ABE 691C Environmental Data Handling  
Autumn 2006 Instructor: Keith Cherkauer  
3 Credit Hours, (The class will meet on Tuesday and Thursday. The class meeting time will be adjusted to accommodate the schedules of the registrants.)

Course Description

Data volumes in the environmental field are increasing with the advent of sensor networks, the increase in the number of high-resolution and multispectral remote sensing images, and the increasing use of distributed models. To most effectively utilize these new and varied data streams, students require a new tool box of skills so that they can handle data in a wide variety of formats, and manipulate large numbers of files or even just a few very large files. This course will educate students in the use, manipulation and analysis of environmental data by introducing them to scripting languages (e.g. c-shell, python), data types (e.g. ASCII, binary, NetCDF), databases (e.g. XML, DBF) and data visualization software (e.g. GMT, ArcMap) as well as techniques for checking data quality, handling time series and spatial data, and filling in missing data. Students will manipulate, check and insert data from a variety of sources, use that data as input to a distributed hydrologic model and analyze model output. The course format will consist of a 1 hour lecture followed by a 1 hour computer lab, twice weekly. All skills learned should be applicable to most computer operating systems, but the majority of work for this class will be done within the Linux environment.

Objectives:

Successful completion of the course will enable the students to:
1. Write scripts or batch files to control tasks requiring multiple steps.
2. Write scripts or programs to process multiple data files.
3. Design quality control programs to check for obvious inconsistencies in data.
4. Design programs that will fill missing or invalid data with a “best guess” value.
5. Identify a variety of file formats and understand how to transfer data into and out of them.
6. Manipulate data from a variety of sources so that it is properly formatted for use with a distributed model.
7. Handle large data volumes and analyze them using summary statistics and plots.

Course Materials

Suggested Texts:
3. Books on C, Fortran, Java, PHP, XML and other topics covered in this class are available and recommended if you plan on digging deeper into a particular package.

**Additional Reference Material:**
1. GMT home page: http://gmt.soest.hawaii.edu/
3. PHP home page: http://www.php.net/
5. HDF home page: http://hdf.ncsa.uiuc.edu/
6. Other material may be handed out in class as required.

**Grading:**

The student’s grade will be determined as follows:

<table>
<thead>
<tr>
<th></th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assignments</td>
<td>320</td>
</tr>
<tr>
<td>Final Project</td>
<td>80</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>400</td>
</tr>
</tbody>
</table>

A ≥ 360, B ≥ 320, C ≥ 280 and D ≥ 240

Point totals required for a specific grade may be lowered at the instructor’s discretion.

**Syllabus:**

1. Introduction to Linux
2. Fundamentals of programming
   2.1. Variables
   2.2. Loops
   2.3. Structures
   2.4. Debugging techniques
3. Introduction to Data Formats and Databases
   3.1. Data Formats
      3.1.1. ASCII files
      3.1.2. Binary files
      3.1.3. NetCDF
      3.1.4. HDF
   3.2. Databases
      3.2.1. DBF
      3.2.2. XML
      3.2.3. Others
4. Programming Languages
   4.1. Shells and text manipulation
      4.1.1. C-shell
      4.1.2. Bat
      4.1.3. Awk
4.2. Scripting languages
   4.2.1. Python
   4.2.2. Perl
   4.2.3. Java
4.3. Compiled languages
   4.3.1. C, C++
   4.3.2. Fortran

5. Plotting Packages
   5.1. Generic Mapping Tools (GMT)
   5.2. NCAR graphics
   5.3. ImageMagick (convert)

6. Converting Formats
   6.1. ASCII to ASCII
   6.2. Binary to ASCII
   6.3. ASCII to Binary
   6.4. ASCII to database

7. Working with temporal data
   7.1. Handling dates
   7.2. Finding missing dates
   7.3. Strategies for finding invalid data
   7.4. Strategies for filling missing time series data

8. Working with spatial data
   8.1. Handling grids
   8.2. Interfacing with ArcGIS
   8.3. Using GMT grids
   8.4. Strategies for finding invalid data
   8.5. Strategies for filling missing spatial data

9. Running model simulations
10. Analyzing large datasets
    10.1. Evaluation using statistics
    10.2. Plotting data
    10.3. Presenting your findings