EE 508/DS 537- Data Science for Conservation Decisions

Fall 2022

Lecture & Lab: Tue + Thu, 2-3:15pm — PSY B37 (64-86 Cummington Mall)

Instructor: Christoph Nolte — CAS 445 (685 Commonwealth Ave)
chrnolte@bu.edu — placeslab.org

Office Hours: Tue + Thu, 5-6 pm + TBD — Sign-up & Zoom link: bit.ly/3iEsCLW

Application of quantitative methods to support conservation decisions. Ecosystem value mapping, systematic conservation planning, policy instrument design, rigorous impact evaluation, decision theory, data visualization. Implementation in state-of-the-art open-source software. Real-life case studies from the U.S. and abroad.

Course Description

This course offers an introductory treatment of quantitative methods and tools used by conservation scientists, policy makers, and land managers to inform decisions on the protection and use of ecosystems.

Growing demand for food, fiber, housing, and energy affects global ecosystems in ways that reduce their ability to generate environmental benefits. Conservation interventions can protect and restore ecosystems and the benefits they provide. However, they often face budget constraints, limited political will, and low willingness of landowners to forgo productive uses.

Conservation decisions thus need to be "smart" to maximize environmental benefits under those constraints. Decision makers regularly face questions such as:

- How can we quantify changes to the spatial distribution of environmental values across a given landscape?
- Given limited budgets, where should conservation occur to maximize benefits (species representation, carbon sequestration, water protection)?
- How can policy instruments, such as ecosystem markets and payments, be configured in ways that deliver the greatest environmental benefits?
- What difference did a given intervention make?

Conservation scientists have developed many methods and computational tools to answer these questions. Some are widely applied to support decisions in governments, non-profits, and academia. We will learn a key subset of these methods and apply them to real-life cases.

The course will be taught using open-source, cross-platform programming languages and software packages, including QGIS, Python, R, and Marxan.

Course Objectives

After taking this course, you will be able to:

- **Identify** questions that are of interest to conservation decision makers and that can be solved using computational methods.
- **Design** analyses to answer these questions, including characterization of the problem, data acquisition, processing, analysis, and presentation of results.
- **Implement** such analyses using state-of-the-art, open-source, and platform-independent (OSX/Windows/Linux) software tools.
- **Evaluate** the strengths and shortcomings of different methods, including a critical assessment of underlying data and theoretical frameworks.
- Efficiently and confidently **access** online information, including public datasets, software documentation and knowledge exchange websites, to train and enhance the capacities in self-directed learning and troubleshooting that are essential for data science careers.

BU Hub Learning Outcomes

- **Quantitative Reasoning II**: you will be trained in using analytical, statistical and computational methods to support decisions in conservation policy, including the formulation and testing of hypotheses, the presentation of results, and the interpretation of findings considering methodological strengths & limitations.
- **Digital/Multimedia**: you will learn to communicate the results of your analyses to a diverse range of decision makers, including through digital visualizations, interactive map products, visual abstracts, reports and presentations.
- **Research and Information Literacy**: you will learn to confidently access open-source datasets, software packages, and knowledge exchange websites in order to conduct both prescribed and self-directed research to inform conservation.

Prerequisites

- Introduction to programming. You need to know basic variable types (e.g., integer, float, string) and control structures (e.g., if, for, while). This content is covered in EE375 and similar classes.
- Introduction to statistics. You need to have been exposed to the basics of probability theory and simple multivariate linear regressions, e.g., be able to explain what a p-value is. EE270, EE516, and related classes fulfill this requirement.
- Introduction to spatial data. You need a basic understanding of the structure of vector and raster data, and the importance of projections. EE465, EE505, and similar classes cover this content.
- Recommended: EE420, or another introduction to optimization.

Instructional Methods and Assignments

This is a "hands on" course designed to help you become familiar with key methods and open-source software tools used to support conservation decisions. Much of your grade will be based on your ability to understand the underlying theories, to implement the methods correctly, and to present results in a form understandable to non-experts. In your final project, you bring those skills together to support a real-life conservation decision of your choice.

Labs (65%)
Four hands-on labs will develop and test your capacity to tackle computational questions in the following areas:

1. Open-source spatial data processing and visualization (QGIS & Python)
2. Systematic conservation planning (Marxan)
3. Conservation cost prediction & policy simulations (Python)
4. Rigorous impact evaluation (R, Matching)

Final Project (25%)
Identify and help solve a real-life conservation problem using one or more of the methods covered in class. This project will be developed over several stages:

1. Prospectus (5%): identify an environmental conservation issue that interests you, then define a decision problem that a donor, government, or nonprofit who also cares about your problem might actually try to solve in real life. This involves defining desirable outcomes in a way that makes them measurable (objective function), decision variables that a decision maker has to choose between (possible actions), as well as a description of the social-ecological system of interest.
2. Project plan (5%): formulate your research question and your research plan in detail. The research plan needs to include all datasets you need, mention the ones you already have access to, and identify a deadline for when you will stop looking for more data.
3. Final paper (15%): write up your problem, methods, and findings in a short report and share it with the class.

Class Participation (10%)
Extra credit for answering your classmates' questions in online discussions.

Course Materials
There is no textbook required for this class. All readings are available online or will be posted in the class folders.

Course Policies
Attendance & Participation: Class participation is beneficial to group learning. Attendance is expected. More than two unexcused absences will affect the participation component of your grade. Please come to class familiar with the day's readings, and ready to engage by asking and answering technical questions, and critically discussing the readings' contents.

Zoom: A Zoom session will be live and recording during class time. This allows participants to drop in remotely if quarantining, sick, or absent for other reasons. I will make these recordings internally available on Blackboard for students that weren't able to attend in person.
https://bostonu.zoom.us/j/834486460?pwd=aGNQbzd0MWFzdGc5dHF5Yi9PazhGQT09

Assignment Completion: Assignments are submitted through the class folders. You can achieve up to 100% (or more) in the first submission. After that, there are second chances: if you figure out how to correct your results, you get back up to two thirds of the missed points.

Late Policy: Assignments need to be submitted by 11:59pm. You have a time bank of 48 hours that you may withdraw from and apply to the submission of any of your assignments (problem sets, project plan, report). This eliminates the need to request extensions and allows you some flexibility in managing your workflow. After you empty your time bank, graded assignments will be penalized by one-third of a letter grade for each day of lateness. If you anticipate difficulties due to documentable extenuating circumstances, please notify me as soon as possible.

Religious Observances: Campus policy regarding religious observances requires that faculty make every effort to reasonably and fairly deal with all students who, because of religious obligations, have conflicts with scheduled exams, assignments or required attendance. Please notify me as soon as possible so that the proper arrangements can be made. For details, consult http://www.bu.edu/chapel/religion/ and http://www.interfaithcalendar.org/

Academic Conduct: All Boston University students are expected to maintain high standards of academic honesty and integrity. It is your responsibility to be familiar with the Academic Conduct Code, which describes the ethical standards to which BU students are expected to adhere and students' rights and responsibilities as members of BU's learning community. All instances of cheating, plagiarism, and other forms of academic misconduct will be addressed in accordance with this policy. Penalties for academic misconduct can range from failing an assignment or course to suspension or expulsion from the university http://www.bu.edu/academics/policies/academic-conduct-code/

Diversity & Inclusion: Diversity enriches all research and education, and is realized only with all voices, views, and perspectives operating within a supportive and respectful community. For this reason, the Department of Earth & Environment, including myself and the students in this course, are committed to fostering diverse, inclusive, and equitable living, learning, and working environments that are supportive and free from violence, harassment, disruption, and intimidation. Further, the Department of Earth & Environment recognizes that creating a safe environment and a culture of respect is the shared responsibility of all members of our community. To ensure an equitable environment that values and respects the unique experiences and perspectives of our community, the Department, including myself and the students in this course, are dedicated to promoting diversity, inclusion, and equity among all members of our departmental community and encouraging open, honest, and compassionate communication. See also: http://www.bu.edu/earth/about/diversityinclusion/

Schedule

Day	Topic	Readings
Tue 9/6	Introduction	
Thu 9/8	Data Science for Conservation Decisions Data visualization & processing in QGIS	Jenkins et al. 2015a, Brown et al. 2015, Jenkins et al. 2015b
Tue 9/13	Programming in `Python`	A Byte of Python (Basics - Functions, Data Structures) PEP 8 Python Style Guide
Thu 9/15	Data analysis with `pandas`	**10 Minutes to pandas** Intro to Data Structures
Tue 9/20	Map making and data processing with `geopandas`	Geopandas User Guide
Thu 9/22	Raster processing with `rasterio` & GDAL	rasterio Quickstart
Tue 9/27	Valuing ecosystem services: key concepts	Goulder & Kennedy 2011
Thu 9/29	Valuing ecosystem services: examples & tools ✔ *Lab 1 due*	
Tue 10/4	Systematic conservation planning: theory, methods, and tools	InVEST User Guide, Game et al. 2013, Watson et al. 2011
Thu 10/6	Systematic conservation planning with Marxan: Jennifer McGowan, Spatial Planning TC, Global Science, The Nature Conservancy [remote]	Ball et al. 2009, Marxan User Manual, Jen's pubs: https://bit.ly/3q5KYLs
Tue 10/11	*No class (Monday schedule after Indigenous Peoples' Day)*	
Thu 10/13	Systematic conservation planning with Marxan: building your own dataset ✔ *Class project: prospectus due*	
Tue 10/18	Accessing and analyzing public datasets [remote]	
Thu 10/20	Conservation policy instruments: an overview of common strategies [remote] ✔ *Lab 2 due (systematic conservation planning)*	
Tue 10/25	Explaining conservation costs with `statsmodels`	
Thu 10/27	Predicting conservation cost with `scikit-learn`: cross-validation	
Tue 11/1	Predicting conservation cost with `scikit-learn`: ensemble methods ✔ *Class project: abstract due*	Background on algorithms (optional): Geurts et al. 2006, Natekin & Knoll 2013
Thu 11/3	Simulating the cost-effectiveness of policy strategies	Newburn et al. 2006
Tue 11/8	Class project workshop	

Thu 11/10	Rigorous impact evaluation: framework and methods ✔ *Lab 3 due (cost prediction, policy simulation)*	Ferraro 2009
Tue 11/15	Rigorous impact evaluation: matching	Ferraro & Hanauer 2014
Thu 11/17	Rigorous impact evaluation: matching exercise I	Intro to R, Ho et al. 2007
Tue 11/22	Rigorous impact evaluation: matching exercise II	
Thu 11/24	*No class (Thanksgiving)*	
Tue 11/29	Rigorous impact evaluation: matching exercise III	
Thu 12/1	Class project workshop / open topic ✔ *Lab 4 due (rigorous impact evaluation)*	
Tue 12/6	Communicating results: interactive mapping	
Thu 12/8	Communicating results: 3D visualizations	
Mon 12/19	✔ *Class project: writeup due*	

remote https://bostonu.zoom.us/j/834486460?pwd=aGNQbzd0MWFzdGc5dHF5Yi9PazhGQT09
Christoph has to attend principal investigator workshops from two NASA programs that fund projects at the PLACES Lab (Land Use/Land Cover Change and Water Resources Applications)

Readings

Ball, I.R., Possingham, H.P. & Watts, M.E. (2009). Marxan and Relatives: Software for Spatial Conservation Prioritization. *Spat. Conserv. prioritization Quant. methods Comput. tools.*

Brown, C.J., Bode, M., Venter, O., Barnes, M.D., McGowan, J., Runge, C.A., Watson, J.E.M. & Possingham, H.P. (2015). Effective conservation requires clear objectives and prioritizing actions, not places or species. *Proc. Natl. Acad. Sci.*, 112, E4342–E4342.

DEFRA. (2008). *An introductory guide to valuing ecosystem services*.

Ferraro, P.J. (2009). Counterfactual thinking and impact evaluation in environmental policy. *New Dir. Eval.*, 2009, 75–84.

Ferraro, P.J. & Hanauer, M.M. (2014). Advances in Measuring the Environmental and Social Impacts of Environmental Programs. *Annu. Rev. Environ. Resour.*, 39, 495–517.

Game, E.T., Kareiva, P. & Possingham, H.P. (2013). Six Common Mistakes in Conservation Priority Setting. *Conserv. Biol.*, 27, 480–485.

Geurts, P., Ernst, D. & Wehenkel, L. (2006). Extremely randomized trees. *Mach. Learn.*, 63, 3–42.

Goulder, L.H. & Kennedy, D. (2011). Interpreting and estimating the value of ecosystem services. In: *Nat. Cap.* Oxford University Press, pp. 15–33.

Ho, D.E., Imai, K., King, G. & Stuart, E. a. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Polit. Anal.*, 15, 199–236.

Jenkins, C.N., Van Houtan, K.S., Pimm, S.L. & Sexton, J.O. (2015a). US protected lands mismatch biodiversity priorities. *Proc. Natl. Acad. Sci. U. S. A.*, 0–5.

Jenkins, C.N., Van Houtan, K.S., Pimm, S.L. & Sexton, J.O. (2015b). Reply to Brown et al.: Species and places are the priorities for conservation, not economic efficiency. *Proc. Natl. Acad. Sci.*, 112, E4343–E4343.

Natekin, A. & Knoll, A. (2013). Gradient boosting machines, a tutorial. *Front. Neurorobot.*, 7.

Newburn, D.A., Berck, P. & Merenlender, A.M. (2006). Habitat and open space at risk of land-use conversion: Targeting strategies for land conservation. *Am. J. Agric. Econ.*, 88, 28–42.
Watson, J.E.M., Grantham, H.S., Wilson, K.A. & Possingham, H.P. (2011). Systematic Conservation Planning: Past, Present and Future. In: *Conserv. Biogeogr.* John Wiley & Sons, Ltd, Chichester, UK, pp. 136–160.

Background Resources

Introductory-level knowledge of programming, statistics, and GIS is a prerequisite for this course. If you want to refresh what you have learned and fill any gaps, work your way through the resources provided below. Because they use the same open-source languages and packages we'll cover in class, they will even give you a head start.

Coding Basics (in Python)

- Introduction to Python (interactive tutorial, recommended)
- A Byte of Python (text)

Regression Basics

- Regression Analysis
- Multiple Regression Analysis

GIS Basics (in QGIS)

- A Gentle Introduction to GIS