Syllabus: Spatial Bioinformatics

Instructors:
Mary Blair (AMNH): mblair1@amnh.org
Rob Harbert (AMNH): rharbert@amnh.org
Peter Galante: pgalante@amnh.org

Credits: **1**

Dates: **May 14-23, 2018** Daily except Saturday and Sunday (8 class days)

Overview

Spatial data and models are ubiquitous in modern comparative biology and ecology due to the vast amount of available data and ever developing modeling methods. This course will focus lectures on a series of "best-practices" in handling and modeling spatial biological data including data-mobilization, bias detection and reduction, geographic projection management, and comparative modeling frameworks. Labs will concentrate on demonstration of best-practices on a range of datasets including student's personal data. The course will culminate in student's working in "hackathon" style working groups to develop a spatial data analysis pipeline to address a question of mutual interest to be posted on appropriate code-sharing repositories.

Learning Objectives:

- To introduce students to concepts and best practices related to spatial representation and analysis of biological data.
- To present the different methods and data analyses associated with spatial representation and analyses of biological data.
- To present examples from comparative biology studies and of questions that can be answered by applying these methods.
- To provide students with tools to apply the methods for their own studies.

Bibliography:

- Guisan, A., Thuiller, W. and Zimmermann, N.E., 2017. Habitat Suitability and Distribution Models: with Applications in R. Cambridge University Press.

- Wegmann, M., Leutner, B. and Dech, S. eds., 2016. *Remote sensing and GIS for ecologists: using open source software*. Pelagic Publishing Ltd.

Software:

Students should come to class with a laptop installed with R, RStudio, and QGIS. Any questions on how best to do this should be directed to the instructors. A small amount of time on the first day will be devoted to making sure that necessary software is working.

Daily schedule:

Lecture/Workshop: 10:00-12:00 Working Lab: 1:30-4:30

Day 1 - Lecture: Introduction to programming

Lab: Unix command line, basic data handling in R, Git

Reading: rOpenSci — https://openresearchsoftware.metajnl.com/article/10.5334/jors.bu/

HW: Watch - https://t.co/lh0sI4jo7B **Identify at least ONE open data/code resource discussed and create a single Powerpoint slide that summarizes the key features.

Day 2 - Lecture: Geography, coordinate systems, projections, georeferencing, GIS data types (Vector, Raster), sampling bias, spatial autocorrelation, and remote sensing

Lab: What's in a GPS coordinate? GPS vs. cell-phone locality information

Reading:

Graham et al., 2004. https://www.sciencedirect.com/science/article/pii/S0169534704002034

Wegmann et al. 2016. Chapter 2 (pgs. 22-39)

HW: Create maps of your coordinates using ggmaps. Try showing your points as individual points and convert to shapefile(s). Plot on different base-maps.

Day 3 - Lecture: Occurrence data and distribution modeling

Lab: Comparative niche modeling. Covering - background/absence sampling, environmental layers, model evaluation and parameterization, niche differentiation and overlap.

Reading: Guisan et al. 2017. Habitat Suitability and Distribution Models in R. Part 1: Sections 3, 4, and 5 (pgs. 21 - 57).

ENMeval paper — https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.12261

Blair et al., 2013 — https://onlinelibrary.wiley.com/doi/full/10.1111/jeb.12179

HW: **Create a distribution model for a species of interest to you using ENMeval.** Make a few slides describing what you did, what model was selected, and how your best model performed.

Day 4 - Lecture: Survey of advanced modeling topics (TBD)

Lab: Advanced modeling exercises. (possibly including: past/future range estimation, biogeographic reconstruction, landscape genetics, landscape analysis, modeling biotic interactions, mechanistic models – may be tailored to class interests)

Readings: TBD; Possible: Galante et al., 2018 - https://onlinelibrary.wiley.com/doi/full/10.1111/ecog.02909, or Harbert et al., 2014 - https://onlinelibrary.wiley.com/doi/abs/10.3732/ajb.1300417

HW: TBD;

Day 5 - Lecture: Literature Discussion of spatial data in Comparative Biology - Survey of 2-4 papers.

Lab: Advanced topics in R programming and an introduction to Git for code sharing and management. Will cover scripting, functions, and documentation. A primer in preparation for workshops and hackathon days.

Reading: The Introduction to "R Packages" by Hadley Wickham - http://r-pkgs.had.co.nz/intro.html

HW: Create an R package skeleton and push to GitHub.

Day 7-8 - BYOD (Bring Your Own Data) Workshops 1-4 pm

Students will bring or find data relevant to their own research (from i.e., Landsat, GBIF, iNaturalist) and will work through practical issues with their data analysis plans. These sessions will also serve as a brainstorming and guidance period for Hackathon project development.

Day 9 - Project/Hackathon day 9AM - 5PM

Students work in small groups a project that will develop a data analysis pipeline to address a question of interest to them. The goal of a hackathon is to create usable software with a specific focus by collaboration between interested parties. Students will develop an R script, an example dataset, and a demonstration. Students will supplement this usable code with a write-up of the supporting literature and target audience for the code. Final code will be shared on the course GitHub to facilitate reuse and further development

Grading/Assignments

Grading: RGGS Comparative Biology students will be graded pass/fall; students enrolled in our partner programs will be assigned either a letter grade or pass/fall as required by their registrar.

Evaluations Basis: Students will be evaluated on attendance (20%), participation in class and on class exercises (40%) and on performance on course projects/writing (40%).

Assignments: All grades in this course are pass/fail only and successful completion of this course will rely heavily on active participation in ALL class activities. A brief homework assignment will be assigned with each Lab on days 1-5. Days 6-7 will be spent workshopping issues with datasets that you are interested in, your participation these days will include bringing data or code in sessions with a targeted workflow and application in mind. Day 8 will be spent developing term projects that will consist of a reproducible and scalable R workflow for the analysis of some kind of spatial data, ideally organized as a draft R library and published on GitHub.

Statement on Academic Integrity: Each graduate student bears the responsibility to observe traditional canons of scholarly discourse, scientific research, and academic honesty. Plagiarism, cheating, and fraud in research will not be tolerated. Accordingly it is expected that students work individually unless specifically instructed to work in groups. The full Academic Integrity policy is in the student handbook.

Course Evaluations: Each student is required to complete an anonymous course evaluation at the end of the term. The course evaluation is a tool for faculty and administrators to improve the student learning experience.