Soil Science 585: Using R for Soil and Environmental Sciences

3 Credits, Spring Semester, Graduate and Undergraduate Level

When: January 23 to May 10, Tuesday and Thursday (9:30-10:45 AM); students need to complete weekly homework to watch the pre-recorded videos prior to the classes; **ONLY** during the lecture hours (Tuesday and Thursday, 9:30-10:45 AM), the instructor will answer the questions and troubleshoot the programming problems with the students.

Instructional mode: remote-synchronously

Instructor: Dr. Jingyi Huang (jhuang426@wisc.edu) ;

website: https://soilsensingmonitoring.soils.wisc.edu/

Office hours: Only available Tuesday and Thursday (9:30-10:45 AM); Via Zoom meetings.

Course designations and attributes: Physical Science, LAS Intermediate level, 50% Graduate credit, Sustainability

Requisites: (STAT 240,301,324,371,STAT/F&W ECOL/HORT 571) and (BSE 301, GEOG/CIV ENGR/ENVIR ST 377, F&W ECOL/ENVIR ST/G L E/GEOG/GEOSCI/LAND ARC 371 or 372) and/or graduate/professional standing;

Other requirements: students should have access to a Windows PC or laptop (Mac or Linux OS not supported) to attend this course.

Reading materials:

Using R for Digital Soil Mapping (By B.P. Malone, B. Minasny and A.B. McBratney, Springer, 2016): Chapters 1–7

Course description:

Data science techniques are increasingly important in soil and environmental science, improving the efficiency and repeatability of data analysis and enhancing fundamental understanding of soil and environmental issues. Various R packages will be introduced and used to analyze and process soil and environmental data collected

using a variety of in situ, ground-based, and remote sensing platforms. R software will be applied to detailed case studies covering soil and environmental data processing, manipulation, and modeling.

This course consists of the following modules:

Module 1: Introduction to R language

This module focuses on the basics of data processing, computation, & visualizing in R. Basic statistical inference & hypothesis testing are performed on soil and environmental datasets collected by students and from an open source platform. Assistance will be provided on processing, modeling, and interpreting the results generated using various R packages.

Module 2: Spatial and temporal analysis in R

This module provides case studies on the use of various R packages for basic GIS and terrain analysis, spatial data manipulation and modeling, spatial sampling, as well as basic time series analysis. The module also focus on the use of machine learning algorithms for mapping soil properties and predicting environmental variables (e.g. groundwater level) over time.

Module 3: Machine learning modeling with soil and environmental data

This module provides case studies on the use of R packages for spectral data processing collected by visible-near infrared (vis-NIR), mid-infrared (MIR), and portable X-ray florescence (XRF) spectrometers and the use of supervised (regression and classification) and unsupervised (clustering) machine learning models for analyzing soil and environmental data collected using *in situ*, ground and remote sensing platforms. Detailed guidance is provided on building spectral libraries for predicting soil properties (e.g. clay content, organic carbon, heavy metals) using machine learning algorithms (e.g., Partial Least Square Regression, Random Forest, Cubist, k-means clustering). The module also provides case studies on basic sensor fusion and data assimilation and demonstrate intelligent laboratory for automatic soil sample analysis in smart agricultural and environmental management.

Module 4: Open-source soil and environmental data management

This module provides case studies on obtaining, processing, and modeling of various national and international open-source soil and environmental data using R software. The open-source datasets include profiles and maps of soil series and properties from USDA-NCSS/NASIS (e.g. SSURGO), soil and groundwater monitoring sites (e.g. USGS-EPA; USDA-SCAN), climate and meteorological data (PRISM/Daymet), and remote sensing soil moisture data (NASA-SMAP). In this module, you have the chance to select one of the following projects for further study and present the results as a poster presentation:

Project 1: Develop a vis-NIR/XRF library for predicting soil properties/elemental data

Project 2: Digital soil mapping of soil properties/series across a selected field/watershed

Project 3: Predicting/hindcasting soil and environmental (e.g., groundwater, surface water) variables over time

Project 4: Use R to improve the analysis of your own research project

Online field trip: Sensing technologies for soil and environmental management

An online field trip is provided on the use of various portable and mobile soil sensors and remote sensing datasets for soil and environmental management applications. You will learn the use of electromagnetic induction, visible near-infrared spectroscopy, and remote sensing for delineating water and nitrogen management zones for site-specific (precision) agricultural management. You will also learn the use of electromagnetic induction and X-ray fluorescence for assessing soil heavy metals across a heavy metal contaminated field.

The course schedule is provided at the end of this syllabus.

How the credit hours are met:

This class meets for two 75-minute class periods each week and expects that students work on course learning activities (reading, writing, programming on assignments, studying, etc.) for about 3 hours out of classroom for every class period.

Learning outcomes:

1. Use R software for basic data manipulation, processing, and visualization.
2. Use R software for basic statistical analysis and hypothesis test.
3. Use R software for spatial and temporal analysis on soil and environmental datasets.
4. Become familiar with the use of electromagnetic induction, portable visible near infrared spectroscopy and portal X-ray florescence spectroscopy for estimating soil and environmental variables in the lab and in the field.
5. Build machine learning models using R software for mapping and predicting soil and environmental variables in space and time.
6. Access, process, and build models using various open-source soil and environmental datasets from USDA, USGS, NASA and other sources.
7. Become competent in the oral presentation.
8. Apply the various R packages to your own research datasets for solving problems in your own disciplines (Graduate students).
9. Explain the social, economic, and environmental dimensions of the sustainability challenges of soil and environmental monitoring and management.
10. Analyze sustainability issues and/or practices on soil and environment using a systems-based approach.

Exams and Grading Rubric

The course evaluation consists of eight lab assignments, one oral presentation, and one project report.

1. Lab assignment (40%):
   In total, eight lab assignments are provided, each of which count for 5 points. You need to use the R codes provided from the lectures to solve the problems. There are one assignment for Module 1, three assignments for Module 2, three assignments for Module 3, and one assignment for Module 4. The submitted R codes will be tested by the instructor. Students will get full points only when the R codes produce the correct answers to the questions. Partial points will be given to those who use the correct R

functions but cannot generate the right answers. **These lab assignments are due one week after the assignments are posted.** A penalty of 10% will be applied for late submission of the assignments.

2. Mini-project (40%), **due on May 2**:

   You need to choose a project from Module 4 or work on other ideas approved by the instructor. It counts for 40 points. The project report consists of the R code and associated dataset, which should be compressed into a Zip file and submitted for evaluation. The instructor will test the codes provided by the students to evaluate the project result. Students will get full points only when the R codes produce the results with no error messages. **The mini-project is due on May 2.** A penalty of 10% will be applied for late submission of the project.

   Undergraduate students:

   The project will be marked based on data preprocessing (30%), model construction (40%), validation and prediction using the models (30%).

   Graduate students:

   The percentages of all the components are data preprocessing (20%), model construction (30%), validation and prediction using the models (30%), and interpretation (20%).

3. 15-min oral competition (20%) **due on May 7 and 9**:

   Students need to present the mini-project result as a 15-min oral presentation. It counts for 20 points. After each presentation, there is a Q&A session of 5 minutes. Each presenter will be evaluated by other students and the instructors.

   The evaluation rubric for **undergraduate students** are the organization of the presentation slides (50%) and clarity of the talk (50%).

   The evaluation rubric for **graduate students** are the organization of the presentation slides (40%), clarity of the talk (40%), and the response to the questions (20%). **Graduate students are required to provide separate sections discussing the modeling results.** The discussion must include the following sections/slides; 1) critically evaluating the models/methods used to generate the results (summarizing the advantages and disadvantages of the models/methods); 2) interpretation the causes of experiments/modeling results using domain science knowledge (e.g. soil and environmental sciences; ecology; geography); and 3) indicating what future work can be done to improve the studies.

   The oral presentation will be held online via Zoom.

The final scores will be converted to letter grades using the following scales:

A 90–100	B 75–80	D 60–65
AB 80–90	C 65–75	F <60

The instructors reserves the right to adjust the point-to-grade scheme to the advantage of the class and to add fractional points based on extraordinary effort to avoid straddling a grade boundary.

Attendance

Students should attend all classes to gain the full benefit of this course.

Regular and substantive student-instructor interaction

Synchronous lectures allow students weekly opportunities for direct interaction with the instructor during scheduled learning sessions. During these lectures, students will work with the instructor to troubleshoot the

programming errors and develop professional R programming skills and critical thinking. Throughout the semester, students submit assignments (R code and presentations), on which the instructor provides individual student feedback on Canvas or by email to improve students' programming skills, the understanding of soil and environmental science knowledge, as well as academic presentation skills. Additionally, discussion forums will be used throughout the semester to gauge student learning and comprehension from assigned readings.

Academic integrity

By enrolling in this course, each student assumes the responsibilities of an active participant in UW-Madison's community of scholars in which everyone's academic work and behavior are held to the highest academic integrity standards. Academic misconduct compromises the integrity of the university. Cheating, fabrication, plagiarism, unauthorized collaboration, and helping others commit these acts are examples of academic misconduct, which can result in disciplinary action. This includes but is not limited to failure on the assignment/course, disciplinary probation, or suspension. Substantial or repeated cases of misconduct will be forwarded to the Office of Student Conduct & Community Standards for additional review. For more information, refer to studentconduct.wiscweb.wisc.edu/academic-integrity/.

Accommodations for students with disabilities

McBurney Disability Resource Center syllabus statement: "The University of Wisconsin-Madison supports the right of all enrolled students to a full and equal educational opportunity. The Americans with Disabilities Act (ADA), Wisconsin State Statute (36.12), and UW-Madison policy (Faculty Document 1071) require that students with disabilities be reasonably accommodated in instruction and campus life. Reasonable accommodations for students with disabilities is a shared faculty and student responsibility. Students are expected to inform faculty [me] of their need for instructional accommodations by the end of the third week of the semester, or as soon as possible after a disability has been incurred or recognized. Faculty [I], will work either directly with the student [you] or in coordination with the McBurney Center to identify and provide reasonable instructional accommodations. Disability information, including instructional accommodations as part of a student's educational record, is confidential and protected under FERPA."
http://mcburney.wisc.edu/facstaffother/faculty/syllabus.php

Diversity & inclusion

Institutional statement on diversity: "Diversity is a source of strength, creativity, and innovation for UW-Madison. We value the contributions of each person and respect the profound ways their identity, culture, background, experience, status, abilities, and opinion enrich the university community. We commit ourselves to the pursuit of excellence in teaching, research, outreach, and diversity as inextricably linked goals. The University of Wisconsin-Madison fulfills its public mission by creating a welcoming and inclusive community for people from every background – people who as students, faculty, and staff serve Wisconsin and the world." https://diversity.wisc.edu/

Course Schedule

Week	Tuesday	Thursday
Jan 23 & 25	Module 1: Introduction	Module 1: Basics of R software
Jan 30 & Feb 1	Module 1: Data manipulation & processing in R	Module 1: Statistical inference & hypothesis testing in R
Feb 6 & 8	Module 2: Basic GIS and data visualization in R	Module 2: Terrain analysis and LiDAR data processing in R
Feb 13 & 15	Module 2: Spatial analysis in R: design-based sampling	Module 2: Spatial analysis in R: model-based sampling
Feb 20 & 22	Module 2: Spatial analysis in R: kriging – Part 1	Module 2: Spatial analysis in R: kriging – Part 2
Feb 27 & Feb 29	Module 2: Spatial-temporal analysis in R Review of Modules 1 and 2 (Tutorial)	Online field trip on proximal soil sensing
Mar 5 & 7	Module 3: Spectral data analysis in R: vis-NIR and MIR spectra: Part 1	Module 3: Spectral data modeling in R: vis-NIR and MIR spectra: Part 2
Mar 12 & 14	Module 3: Spectral data analysis and modeling in R: portable XRF spectra	Module 3: Multi-sensor fusion for delineating soil and environmental management zones
Mar 19 & 23	Module 3: Digital soil mapping: towards a universal soil classification system	Module 3: Digital soil mapping: digital soil morphometrics
Mar 26 & 28	Spring Recess	Spring Recess
Apr 2 & 4	Module 3: Digital soil mapping: soil properties (continuous data)	Module 3: Digital soil mapping: soil types (categorical data)
Apr 9 & 11	Module 4: Open-source soil and environmental data management in R: agricultural, hydrological, and environmental data	Module 4: Open-source soil and environmental data management in R: climate and remote sensing data
Apr 16 & 18	Module 4: Preparing final projects on selected topics.	Module 4: Preparing final projects on selected topics.
Apr 23 & 25	Module 4: Preparing final projects on selected topics.	Module 4: Preparing final projects on selected topics.
Apr 30 & May 2	Module 4: Preparing final projects on selected topics.	Module 4: Preparing final projects on selected topics.
May 7 & 9	Oral presentation on projects.	Oral presentation on projects.

Dr. Jingyi Huang | Soil Sensing and Monitoring Lab
Department of Soil Science | University of Wisconsin – Madison | jhuang426@wisc.edu