March 2010

Patterns in Bird Migration Phenology Explored through Data Intensive Analysis and Visualization

Steve Kelling[1], Daniel Fink[1], Suresh Santhana Vannan[2], Kevin Webb[1], Robert Cook[2], William Michener[3]

Jeff Morisette[4], Claudio Silva[5], Tom Dietterich[6], Patrick O'Leary[7], Alyssa Rosemartin[8], and Damian Gessler[9]

[1]Lab of Ornithology, Cornell University, [2]Oak Ridge National Laboratory, [3]University of New Mexico, [4]U.S. Geological Survey, [5]University of Utah,
[6]Oregon State University, [7]Idaho National Laboratory, [8] USA National Phenology Network, [9]I-plant, University of Arizona

Introduction: Goals of DataONE

DataONE (Observation Network for Earth) is a five-year project (W.M. Michener, PI, University of New Mexico) whose goal is to ensure the preservation and access to multi-scale, multi-discipline, and multi-national data and to provide an open, persistent, robust and secure access to well-described and easily discovered data.

To further those goals, DataONE has established working groups that serve as catalysts to achieve specific goals for DataONE. For example, the Scientific **Exploration, Visualization, and Analysis (EVA)** working group was established to develop data intensive exemplars that serve as the foundation DataONE analysis and visualization functionality.

EVA Working Group: Objectives and Approach

Data synthesis
- Integrate highly diverse data
- Share data across a broad group of researchers

Extract knowledge
- Exploratory analysis to detect empirical patterns
- Confirmatory analysis to quantify the support for hypotheses

Develop Tools for exploration, visualization, and analysis
- Powerful, flexible, and efficient exploratory (model creation) tools
- Analyses and visualizations dynamically incorporated into reports

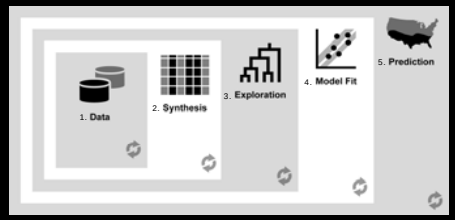
In formulating this EVA working group, DataONE recognizes the need for community involvement in the creation of data intensive procedures for the domain sciences. This working group consists of domain scientists (Ecology), computer scientists, visualization and workflow experts, and application developers.

The strong desire and willingness of the Working Group members to contribute to EVA will ensure the development of data intensive science processes that are well suited to the needs of scientific research and will be applied to the core cyber-infrastructure of DataONE.

EVA: First Exemplar

The first exemplar will be to understand how bird migration patterns change through time and examine the various environmental drivers associated with the timing, direction, and speed of migration. The relationships will be captured in a species distribution model and used to forecast distributional changes and inform conservation and land use policy.

Workflow



An iterative data-intensive workflow, using VisTrails: (1) Data are gathered from various sources and (2) synthesized; (3) multiple exploratory analyses characterize the data and suggest approaches to modeling; (4) model fitting leads to (5) predictions and hypothesis-testing. At any step, it may be necessary to back-track, taking some other route forward.

VisTrails [http://www.vistrails.org]

Data Synthesis

Obtain bird occurrence, environmental, climate, and land cover data and place into a common format for easy analysis

Bird Occurrence
- eBird (time, location, presence/absence)

Land Cover / Land Use
- National Land Cover Data (NLCD2001)
- Landscape fragmentation metrics
- Anthropogenic features

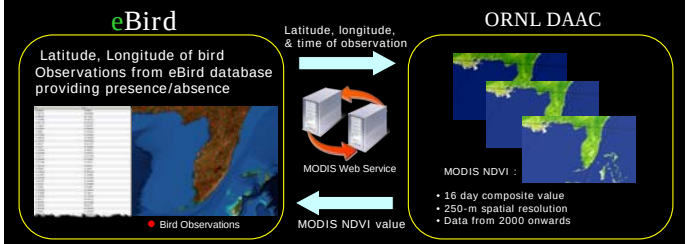
Phenology
- MODIS NDVI

Elevation

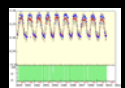
Climatology
- Prism

MODIS Data Processing

Extract MODIS NDVI for each bird observation location using ORNL DAAC's MODIS Subsetting Web Service.



- Observations from ~350,000 locations used in the extraction
- More than 15 million NDVI/EVI values extracted
- More than 30 million calls to the MODIS Web Service



MODIS NDVI bands extracted:
- 250m_16_days_EVI
- 250m_16_days_NDVI
- 250m_16_days_VI_Quality
- 250m_16_days_pixel_reliability

NDVI/EVI from ~65 composite dates extracted, primarily from 2008

eBird Web Page: http://ebird.org/content/ebird/

ORNL DAAC Web Service: http://daac.ornl.gov/modiswebservice

Spatio-Temporal Exploratory Model (STEM)

STEM was developed to analyze spatiotemporal distributions at broad-scale with fine resolution. A multi-scale strategy is used to differentiate between local and global-scale spatiotemporal structure. A user specified species distribution model, $f_i(X, s, t)$, accounts for spatial and temporal patterning at the local level. These local patterns are then allowed to "scale up" via ensemble averaging to larger scales. This makes STEM especially well suited for exploring distributional dynamics arising from a variety of processes.

$$F(X,s,t) = \frac{1}{n(s,t)} \sum_{i=1}^{n} f_i(X,s,t) I(s,t \in \theta_i)$$

Quantitative Intuition

Statistical Experimental Design:
Block over restricted ST extents to control ST variance

Bagging: average out local-level variation in predictions

Ecological Intuition

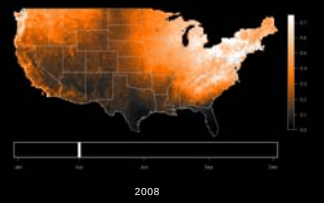
Local predictor-response learning. No "long-range" learning.

"Local" averaging allows large-scale patterns to emerge from the local-scale

Results: Estimated Occurrence Surfaces

Song Sparrow (*Melospiza melodia*)

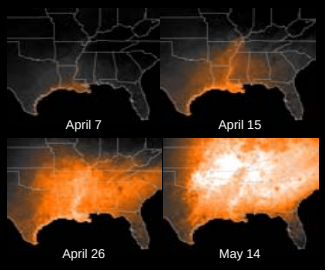
Predicted probability of occurrence, controlling for observation effort (April 1, 2008)



Results: Spring Migration Detail

Indigo Bunting (*Passerina cyanea*)

- Model predictions show population timing and location through migration
- Early April – Concentration on Gulf of Mexico Coast
- Mid April – Concentration along Mississippi River valley
- Mid May – Breeding distribution

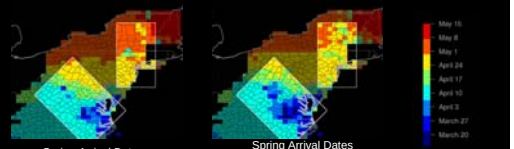


Results: Environmental Cues of Migration, NDVI

Questions
1. Is Red-eyed Vireo migration timing associated with NDVI?
2. If so, how is migration timing, direction, and speed affected by NDVI?

Experiment
1. Fit STEM with NDVI predictor
2. Advance "Greening Date" by 14 days
3. Plot "Spring Arrival Dates" as the first date when predicted probability of occurrence > 5%

Red-eyed Vireo (*Vireo olivaceus*)



Products and Next Steps

1. Reference data set of bird occurrence data and environmental characteristics, including NDVI
2. VisTrails Workflow
3. Bird Migration Atlas of North America (Interactive Web Tool)
4. Generic Tool Set for data processing, visualization, and analysis
5. Manuscripts on Data Intensive Science Process for Invasive Species Modeling

Potential Next EVA topics: Plant invasive species and allometry

Acknowledgements

DataONE is funded by the NSF Office of Cyberinfrastructure.

While DataONE is providing the vehicle for the EVA working group, the majority of support for this work is provided through the members of the working group.

These include: NSF CISE Pathways Computational Sustainability 0832782, Leon Levy Foundation, and NASA

Learn more about the DataONE at https://dataone.org