Modelling prey-predator interactions

Predator-prey relationships can be important for conservation, including the interactions between tigers and their ungulate prey. There is a good deal of ecological theory about these relationships but not much hard evidence from the field. We don't (yet) have much data on tigers. However, at Isle Royale NP, Michigan, which is an isolated island in Lake Superior, the moose and wolf populations and kill rates have been monitored continuously since 1971. We can use these data to test a range of theoretical models in which kill rates are related to prey density, predator density, or the ratio between them.

Download the file "wolf_kill_rates.xls" from the wcsmalaysia.org web site.

Open the file and go to the 'Prey dependent' worksheet.

Estimates of moose (*N*) and wolf (*P*) abundances are given for each year. In most years, there were several packs of wolves, and separate kill rates (*K*) are given for each pack. The kill rate is expressed as number of kills per wolf per month.

Modeling in MS Excel®

For this exercise you will need the **Solver** add-in in Excel. In Excel go to the 'Tools' menu and look for 'Solver...'. If it is not there, click on 'Add-ins...' and make sure that the "Solver Add-in" item is checked. You may need to install the add-in, for which you may require your original MS Office installation discs.

Other spreadsheet programs may eventually have equivalent functionality, but I did not find it when I last check the web sites. Note that 'Goal Seek' in Excel, StarOffice and OpenOffice is **not** equivalent to 'Solver'.

Excel works well enough for this exercise, but do not rely on it for statistical analysis where the results could have financial or health implications. That includes your reputation as a scientist! For details of issues with Excel go to www.mis.coventry.ac.uk/~nhunt/pottel.pdf

Prey-dependent model

We'll set up three models and compare them with AIC. The first takes a while to set up in Excel, but once that's done, the others will be very easy.

The first theory proposes that kill rate depends only on prey abundance. In its simplest form that would be:

$$K = aN$$

where *a* is a constant representing the proportion of prey that a predator encounters. Most predators take time to 'handle' prey – to consume and digest the carcase – and this reduces the time spent hunting and hence the number of prey encountered. A more realistic model is therefore:

$$K = \frac{aN}{1 + ahN}$$

where *h* represents the handling time. We'll calculate **Maximum Likelihood Estimates** of *a* and *h* given the data on moose and observed kill rates.

In column J of the spreadsheet are plausible starting values for *a* and *h* (*a* = 0.001 and *h* = 0.5); we'll adjust these later, but they will give us something to check our work[1]. We can now calculate the

[1] We could leave these blank, but the predicted values would then all be zero.

predicted value of the kill rate, *K*, for the first row of the worksheet. In Excel's language, the equation above translates to:

= $J$2 * B2 / (1 + B2 * $J$2 * $J$3)

where $J$2 and $J$3 are the starting values of *a* and *h* and B2 is the number of moose (*N*). (If you are not familiar with the "$" sign here, check Excel help for "absolute reference".)

Enter the formula above in cell E2. Then copy the formula down the column to cells E2 to E95.

In the next column, calculate the difference between the observed and predicted kill rates. Save.

We will assume that the residuals (ie. the difference between the observed kill rate and the model prediction) are normally distributed with a mean of zero, and we will use Excel's NORMDIST function to calculate the probability of each value. But for that we need an estimate of the standard deviation (SD or 'sigma'), which we'll put in cell J7.

In cell J7 put the formula = STDEVP (F2:F95).

Notice, by the way, that we're using STDEVP, not STDEV: in this case we want the standard deviation calculated with *n* degrees of freedom, not $n - 1$, because we don't have to estimate the mean from the sample – we know it will be zero when the model fits properly.

In cell G2 put = NORMDIST (F2, 0, $J$7, FALSE); the FALSE at the end tells Excel that we don't want the cumulative distribution. Copy the formula down the column.

In the next column put the natural log of the likelihood, which uses the LN function (not the LOG function). Save.

The value in cell G2 is the likelihood of the model with the starting values of *a* and *h*, given the data for moose and kill rate for 1971. To find the likelihood given all the data, from 1971 to 2001, we must multiply all the values in column G. This will be a tiny number and Excel will treat it as zero (try it if you like, the Excel function is PRODUCT). Adding the logs is equivalent to multiplying the numbers, and the maximum log likelihood corresponds to the maximum likelihood: so we work with log likelihood. We now need the total likelihood for the starting parameters.

In cell J8 put the total of column H, the log likelihood, using SUM(H2:H95). Save.

We're now ready to adjust the values of *a* and *h* to maximize the total log likelihood for the model.

Go to Tools > Solver… on Excel's main menu to start Solver. (This is not the same as Goal Seek…!)

- Set Target Cell: $J$8; Solver uses absolute references, so if you type "J8" it will change it to "$J$8".
- We want Solver to find the 'Equal to: Max' value.
- By Changing Cells: $J$2:$J$3, ie. the values of the parameters *a* and *h*.

Click the 'Solve' button.

The Solver Result box opens; check that 'Solver found a solution…'[2].

Select 'Keep Solver Solution' and press 'OK'. Save.

If the prey-dependent model is correct, the kill rate should be:

$$\frac{0.0015 \times N}{1 + 0.0015 \times 0.42 \times N}$$

The last step is to calculate the AIC for this model, so that we can compare it with other models. The authors of the original paper used AICc, AIC with a small sample correction. With *n* observations and *k* parameters, the formula is:

[2] Sometimes Solver fails to find a solution or returns implausible results. This is often due to the starting values and it may work better to allow Solver to choose its own starting values. To do this simply delete the starting values in cells J2 and J3.

$$AICc = -2\log(likelihood) + 2k + \frac{2k(k+1)}{n-k-1}$$

There are n = 94 observations, and we estimated three parameters from the data (a, h and sigma) so k = 3.

Put 94 in cell J5 and 3 in J6.

Now enter the formula for AICc in cell J9. It's a bit messy, so be careful:

= -2 * J8 + 2 * J6 + 2 * J6 * (J6 + 1) / (J5 - J6 - 1)

We'll come back to AICc when we have calculated the other models.

Predator-dependent model

This theory says that kill rate also depends on predator abundance. The formula is:

$$K = \frac{aN}{bN + P - c}$$

where a, b and c are constant parameters. We'll estimate values for these parameters as we did before.

Make a new copy of the 'Prey dependent' worksheet and name it 'Predator dependent'. (To do this, right-click on the tab at the foot of the worksheet and select 'Move or Copy…'. Then check the box next to 'Create a copy' and click 'OK'. To rename the new worksheet, right-click on the tab and select 'Rename'.)

Change the parameter names in cells I2, I3 and I4 to 'a', 'b' and 'c'.

In cells J2, J3 and J4 put in the starting values:
a = 0.03 *b* = 0.02 *c* = 7
Also change the value for *k* in cell J6 to 4, as we now have one more parameter.

In cell E2 change the formula for the model prediction to the predator-dependent equation:

= $J$2 * B2 / ($J$3 * B2 + C2 - $J$4)

Copy the formula down the column. Save.

That's all we need to do to set up the new model. Now use Solver to get MLE's for the parameters.

In Solver, the target cell is <u>$J$8</u> as before, and '<u>Equal to: Max</u>' is the same. We now have 3 parameters so the cells to change are <u>$J$2:$J$4</u>.

Ratio-dependent model

This theory says that kill rate also depends primarily on the ratio of prey to predators. The formula is:

$$K = \frac{aN}{P + ahN}$$

where a, h are constants. This is very similar to the first model, but the parameter a has a different interpretation.

Make a new copy of the **first** worksheet, called 'Prey dependent', and name it 'Ratio dependent'.

Change the starting values in cells J2 and J3 to: a = 0.05 h = 0.5

Edit the formula in cell E2 to include the number of predators in cell C2:
= $J$2 * B2 / (C2 + B2 * $J$2 * $J$3)
and copy down the column. Save.

The settings for Solver are the same as for the first model.

We're now ready to compare the three models.

Comparing the three models

The worksheet labelled 'AICc' summarizes the AICc values for the three models.

Transfer the AICc values in cell J9 for each model to the 'AICc' worksheet: Copy the value, then in the 'AICc' worksheet select the target cell and use Edit > Paste Special... and select 'Values'. (If you just Copy and Paste, Excel will copy the **formula** in cell J9, not the value.)

Sort the values by AICc.

In the next column, calculate 'deltaAICc', the difference between AICc for the particular model and the minimum AICc for the whole set.

Finally calculate the Model Likelihood: $e^{-\text{deltaAIC}/2}$

In cell D2 put the formula: = EXP(-C2/2). Copy to the rest of the column. Save.

Which of the three models is the best according to AICc?

Are any of the other models still plausible?

Conclusion

These data were analysed by Vucetich et al (2002). They tried a total of 15 different models, with several variations on the three different types; the three we looked at here are among the best in each group.

Vucetich et al also checked model fit, and found that even the best model only explained 1/3 of the variation in kill rates. That means that there are other factors – besides prey and predator abundance – which have a major effect on kill rates.

Hobbs and Hilborn (2006) suggest this as an interesting case study, and provide the data in the Supplementary Material for their paper.

Take home points

- A **model** is a precise mathematical expression of a theoretical idea which gives an 'expected' or 'predicted' value for comparison with the observed value. Models usually involve one or more **parameters**.
- The difference between observed and predicted values is used to calculate the likelihood of parameter values, so **Maximum Likelihood Estimates** can be found.
- When maximized, the likelihood is used to calculate **AIC** (or AICc or similar statistic). AIC can be used to compare **different models**, provided the **same data** are used.
- The **best** model has the **lowest** AIC. Models within 2 units of the best are also good candidates, ie. there is uncertainty about which is the best.

Literature cited:

Hobbs, N T; R Hilborn. 2006. Alternatives to statistical hypothesis testing in ecology: a guide to self teaching. *Ecological Applications* **16**:5-19.

Vucetich, J A; R O Peterson; C L Schaefer. 2002. The effect of prey and predator densities on wolf predation. *Ecology* **83**:3003-3013.

More information on the wolves of Isle Royale is at http://www.isleroyalewolf.org.