Tools for Integrating Remote Sensing and Organismal Occurrence Data Streams

NASA BEF Breakout Round 4: Trainings and Tutorials
22 May 2019
3.50-4.30pm & 4.45-5.25pm

While you are waiting:
1) Create an account at auth.mol.org/register
2) Go to tiny.cc/2019BEF
Tools for Integrating Remote Sensing and Organismal Occurrence Data Streams

AIST-16-0092

Walter Jetz (PI, Yale University)
Adam Wilson (Co-I, State University of New York, Buffalo)
Robert Guralnick (Co-I, University of Florida, Gainesville)

While you are waiting:
1) Create an account at auth.mol.org/register
2) Go to tiny.cc/2019BEF
Tools for Integrating Remote Sensing and Organismal Occurrence Data Streams

While you are waiting:
1) Create an account at auth.mol.org/register
2) Go to tiny.cc/2019BEF
AIST Program Objectives

- Reduce the risk, cost, size, and development time for Earth science space-based and ground-based information systems,
- Increase the accessibility and utility of science data, and
- Enable new observation measurements and information products.
80% of a data scientist’s time is spent finding, cleansing, and organizing data, leaving only 20% to actually perform analysis (IBM 2019)
Environmental Niche Modeling

Niche model in “Environmental Space”

“Species Distribution Modeling”. Google Scholar: 3.5m results (74k since 2018)!

Many ways to use these data (maxent and friends)

tiny.cc/2019BEF
Integrating species occurrence and environmental data is not trivial.

Spatio-temporal grain and uncertainties
- Occurrence record
- Environmental data

tiny.cc/2019BEF
Integrating species occurrence and environmental data is not trivial.

Example eBird transect

Spatiotemporal Uncertainties (location error)

Grain of Observation

American Badger
Taxidea taxus

Least Concern (IUCN 2016)
Motivations for cloud-based environmental annotation

1. Massive data of varying type
   a. Digitized specimen records
   b. Camera traps data
   c. GPS tags (ICARUS)
   d. Citizen science records

1. Observation-level uncertainties and grains (space and time)
   Need to be accounted for in annotation

1. Need for dynamic annotation
   a. Temperature on the day of observation?
   b. Long-term mean conditions?
   c. Environmental Anomalies?

tiny.cc/2019BEF
AIST Project Overview

NASA Technology Readiness Level (TRL)

This Project
Computational Infrastructure

Front End (GUI)

Back End
Today’s Tutorial

Data Upload
Upload a sample species occurrence dataset

Annotation
Annotate it with various environmental data

Data Download
Download the annotated data for further processing

Survey
Complete a survey about the design and utility of this application

tiny.cc/2019BEF
Open the workshop website and scroll down to “Explore the Prototype”
Project Overview: Data Input
A simple, step by step upload system
https://mol.org/upload-dev

You don't have any datasets yet.

Upload Data

tiny.cc/2019BEF
Upload - Locate local file

tiny.cc/2019BEF
Match Columns

We have detected columns that may match Map of Life field names. Please verify that column names were matched correctly using the drop-down menus below.

Match "scientificName" to:
- scientificName

Match "family" to:
- family

Match "decimalLatitude" to:
- decimalLatitude

Match "decimalLongitude" to:
- decimalLongitude

Match "coordinateUncertaintyInMeters" to:
- coordinateUncertaintyInMeters

Use example dataset available on website
Upload - Matching Successful, move to finalization
Uncheck permissions so the data isn’t published.
Wait until the data are processed
If you don’t see the “Annotate” button, confirm you are at mol.org/upload-dev
Environmental data availability increasing rapidly

Current:
- MODIS EVI
- CHELSA

Planned:
- Radar-derived forest cover
- Evapotranspiration
- Global 30m DEM
- Surface Water
May have to adjust x-axis by selecting dates of interest (post-2000)
Roe Deer
*Capreolus capreolus*

Or select any desired mammal species.
Environmental Annotator - LST (Day)

Roe Deer
Environmental Annotator

Enter a species name: Taxidee taxus

Colour by axis: Annual Mean Precipitation

Choose X Axis: Mean

Choose Y Axis: Mean

Temporal | Bivariate

Download
Phase II

Characterize the available habitat
Background points

Environmental data availability
Increased options for environmental data products

Spatial and temporal aggregation
Increase the options for spatial and temporal aggregation

Exploratory Data Analysis and Variable Selection
Preliminary comparisons of environmental datasets to guide user through variable selection
## Products to be included in Phase II

<table>
<thead>
<tr>
<th>Type</th>
<th>Product</th>
<th>Platform / Sensor / Product</th>
<th>Archive¹</th>
<th>Resolution²</th>
<th>Domain³</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Temporal</td>
<td>Spatial</td>
<td>Temporal</td>
</tr>
<tr>
<td>Terrestrial</td>
<td>Surface Reflectance</td>
<td>Landsat 4-8</td>
<td>D</td>
<td>E</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MODIS (MCD43A2)</td>
<td>D</td>
<td>G</td>
<td>Daily</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ASTER</td>
<td>D</td>
<td>G</td>
<td>16-day</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sentinel 2</td>
<td>D</td>
<td>G</td>
<td>10-day</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Airbus SPOT (OneAtlas)</td>
<td>D</td>
<td>Once</td>
<td>1.5m</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Airbus Pleiades (OneAtlas)</td>
<td>D</td>
<td>Once</td>
<td>0.5m</td>
</tr>
<tr>
<td></td>
<td>Vegetation Indices (NDVI/EVI)</td>
<td>MODIS (MOD13)</td>
<td>G</td>
<td>8-day</td>
<td>250m</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Descartes MODIS 16-day VI</td>
<td>D</td>
<td>16-day</td>
<td>250</td>
</tr>
<tr>
<td></td>
<td>Land Cover Dynamics</td>
<td>MODIS (MCD12Q1)</td>
<td>D</td>
<td>G</td>
<td>Annual</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MODIS (MCD12Q2)</td>
<td>D</td>
<td>G</td>
<td>Annual</td>
</tr>
<tr>
<td></td>
<td>3D Structure</td>
<td>Sentinel-1 SAR</td>
<td>D</td>
<td>G</td>
<td>12 days</td>
</tr>
<tr>
<td></td>
<td>Elevation</td>
<td>Airbus</td>
<td>E</td>
<td>Once</td>
<td>~24m</td>
</tr>
<tr>
<td>Freshwater</td>
<td>Surface Water</td>
<td>JRC Global Surface Water</td>
<td>G</td>
<td>Monthly</td>
<td>30m</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MOD44W Surface Water</td>
<td>G</td>
<td>Annual</td>
<td>250m</td>
</tr>
<tr>
<td>Marine</td>
<td>World Ocean Atlas</td>
<td>Temperature, Salinity, Oxygen; others</td>
<td>E</td>
<td>Decadal</td>
<td>25km</td>
</tr>
<tr>
<td></td>
<td>NOAA OI SST</td>
<td>Sea Surface Temperatures</td>
<td>E</td>
<td>Daily</td>
<td>25km</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sea Surface Temperatures, Topography, Colour</td>
<td>D</td>
<td>G</td>
<td>1-2 days 300m - 1km</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MODIS Aqua</td>
<td>G</td>
<td>Daily</td>
<td>1km</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MODIS Aqua</td>
<td>G</td>
<td>Daily</td>
<td>1km</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HYCOM + NCODA</td>
<td>E</td>
<td>~10km</td>
<td>2008</td>
</tr>
<tr>
<td>Climate</td>
<td>Various Climate Variables</td>
<td>TerraClimate</td>
<td>G</td>
<td>Monthly</td>
<td>1km</td>
</tr>
<tr>
<td></td>
<td></td>
<td>³CHELSA / EarthEnv</td>
<td>G</td>
<td>Daily</td>
<td>1km</td>
</tr>
<tr>
<td></td>
<td></td>
<td>³CHELSA / EarthEnv</td>
<td>G</td>
<td>Daily</td>
<td>1km</td>
</tr>
<tr>
<td></td>
<td>Growing Degree Days</td>
<td>CHELSA Derived</td>
<td>G</td>
<td>Monthly</td>
<td>1km</td>
</tr>
<tr>
<td></td>
<td>Land Surface Temperature</td>
<td>MODIS (M*D11A1/A2)</td>
<td>D</td>
<td>G</td>
<td>8-day</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ASTER</td>
<td>D</td>
<td>G</td>
<td>16-day</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EarthEnv</td>
<td>G</td>
<td>Monthly</td>
<td>1km</td>
</tr>
</tbody>
</table>
We want your feedback! Please fill out the survey!

Please provide feedback on the current prototype and offer suggestions about what else this tool should do.

Direct survey link here: https://forms.gle/pxu69giSBmUvwYDMA
Thank you for your attention!
- MOL has a well developed access control system
- One of the biggest challenges is managing taxonomy/species names
- Taxonomy management - nearly a million names and synonymies
- Prototype tools to help harmonize names
Satellite and model-derived 1km environmental data at varying temporal resolutions

<table>
<thead>
<tr>
<th>Source</th>
<th>Daily</th>
<th>Monthly</th>
<th>Annual</th>
<th>Climatology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enhanced Vegetation Index</td>
<td>MODIS</td>
<td>✔️</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Land Surface Temperature</td>
<td>MODIS</td>
<td>✔️</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Air Temperature (Daily Min, Max, Mean)</td>
<td>CHELSA</td>
<td></td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>Precipitation (Daily Min, Max, Mean)</td>
<td>CHELSA</td>
<td></td>
<td></td>
<td>✔️</td>
</tr>
</tbody>
</table>

Available in the current prototype