Diatom predictive model for Blue Mountains region, Oregon

Mark deVisser
Robin Fladebo
Nadia Gillett
Jack Holmberg

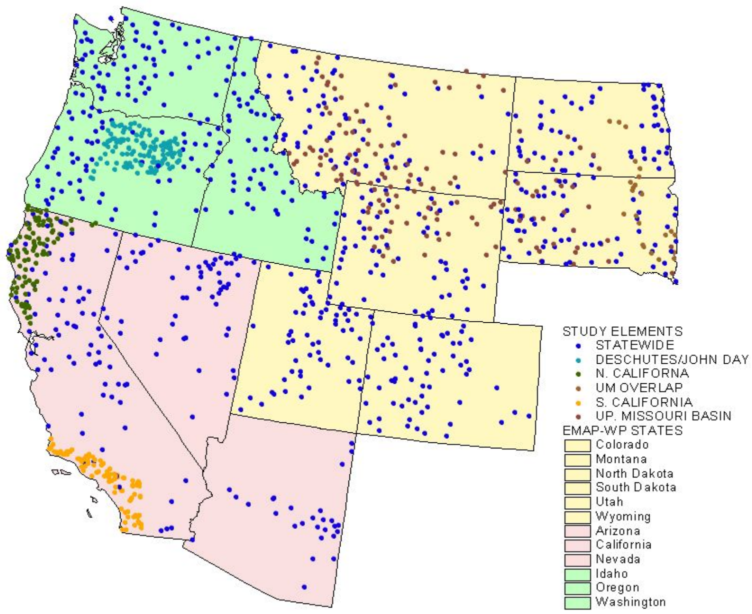
GIS II, Winter 2006

Introduction

Environmental Monitoring and Assessment Program (EMAP)

- EPA
- Assess freshwater quality in the US
- Collect physical, chemical and biological data
- Western EMAP 2000-2004
- Rivers/ streams in 12 Western states
- 400-500 periphyton samples per year

PRIMARY CANDIDATE SAMPLING SITES: 2000-2003



Introduction

Objective:

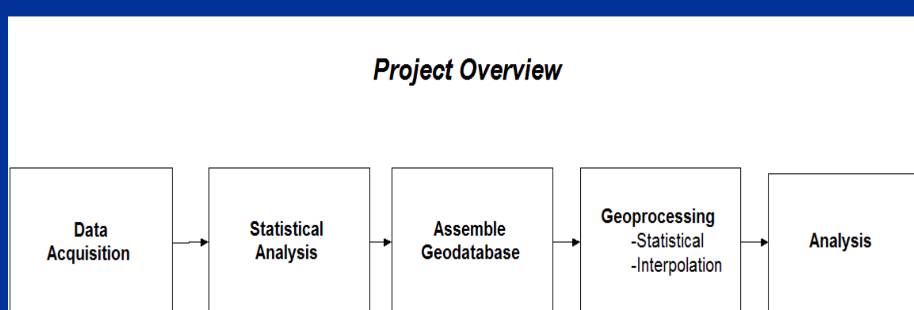
- Predict species occurrences in relationship to their most important environmental variables

Material and Methods

Blue Mountain sites

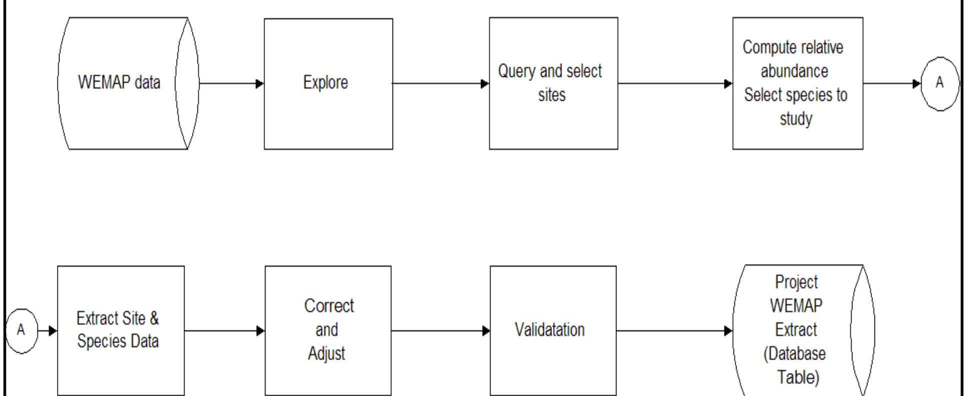
- WEMAP 2000, 2001, 2002
- 71 diatom samples
- 1-5 order streams
- Environmental (physical habitat, chemical variables) and spatial data
- Oregon Geospatial Data Clearinghouse
- ArcGIS
- Excel

Material and Methods
Flowchart



Material and Methods

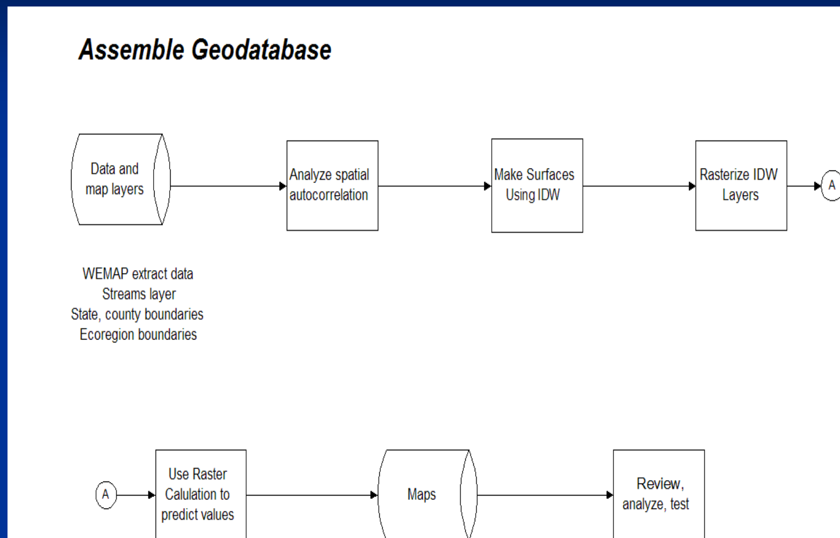
Compiling data for our project



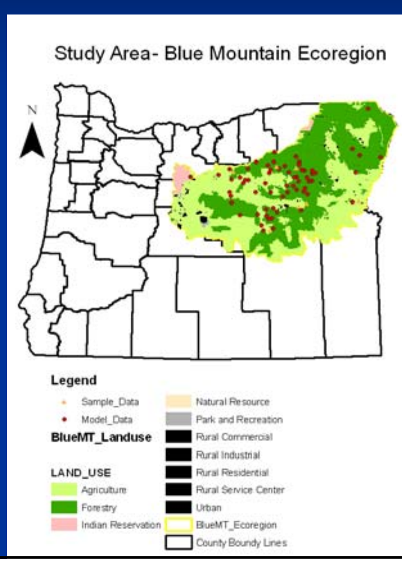
Material and Methods
Statistics

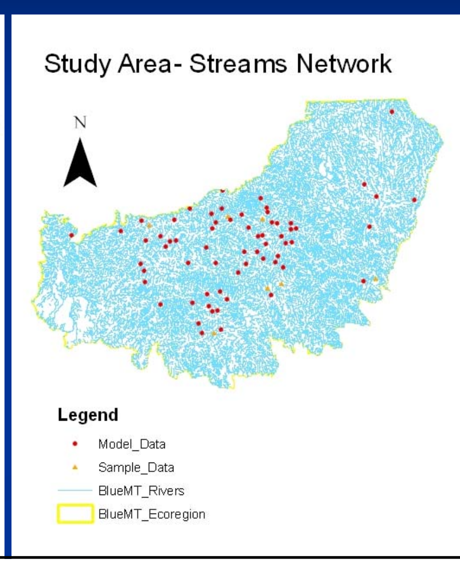
- Select test sites: 10% of 71 sites
- Select species and variables to analyze
  - Discovery: compute relative abundance
  - Correlation: species abundance and environmental variables; determine which variables are significant
  - Adjust: Choose one representative variable in cases where environmental variables are highly correlated with others
- Run descriptive statistics for selected environmental variables
  - Transform as necessary
- Run regression (simple linear or multiple stepwise)
- Test formulas

Material and Methods



Material and Methods
Map of the study region

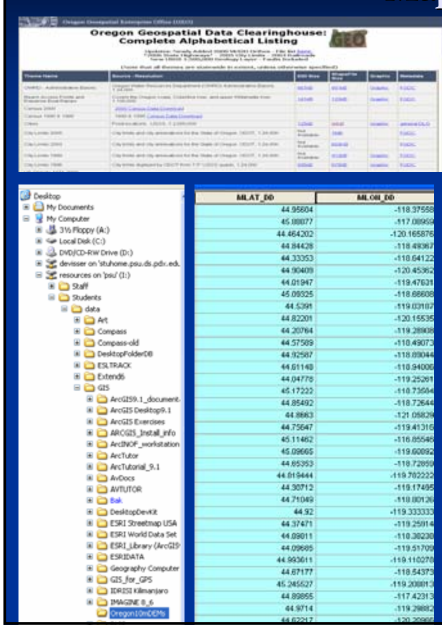


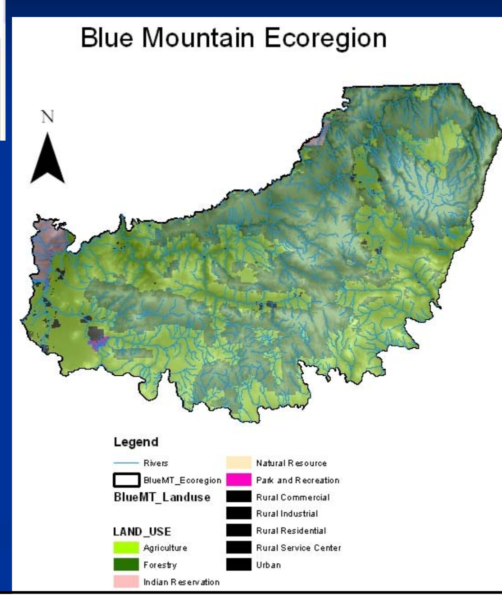


Material and Methods
GIS Methods

- Assembling data
- Autocorrelation (Moran’s I)
- IDW layers for each variable
- Rasterized the IDW layers
- Raster calculator to compute the predicted values

Material and Methods
Map Design



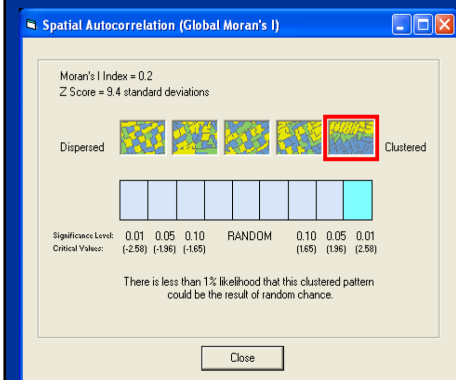


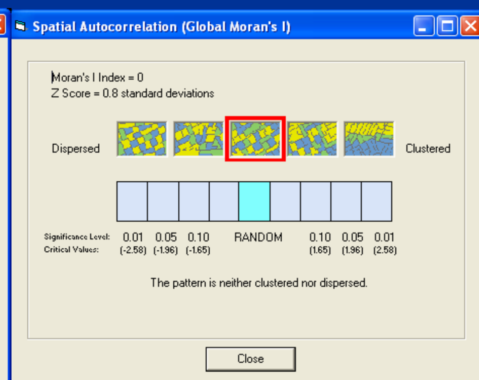
Material and Methods
GIS Methods

Autocorrelation (Moran's I)

SA stats for conductivity

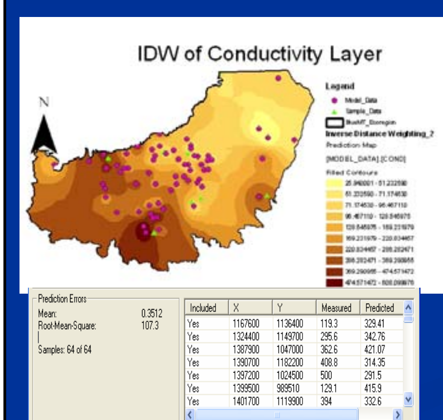
SA stats for log10 of conductivity



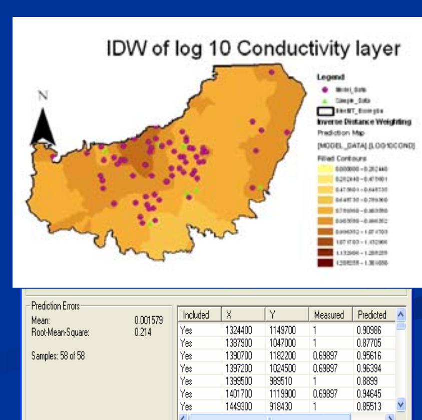


Material and Methods
IDW Layers of Conductivity

- Inverse Distance Weighted layers were made for both transformed and non transformed variables in order to further test the effects of the normalization process.



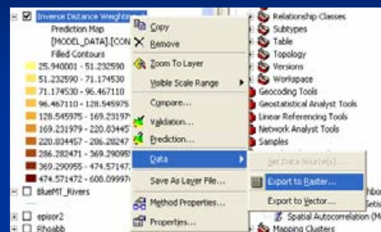
Included	X	Y	Measured	Predicted
Yes	1167600	1136400	119.3	329.41
Yes	1324400	1149700	295.6	342.76
Yes	1387900	1047000	362.6	421.07
Yes	1390700	1182200	408.8	314.35
Yes	1397200	1024500	500	291.5
Yes	1399500	989510	129.1	415.9
Yes	1401700	1119900	394	332.6

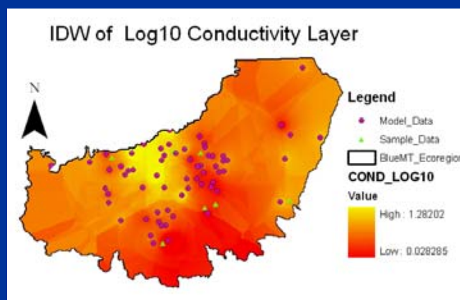


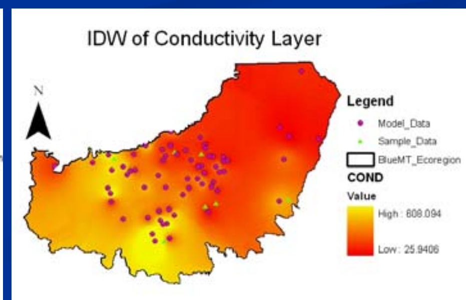
Included	X	Y	Measured	Predicted
Yes	1324400	1149700	1	0.90986
Yes	1387900	1047000	1	0.87705
Yes	1390700	1182200	0.69897	0.95616
Yes	1397200	1024500	0.69897	0.96394
Yes	1399500	989510	1	0.8899
Yes	1401700	1119900	0.69897	0.94645
Yes	1449300	918430	1	0.85513

Material and Methods
Rasterized IDW Layers of Conductivity

- In order to use the IDW Layers in a model in ArcMap, they must be rasterized in order to be entered into the Raster Calculator in Spatial Analyst.



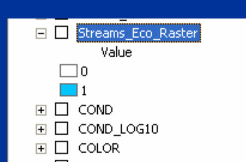


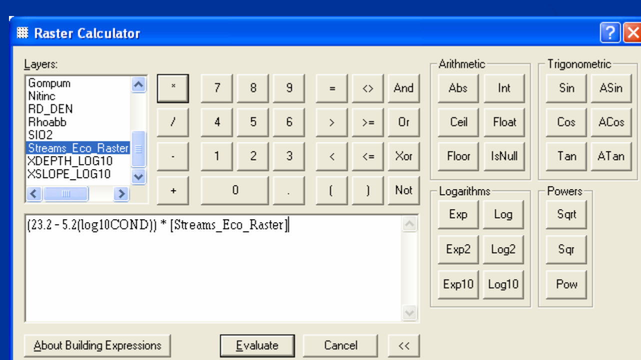


Material and Methods
Created Prediction Layers Using Raster Calculator

- We entered our regression models into Raster Calculator to create prediction layers of the relative abundance of the 5 species.

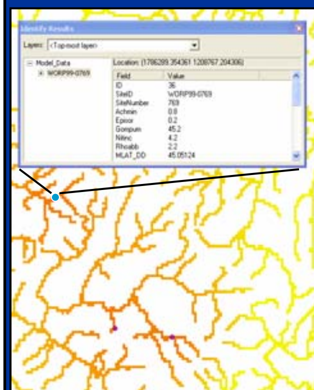
We multiplied the regression model by the Streams_Eco_Raster layer to only predict values within the stream network.



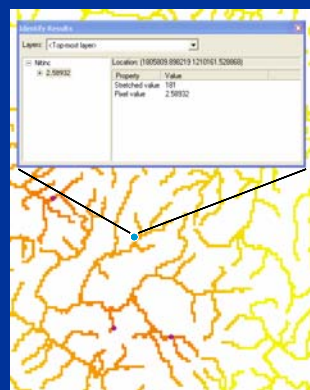


Material and Methods
Layout of Prediction Layers

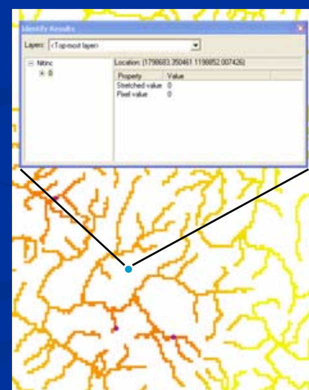
Sample Point



Predicted Point on the Stream Network



Sample Point not on the Stream Network



Results

- **Most significant variables:**

LAT_DD latitude in decimal degrees
COND* specific conductance (uS/cm)
COLOR* true color (dissolved organic material)
SiO2 silica (mg/L)
XC* riparian veg canopy cover
XSLOPE* channel slope-reach mean (%)
XDEPTH* thalweg mean depth (cm)
RD_DEN road density (m/ha) (TOT_RD/AREA_WS)
CHL_M2* chlorophyll (ug/m2)

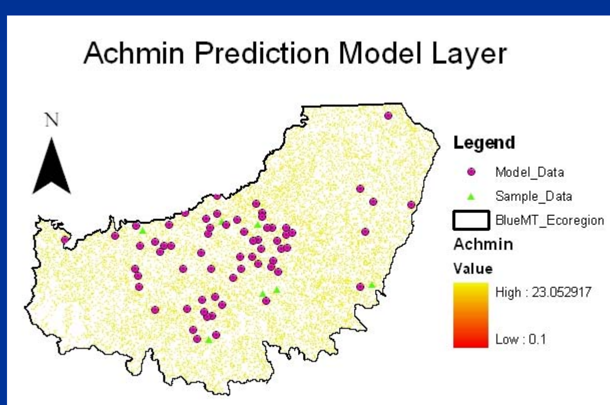
* log10 transformed variables

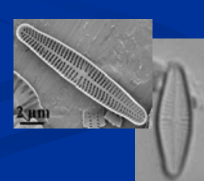
Results
Regression Formulae

Species	Formula	R^2
R_Achnanthidium minutissimum …	23.2 - 5.2(log10COND)	0.02
R .Epithemia sorex Kützing	-7.68 + 6.12(log10COLOR) + 20.37(log10CHL_M2)	0.18
R. Gomphonema pumilum …	30.58 - 7.6(log10XDEPTH)- 7.74(log10COND)	0.16
R. Nitzschia inconspicua Grunow	1.79 + 0.163(RD_DEN)	0.05
R. Rhoicosphenia abbreviata …	-0.789 + 0.368(SIO2) + 6.133(log10_XSLOPE) - 0.319(RD_DEN)	0.24

Results
Achnanthidium minutissimum (Kützing) Czarnecki

Relative abundance = 23.2 - 5.2(log10COND)
$R^2 = 0.02$

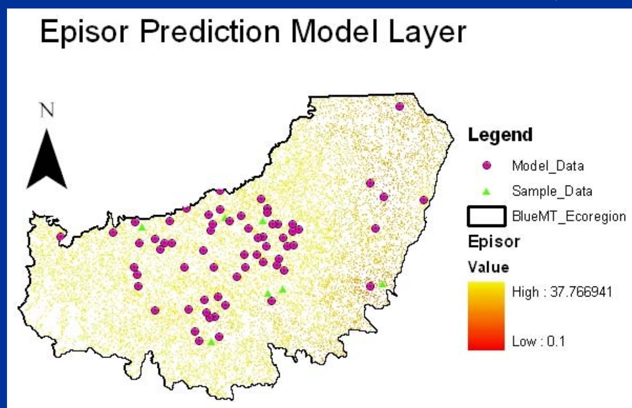


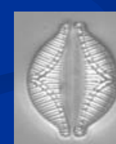


Results
Epithemia sorex Kützing

Relative abundance =-7.68 + 6.12(log10COLOR) + 20.37(log10CHL_M2)
$R^2 = 0.18$

	Log10COLOR	Log10CHL_M2
R Square	0.114935	0.178822

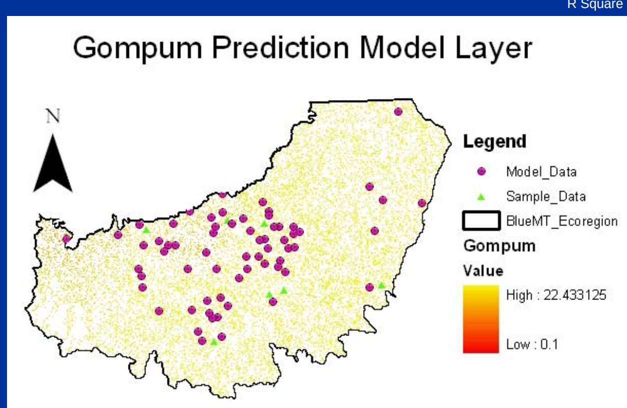


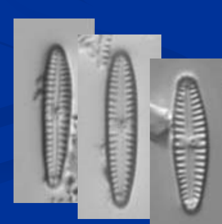


Results
Gomphonema pumilum (Grunow) Reichardt et Lange-Bertalot

Relative abundance = 30.58 - 7.6(log10XDEPTH)- 7.74(log10COND)
$R^2 = 0.16$

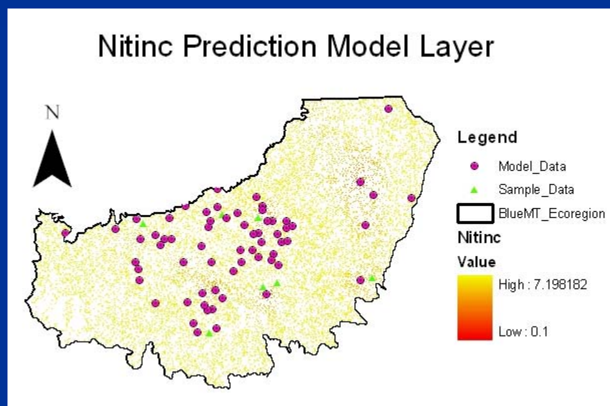
	Log10XDEPTH	Log10COND
R Square	0.08865	0.158474

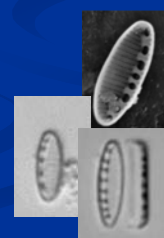




Results
Nitzschia inconspicua Grunow

Relative abundance = 1.79 + 0.163(RD_DEN)
$R^2 = 0.05$

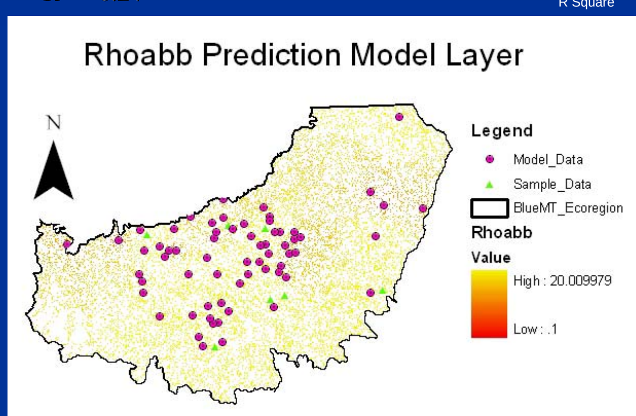


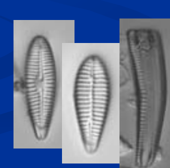


Results
Rhoicosphenia abbreviata (Agardh) Lange-Bertalot

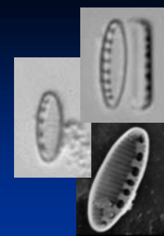
Relative abundance = -0.789 + 0.368(SIO2) + 6.133(log10_XSLOPE) - 0.319(RD_DEN)
$R^2 = 0.24$

	SIO2	XSLOPE	RD_DEN
R Square	0.099853	0.193705	0.243737





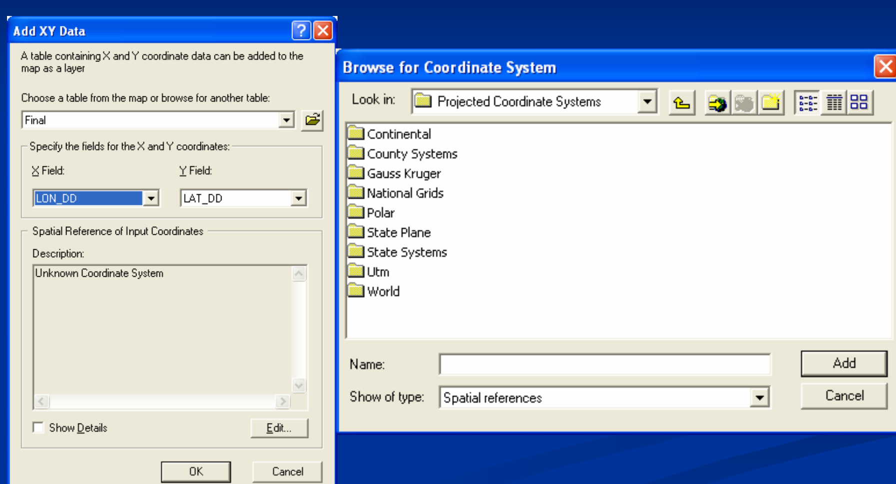
Results
Testing formulae



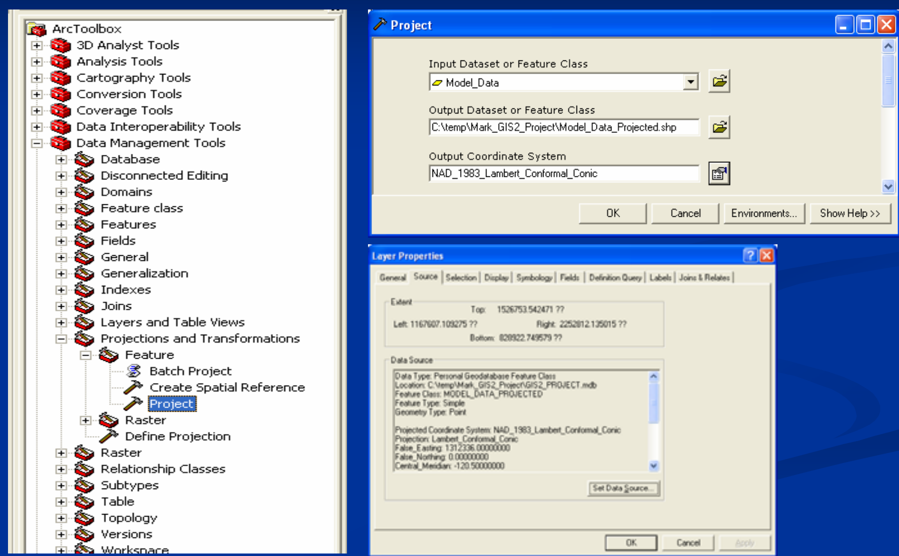
Nitzschia inconspicua Grunow

Site	Actual	Predicted	Variance	% Error
523	1	2.81	1.81	181.04%
607	4.7	4.52	-0.18	-3.86%
691	7.2	6.54	-0.66	-9.15%
728	2	2.21	0.21	10.69%
728	2	2.21	0.21	10.69%
856	1.2	3.86	2.66	221.40%
924	7.2	6.78	-0.42	-5.89%

Results
Problems: Projections

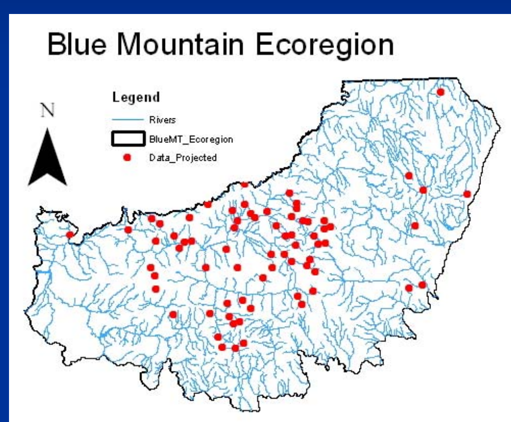


Results
Problems: Projections



Results
Problems: Number of Sample Sites

- In order to create a more accurate regression model a larger number of points would be needed from the region.



Conclusions

- Very good GIS exercise
- Good predictive power of the model
- Challenges in variable selection for biological data
- Difficulty in modeling natural systems
- Keep trying

Thank You!