Spreadsheet Setup Using Biomapper 3 and Microsoft Excel For Comparison of R² Values
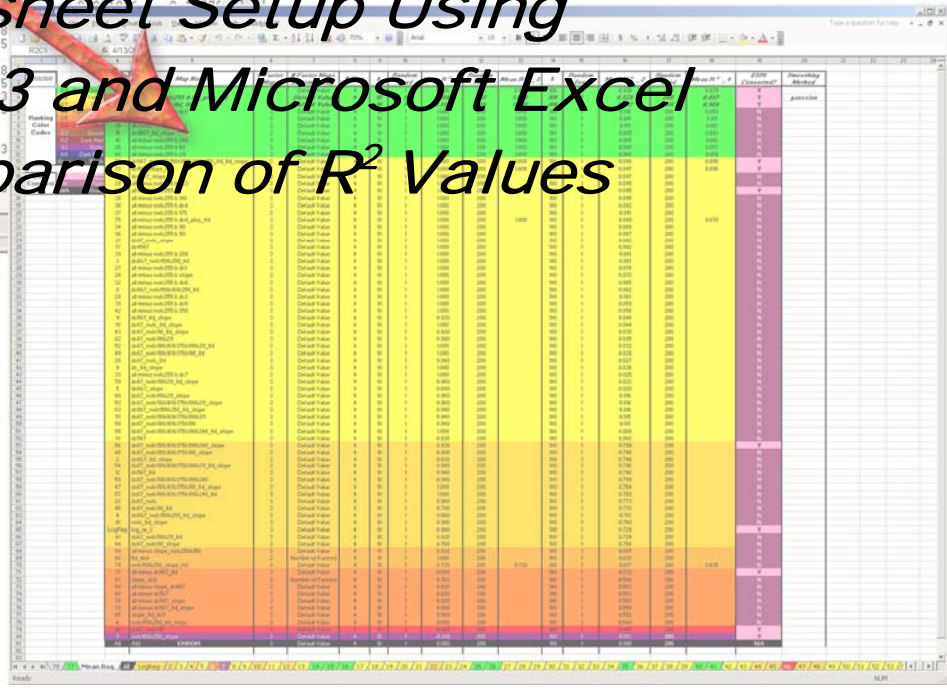
Created by: Emil Peter IV
Created for: Karner Blue Butterfly Habitat Modeling Project
Forest Landscape Ecology Lab
Department of Forest Ecology & Management
University of Wisconsin – Madison
120 Russell Labs, 1630 Linden Lane
Madison, WI 53706
April 11, 2006

As suggested in the document, ***Using Biomapper 3 to Create Habitat Models: A Step-By-Step Guide***, using a spreadsheet program to help check your R^2 values is very useful. It allows to you see all of the values you need in order to help determine which of your models is the best or worst choice for your data. Also as mentioned before, as R^2 gets closer to 1.0, the model is getting better. By better, it is meant that your model is predicting the presence or absence points more accurately. The further your R^2 value is from 1.0, the worse your model is at predicting presences or absences dependably. It was also mentioned that if your R^2 value is 1.0, you need to increase the number of random seeds or increase your k value, because 1.0 is possible only in a perfect world. While it would be nice for your model to be perfect, deep down we all know that nature can not be perfectly predicted, but it can be rather well guessed.

Since the data in the **Results** window of BioMapper is space delimited, copying and pasting the data into Excel makes all of the values fall into individual cells. If all of the values are in individual cells then it makes it much easier to read and follow. Which of these would you rather try to find your data in, this?

```
area-adjusted frequencies
Repl. Bin 1 Bin 2 Bin 3 Bin 4 Binned: Lin.      Qual. Rs      P(Rs=0)      Continuous: Lin.  Qu
1      0.17343    2.8383    0.75162    2.5329    0.047529    0.1349      0.4    0.6    0
2      0.22318    2.0313    0.92318    2.0504    0.060259    0.12356     0.8    0.2    0
3      0.32237    1.4165    0.65316    3.0757    0.12033     0.37009     0.8    0.2    0
4      0.10024    2.6062    0.75518    2.9862    0.057223    0.17088     0.8    0.2    0
5      0.024797   3.0692    0.75114    2.7741    0.048791    0.14975     0.4    0.6    0
6      0.22318    2.1222    0.75137    2.7741    0.067306    0.18672     0.8    0.2    0
7      0.3007     2.2583    0.85891    2.3158    0.053162    0.12311     0.8    0.2    0
8      0.099192   2.1246    1.0451     2.5329    0.080023    0.20269     0.8    0.2    0
9      0.17359    2.0265    0.77914    2.7138    0.07304     0.19822     0.8    0.2    0
10     0.20047    1.4464    0.69311    2.6205    0.10908     0.28585     0.8    0.2    0

Mean:  0.18412    2.194 0.79619    2.6376      0.071674    0.19458     0.72   0.28   0      0
SD:    0.091821   0.53573    0.11599     0.30313     0.025046    0.078571    0.16865        0.
```

Or this?

|	area-adjusted frequencies														
Repl.	Bin 1	Bin 2	Bin 3	Bin 4	Binned: Lir	Qual.	Rs	P(Rs=0)	Continuous	Qual.	Rs	P(Rs=0)	Res	AVI	CVI
1	0.17343	2.8383	0.75162	2.5329	0.047529	0.1349	0.4	0.6	0	0	0	1	1.6129	0.67708	0.1856
2	0.22318	2.0313	0.92318	2.0504	0.060259	0.12356	0.8	0.2	0	0	0	1	1.6129	0.58333	0.16237
3	0.32237	1.4165	0.65316	3.0757	0.12033	0.37009	0.8	0.2	0	0	0	1	1.6129	0.73958	0.2479
4	0.10024	2.6062	0.75518	2.9862	0.057223	0.17088	0.8	0.2	0	0	0	1	1.6667	0.76842	0.26116
5	0.024797	3.0692	0.75114	2.7741	0.048791	0.14975	0.4	0.6	0	0	0	1	1.5385	0.71875	0.22706
6	0.22318	2.1222	0.75137	2.7741	0.067306	0.18672	0.8	0.2	0	0	0	1	1.7241	0.71875	0.22716
7	0.3007	2.2583	0.85891	2.3158	0.053162	0.12311	0.8	0.2	0	0	0	1	1.4706	0.67368	0.18232
8	0.099192	2.1246	1.0451	2.5329	0.080023	0.20269	0.8	0.2	0	0	0	1	1.6129	0.77083	0.27915
9	0.17359	2.0265	0.77914	2.7138	0.07304	0.19822	0.8	0.2	0	0	0	1	1.5385	0.69792	0.23106
10	0.20047	1.4464	0.69311	2.6205	0.10908	0.28585	0.8	0.2	0	0	0	1	1.5873	0.56842	0.22864
Mean:	0.18412	2.194	0.79619	2.6376	0.071674	0.19458	0.72	0.28	0	0	0	1	1.5977	0.69168	0.22324
SD:	0.091821	0.53573	0.11599	0.30313	0.025046	0.078571	0.16865	0.16865	0	0	0	0	0.070672	0.06946	0.036649

Agreed, the bottom is easier to read and find your values in more efficiently. The following steps will guide the reader to an easy to read and follow spreadsheet.

Initial Steps…

⇒ Open Microsoft Excel and save the workbook as something meaningful like **Stats.xls** or **BioMapper_Modeling_Statistics.xls**. Whatever you like so long as the name is recognizable to you.

⇒ Open BioMapper 3.
  o Click **OK** at the splash screen.

⇒ Open your project.

⇒ Once you have run all of the map verification steps and have created a number of models you are happy with, you are ready to extract your values.

Data Extraction…

⇒ Select all of the text in your **Results** window by holding the **shift** key and dragging your mouse cursor to the top of the screen or by clicking your mouse cursor at the bottom of all of the text, scrolling up to the top, holding the **shift key**, and then click your mouse cursor at the top of the screen. All of the text should be highlighted like so:
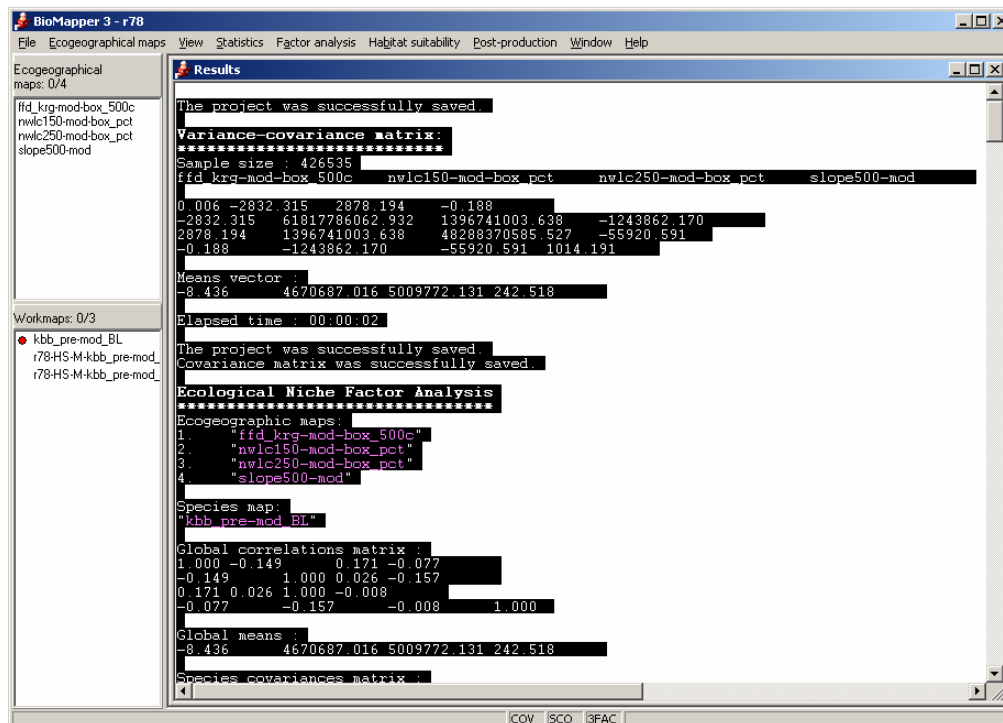
⇒ Pressing the **ctrl + c** keys together will copy your selected text.

⇒ Now return to Excel and in the first cell at the top left (your left, not the monitor's) right-click your mouse and select **paste**.
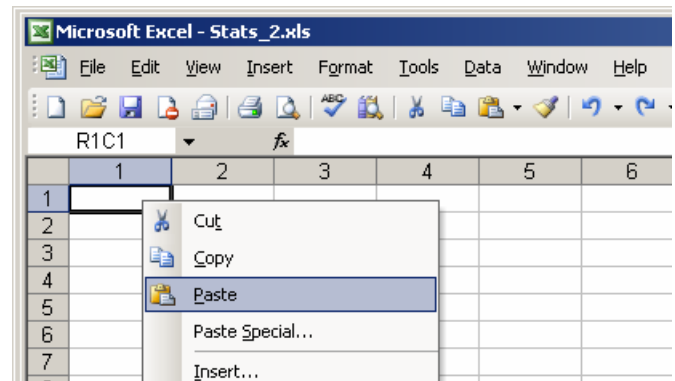
⇒ All of your selected text will then be placed into your worksheet. If you scroll down you will see the rest of your data.
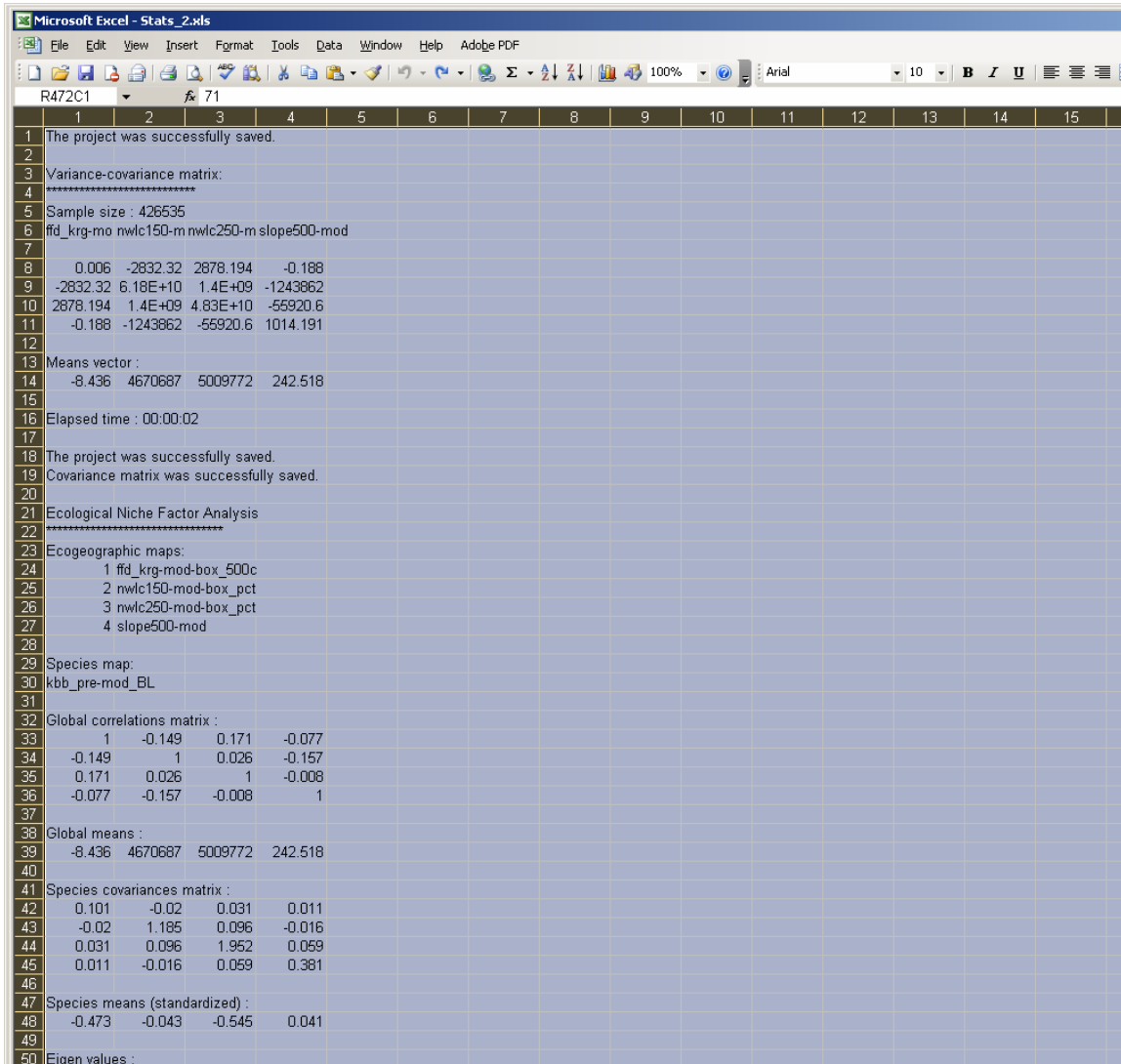
⇒ You can rename the worksheet by double-clicking in the tab at the bottom of your Excel window and typing in a name for your sheet. This sheet is named **78** because it is the 78th iteration of the model.
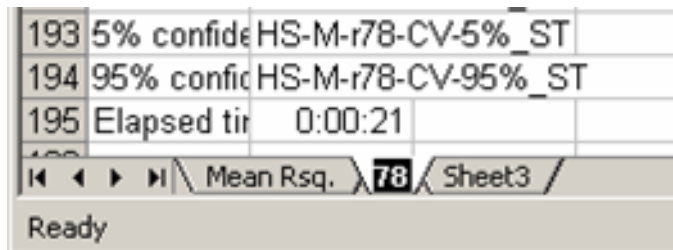
⇒ Save your workbook. ***Save your workbook often as you move through these steps.***

⇒ Now that you have all of your text pasted into your worksheet you are now ready to extract the useful parts, but first you need a place to put them.

R^2 Values Table...

⇒ In one of the other worksheets Excel opens when you create a new document, rename it **Rsq.** or some other meaningful name so you know which worksheet tab has your table of R^2 values.

| 52 |
| 53 |
| 54 |
| 55 |

◄ ◄ ► ►ǀ \ **Mean Rsq.** ╱ 78 ╱ Sheet

Ready
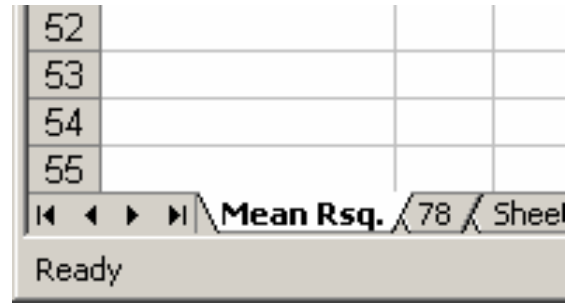
⇒ This table has all of the spaces for the **Sheet**, **Map Name**, the **Number of Factor Maps**, how many **Bins** the map was classified or reclassified into, the **k Value** for the map, the number of **Random Seeds** used in validation, and a column for the **Mean R^2 Value**.

Sheet	Map Name	# Factor Maps	# Factor Maps determined by:	# Bins	k	Random Seed	Mean R^2_1
78	nwlc150&250_slope_ffd						

There are also spots for more **k Values** and **Random Seeds**, as well as **Mean R^2 Values** for further validation and more random seeds. Also included are columns for whether or not the grid was converted in **ESRI format** and which **Smoothing Method** was used. These two issues are outlined in a different guide, but they are pointed out here for consideration if you are using ArcGIS for analysis or mapping instead of Idrisi. Formatting of the table (i.e. the highlighting of boxes, font sizes, bolding or italicizing) is up to the user.

Random Seed	Mean R^2_3	k	Random Seed	Mean R^2_2	Random Seed	Mean R^2_4	ESRI Converted?	Smoothing Method
							N	
							N	

⇒ Also in this table, there is a legend for the color coding of different R^2 values and Conversion Values. This makes it easier to glance at the

Ranking Color Codes	0.9	Bright Green
	0.8	Yellow
	0.7	Gold
	0.6	Light Orange
	0.5	Orange
	0.4	Red
	0.3	Brown
	0.2	Dark Red
	0.1	Violet
	0.0	Dark Blue
	Error	Black

ESRI Conversion	Y	Rose
	N	Plum
	N/A	Black

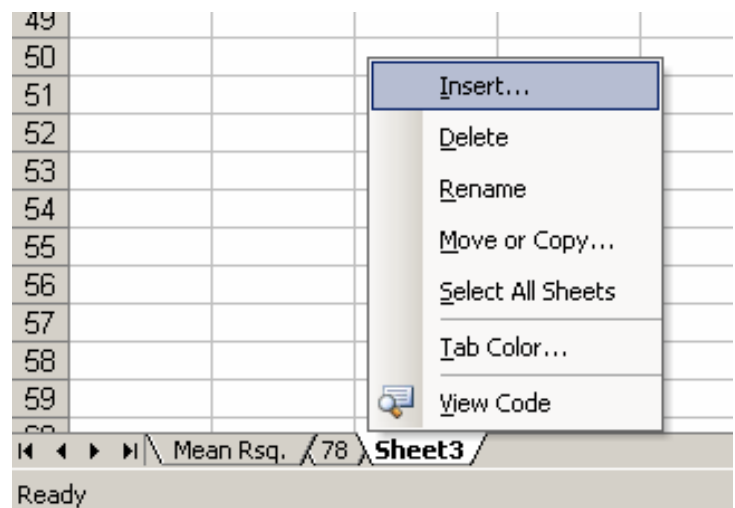table and find which values are best or worst, converted or not. More about this in a minute.

⇒ By filling in each of these columns with the proper values, we can easily compare models by any of the values associated with them. Remember: as you build more models, in order to compare them reliably by their R^2 values, the other variables need to be consistent. If you change the **number of bins** or the **number of factor maps** you calculated, then you have changed your model and cannot say that it is necessarily better than a model with different values in those places.

⇒ Go to the worksheet you pasted all of your values from BioMapper in and scroll down until you find your **Area-Adjusted Frequencies** section. This is the output from your **cross-validation** method and contains your R^2 value. Your R^2 value is in the column labeled Rs. This is just BioMapper's notation for R^2.

area-adjusted frequencies

Repl.	Bin 1	Bin 2	Bin 3	Bin 4	Binned: Lir	Qual.	Rs	P(Rs=0)	Continuous Qual.	Rs	P(Rs=0)		Res	AVI	CVI
1	0.17343	2.8383	0.75162	2.5329	0.047529	0.1349	0.4	0.6	0	0	0	1	1.6129	0.67708	0.1856
2	0.22318	2.0313	0.92318	2.0504	0.060259	0.12356	0.8	0.2	0	0	0	1	1.6129	0.58333	0.16237
3	0.32237	1.4165	0.65316	3.0757	0.12033	0.37009	0.8	0.2	0	0	0	1	1.6129	0.73958	0.2479
4	0.10024	2.6062	0.75518	2.9862	0.057223	0.17088	0.8	0.2	0	0	0	1	1.6667	0.76842	0.26116
5	0.024797	3.0692	0.75114	2.7741	0.048791	0.14975	0.4	0.6	0	0	0	1	1.5385	0.71875	0.22706
6	0.22318	2.1222	0.75137	2.7741	0.067306	0.18672	0.8	0.2	0	0	0	1	1.7241	0.71875	0.22716
7	0.3007	2.2583	0.85891	2.3158	0.053162	0.12311	0.8	0.2	0	0	0	1	1.4706	0.67368	0.18232
8	0.099192	2.1246	1.0451	2.5329	0.080023	0.20269	0.8	0.2	0	0	0	1	1.6129	0.77083	0.27915
9	0.17359	2.0265	0.77914	2.7138	0.07304	0.19822	0.8	0.2	0	0	0	1	1.5385	0.69792	0.23106
10	0.20047	1.4464	0.69311	2.6205	0.10908	0.28585	0.8	0.2	0	0	0	1	1.5873	0.56842	0.22864
Mean:	0.18412	2.194	0.79619	2.6376	0.071674	0.19458	0.72	0.28	0	0	0	1	1.5977	0.69168	0.22324
SD:	0.091821	0.53573	0.11599	0.30313	0.025046	0.078571	0.16865	0.16865	0	0	0	0	0.070672	0.06946	0.036649

⇒ Copy and paste your R^2 value into the appropriate cell for that iteration. Remember: If you cross-validated more than one time with different **k values** or **random seeds**, you will need multiple fields for your multiple R^2 values and thus, will repeat this step until you have all of your values placed in the table.

Sheet	Map Name	# Factor Maps	# Factor Maps determined by:	# Bins	k	Random Seed	Mean R^2_1
78	nwlc150&250_slope_ffd	3	Default Value	4	10	1	0.720

⇒ Once you have all of your values filled in, you can move onto the next iteration of your model. By right-clicking on the tabs at the bottom of the Excel screen, you can select *Insert…* and insert another worksheet when prompted. Repeat the steps used to get to this point for your new worksheet.

49	
50	
51	Insert...
52	Delete
53	Rename
54	
55	Move or Copy...
56	Select All Sheets
57	
58	Tab Color...
59	View Code

◄ ◄ ► ►│ \ Mean Rsq. / 78 \ Sheet3 /

Ready

⇒ Filling in the remainder of the table is a matter of repeating the previous steps until all of your iteration data has been migrated into Excel and you can compare their R^2 values.

⇒ To facilitate the viewing of your data values, a color coded legend may be helpful to color code your values. If you set up a color code for your table, you can sort or organize the table by R^2 values and look at which sheets (iterations of your model) are ranked highest.

Ranking Color Codes	0.9	Bright Green
	0.8	Yellow
	0.7	Gold
	0.6	Light Orange
	0.5	Orange
	0.4	Red
	0.3	Brown
	0.2	Dark Red
	0.1	Violet
	0.0	Dark Blue
	Error	Black
ESRI Conversion	Y	Rose
	N	Plum
	N/A	Black

Ranking Your Models…

⇒ To do this, fill in your table completely and color code each of the model iterations according to your legend.

⇒ Next, highlight all of the values in your table by dragging from the top corner cell of your table all the way to the bottom opposite corner cell. In the example table, the top corner cell is named **Sheet** and the opposite bottom corner is empty.

⇒ From the *Data* menu, select *Sort…*
⇒ The **Sort…** dialogue will come up and look similar to this:
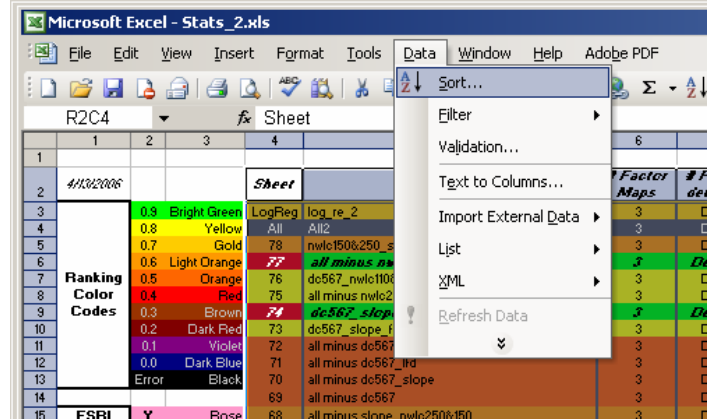
⇒ Using the dropdown menus, select the names of the columns you want to sort by. Sorting by your R^2 values is and selecting the **Descending** radio button will sort the values from **Highest to Lowest**. Since the closer R^2 is to 1.0, this would be the recommended way of sorting the values. If you do not see your column names appear in the dropdown menu and only letters or numbers, make sure you selected your header row in your worksheet, and then make sure to select the **Header Row** radio button at the bottom of the dialogue to include your header row. It should select itself by default, but on occasion it will not and you will have to click it.

⇒ Click **OK** to sort.

⇒ Your data should be sorted and if you color coded, will look similar to this:

⇒ The values at the top of the list are closest to 1.0 and should therefore be the better models at predicting habitat suitability. In this case, the 2nd and 3rd models are the recommended models to look at due to one variable.

Tab Colors…

⇒ The tabs in this example have been color coded to agree with their iterations in the table for ease in finding them at the bottom of the screen as there are more than 80 models in this example. Color coding the tabs is not a difficult task and makes for a quick way to organize your worksheets.

⇒ To color your tabs, right-click on the tab and choose *Tab Color…*
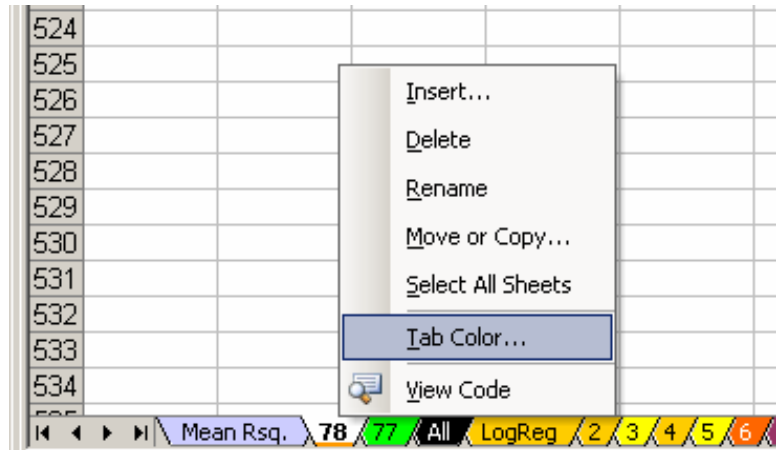
⇒ The **Format Tab Color** dialogue box will appear and you can then choose which color you would like your tab to be.

Also and maybe most importantly…

It cannot be stressed enough to save your work when working with Microsoft Excel, BioMapper or any other program where you have done a lot of work. If the program you are working with encounters a problem, and it will, the problem *may* be fatal. Chances are very good that the problem *is* fatal and the program will need to be shut down and restarted at which point you will wish you had saved your work before taking that last step or making that last calculation.