EE 508 / DS 537- Data Science for Conservation Decisions

Fall 2023

Lecture & lab: Tue + Thu, 2 – 3:15pm — CGS 115 (871 Commonwealth Ave)

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

Office hours: Mon, 11am – 12pm (Zoom) — Sign-up & Zoom link: bit.ly/3Qywd2o
Tue + Thu, 4 – 5pm (CAS 445)

> 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 introduction to quantitative methods and tools used by conservation scientists, policy makers, and land managers to inform decisions on land protection and use.

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, but usually face budget constraints and 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 our interventions make?

Conservation scientists have developed a range of methods and computational tools to answer these questions. Some are widely applied to support decisions in governments, non-profits, and academia. Here, you will use a few important ones and apply them to real-life cases.

All labs can be implemented in open-source, cross-platform programming languages and software packages. We will mostly use Python, but also QGIS, 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 in state-of-the-art open-source software tools that work on all major operating systems (Mac, Windows, and Linux).
- **Evaluate** the strengths and shortcomings of different methods to find solutions, including a critical assessment of underlying data and theoretical frameworks.
- Efficiently and confidently **access** online information – public datasets, software documentation, and knowledge exchange websites –, to improve your troubleshooting skills and to speed up your future learning as a spatial data scientist.

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

This course has two main prerequisites:

- **Introduction to programming**. To enjoy this class, you need the following:
    - a clear grasp of basic data types (e.g., integer, float, string, list, set, dictionary),
    - applied knowledge of control structures (`if` conditions, `for` loops, functions),
    - experience with developing small coding projects (~1000 lines), and
    - experience with troubleshooting (e.g., how to search errors for clues).

  There are many ways to obtain these skills, and any programming language is okay. **EE 375** and most introductory coding classes have you covered, as will most boot camps and self-teaching resources. If you are uncertain, just send me a quick email or come talk to me after class. There will be a series of smaller, graded coding tasks in the first weeks of the course to help you gauge whether the speed of the class works for you.
- **Introduction to statistics**. You need to have completed sufficient prior coursework in statistics to explain the concept of a p-value and to interpret coefficient estimates in simple multivariate linear regression models (as in: $y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$). **EE 270, EE 516**, and many other classes at BU fulfill this requirement.

You will find this class much easier if you have prior knowledge in the following two areas. These are not required, however: students with no prior knowledge can pick it up quickly.

- **Spatial data**. Working with spatial data requires a basic understanding of vector and raster data and the ability to work with geographic projections. **EE 465**, **EE 505**, and other classes cover this. We will go over essentials at the beginning of the course.
- **Optimization** (algebraic or computational). Experience with defining objective functions, constraints, and decision variables will come in useful when we wrap our heads around complex yet solvable decision problems. **EE 545** is one of the classes that teaches this approach to think analytically. We will go over essentials at the beginning of the course.

Course experience and grading

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 is 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 will bring those skills together to support a spatial policy design 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, 15%)
2. Systematic conservation planning (Marxan, 20%)
3. Conservation cost prediction & policy simulations (Python, 20%)
4. Rigorous impact evaluation (R, Matching, 10%)

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 a policy 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:
    a. defining desirable outcomes so they can be measured (objective function)
    b. decision variables that one has to choose between (possible actions)
    c. a description (a simple model) of the social-ecological system of interest.
2. Project plan (5%): formulate your policy / decision question and your analytical plan. The project plan lists all datasets you need, mentions the ones you have access to, and identifies 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%)

You will get extra credit for answering your classmates' questions on Piazza.

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 generally expected. More than two unexcused absences will affect the participation component of your grade. Please come to class familiar with the day's readings, ready to engage by asking and answering technical questions, and critically discussing the readings' contents. Please review the labs before we discuss them in class and start implementing them as you read, until you get stuck somewhere.

Coding buddies
Coding together can be an effective way to build skills. If you prefer to work alone (as most previous students) you are welcome to do so. You can also opt into working on each lab with a "coding buddy": a (rotating) peer with whom you work on the labs together and share a grade for a single submission. You can organize your collaboration however works best for you, but the expectation is that you take turns as navigator and driver. For Lab 1, let me know if you already have someone you hope to work with, or if you would like to join our first test of a random allocation process that we come up with as a group. Please share your preferences with me by Sep 8 (chrnolte@bu.edu).

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

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 and 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: https://www.bu.edu/earth/diversity-and-inclusion/

Schedule

Day	Topic	Readings
Tue 9/5	Introduction	
Thu 9/7	Data Science for Conservation Decisions Data visualization & processing in QGIS	Jenkins et al. 2015a**,** Brown et al. 2015**,** Jenkins et al. 2015b
Tue 9/12	Programming in `Python` ✔ *Lab 1-1 due*	A Byte of Python (Basics - Functions, Data Structures) Code styles: black, PEP8
Thu 9/14	Data analysis with `pandas`	10 Minutes to pandas Intro to Data Structures
Tue 9/19	Map making and data processing with `geopandas` ✔ *Lab 1-2 due*	GeoPandas user guide
Thu 9/21	Raster processing with `rasterio` & GDAL ✔ *Lab 1-3 due*	rasterio: Python Quickstart
Tue 9/26	Valuing ecosystem services: key concepts ✔ *Lab 1-4 due*	Goulder & Kennedy 2011

Thu 9/28	Valuing ecosystem services: examples & tools ✔ *Lab 1-5 due*	
Tue 10/3	Systematic conservation planning: theory, methods, and tools ✔ *Please schedule 15min with me this week to explore ideas for your final research project.*	InVEST user guide, Game et al. 2013, Watson et al. 2011
Thu 10/5	Systematic conservation planning with Marxan	Ball et al. 2009, Marxan User Manual
Tue 10/10	*No class: Monday schedule after Indigenous Peoples' Day* *(✔ Last day to drop course without a "W" grade)*	
Thu 10/12	Systematic conservation planning with Marxan: building your own dataset ✔ *Class project: prospectus due*	
Tue 10/17	Accessing and analyzing public datasets	
Thu 10/19	Conservation policy instruments: an overview of common strategies ✔ *Lab 2 due (systematic conservation planning)*	
Tue 10/24	Explaining conservation costs with `statsmodels`	
Thu 10/26	Predicting conservation cost with `scikit-learn`: cross-validation	
Tue 10/31	Predicting conservation cost with `scikit-learn`: ensemble methods	Background on algorithms (optional): Geurts et al. 2006, Natekin & Knoll 2013
Thu 11/2	Simulating the cost-effectiveness of policy strategies	Newburn et al. 2006
Tue 11/7	Class project workshop	
Thu 11/9	Rigorous impact evaluation: framework and methods ✔ *Lab 3 due (cost prediction, policy simulation)*	Ferraro 2009
Tue 11/14	Rigorous impact evaluation: matching	Ferraro & Hanauer 2014
Thu 11/16	Rigorous impact evaluation: matching exercise I ✔ *Class project: abstract due*	Intro to R, Ho et al. 2007
Tue 11/21	Rigorous impact evaluation: matching exercise II	
Thu 11/23	*No class: Thanksgiving*	
Tue 11/28	Rigorous impact evaluation: matching exercise III	
Thu 11/30	Class project workshop / open topic ✔ *Lab 4 due (rigorous impact evaluation)*	
Tue 12/5	Communicating results: interactive mapping	
Thu 12/7	Communicating results: 3D visualizations	
Tue 12/12	Open topic	
Mon 12/15	✔ *Class project: writeup due*	

Readings

Ball, I.R., Possingham, H.P. & Watts, M.E. (2009). Marxan and Relatives: Software for Spatial Conservation Prioritization. *Spatial Conservation Prioritization: Quantitative Methods and Computational Tools. Oxford University Press.*

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)

- A Byte of Python
- Or enter the search term "introductory python tutorial" on your trusted search engine or streaming platform and you will find excellent tutorials for your level within seconds.

Regression basics

- [Regression Analysis]
- [Multiple Regression Analysis]

GIS basics (in QGIS)

- [A Gentle Introduction to GIS]