Analyzing Multidimensional Scientific Data in ArcGIS

John Fry
Solution Engineer, Esri
jfry@esri.com
Fishing Performance Surface

ESRI Federal User’s Conference

2/10/2015

Christopher Henton

Christopher.henton.ctr@navy.mil

228-688-5314
Background

• **Naval Oceanographic Office**
  – NOAA for the NAVY
  – Survey ships
  – Oceanographic models

• **Requirement:**
  – Various customers have requested a performance surface model to predict ideal locations for fishing based on biological and environmental data.

• **GOAL:**
  – To identify potential fishing zone (PFZ) forecasting.
  • “**Best Areas**” where fish could be!
Species Environmental Envelope (HSPEN):

- In Aquamaps, this represents a “response curve describing the habitat usage of a species or its preferences with respect to specific environmental parameters”;
- Environmental parameters include:
  - Depth (m)
  - Surface Temperature (°C)
  - Surface Salinity (PSU)
  - Primary Production (mgC·m⁻²·day⁻¹)
  - Sea Ice Concentration (% cover)
  - Distance to Land (km);
- The environmental envelope is presented as a range of values for each parameter: Minimum, Preferred minimum (Minₚ), Preferred maximum (Maxₚ), Maximum.

Source: K. Reyes, AquaMaps: Algorithm and Data Sources for Aquatic Organisms
Normalization

- **In Aquamaps, this represents a “response curve describing the habitat usage of a species or its preferences with respect to specific environmental parameters”;**

- **Environmental parameters include:**
  - Depth (m)
  - Surface Temperature (°C)
  - Surface Salinity (PSU)
  - Primary Production/Chlorophyll (mgC·m⁻²·day⁻¹)
  - Sea Ice Concentration (% cover)
  - Distance to Land (km)

- **The environmental envelope is presented as a range of values for each parameter: Minimum, Preferred minimum (MinP), Preferred maximum (MaxP), Maximum.**

<table>
<thead>
<tr>
<th>Species Envelope (HSPEN)</th>
<th>Min</th>
<th>Pref Min (10th)</th>
<th>Pref Max (90th)</th>
<th>Max</th>
<th>Common Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depth (m)</td>
<td>0</td>
<td>2</td>
<td>11</td>
<td>20</td>
<td>Abe’s flyingfish</td>
</tr>
<tr>
<td>Temperature (°C)</td>
<td>17.99</td>
<td>26.37</td>
<td>29.15</td>
<td>33.35</td>
<td>Abe’s flyingfish</td>
</tr>
<tr>
<td>Salinity (psu)</td>
<td>28.81</td>
<td>32.27</td>
<td>35.81</td>
<td>36.79</td>
<td>Abe’s flyingfish</td>
</tr>
<tr>
<td>Primary Production</td>
<td>243</td>
<td>332</td>
<td>1577</td>
<td>2059</td>
<td>Abe’s flyingfish</td>
</tr>
</tbody>
</table>

\[ a + (x - A) \times (b - a) / (B - A) \]
Normalization Examples

- **Temperature (°C)**: 17.99, 26.37, 29.15, 33.35
- **Depth (m)**: 0, 2, 11, 20

Abe's flyingfish

Abe's flyingfish
Probabilities of Occurrence

“The relative probabilities of occurrence are presented as probability values ranging between 0.00-1.00. A probability of occurrence is first calculated for the each of the individual environmental predictors.”

\[ P_c = P_{bathymetry_c} \times P_{temperature_c} \times P_{salinity_c} \times P_{primary~production_c} \]

“This multiplicative approach allows each environmental predictor to act as a “knock-out” criterion.”

Multidimensional Processing

\[ S_s = D_i \times Chl_S \times \sum_D (T_D \times S_D + F_D) \]

\[ F_S = \frac{\sum S_s}{F_S_{(max)}} \]

\[ P_O = \frac{\sum (F_S)}{P_O_{(max)}} \]

**Definitions:**
- **\( S_s \):** Species Sum
- **\( Chl_S \):** Chlorophyll at surface
- **\( S_D \):** Salinity at depth
- **\( D_i \):** Distance from Shore
- **\( T_D \):** Temperature at depth
- **\( F_D \):** Frontal boundaries at depth
- **\( F_S \):** Fish Family Sum
- **\( P_O \):** Probability of Occurrence
**Experiment**

- **Environmental Factors**
  - *Salinity (PSU): by depth / forecasted time*
  - *Temperature (°C): by depth / forecasted time*
  - *Chlorophyll (µg/L): Remote Sensing (monthly)*
  - *Distance from Shore*
  - *Depth*
  - *Frontal Boundaries (MGET/Modeled – Not included yet)*

- **Fish Habitat Preferences (AquaMaps)**

- **Maritime Boundaries, EEZ, Fish Closure Areas, Marine Protected Areas (layers)**

- **10 Fish Families**
  - **78 Species**

- **Raw Data**
  - Temp
  - Salinity
  - Chlorophyll
  - Fish Preference

- **Normalize / Fuse**

- **Output**
  - HTML
  - Shapefile
  - KML/Image

Naval Oceanography
### Data Management

- **78 Species**
- **96hr forecast**
- **200 Depths**
- **Water Temp & Salinity**

**Results**
- **5,331,200 Raster Layers!!!!!!!!!**
- ~**1TB**
- ~**5 Days**

**Equation:**

\[
(T_m \times D_r) + [(T_m \times D_r) \times (N_p) \times (F_s)] + [(T_m \times D_r) \times (F_s)] + (T_m \times D_r) \times E_p = R_l
\]

\[
[(34 \times 200) + (34 \times 200) \times (4) \times (78)] + [(34 \times 200) \times (78)] + (34 \times 200) \times 2 = 5,331,200 \text{ Raster Layers}
\]

- **\(T_m\)** = Time Range
- **\(D_r\)** = Depth Range
- **\(N_p\)** = Normalized Preference
- **\(F_s\)** = Fish Species
- **\(E_p\)** = Environmental Parameter
- **\(R_l\)** = Raster Layers

**Extraction**
- **Environmental Normalization**
- **Depth Normalization**
- **\(F_s\) Sum**

**Results**
- **5,331,200 Raster Layers!!!!!!!!!!**
- ~**1TB**
- ~**5 Days**
Summary

• Multidimensional problems are complex and require new ways of processing and thinking.
  – Data Management
  – Analytics
  – Visualization

• Broader temporal scales
  – Climatology
  – Predictive (Months/Years/Decades)
REFERENCES


• K. Reyes, AquaMaps: Algorithm and Data Sources for Aquatic Organisms


Additional Reading:


• Josephine “Skit” Rius-Barile. AquaMaps - behind the scenes (http://www.aquamaps.org/main/home.php)


Outline

- ArcGIS and Scientific Data
- Multidimensional Tools
- Ingest and aggregation
- Visualization and Analysis
- Services, Ready-to-Use Maps, Web Applications
Communicating Science to Decision Makers

“Winds in the Area of Interest will be sustained at 20kts with gusts to 35kts.”

“You can conduct X operation at these locations, but not at these locations.”

“What does that mean???”

“Will my mission be a success?”

“What platform can I use?”

“Waves in the Area of Interest will between 2.5 m and 3 m.”
Communicating Science to Decision Makers

“Reduction in upwelling and mixing not beneficial in these locations. Stratification occurring at these locations”

“What does that mean???”

“Will it be too costly to fix?”

“What techniques should I use?”

“Pollutant runoff in surrounding regions needs to be addressed”

“Fish yields will be low in these areas”
Visualize Scientific Data Clearly

http://dtc-sci01.esri.com/DeadZoneStoryMap/

Dead Zones in our Oceans

Dead zones or hypoxia, refers to a reduced level of oxygen in the water that either kills fish and other marine life or, if they are mobile such as fish, leave the area.

2 Stratification: Water Temperature

Stratification of ocean water is a naturally occurring phenomenon that is important to the structure, circulation and productivity of the oceans. The formation of vertical stratification in the water column is a consequence of water masses with different densities. Water density is strongly influenced by temperature and salinity; with less dense, warmer surface waters floating on top of denser, colder waters.

This example is displaying water temperature over depth and time. Areas in blue are waters that are cool in temperature and red are areas that are warmer. These datasets are created in a scientific data format known as netCDF. Researchers, Scientists, GIS Analyst, and policy makers need to get their hands on these datasets and be able to overlay additional geospatial information or perform analysis, to understand our environment.

http://centerforoceansolutions.org/climate/impacts/ocean-warming/water-column-stratifi/

3 Stratification: Salinity
Scientific Data

- Oceanographic data
  - Sea temperature
  - Salinity
  - Ocean current

- Meteorological data
  - Temperature
  - Humidity
  - Wind

- Land
  - Soil moisture
  - NDVI
  - Land cover
1. Latitude (y)
2. Longitude (x)

2D
1. Latitude (y)
2. Longitude (x)
3. Height (z)
1. Latitude (y)
2. Longitude (x)
3. Time (t)
NetCDF WMS with n Dimensions

This example shows the new capability within 10.2.1 to serve out a NetCDF file as a WMS with time and an n dimension such as depth.

1. Latitude (y)
2. Longitude (x)
3. Depth (d)
4. Time (t)
Scientific Data in ArcGIS
Scientific formats in ArcGIS 10.3

- **GRIB** (General Regularly-distributed Information in Binary form)
  - A gridded format used operationally worldwide by most meteorological centers.

- **netCDF** (Network Common Data Form)
  - A self-describing, machine-independent data format that support the creation, access, and sharing of array-oriented scientific data.

- **HDF** (Hierarchical Data Format)
  - A scientific format that supports a proliferation of different data models, including multidimensional arrays, raster images, and tables.

Variable / Dimension
- Temperature / Time
- Salinity / Depth
- Pressure / Altitude
CF Convention

• Climate and Forecast (CF) Convention
  - [http://cfconventions.org/](http://cfconventions.org/)

• Initially developed for:
  - Climate and forecast data
  - Atmosphere, surface and ocean model-generated data
  - Also for observational datasets

• CF is now the most widely used conventions for geospatial netCDF data. Current version 1.6

• According to CF conventions, variables representing time must always explicitly include the units attribute.

• You can use Compliance checker utility to check a netCDF file.

• CF-Convention Compliance Checker for netCDF Format
Using Scientific Data in ArcGIS

Behaves the same as any layer or table

• Display
  o Same display tools for raster and feature layers will work on multi-dimensional netCDF raster and netCDF feature layers.

• Graphing
  o Driven by the table just like any other chart.

• Animation
  o Multi-dimensional data can be animated through time dimension

• Analysis Tools
  o Will work just like any other raster layer, feature layer, or table. (e.g. create buffers around netCDF points, reproject rasters, query tables, etc.)
Ingesting Scientific data in ArcGIS

- Directly reads netCDF file using
  - Make NetCDF Raster Layer
  - Make NetCDF Feature Layer
  - Make NetCDF Table View

- Directly reads HDF and GRIB data as raster

Multidimension Supplemental Tools
Download from ArcGIS Online
Changing Time Slice

![Image of map with grid and time slice options]

Time = 1

In the map, the time slice for the variable 'time' is set to 12/31/1995. The grid on the right shows the dimension values with 'time' set to 12/31/1995.
Spatial and Temporal Analysis

- Several hundreds analytical tools available for raster, features, and table
- Temporal Modeling
  - Looping and iteration in ModelBuilder and Python
Working with Multidimensional Data

- Direct Ingest
- Analysis
- Geoprocessing Models
- Predictive Analysis Toolset
N-dimensional capability in ArcGIS 10.3

- Observations
- Modeling
  - Processing on-the-fly Raster Functions
  - Mosaic Data Set
- Native METOC Data Formats
  - netCDF
  - GRIB
  - HDF

Visualization
- Server
- ArcGIS Desktop

Dissemination

Exploitation
Raster Mosaic Dataset for N-dimensional Data (ArcGIS 10.3)

Resolves Traditional Raster Data Management Issues

Processing Time
Reduced exponentially

Overlapping Imagery
Pick “Best” raster

Disparate Datasets
Handle large NoData areas

Data Quality
Reduces resampling

Storage
Reduces storage by removing redundancy

Multi-resolution Data
No need to sample up or down

Maintenance
Add Rasters as required

Maintain Metadata
Appended directly upon registration
Scientific data support in Mosaic Dataset

- Supports netCDF, HDF and GRIB
  - Spatial Aggregation
  - Temporal Aggregation
  - On-the-fly analysis
- Accessible as Map Service
- Accessible as Image Service
- Supports direct ingest
- Eliminates data conversion
- Eliminates data processing
- Improves workflow performance
- Integrates with service oriented architecture
Multidimensional Mosaic Datasets

Aggregate (mosaic) spatial, time, and vertical dimensions

- Raster Types for netCDF, HDF & GRIB
- Define variables when adding Rasters
- Each Row is a 2D Raster with variables and dimension values
- Define on-the-fly processing
- Serve as Multidimensional Image Service
  Map Service
  WMS

Define variables when adding Rasters
- Each Row is a 2D Raster with variables and dimension values
- Define on-the-fly processing
- Serve as Multidimensional Image Service
  Map Service
  WMS

About the NetCDF variables property page

Options:
- Select All
- Unselect All
- Remove Unselected

Define variables:
- Name: water_temp
  Type: Water Temperature
  Description: (time=25, depth=40, lat=561, lon=401)
- Name: surf_el
  Type: Water Surface Elevation
  Description: (time=25, lat=561, lon=401)
- Name: tau
  Type: Tau
  Description: (time=25)

Extract original dimension values

Table:

<table>
<thead>
<tr>
<th>OBJ</th>
<th>Name</th>
<th>Variable</th>
<th>Standard Time</th>
<th>Standard Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>&lt;Raster</td>
<td>water_temp:0</td>
<td>5/17/2013</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>&lt;Raster</td>
<td>water_temp:1</td>
<td>5/17/2013</td>
<td>-2</td>
</tr>
<tr>
<td>3</td>
<td>&lt;Raster</td>
<td>water_temp:2</td>
<td>5/17/2013</td>
<td>-4</td>
</tr>
<tr>
<td>4</td>
<td>&lt;Raster</td>
<td>water_temp:3</td>
<td>5/17/2013</td>
<td>-6</td>
</tr>
<tr>
<td>5</td>
<td>&lt;Raster</td>
<td>water_temp:4</td>
<td>5/17/2013</td>
<td>-8</td>
</tr>
</tbody>
</table>
Mosaic Dataset - Visualization

- Visualize variables at any time and vertical dimension
- Visualize flow direction and magnitude variables
- Visualize temporal change of a variable
Visualization of Raster as Vectors

- New Vector Field renderer for raster
  - Supports U-V and Magnitude-direction
  - Dynamic thinning
  - On-the-fly vector calculation

- Eliminates raster to feature conversion
- Eliminates data processing
- Improves workflow performance
Modeling with Raster Function Templates (RFT)

- A scientific model = a raster function template

- Windchill = $35.74 + 0.6215T - 35.75(V^{0.16}) + 0.4275T(V^{0.16})$

- RFT
  - Created using
  - Contains Query, GroupName, Variables
  - Set Mosaic dataset properties
  - Used during adding raster data
Fish suitability

- New Dimensions in ArcMap 10.3 Blog Post
- Shown at 2014 UC Plenary
Using the Mosaic Dataset for Analysis

- Creating Mosaic Datasets
- Adding functions to mosaic datasets
Create Space-Time Cube & Emerging Hot Spot Analysis
Using ArcGIS Pro with Multidimensional data

- Time Aggregation of Multi-dimensional data
- Update a Mosaic Dataset
Sharing Multidimensional Datasets

2 Ways to Serve Data

• Directly from individual raster dataset
• From mosaic dataset (requires image extension)
Publishing a WMS on ArcGIS Server

• Enable WMS capabilities on Service Editor or Manager
Ready-To-Use Analysis Services

- Esri hosted analysis on Esri hosted data
  - Simplify job of GIS Professionals
  - Can be used in models and scripts just like any other tool
  - Extend spatial analysis to a much broader audience
  - Available in Desktop or as REST service
Ready-to-Use Scientific Data Maps

- ArcGIS Online
- GEOSS Water Services:
  - GLDAS Soil Moisture Data
    - Evapotranspiration
    - Soil Moisture
    - Snow Pack
Web Application Samples - GitHub

- NetCDF MultiDimensional Template with ArcPy-NetCDF-Web-Charting-Tools
- WMS Multi Dimensional Esri Viewer
Tell the story of your scientific data – Create Story Maps

- Storymaps.esri.com

Dead Zones in our Oceans
http://dtc-sci01.esri.com/DeadZoneStoryMap

Connecticut Department of Energy
http://storymaps.esri.com
Web Resources
Things to Consider…

• Embrace the Common Data Model (netCDF, HDF etc.)
  • Use Data and metadata standards (OGC, CF etc)

• Provide “mechanism” so that we can access scientific data using a single set of APIs….
  • and can expect data to be CF complainant

• Make your data “spatial” (by specifying geographic or a projected coordinate system)
  • Clearly define workflow and requirements
  • Create sample tools / templates where possible –
    • ArcGIS Online / Solutions Pages / GitHub
Scientific Data Resources

- Esri Raster Function Templates GitHub - https://github.com/Esri/raster-functions
- Working with multidimensional data

Help:

Don’t forget to complete a session evaluation form!
Print your customized Certificate of Attendance!

Printing stations located on L St. Bridge, next to registration
GIS Solutions EXPO, Hall D

Monday, 12:30pm – 6:30pm
Tuesday, 10:45 AM–4:00 PM
- Exhibitors
- Hands-On Learning Lab
- Technical & Extended Support
- Demo Theater
- Esri Showcase
Networking Reception:

National Museum of American History

Tuesday, 6:30 PM–9:30 PM
Bus Pickup located on L Street
Interested in diving deeper into Esri technology?

Add a day to your Fed GIS experience and register to attend the Esri DevSummit Washington DC. Stop by the registration counter to sign up.
Understanding our world.