# 16. Data analysis in DISTANCE – Part 2

## 2. Extending the scope of your analysis (continued)

We can now look at:

- Analysing different colours separately
- Adding covariates to the detection function model

### *2.3 Analysing different colours separately*

Now try the above for each colour separately. You need to create new a **Data filter** for each colour (and each truncation).

To select just one colour, e.g. blue, under the **Data selection** tab, click on the + button on the right to add a selection criterion. Under **Layer type** select **Observation**. Double click under **Selection criteria** and type **Colour = 'B'**. Only observations of blue plugs will be analysed. Give your new Data filter a **Name**, e.g. "Blue no trunc", and click **OK**.

Choose your Model definition with both Half-Normal and Hazard-rate. Give your Analysis a new name.

Try truncating the perpendicular distance data.

Try this with all three colours: **Colour = 'R'** for red and **Colour = 'G'** for grey.

Store your analyses systematically in different **Sets**.

**Question 3**: Which are the best models so far for each colour?

(a) Blue

(b) Red

(c) Grey

### *2.4 Adding covariates to the detection function model*

Detection probability is likely to vary for a variety of reasons, e.g. sighting conditions, observer, group size, the cue given by the animals. If we have data on these factors we can include covariates representing them in the detection function to see if they improve model fit.

For one of the colours, highlight your currently favoured analysis in the left panel of the **Project Browser** (e.g. "Blue no trunc HN"), select **New Analysis** and bring up the Analysis Details screen again. Select a **New Model definition**. This time we need to change the **Analysis Engine** at the top to **MCDS – Multiple covariate distance sampling**.

Select the **Estimate** tab and set your stratification options as before (see section 1.2).

Under the **Detection function** tab, select the **Models** and **Adjustment terms** tabs to check your model(s).

Adding Team as a covariate

Now explore whether adding Team as a covariate improves the detection function. Select the **Covariates** tab. Click on the + button on the right to add a **Detection function covariate**.

Select **Sample** under **Layer type containing covariate**. Then select **Team** under **Field name of covariate**. Check the **Factor** box because Team is a factor covariate not a measurement.

Under the **Cluster size** tab, check your size-bias regression options, as before (see section 1.2).

Give the Model definition a new name e.g. "Half-normal + Team". Click OK, give your analysis a new name, e.g. "Blue no trunc HN + Team", and run the analysis.

The **Results** now include pages showing how the detection function varies with Team.

**Question 4**: Does the detection function vary much with Team?

Does adding Team as a covariate improve the model fit?

Adding Observer as a covariate

Now create a new Model definition to try Observer. This time you need to select **Observation** under **Layer type containing covariate**. Again, check the **Factor** box, give the Model definition a new name, give your Analysis a new name and run the analysis.

The **Results** now include pages showing how the detection function varies with Observer. Some of these will have small sample sizes.

**Question 5**: Does the detection function vary with Observer?

Does adding Observer as a covariate improve the model fit?

Do this for all three colours to find the best detection function for Blue, Red and Grey.

**Question 6**: Which are the best models for each colour?

(a) Blue

(b) Red

(c) Grey

### *2.5 Including group size as a covariate*

DISTANCE calculates the estimates in a different way if group (cluster) size is a covariate because it is not appropriate to let group size influence detection probability and then simply multiply density of groups by mean group size to get density of individuals. Instead, the density of groups and density of individuals are both estimated directly using something called a Horvitz-Thompson estimator. Group size is then estimated as density of individuals divided by density of groups.

**Note**: If you include group (cluster) size as a covariate, DISTANCE is unable to calculate a stratified estimate so you will be limited to an unstratified estimate.

Another thing that is different with using group (cluster) size as a covariate is that the variance of the estimates needs to be estimated using a bootstrap procedure. This is set up under the **Variance** tab in the **Model definition**.

Try one (or more) of your models with group size as a covariate.

The Results pages will include figures of the fitted detection functions for a selection of group sizes. Check that these make sense – you would expect detection probability for larger groups to be greater than for smaller groups, especially at greater perpendicular distances.

### *2.6 Using Bootstrap to estimate variance*

To estimate variance using a bootstrap procedure, select the **Variance** tab under **Model Definition Properties**. Check the **Select non-parametric bootstrap** box under **Bootstrap variance estimate**. Leave the **Resample samples** box checked (this means that transects will be used as sampling units). Under **Bootstrap options**, set the Number of resamples to 100.

Save your Model Definition and Run your analysis.

The **Results** tab should now have additional summary pages at the end for the bootstrap results.

| **Question 7**: Does including group size as a covariate improve the fit to any of the models? |
|---|
| |

You can estimate variance (and therefore the SE and CV) and confidence intervals using the bootstrap for any model. Try with a model without group size as a covariate so you can compare the results for the analytic and bootstrap 95% confidence intervals.

You can add columns to the **Results Browser** for **Nb**, **N CV b**, **N LCL b** and **N UCL b** to show the abundance results for the bootstrap.