

# LABORATORY 8

## Confronting Ecological Models with Data

Biol534, Fall 2011

### 1 Introduction

In class we talked about several ways in which modelers must consider and use empirical data. These include parameter estimation, function selection, forcing or exogenous variables, and model evaluation. In this laboratory, we will practice (1) using a non-linear least squares method to fit parameters to a proposed function that describes the process and (2) using empirical observations of an exogenous variable to drive the system dynamics. These are eminently practical skills for mathematical modelers and data analysis in general.

### 2 Importing Data from Text Files

The first new issue you will encounter is that we have data stored as space delimited text files that we need to enter into R. This required using the *read.table* command. If my data file is called "data.txt" and the file is in my current working directory of R, then I would use the command as "*d = read.table("data.txt")*". The R variable *d* is now a data frame in which the columns are variables and rows are observations. For example, if my text file contains

```
1 50 2
2 51 6
3 52 5
4 53 4
```

then, *d\$V1* is the first column of data.

It is convenient to add column headers to your data when it is store as an ascii text file like this because it helps you remember what the variables were. If our data file has variable names on the first line, we will want to modify our command to read "*d = read.table("data.txt", header=TRUE)*". When you use this command on the data for this exercise, use the *str(d)* command to see how this changes the data frame.

The *read.table* command assumes that your observations are space delimited, but your data may delimited by other markers like commas or tabs. You can add a parameter to the *read.table* command to specify the delimiter. For example, "*d = read.table("data.txt", header=TRUE, sep=",")*" specifies that the delimiter is a comma.

It is good practice to store your data in flat plain ascii text files. This can generally be read into a wide variety of analytical software including SAS, Matlab, and R and it is a non-proprietary format. Many biologists store there data in Excel. This can be convenient, but

it often requires extra steps to extract the data for analyses beyond Excel, and you cannot open or read your data without that exact version of the Microsoft product. This might seem like a reasonable trade off today, but how much scientific data was lost because it was stored in Lotus 1-2-3 spreadsheets? Further, if you store your data in text files and perform your analysis with script based software like R you are less likely to accidentally delete or modify a cell or column of your raw data. You can import and export data from Excel as ascii text files if needed. If you just cannot give up Excel, there is a package called “xlsx” that will let you read data in directly from Excel.

### 3 Fitting Parameters Using Non-Linear Least Squares

Your first task for this laboratory is to complete Exercise 9.7 in Ellner and Guckenheimer’s Dynamic Models in Biology (2006, Chapter 9, p. 297). The exercise reads as follows:

Download the data from Figure 9.5 from the book’s web site, and write a script to find least-squares parameter estimates for  $V$  and  $K$  on the untransformed scale (you should find that you duplicate the values above) and again using power transformation with  $\beta = 0.5$ . Do your parameter estimates for  $\beta = 0.41$  indicate that trial-and-error choice of  $\beta = 0.5$  was close enough?

The full text of Chapter 9 is posted on the biol534 course website as is the EggRatio.dat data set.

I believe that Ellner and Guckenheimer had in mind that you would actually write a script to solve the least squares problem. This might be instructive, but R has a built in algorithm for solving non-linear least squares problems called *nls*. We will use this function. Please review the help file for this function.

In general, an *nls* call will look like the following “fit j- nls(model,data,starting.values)”. When using *nls*, you must specify initial parameter values to try. Again, see the documentation for more details as well as your previous course notes.

### 4 Exogenous Data

The second task of this lab is to create a model in which the dynamics are driven by exogenous variables. For this task please use the Ross Sea data set posted on the class website.

Lets start by creating a lotka-volterra predator-prey model to simulate predation dynamics. Recall, that the equations for this model are

$$\frac{dN_1}{dt} = (\mu - d)N_1 - cN_1N_2 \quad (1)$$

$$\frac{dN_2}{dt} = bcN_1N_2 - dN_2 \quad (2)$$

Please select a set of parameters for the nominal simulation.

Now lets imagine that our prey is a phytoplankton growing in the Ross Sea. We would expect that its growth rate may depend on temperature. Jørgensen and Bendoricchio (2001) provide several possible ways of modelling the effect of temperature, but here lets select a relatively simple one,  $e^{a*T}$ . We will assume that  $a = 0.06933$ , which describes the non-linear effect of increasing temperature on the growth rate. Thus, lets define  $\mu = \mu_{max} * e^{a*T}$ .

Now you will need to use the spline function to interpolate values to use this in your model. See the “connecting models to data” lecture notes for how to do this. Then, see if you can use a monod function to model the effect of light (PUR in the data file) on the phytoplankton growth. How do these simulations compare to your nominal simulation?

## **5 Reporting Your Work**

Please write a brief ( 1 page) summary of your findings in this laboratory.