



JOHNS HOPKINS
BLOOMBERG
SCHOOL of PUBLIC HEALTH

Department of Biostatistics

BIOSTATISTICS SEMINAR

eBird: Statistical Models for Online Crowdsourced Bird Data

Daniel Fink, df36@cornell.edu, Cornell Lab of Ornithology, USA

More than ever, research to identify the environmental drivers that shape species' distributions is needed to manage and conserve earth's natural systems. However, obtaining this knowledge is challenging because: 1) species' distributions vary dramatically through time and space, 2) detailed species observation data are difficult to collect and organize across sufficiently large scales, and 3) conventional analytical methods have not been effective for facilitating spatiotemporal pattern discovery with sparse, noisy data and highly variable ecological signals. The goal of our research program has been to advance data intensive ecology to meet these challenges and improve our understanding of the broad-scale dynamics of bird migrations.

We have assembled bird observation data collected by the citizen science project, eBird (<http://www.ebird.org>), with local-scale environmental covariates such as climate, habitat, and vegetation phenology. This data set contains over 600,000 observations made between 2004-2009 at over 107,000 unique locations within the continental U.S. eBird is unique among broad-scale bird monitoring projects because it collects data year-round. The modeling challenge to study continent-wide inter-annual migrations of North American birds is to develop a method that can adapt to and explore the wide variety of species' intra-annual spatio-temporal dynamics. To do this we developed a simple mixture model for non-stationary spatiotemporal processes, SpatioTemporal Exploratory Models (STEMs). At the local scale a base model (decision trees) associates predictors and response. The mixture of base models is defined across an ensemble of uniformly distributed support sets that cover the spatiotemporal extent of analysis. This uniform ensemble can adapt to a wide variety non-stationary spatial and temporal processes so the user does not need to specify the dynamics. This approach can be computationally expensive depending on the number and efficiency of the base models.

Using a startup allocation from the TeraGrid we calculated weekly species distribution estimates for over 100 bird species, including several species of conservation concern, for the 2011 State of The Birds Report, an annual national conservation report (Department of the Interior). Visualizations from these models provide an empirical view of population-level movement. This is essential information to advance ecology and develop more comprehensive science-based management strategies. Ecologists are using these results to study how local-scale ecological processes vary across a species range, through time, and between species. Challenges lie ahead in scaling these analyses as well as the discovery and quantification of complex spatiotemporal patterns.

**The Johns Hopkins Bloomberg School of Public Health
Department of Biostatistics, Wednesday, November 2, 2011
Room W2030 School of Public Health, 4:00-5:00pm (Refreshments: 3:30)**

For disability access information or listening devices, please contact the [Office of Support Services](http://www.jhsph.edu/SupportServices) at 410-955-1197 or on the Web at www.jhsph.edu/SupportServices. EO/AA