveys and uses these to work out the statistics for on-effort and total trackline length. Run your designs and in the **Design Details Results** tab that opens review the statistics to decide on suitable systematic line spacing.

## Automated generation of new surveys

To see an example survey, go back to the **Design Details** window for your selected design click **Run** again this time choosing the "Generate a new Survey" option. The second page of the survey results displays a map of the survey region with the systematic lines superimposed. You can add this map to the **Map browser** and manipulate it there by clicking on the 6<sup>th</sup> button on the Survey map results page.