BIO 401-01 / BIO 598-02: GIS for Bio & Enviro Sciences

Instructor info

Dr. Giuseppe Amatulli (amatullig@wcsu.edu)

Giuseppe is a lead scientist in GeoComputation for spatio-temporal data analysis. He is an accredited GIS data expert with deep expertise in spatial modelling and coding with open source software for environmental applications. His current focus is on hydrological modelling at a global scale, in addition to researching species distribution under climate change scenarios.
Giuseppe has a breadth of experience in GIS technologies, remote sensing, informatics, cluster and parallel computing, and statistics. He is proficient in applying complex modelling techniques to automate the analysis of high-resolution data under Linux-based platforms.
Giuseppe holded a research scientist position at Yale University in the US, and has previously worked at the European Commission Joint Research Centre, and the University of Zaragoza. His coverage of the forestry and environmental sectors entails wildfire occurrence risk and pattern recognition, ecological shifts under climate geo-engineering scenarios. Notably, he runs geocomputation training courses worldwide on the latest data programming techniques.
Giuseppe has a PhD from the University of Basilicata in Italy, an MSc in Geo-Information Science from Wageningen University, and MSc in Forestry from Bari University. His time away from coding is spent leading adventure trips in remote locations, which involve canyoning, caving, rafting and hiking.

Invited speaker

Dr. Longzhu Shen (lqshen10@gmail.com)

Longzhu is a quantum chemist and mathematical modeller. He obtained his PhD in Chemistry at Carnegie Mellon University with focus on density functional theory. After that he extended his research to couple quantum mechanical principles and machine learning methodologies to address biogeochemical challenges at Yale university. He charted the distribution of chemical nutrients in freshwater systems in the USA and engineered the first two-way chemical design diagram with aim of reducing their lower toxicity potentials. Mostly recently, he worked on modelling viral evolution as a research scientist at Cambridge University. He is particularly interested in addressing cross-disciplinary questions including but not limited to geoinformatics, theoretical chemistry, computational toxicology and mathematical biology. He did so via investigating the underlying mathematical connections between seemingly distant fields. Dr. Shen likes to describe himself as a scientific programer. He is capable to coding with multiple languages to achieve the modelling goal. He has experience with all the main-stream linux distributions and BSD systems. During his spare time, Dr. Shen enjoys hiking, biking and travelling.

What is GIS anyway?

Geographic Information Systems (GIS) are powerful systems for organizing, analyzing and visualizing data that have spatial locations, and they are amazing! GIS and related technologies (like GPS) represent an enormous advancement in our ability to make sense of space, place, and geographic relationships. They have revolutionized many academic fields and professions, including those relating to ecology and environmental science. Understanding GIS is a valuable job skill. The U.S. Dept. of Labor has identified three areas that will dominate the U.S. economy and job growth in coming decades: biotech, nanotech, and geotech. The availability of small, cheap, precise GPS chips has made it possible for everything from cell phones to bulldozers to become spatially aware. GIS technology allows this awareness to be connected to other information sources via the Internet and other networks, opening tremendous possibilities and opportunities that are fundamentally changing the ways we live and work.

Course description

In this course, students will learn GIS concepts by using open source software in a Linux environment, opening in this way new horizons in the use of the outstanding power processing routines. We will guide the students who have never used a command line terminal to a stage where they are able to understand and use very advanced open source data processing routines. Our focus is to provide attendees the tools and competencies to continue developing their skills independently. This heuristic approach allows participants to continue progressing and improving in an ever-evolving technology environment. The acquired skills will be beneficial, not only for GIS related application, but also for general data processing and applied statistical computing in a number of fields. These essentially lay the foundation for career development as a data scientist.

Academic program:

The proposed course intends to provide students with the opportunity to develop crucial skills required for advanced spatial data processing. Throughout the full semester students will focus on developing fundamental and independent-learning skills in advanced data processing - a field that is continuously evolving with the availability of increasingly complex data and ongoing technological advances. A diverse set of complementary and sometimes overlapping tools will be presented for an overview of the universe of open source softwares available for spatial data processing. We demonstrate their strengths, weaknesses and key features for various data processing objectives (ex.: modelling, data filtering, queries, GIS analyses, graphics or reporting) and data types. Specifically, we guide students in using these tools and software and assist them along the steep curve of learning, command-line programming. We focus our training on helping students to develop independent learning skills and to find online help, solutions and strategies, in order to fix bugs, and independently progress with complex data processing problems.

The Academic Programme is divided into 3 main areas of study:

Lectures: (30 min to 1 hour each) Students take part in a series of lectures introducing the basic functioning of tools, theoretical aspects or background information needed for a better understanding of concepts that are subsequently applied in data processing.
Tutorials: Students are guided during hands-on sessions where trainers perform data analyses on real case study datasets, allowing the former to replicate the procedures on their own laptops.
Assignments: In addition to tutorials and lectures, assignments will be given in order to reinforce the newly learned data processing skills. Assignments are designed to equip students with the confidence and resources to become independent learners and to effectively address the demands of advanced spatial-data processing.

Assignments topics are addressed according to their programming knowledge-level, mathematical and statistical skill, GIS and Remote Sensing competencies. The exercises and examples provided are cross-disciplinary: forestry, landscape planning, predictive modelling and species distribution, mapping, nature conservation, computational social science and other spatially related fields of study. Furthermore, these case studies can be viewed as template procedures and easily adapted to be applied to different thematic challenges across disciplines.

Learning objectives

This course will enable students to further develop and enhance their spatio-temporal data processing skills. Most importantly, it will endow them with proficiency in a fully-functional open source operating system with all the requisite software tools. With continuous practice through the weeks, students will become familiar with command lines and cover numerous topics, including:

- Learning a large suite of existing tools and knowing which ones to employ for project-specific applications.
- Acquiring confidence in using several command line utilities for spatial data processing and with the Linux operating system.
- Developing data processing skills; and understanding data types, data modelling and data processing techniques.
- Independent learning, critical thinking and efficient data processing.

Upon completion of the course, you should be able to:

- Apply the process of science, as demonstrated by conducting, analyzing, and interpreting findings related to GIS project of each student's own design.
- Use quantitative reasoning by identifying appropriate statistical/spatial analysis approaches for independent GIS projects.
- Clearly convey your understanding of environmental phenomenons through written and oral communication, as demonstrated by giving a formal presentation of your GIS project in a forum attended by faculty and peers.

Required course materials

Note-taking and Organization: To be successful in this course, you will be required to take excellent notes and manage your written and electronic data in a highly organized way. Your curiosity in searchin new methods, commands and scripting procedures will be fundamental to score high notes.
PC and data Storage: You must have a laptop with **40 GB of free disk space and 8GB of RAM**. Although you may wish to use WCSU computers, we encourage you to use your personal laptop, in order to get familiar with your own OS and be ready for using it also in other disciplines
Text: There is no required text in particular for this course. We will provide you some readings from various GIS books as well as papers from the primary literature. The readings will be handled in the syllabus along the semester
Class materials: All the class material will be presented from www.spatial-ecology.net site and in particular under the Jekyll documentation site http://spatial-ecology.net/docs/build/html/index.html. Additional material for a deep immersion in coding for GIS and RS application can be found in http://spatial-ecology.net/dokuwiki/doku.php
Course requirements: Course participants should have basic computer skills and a strong desire to learn GIS using open source tools. We expect participants to have a specific interest in geographical data analyses, and prior knowledge in basic calculus and statistics.

Class attendance, participation and engagement.

You begin each class with 5 points for attendance and participation. They are yours to keep unless you arrive late, miss classes, show up unprepared, or fail to take a proactive role in class work. Attendance is crucial to this class and it is extremely important that you attend every class to stay current on class topics. You will find it very difficult to catch up in this class if you fall behind. Besides, active participation and engagement is more beneficial than passive attendance. Therefore, make questions, be proactive in searching for alternative command solutions and find out and test new commands. The success of a good written code is driven by the curiosity of testing new processing techniques and seeing the final results. Therefore, be curious and enforce your critical-thinking in an hands-on experience. You will get several “syntax errors”, but do not get frustrated, try and re-try and search for a solution on the internet. Same or similar errors are likely to have occurred already, so you just need to seek for the solutions.
Coding requires time, and when you are tired you do not see the errors and 2 min of debugging can become 2 hours, thus plan ahead! DO NOT WAIT until the night before a due date to begin your work for this class. If you must be absent for illness, emergency, military service, jury duty, or other legitimate reasons, you will be permitted to complete the in-class assignment (under a strict deadline), but it is impossible to make up missed quizzes. You must provide documentation for your absences no later than the next class meeting following your absence(s). Please let me know about upcoming absences as far in advance as possible so I can help you plan accordingly.

Participation in class is expected. It is also possible to earn negative participation points for behavior or activity that interferes with the conduct of class, is distracting to your peers, or otherwise interferes with their progress or success. This includes, but is not limited to:

- Email, IM, web browsing or other inappropriate computer use during class activities

- Tardiness or leaving class (except during assigned breaks)
- Inappropriate talking or other class disruptions
- Modifying, deleting, manipulating, or otherwise messing with anybody else's files
- Leaving your computer station in disarray
- Arriving unprepared for class

Blackboard

Blackboard is used extensively in this course to provide you access to course materials, assignments, grades and announcements. Please make sure you have access to Blackboard and that you check here regularly for important course information and updates. Weather-related updates about class (e.g., what to do if we have school cancellations or delays) will be posted to Blackboard Announcements. All written assignments must be submitted through Blackboard in the format requested (Word of PDF, etc.) in each assignment's instructions, unless otherwise instructed. In case of coding procedure the assignment must be submitted through git (see 9 Topic) For all assignments, late submissions will be automatically penalized at a rate of 10 points per calendar day. This will be strictly enforced, and the time-stamp on Blackboard will be used to determine the time of submissions. Emailed assignments will **not** be accepted.

Office hours

Students that need support or additional explanation can talk directly with trainers during the office hours. Office hours will holded remotely, and will be set for 2 hours a week: every Tuesday/Thursday at 4-5pm. The platform for office hours will be selected with the students during the first day.

Grades

Your final grade is composed of four elements, in class work, homework, research project and final scientific presentation. Attendance and participation in class is also crucial to your success. The percentage of your final grade determined by each type of assignment is detailed in the table below:

Grading factors	**Percentage contribution**
Homework assignments	20%
Class attendance	5%
Class participation and engagement	10%
Use of git for handling assignments	5%
Final project Objectives and results Scripting procedure Presentation	 15% 35% 10%

Total	100%

Homework assignment grading will assigned considering several factors

1) Scripting procedure
    a) Use of new commands 30%
    b) Complexity of script (use for loop, nested command and function) 30%
    c) Coding efficiency (no reinventing the wheel) 10%
    d) Clarity and reproducibility, readability, 20%
    e) Find out the correct answer 10%
2) Theory foundation
    a) Clear understanding of the key concepts 40%
    b) Ability of logic derivations and mathematical operations 35%
    c) Mastery of critical/analytical thinking 25%
3) Numerical challenges
    a) Understanding the basis for numerical analysis 40%
    b) Familiarity with numerical algorithms and their applications 30%
    c) Uncertainty and error analysis 30%

If you are struggling in class, please make time to see me as soon as trouble arises. Do not wait until the last weeks of the semester to seek help. There is no grading curve. Letter grades will be assigned according to the following scale:
A = 100 – 93; A- = 92.9 – 90; B+ = 89.9 – 87; B = 86.9 – 83; B- = 82.9 – 80; C+ = 79.9 – 77; C = 76.9 – 73; C- = 72.9 – 70; D+ = 69.9 – 67; D = 66.9 – 63; D- = 62.9 – 60 = F = Below 60.

Academic honesty

Students are responsible for reading and understanding WCSU's policy on academic honesty. Cheating and/or other violations of WCSU's Academic Honesty Policy (including but not limited to: fabrication, tampering, plagiarism) will not be tolerated. Failure to abide by policy standards may result in no credit for the submitted work, an "F" for the course, and suspension/expulsion from the University. Written assignments will be screened by antiplagiarism software. No matter how stressed you may feel, by violating the Academic Honesty Policy you are not just cheating in class, you are cheating yourself. Think twice before deciding to take this road. You review WCSU's Academic Policy here: http://catalogs.wcsu.edu/ugrad1819/academic-services-procedures/

Spring 2020 Class Schedule

We will meet Monday and Wednesday 9:30am-12:15pm, along the 2021 spring semester, following the below schedule. Enter in the virtual class 10 minute earlier to allow troubleshooting and also for a free chat among us. We will have a 10-15 minutes break during the class.

WEBEXLINK https://wcsu.webex.com/wcsu/j.php?MTID=mea6dadec39012e113b69f749c88cceb7

WEEK 1

Wed, January 27th: GA, LS
1 Topic: Introduction, teachers and student presentation, course presentation and overview, GIS application.
Assignment: Download the OSGeoLive, following
http://spatial-ecology.net/docs/build/html/VIRTUALMACHINE/00_Setting_OSGeoLive_for_for_Spatial_Ecology_course.html
Youtube video Lecture "Introduction to GIS" https://youtu.be/jgbTosOPU-U

WEEK 2

Mon, February 1st: GA
2 Topic: OSGeoLive Installation. Working with a Linux environment and with the sw in the OSGeoLive.
http://spatial-ecology.net/docs/build/html/VIRTUALMACHINE/00_Setting_OSGeoLive_for_for_Spatial_Ecology_course.html

Assignment: Get familiar with the OSGeoLive, explore the GIS and RS software in the VM;
follow youtube video Lectures "Open software for GIS and RS"
https://www.youtube.com/watch?v=rB_qnTE-W9c&ab_channel=OpenSourceGISBlog
https://www.youtube.com/watch?v=UraLRlZ4-Gg&ab_channel=FOSS4G
Read: GIS Fundamentals: A First Text on Geographic Information Systems, Chapter 1

Wed, February 3rd: GA
3 Topic: Working with a Linux environment. GIS and Remote Sensing sw in the OSGeoLive.
Presentation: GIS and Remote Sensing application.
Useful link
https://www.osgeo.org ; https://www.osgeo.org/projects/ ;
https://maptimeboston.github.io/leaflet-intro/ ; https://leafletjs.com/ ; https://geonode.org/
https://www.postgresql.org/ ; http://www.rasdaman.org/
https://grass.osgeo.org/ ; https://qgis.org/en/site/
https://www.openstreetmap.org ; https://caltopo.com/ ; https://ridewithgps.com/
https://jupyter.org/ ; https://www.r-project.org/ ; https://www.orfeo-toolbox.org/
http://geoserver.org/ ;
Assignment: Get familiar with the OSGeoLive, explore GIS and RS sw and do also other tasks;
Read: GIS Fundamentals: A First Text on Geographic Information Systems, Chapter 2
Read: Principles of Remote Sensing, Chapter 1 and 2.

WEEK 3

Mon, February 8th: GA
4 Topic: QGIS software. Cartographic representation: vectors and rasters. File formats. Data structure, satellite images, satellite products.

https://docs.qgis.org/3.10/en/docs/
Assignment: Download vector and raster data, load in QGIS and get familiar with the visualization. Be ready to talk for a minute concerning the geodata that you found and the kind of visualization that you did in QGIS. Also add a few words on the project that you are planning to develop for the final project.

Wed, February 10th
5 Topic: Vector (buffer, intersect, union, clip, dissolving, reprojection) and raster (resampling, masking, clumping, reclassify, cropping, mosaiching, calculation) analysis in the QGIS software.
https://docs.qgis.org/3.10/en/docs/
Assignment: Get familiar with vector and raster analysis in QGIS with the data that you download. Be ready to talk for a minute concerning the analysis that you did in QGIS.

WEEK 4

Wed, February 17th: GA
6 Topic: Geographic coordinates and projection in the QGIS software.
https://docs.qgis.org/3.16/en/docs/gentle_gis_introduction/coordinate_reference_systems.html
Assignment:
Read: GIS Fundamentals: A First Text on Geographic Information Systems, Chapter 3

WEEK 5

Mon, February 22nd: GA
7 Topic: Linux bash language. Working with text file.
http://spatial-ecology.net/docs/build/html/BASH/03_bashintro_osgeo.html
http://spatial-ecology.net/docs/build/html/BASH/03_bashinter_osgeo.html

Wed, February 24th: GA
8 Topic: Linux bash language. Use of variables and for loop in bash, if statement.
http://spatial-ecology.net/docs/build/html/BASH/03_bashinter_osgeo.html

WEEK 6

Mon, Mar 1st: GA
9 Topic: Linux bash language. Use of variables and for loop in bash, if statement.
http://spatial-ecology.net/docs/build/html/BASH/03_bashinter_osgeo.html
Assignment: see black-board

Wed, Mar 3rd: GA
10 Topic: GDAL/OGR for raster/vector operations (cropping, reprojection, map algebra, resampling, buffer, vrt, etc)

http://spatial-ecology.net/docs/build/html/GDAL/01_gdal_osgeo.html

Assignment: see black-board
From https://www.youtube.com/channel/UCrWEjubnm0HenqzOQgAVScw follow the GDAL Tutorial #1,#2,#3,#4,#5

WEEK 7

Mon, Mar 8th: GA
11 Topic: PKTOOLS for raster/vector operations (masking, composite, filtering, histogram, zonal statistic, etc)
Assignment: ???

Wed, Mar 10th: GA
12 Topic: Introduction to GRASS (grassdb, location, mapset, region, raster and vector operations)
Assignment: ???

WEEK 8

Mon, Mar 15th: LS, GA
13 Topic: Digital Elevation Model: theory and foundation. Use of GRASS for hydrologic analysis
Assignment: ???

Mon, Mar 17th: LS
14 Topic: Introduction to python (data types, tuples/lists/dictionaries, for loops, conditional statements)
Assignment: ???

WEEK 9

Mon, Mar 22sd: GA
15 Topic: panda for table analysis and plotting in python (dataframe properties, descriptive statistics)
Assignment:

Mon, Mar 24th: LS
16 Topic: geopanda for vector analysis (reading vectors, clipping, intersect, extract, geocoding)
Assignment:

WEEK 10

Mon, Mar 29th: GA
17 Topic: map projections geopanda (understanding projection, reprojection)
Assignment: ???

Mon, Mar 31st: LS
18 Topic: rasterio for raster analysis (reading raster, map algebra, cropping, masking, zonal statistic)
Assignment: ???

WEEK 11

Mon, Apr 5th: LS
19 Topic: python for QGIS (raster and vector functions, creating plugins in QGIS)
Assignment: ???

Mon, Apr 7th: LS
19 Topic: Introduction of probability (distribution, transformation, linear/multiple regression, dealing with dresiduals)
Assignment: ???

WEEK 12

Mon, Apr 12th: LS
19 Topic: geostatistic 1 (spatial interpolation, spatial autocorrelation, kriging, spatial autoregressive model)
Assignment: ???

Mon, Apr 14th: LS
19 Topic: geostatistic 2 (bayesian estimations, markov processing)
Assignment: ???

WEEK 13

Mon, Apr 19th: LS
19 Topic: machine learning 1 (machine learning theory, unsupervised learning algorithms, such as clustering and pattern recognition)
Assignment: ???

Mon, Apr 21st: LS
19 Topic: machine learning 2 (supervised learning algorithms such as ensemble learnings, neural networks and graph theories)
Assignment:

WEEK 14

Mon, Apr 26th: GA, LS
19 Topic: Class review and problem solving for final project scripting procedure. Here the students can ask questions for special topics or scripting procedure methodologies/commands.

Assignment:

Wd, Apr 28th: GA, LS
19 Topic: Class review and problem solving for final project scripting procedure. Here the students can ask questions for special topics or scripting procedure methodologies/commands.
Assignment:

WEEK 15

Mon, May 3rd: GA, LS
19 Topic: Student presentation final project
Assignment:

Wed, May 5th: GA, LS
19 Topic: Student presentation final project
Assignment:

Continue with bash script. Introduction to Version Control and the use of git (clone, add, commit, push)