Watershed Sciences 4930 & 6920
GEOGRAPHIC INFORMATION SYSTEMS

WEEK THREE—
INTRODUCTION TO
GEOPROCESSING &
UNCERTAINTY IN GIS...

Joe Wheaton
UTAH AGRC SGID – JUST TO MAKE SURE

• SGID (State Geographic Information Database) Data & Map Servers
• Alternative to ‘Add Basemap’ from ESRI

SGID Data Services

Online Base Maps
Fast and free online map services from AGRC featuring local data.

SGID Image Server
Utah aerial photography without the need to download data.

SGID ArcSDE Database Server
Direct connect to AGRC’s ArcSDE database, the most current data.

SGID Raster Discovery Application
Download aerial photography, LiDAR, DEMs, contour lines and USGS topographic maps via the Utah Raster Discovery App.

Download SGID Data Via FTP
Download data directly through AGRC’s FTP site.

UTAH AGRC

Need a quick, fast base map for your ArcMap project?

AGRC provides several multi-scale base map service options that deliver pre-rendered base maps. Base map native projected coordinate system is UTM Zone 12N, NAD83.

Instructions for ArcMap 9.3 or 10 users:

1. In ArcMap, click the Add Data button
2. At the top of the Add Data window, set the Look In: pick list to point at GIS Servers
3. Select Add ArcGIS Server and then choose the Use GIS Services option
4. Set the Internet Server URL to: http://mapserv.utah.gov/arcgis/services
5. Click Finish and then double click the newly added item “arcgis on mapserv.utah.gov”
6. Select the base map(s) of your choice from the list below:
   - UtahBaseMap-Hillshade
   - UtahBaseMap-Hybrid (streets and other vector layers atop imagery, )
   - UtahBaseMap-Imagery
   - UtahBaseMap-Lite (gray, muted tone)
   - UtahBaseMap-Terrain (hillshade, streets, parks, forest, water, etc)
   - UtahBaseMap-Topo (multiscale mosaic of USGS topo maps)
   - UtahBaseMap-Vector (streets, land ownership, and boundaries)
7. Click Add or OK

Web developers can embed any of these base map services in web-based and internet-connected applications.

More information about the base maps and for application developers.
HOUSEKEEPING

• I’ve got nothing else...
• Any questions?
TODAY’S PLANS… 1st GEOPROCESSING

I. Geoprocessing
II. ModelBuilder
III. Batch Processing
   I. A right-click away
   II. Scripting

TODAY’S OTHER PLAN… UNCERTAINTY IN GIS

I. Uncertainty & Error
II. GIS Errors
III. Error Propagation
IV. All Bad?
V. Summary of GIS Fundamentals
WHAT IS GEOPROCESSING?

An ESRI term...

Introduction to Topic

From ESRI's Help:

"Geoprocessing is for everyone that uses ArcGIS. Whether you’re a beginning user or a pro, geoprocessing will become an essential part of your day-to-day work with ArcGIS.

The fundamental purposes of geoprocessing are to allow you to automate your GIS tasks and perform spatial analysis and modeling. Almost all uses of GIS involve the repetition of work, and this creates the need for methods to automate, document, and share multiple-step procedures known as workflows. Geoprocessing supports the automation of workflows by providing a rich set of tools and a mechanism to combine a series of tools in a sequence of operations using models and scripts...." - ...READ MORE.

Why we’re Covering it

Geoprocessing is what helps make you efficient in your spatial analyses. Whether its combining multiple tools into one using model builder, batch processing a large quantity of data, writing some simple Python scripts, or leveraging the functionality of the Results window, geoprocessing is what can help you move from a novice GIS user, to a highly proficient power user.

Learning Outcomes

This topic supports primary learning outcomes 1, 2, 3 and 4 for the course.
TYPES OF GEOPROCESSING IN ARCGIS

- Tools
- Models
- Scripts
- Add-Ins
- Plug-Ins

These differ from some of the Toolbars you’ve used...
PLACES TO GET GEOPROCESSING TOOLS

- Toolbox
- ArcGIS Resource... Gallery
- From Specific Places...
LOTS OF WAYS TO ACCESS A TOOL

1. Click the tool name in the Search window.
2. Enter values for the tool's parameters.
3. Or double-click the tool in the Catalog window.
4. Click OK to execute the tool.
WHEN A TOOL RUNS...

• Be patient...
• Two places to track status:
  – Progress bar
  – Results Window!
• Use that results window!
GEOPROCESSING RESULTS WINDOW

- A Plethora of Useful Information...
- Keeps track of all past geoprocessing commands!
- Also... double click any...
- Troubleshoot
There are four levels of environment settings:

1. **Application level** settings are the default settings that will be applied to any tool when it is executed.

2. **Tool level** settings are applied to a single run of a tool and override the application level settings.

3. **Model level** settings are specified and saved with a mode and override tool level and application level settings.

4. **Model process level** settings are specified at the model process level, are saved with the model, and override model level settings.
**Tool environment settings**

Tool environment settings inherit from application environment settings: when you open a tool's dialog box and click the **Environments** button, the application environment settings are used as the initial values for the tool's environment settings.

**Note:** Tool environment settings only apply to the current run of the tool and do not update the application environment settings.

---

[Link to official documentation](http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#/Setting_geoprocessing_environments/001w0000003s0000)
EVERY TOOL HAS ENVIRONMENT SETTINGS

- Tools validate parameter values as you enter them...
- They also can override environment settings
BRING THEM UP... & EXPAND TO USE
Environment settings specified in this dialog box are values that will be applied to appropriate results from running tools. They can be set hierarchically, meaning that they can be set for the application you are working in, so they apply to all tools; for a model, so they apply to all processes within the model, or for a particular process within a model. Environments set for a process within a model will override all other setting, and environments set for all processes in a model will override those set in the application.

Changing the default settings that will be used is a prerequisite to performing geoprocessing tasks. You may only be interested in analyzing a small piece of a geographic area, such as changing the extent for results, or you may want to write all results to a specific location (for example, changing the current workspace or the scratch workspace).
PROCESSING EXTENT -> ENV. SETTINGS

- This is how you force grid concurrency
- Extent -> Limits
- Snap... aligns
RECALL MASKED EXTENTS

- Rasters that have the same masked extents simply have the same nodata cells.
- The mask can be derived from a polygon or a raster.
- A concurrent raster mask is the most accurate!
WHERE TO LEARN MORE ABOUT GEOPROCESSING

Additional Resources

Geoprocessing in ArcGIS

• Geoprocessing ArcGIS Resource Center - This is the main hub of information on Geoprocessing and getting yourself up to speed.
• The Geoprocessing Forum - Trying to do something with Geoprocessing and having trouble? Try posting a thread or searching this forum.
• Geoprocessing Help - An overview of Geoprocessing in ArcGIS
• Using the Results Window - An essential new feature of Geoprocessing in ArcGIS

Batch Processing

• A quick Tour of Batch-Processing - ArcGIS 10 Help topic

Finding Existing Tools and Scripts

• Geoprocessing Model and Script Tool Gallery - A hub of tools you can download and install in ArcGIS that other users have made. Very useful!
• See also this week's lab Tutorial Topic 1

Geoprocessing with Model Builder

• Designing and Building Geoprocessing Tools - This video walks you through the basics of building Geoprocessing Tools using Model Builder in ArcGIS 10.
• A quick tour of creating tools with ModelBuilder - ArcGIS 10 Help
• Tutorial for Building Tools with Model Builder - ArcGIS 10 Help Tutorial

Add-Ins

• Add-Ins for ArcGIS 10 - Add-Ins are a new feature in ArcGIS. This video describes them.
• Add-Ins Blog

Plug-Ins

Plug-Ins are typically tool-bars that are installed externally and produced by someone other than ESRI, to work

• <- Links on our website
• Plus,
  – Other Classes
  – ESRI Courses
  – ESRI Tutorials

http://gis.joewheaton.org/topics/geoprocessing#TOC-Additional-Resources
TODAY’S PLAN...

I. Geoprocessing

II. ModelBuilder

III. Batch Processing
   I. A right-click away
   II. Scripting
THE MODELUILDER INTERFACE

- ModelBuilder window: where you edit, test and run models

MODEL ELEMENTS:

- **Variables**: Variables are elements in a model that hold a value or a reference to data stored on disk. There are two types of variables:
  - **Data**: Data variables are model elements that contain descriptive information about data stored on disk. Properties of data that are described in a data variable include field information, spatial reference, and path.
  - **Values**: Value variables are values such as strings, numbers, Booleans (true/false values), spatial references, linear units, or extents. Value variables contain anything but references to data stored on disk.

- **Connectors**: Connectors connect data and values to tools. The connector arrows show the direction of processing. There are four types of connectors:
  - **Data**: Data connectors connect data and value variables to tools.
  - **Environment**: Environment connectors connect a variable containing an environment setting (data or value) to a tool. When the tool is executed, it will use the environment setting.
  - **Precondition**: Precondition connectors connect a variable to a tool. The tool will execute only after the contents of the precondition variable are created.
  - **Feedback**: Feedback connectors connect the output of a tool back into the same tool as input.
• Model elements have 3 states
• Not ready to run (parameters aren’t defined)
  Usually referred to as Derived Data
• Ready to run (all elements are colored)
• Already run (elements are colored and shaded)
TODAY’S PLAN...

I. Geoprocessing
II. ModelBuilder

III. Batch Processing
   I. A right-click away
   II. Scripting

IV. Python Scripting
V. Efficiency
BRUTE FORCE...

• If you do something once or twice, brute force may suffice...

• If you start doing it more than that, make yourself a tool
  – Limits opportunity for sloppy mistakes
  – Makes
  – If good enough, share it with others...
BATCH PROCESSING

- I got to do this same thing 100 times
- AGHHH
- Brute force or Batch Process?
- Use right-click in Arc Toolbox and click ‘batch’
- Write a script that loops through same thing (input varies)
TODAY’S PLAN...

I. Geoprocessing
II. ModelBuilder
III. Batch Processing
   I. A right-click away
   II. Scripting
ANY GEOPROCESSING TASK:

• That you think you want to do more then once...
• Just right click...
THEN .... FILL OUT BATCH GRID

- Add as many rows as you want (batch)...
- Double click on each cell to fill out...
- Copy and Paste from above.
TODAY’S PLAN...

I. Geoprocessing
II. ModelBuilder
III. Batch Processing
   I. A right-click away
   II. Scripting
SCRIPTING IN GIS

• Sequences of GIS operations that can be stored and shared
  • Native languages: Arc Macro Language, Avenue
  • Software independent languages such as Visual Basic for Applications, Perl, or Python
  • a model can be written and executed as a script

• Scripts can be manipulated visually
  • e.g., through ESRI's ModelBuilder
PROGRAMMING?

- Lots of languages (syntax is critical to speak)
- Loops (for or do)
- Conditional (if, then, else)
Simple Loops

```
SQL> declare
  2    counter number := 0;
  3  begin
  4    LOOP
  5      counter := counter + 1;
  6      dbms_output.put_line(counter);
  7      EXIT WHEN counter = 5;
  8  END LOOP;
  9 end;
10 /
```

PL/SQL procedure successfully completed.

Use the EXIT statement to exit loop

Initialize counter!
BUT I DON’T KNOW HOW TO CODE?

• Three simple tricks to learning:
  1. Build a model (visually) and then ‘Export -> To Python Script…’
  2. Try Geoprocessing Results -> ‘Copy as Python Snippet…’
  3. Python scripting window and help!
1. OPEN IT UP... AND READ IT...

• Try and run it at command prompt!
2. COPY AS PYTHON SNIPPET

• The snippet shows the syntax for any geoprocessing command you just ran...

```python
# Replace a layer/table view name with a path to a dataset (which can be a layer file) or create the layer/table view within the script
# The following inputs are layers or table views: "ShearZones", "Velocity_All"
arcpy.gp.zonalStatistics_sa("ShearZones","OBJECTID","Velocity_All","C:/Users/Joe Wheaton/Documents/ArcGIS/Default.gdb/ZonalSt_Shear","MEAN","DATA")
```
3. PYTHON WINDOW

• Start typing... & let auto-complete help

• Then tab select then tab... Type dot `.'
3. PYTHON WINDOW

- Start typing first few letters of command...

![Python window with `arcpy.cli` typed and tab selections]

- Then tab select then tab... Type ',' & fill out:

![Python window with `arcpy.Clip_analysis` function call and parameters]
3. PYTHON WINDOW

• Then run it...

```python
>>> arcpy.Clip_analysis
("Points","ShearZones","dummy.shp")
<Result 'C:\Users\Joe Wheaton\Documents\ArcGIS \dummy.shp'>
```

• It runs, shows you status, saves to geoprocessing results, and adds to display...
JUST PLAY....

• If you really want to learn this stuff:
  – Find a problem you need to solve, have a play, try to bolt pieces together, try to get it to work
  – Lots of help and forums and examples to draw off of

• OR:
  – Take Python Class in Fall (WILD 6900) – 5 week espresso course
  – Take Ethan White’s BIOL 4040/6040 – ‘Python Programming for Biologists’ in Fall
TODAY’S PLAN...

I. Geoprocessing
II. ModelBuilder
III. Batch Processing
   I. A right-click away
   II. Scripting
TODAY’S PLAN...

I. Geoprocessing
II. ModelBuilder
III. Batch Processing
   I. A right-click away
   II. Scripting
IV. Python Scripting
V. Efficiency
I. Uncertainty & Error
II. GIS Errors
III. Error Propagation
IV. All Bad?
V. Summary of GIS Fundamentals
“But there are also unknown unknowns: the ones we don't know we don't know.”
- Donald Rumsfeld

“It’s not the things you don’t know that matter, it’s the things you know that ain’t so.”
- Will Rogers
UNCERTAINTY...

Lack of sureness about something... NOT a lack of knowledge.

To the general public and decision makers:
- Sign of weakness
- Like saying you don’t know anything
- Confusing

To you and I (scientists):
- A statement of knowledge
- Useful information
- Full-employment act
MORE CONSTRUCTIVE DEFINITION

- Considered in terms of sources
- Provides a rationale for treating different sources differently

Figure Adapted from Van Asselt and Rotmans (2002): http://dx.doi.org/10.1023/A:1015783803445
• Uncertainty does not equate to a lack of knowledge
• A statement of uncertainty is not a sign of weakness... it is useful information
• ‘What in life is worth having that you didn’t have to take a risk to get?’ – Mike Clark

Figure from Wheaton et al. (2008)
TODAY’S PLAN…

I. Uncertainty & Error
II. GIS Errors
III. Error Propagation
IV. All Bad?
V. Summary of GIS Fundamentals
GIS DATA ACCURACY

- **Accuracy** is how close an observation (or GIS data layer) is to the *truth*
- **Error** is the measure of how far a measure or observation deviates from the truth
- Many different ways to have errors or blunders
- Are they same?
OTHER WAYS TO BE WRONG...

- Spatial data accuracy issues:
  
a) Positional accuracy

b) Attribute accuracy

c) Logical consistency

d) Completeness
FOUR END MEMBERS

- Positional accuracy of intersection of two freeways

From Chapter 14 of Bolstad (2008)
PRECISION OR RESOLUTION

NOT THE SAME AS ACCURACY!

Precision: the exactness of measurement or description
- the “size” of the “smallest” feature that can be displayed, recognized, or described

- For raster data, it is the size of the pixel (resolution)
- For vector point data, it is the point density

- resolution and positional accuracy
  - you can see a feature (resolution), but it may not be in the right place (accuracy)
  - higher accuracy generally costs much more to obtain than higher resolution
HOW POSITIONAL ACCURACY IS CALCULATED

• All you need is measured coordinates and ‘true’ coordinates
• The lower the error distance, the more accurate...

\[
\text{error distance} = \sqrt{(x_t - x_d)^2 + (y_t - y_d)^2}
\]

From Chapter 14 of Bolstad (2008)
IMPLICATION OF ERROR DISTRIBUTIONS

- How would I get a plot like this?
- If we take 95% of the error...
- With same mean, but different distributions, implications are quite different...

From Chapter 14 of Bolstad (2008)
HOW TO CALCULATE THOSE POSITIOINAL ERRORS

• Find, define or assume *true* values
• Find values of layer to calculate errors for
• Create error field
• Plug and chug
• THIS IS NOT TECHNICALLY CORRECT

From Chapter 14 of Bolstad (2008)
A CLOSER LOOK

From Chapter 14 of Bolstad (2008)

• 95% of the data... depends on distribution shape...

\[
e = \sqrt{(x_\cdot - x_*)^2 + (y_\cdot - y_*)^2}
\]
PUT IT ALL TOGETHER...

• Simple excel or field calculator exercise?

• How would you do it?

<table>
<thead>
<tr>
<th>ID</th>
<th>x (true)</th>
<th>x (data)</th>
<th>x difference</th>
<th>(x difference)^2</th>
<th>y (true)</th>
<th>y (data)</th>
<th>y difference</th>
<th>(y difference)^2</th>
<th>sum x diff^2 + y diff^2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12</td>
<td>10</td>
<td>2</td>
<td>4</td>
<td>288</td>
<td>292</td>
<td>-4</td>
<td>16</td>
<td>20</td>
</tr>
<tr>
<td>2</td>
<td>18</td>
<td>22</td>
<td>-4</td>
<td>16</td>
<td>234</td>
<td>228</td>
<td>6</td>
<td>36</td>
<td>52</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>12</td>
<td>-5</td>
<td>25</td>
<td>265</td>
<td>266</td>
<td>-1</td>
<td>1</td>
<td>26</td>
</tr>
<tr>
<td>4</td>
<td>34</td>
<td>34</td>
<td>0</td>
<td>0</td>
<td>243</td>
<td>240</td>
<td>3</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>5</td>
<td>15</td>
<td>19</td>
<td>-4</td>
<td>16</td>
<td>291</td>
<td>287</td>
<td>4</td>
<td>16</td>
<td>32</td>
</tr>
<tr>
<td>6</td>
<td>33</td>
<td>24</td>
<td>9</td>
<td>81</td>
<td>211</td>
<td>215</td>
<td>-4</td>
<td>16</td>
<td>97</td>
</tr>
<tr>
<td>7</td>
<td>28</td>
<td>29</td>
<td>-1</td>
<td>1</td>
<td>267</td>
<td>271</td>
<td>-4</td>
<td>16</td>
<td>17</td>
</tr>
<tr>
<td>8</td>
<td>7</td>
<td>12</td>
<td>-5</td>
<td>25</td>
<td>273</td>
<td>268</td>
<td>5</td>
<td>25</td>
<td>50</td>
</tr>
<tr>
<td>9</td>
<td>45</td>
<td>44</td>
<td>1</td>
<td>1</td>
<td>245</td>
<td>244</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>110</td>
<td>99</td>
<td>11</td>
<td>121</td>
<td>221</td>
<td>225</td>
<td>-4</td>
<td>16</td>
<td>137</td>
</tr>
<tr>
<td>11</td>
<td>54</td>
<td>65</td>
<td>-11</td>
<td>121</td>
<td>212</td>
<td>208</td>
<td>4</td>
<td>16</td>
<td>137</td>
</tr>
<tr>
<td>12</td>
<td>87</td>
<td>93</td>
<td>-6</td>
<td>36</td>
<td>284</td>
<td>278</td>
<td>6</td>
<td>36</td>
<td>72</td>
</tr>
<tr>
<td>13</td>
<td>23</td>
<td>22</td>
<td>1</td>
<td>1</td>
<td>261</td>
<td>259</td>
<td>2</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>14</td>
<td>19</td>
<td>24</td>
<td>-5</td>
<td>25</td>
<td>230</td>
<td>235</td>
<td>-5</td>
<td>25</td>
<td>50</td>
</tr>
<tr>
<td>15</td>
<td>76</td>
<td>80</td>
<td>-4</td>
<td>16</td>
<td>255</td>
<td>260</td>
<td>-5</td>
<td>25</td>
<td>41</td>
</tr>
<tr>
<td>16</td>
<td>97</td>
<td>108</td>
<td>-11</td>
<td>121</td>
<td>201</td>
<td>204</td>
<td>-3</td>
<td>9</td>
<td>130</td>
</tr>
<tr>
<td>17</td>
<td>38</td>
<td>43</td>
<td>-5</td>
<td>25</td>
<td>290</td>
<td>288</td>
<td>2</td>
<td>4</td>
<td>29</td>
</tr>
<tr>
<td>18</td>
<td>65</td>
<td>72</td>
<td>-7</td>
<td>49</td>
<td>277</td>
<td>282</td>
<td>-5</td>
<td>25</td>
<td>74</td>
</tr>
<tr>
<td>19</td>
<td>85</td>
<td>78</td>
<td>7</td>
<td>49</td>
<td>205</td>
<td>201</td>
<td>4</td>
<td>16</td>
<td>65</td>
</tr>
<tr>
<td>20</td>
<td>39</td>
<td>44</td>
<td>-5</td>
<td>25</td>
<td>282</td>
<td>278</td>
<td>4</td>
<td>16</td>
<td>41</td>
</tr>
<tr>
<td>21</td>
<td>94</td>
<td>90</td>
<td>4</td>
<td>16</td>
<td>246</td>
<td>251</td>
<td>-5</td>
<td>25</td>
<td>41</td>
</tr>
<tr>
<td>22</td>
<td>64</td>
<td>56</td>
<td>8</td>
<td>64</td>
<td>233</td>
<td>227</td>
<td>6</td>
<td>36</td>
<td>100</td>
</tr>
</tbody>
</table>

Sum 1227
Average 55.8
RMSE 7.5
NSSDA 12.9

From Chapter 14 of Bolstad (2008)
WHAT ABOUT POSITIONAL ACCURACY OF SHAPES AS OPPOSED TO VERTICIES?

- Compare true line location to various representations of actual to define epsilon band...

From Chapter 14 of Bolstad (2008)
MEASUREMENT OF POSITIONAL ACCURACY

• Usually measured by root mean square error: the square root of the average squared errors

\[ RMSE = \sqrt{\frac{e_1^2 + e_2^2 + e_3^2 + \cdots + e_n^2}{n-1}} \]

where \( e_i \) is the distance (horizontally or vertically) between the true location of point \( i \) on the ground, and its location represented in the GIS.

• Loosely we say that the RMSE tells us how far recorded points in the GIS are from their true location on the ground, on average.

• More correctly, based on the normal distribution of errors, 68% of points will be RMSE distance or less from their true location, 95% will be no more than twice this distance, providing the errors are random and not systematic (i.e., the mean of the errors is zero)
• Manual digitizing
  – significant source of positional error (roads, streams, polygons)

• Source map error
  – scale related generalization
  – line thickness

• Operator error
  – under/overshoot
  – time related boredom factor
ERROR – OUT OF DATE

• Belvue Washington... At one time it was ‘right’

From Chapter 14 of Bolstad (2008)
IMPRECISE AND VAGUE
MIXED UP

UNCERTAINTY DUE TO VARIABILITY

Inexactness

HNETIC

Inherent Randomness

VALUE DIVERSITY

Geopolitical

Behavioral Diversity

SOCIETAL RANDONOMES

Technological Surprises

UNRELIABILITY

UNCERTAINTY DUE TO LIMITED KNOWLEDGE

STRUCTURAL UNRELIABILITY

Inexactness

Investing Observations & Measurements

Practical

Indeterminate

Inductive Ignorance

Predictable Ignorance

Increasing Uncertainty

505.9

238.4

238.4

505.9
JUST WRONG
City of Sapporo, Japan

GENERALIZATION
UNCERTAINTY IN ANALYSIS

• Just cause you think it will work does not guarantee success—Always LOOK at the results of your analysis!
  – What would a certain combination of inputs result in?
  – How is that likely to change across all inputs?

• Functional REDUNDANCY:
  – There is almost always another (often faster) way of performing any analysis
  – Should produce the *same* result... try it?
DATA QUALITY:  HOW GOOD IS YOUR DATA?

• Scale
  – Can be an output issue; at what scale do I wish to display?
  – Analyses are only as good as the coarsest input

• Precision or Resolution
  – the exactness of measurement or description
  – Determined by input; can output at lower (but not higher) resolution

• Accuracy
  – the degree of correspondence between data and the real world
  – Fundamentally controlled by the quality of the input

• Lineage
  – The original sources for the data and the processing steps it has undergone

• Currency
  – the degree to which data represents the world at the present moment in time

• Documentation or Metadata
  – data about data: recording all of the above

• Standards
  – Common or “agreed-to” ways of doing things
  – Data built to standards is more valuable since it’s more easily shareable
ERROR HANDLING 101

• Awareness
  – knowledge of types, sources and effects

• Minimization
  – use of best available data
  – correct choices of data model/method

• Communication
  – to end user via metadata, honest and thorough reporting of uncertainties
TODAY’S PLAN…

I. Uncertainty & Error
II. GIS Errors
III. Error Propagation
IV. All Bad?
V. Summary of GIS Fundamentals
ERROR PROPAGATION

• Methods for assessing the effects of known degrees of error in a model's inputs
  • Producing measures of confidence in model outputs
  • Normally by simulation
DEM DIFFERENCING

Simple method of quantifying spatial variations in change in storage terms of a sediment budget.

\[ Q_{bi} - Q_{bo} = (1 - \eta) \frac{dV_b}{dt} \]

Where:
- \( Q_{bi} \) is the volumetric rate of bed material transport
- \( Q_{bo} \) is the volumetric rate of bed material transport
- \( \eta \) is the porosity of bed material
- \( dV_b/dt \) is the volumetric rate of change of the bed material


© Wheaton 2008
MINIMUM LEVEL OF DETECTION

- Distinguish those changes that are real from noise
- Use standard Error Propagation
- Errors assumed to be spatially uniform, but can vary temporally

\[
\delta(z) = \sqrt{\left(\delta(z)_{DEM\ old}\right)^2 + \left(\delta(z)_{DEM\ new}\right)^2}
\]

E.g. \(\delta(z) = \sqrt{(10)^2 + (20)^2} = 22.36\)

22.36 cm \(\approx\) 8.8 in

See
- Lane et al (2003): ESPL
HOW DOES A $\text{minLoD}$ GET APPLIED?

- You take original DoD, and remove all changes $\leq \text{minLoD}$
- For example +/- 20 cm
- How would you do that?
- What is the assumption here?
HOW COULD I REPRESENT AS PROBABILITY?

- Using inferential statistics, we’ll calculate a t-score
- $\sigma_{DoD}$ is the characteristic uncertainty
  - In this case $\sigma_{DoD} = \min \text{LoD}$

\[
t = \frac{|Z_{DEM_{new}} - Z_{DEM_{old}}|}{\sigma_{DoD}}
\]

- Just the ratio of actual change to $\min \text{LoD}$ change
- Assuming two-tailed test, t is significant at:
  - 68% confidence limit when $t=1$
  - 95% confidence limit when $t=1.96$
Even when $\min$ LoD is spatially constant, probability varies in space... why?
APPLY FIS ON CELL BY CELL BASIS

FIS Input 1 (Point Density) + FIS Input 2 (Slope) + FIS Input 3 (GPS Quality) = FIS Surface (El. Unc (m))

Both FIS Surfaces Combined to DoD Probability
SENSITIVITY OF THRESHOLD?

A graph shows the sensitivity of the threshold with different confidence intervals (CI). The x-axis represents the CI threshold, ranging from 0.5 to 1, and the y-axis represents the volume of sediment in cubic meters (m³). The graph compares unthresholded DoD with 50%, 68%, 95%, and 99% confidence interval thresholds. The legend indicates different colors for DoD (m) with thresholds ranging from > -2.5 to 1.0.

A. Unthresholded DoD
B. 50% Confidence Interval Threshold
C. 68% Confidence Interval Threshold
D. 95% Confidence Interval Threshold
E. 99% Confidence Interval Threshold

The diagrams illustrate the spatial distribution of sediment volume with varying thresholds, with colors representing different DoD values.
TODAY’S PLAN...

I. Uncertainty & Error
II. GIS Errors
III. Error Propagation
IV. All Bad?
V. Summary of GIS Fundamentals
These contrasting philosophical approaches to dealing with uncertainty are rarely explicitly identified.
REDUCE UNCERTAINTY

- Uncertainty is a nuisance
- It should be constrained wherever possible
- Unquantifiable uncertainty difficult or impossible to constrain

Figure from Wheaton (2004)
COPE WITH UNCERTAINTY

- Fuller appreciation of types of uncertainty
- Uncertainty still viewed as a nuisance
- Acceptance of uncertainty as a given
- Explicit link to adaptive management

Figure from Wheaton (2004)
EMBRACE UNCERTAINTY

- Uncertainty seen as useful information
- Explicit recognition of uncertainty sources
- Use of natural variability as an opportunity
- Explicit linked to adaptive management
TRANSFORM UNCERTAINTY TYPES

- Central to embracing uncertainty
- Many examples of *structural uncertainties* & *uncertainties due to variability* can be transformed (and thereby reduced) to *unreliability uncertainties*

Figure Adapted from Van Asselt and Rotmans (2002): [http://dx.doi.org/10.1023/A:1015783803445](http://dx.doi.org/10.1023/A:1015783803445)
HOW TO COMMUNICATE UNCERTAINTY WITHOUT SOUNING LIKE A QUACK?

- Know the **audience** (general public vs. peers)
- Complete transparency of **source** and **type** of uncertainties
- Relate **significance** in terms of audience’s criteria
- Clear identification of uncertainties leading to **risks** versus **opportunities** versus both
- Distinguish between **transformable** uncertainties & total unknowns (e.g. irreducible ignorance)
- Highlight **tradeoff** between cost of **knowing more** and taking **acceptable risks**
## MIX OF COMMUNICATION OPTIONS

<table>
<thead>
<tr>
<th>Method</th>
<th>Appropriate For</th>
</tr>
</thead>
<tbody>
<tr>
<td>Qualitative Description</td>
<td>Unquantifiable and/or unquantified uncertainties</td>
</tr>
<tr>
<td>Probabilities</td>
<td>Expressions of confidence or likelihood</td>
</tr>
<tr>
<td>Measures of Variance</td>
<td>Uncertainties due to variability</td>
</tr>
<tr>
<td>Upper &amp; Lower Limits (+/-)</td>
<td>Well constrained uncertainties due to inexactness</td>
</tr>
<tr>
<td>Fuzzy Numbers</td>
<td>Uncertainties due to vagueness and ambiguity</td>
</tr>
<tr>
<td>Scenarios &amp; Conceptual Models or Simulation Models</td>
<td>Uncertainty about future (gets away from actual prediction)</td>
</tr>
<tr>
<td>Definition of Plausible Outcomes</td>
<td>Structural &amp; Variability Uncertainties Leading to Predictive Uncertainty</td>
</tr>
</tbody>
</table>
Lab 3: Reproducing Maps – Geologic Map

- Teach you how to manipulate display properties and symbology to reproduce a map in as close as possible a fashion to the original.
- Teach you how to extract summary statistics and data from existing data.
Reading for Tuesday, January 31st
posted 19 minutes ago by Joe Wheaton

Finish reading the following by Tuesday, January 31, 2011 (before lecture):

- Skim Chapters 6 & 7 of Bolstad (2008)
  - Just skim through everything and focus on the big picture concepts about what are the primary sources of GIS data... don't get bogged down in details
- Also, have a look through and read of the links and ESRI help topics below:

Data Sources

- GIS Data Links - This is a page I maintain with some links to downloadable GIS Data sources
- GIS Class LibGuide - This is a custom page our librarian has put together just for this class!
- Datasets Provided with ArcGIS - These are known as ESRI Data and Maps
- A quick tour of ArcCatalog - ArcCatalog is the place to manage and browse your GIS data with ArcGIS
- Adding data from ArcGIS online - There are wealth of great basemap layers you can now load into your maps without having to download anything! Very convenient!

Metadata & Displaying Data

- What is metadata?
- A quick tour of metadata
- Editing Metadata (in ArcGIS)