PLSCS/NTRES 6200
"Spatial Modelling and Analysis for agronomic, natural resources, environmental and regional studies"
Spring semester 2023
Course information

D G Rossiter

Cornell University, New York State College of Agriculture & Life Sciences
School of Integrative Plant Sciences, Section of Soil & Crop Sciences

January 24, 2023

Orientation to course

1. for whom?
2. objectives
3. instructor
4. method
5. assignments & grading
6. project
7. schedule
8. acccessibility and diversity statements
9. some key references

For whom?

- **Graduate** and **senior undergraduate** students who want to advance their understanding of geographic information science and technology applied to problems in the agronomic, natural resources, environmental or regional studies sciences.
  - Student should have followed an undergraduate-level or introductory graduate-level courses in (non-spatial) inferential statistics.
  - Helpful but not necessary: remote sensing, computer programming, GIS theory and practice.
- Especially useful for students starting a **graduate research project** that uses **spatial information** in agronomy, soil science, natural resource management, regional science, hydrology, ecology ...

Learning outcomes

- Student is able to analyze **complex spatial problems** with appropriate **theory**, **statistical methods** and **computational tools**
  - No "silver bullet", each problem must be analyzed on its own terms
  - Always use **domain knowledge** and consider **application requirements**
- Student is able to organize data analysis as **reproducible research** ("literate data analysis")
- Student is competent to process and visualize spatial data; with emphasis on **open-source computer programs** and **publically-available data**
  - Most labs use the the **R environment** for data analysis, (spatial) statistical computing and visualization,
  - One lab uses **GeoDA**; one lab uses **Google Earth Engine**

Instructor

- Adjunct Professor CALS (12th year)
  - Graduate faculty *Soil & Crop Sciences* ‘Environmental Information Systems’ concentration
  - Graduate faculty *Regional Science* ‘Environmental Studies’ concentration
- Guest Researcher (*Gastmedewerker*), ISRIC–World Soil Information, Wageningen (NL) (8th year)
- Visiting Professor (客座教授), Chinese Academy of Sciences, Soil Science Research Institute Nanjing (中国科学院南京土壤研究所) (10th year – postponed past three years due to COVID-19 restrictions)
- Visiting Professor (讲座教授), Nanjing Normal University, School of Geography (南京师范大学地理学学院) (7th year – postponed past three years due to COVID-19 restrictions)
- Retired from University of Twente (NL), Faculty of Geoinformation Science & Earth Observation (after 17 years)

Instruction method

- Graduate-level course, emphasis is on **guided self-instruction**, including critical reading of primary literature, lab. self-paced tutorials
- Lectures are **overviews / orientations** to introduce **fundamental spatial analysis concepts and methods**
- Emphasise ability to read, understand and apply **methods in journal papers**, advanced **textbooks** and **reference** books; necessary in graduate career
- Student **project** allows you to apply ideas and techniques of spatial analysis from this course or elsewhere to your research
- First nine (of 14) lab. periods are tutorial exercises on **methods** and **computation**; example applications to illustrate these; other lab. periods for project work and presentation.

Assignments

- Weeks 1-9: nine (**9**) set **computer lab. exercises**, small hand-in assignment; due before Thurs. of following week (to allow in-class discussion before next lab)
- *(Spring break)*
- Weeks 11-14 four (**4**) short **question sets** based on reading; due before before Tues. of following week
- Week 15: no assignment

Grading

- Course is **4 credits**
- **Nine** lab. exercise assignments: graded 0–3; 45% of final
  - 0 = not submitted; 1 = weak attempt; 2 = some incorrect answers or procedures; 3 = (close to) perfect
- **Four** question sets: graded 0–3; 15% of final
- Project: 40% of final (breakdown: 80% project, 10% reporting, 10% presentation/discussion)
- Grading on absolute scale; Letter equivalents:

≥ 97.5	A+	≥ 92.5	A	≥ 90.0	A-
≥ 87.5	B+	≥ 82.5	B	≥ 80.0	B-
≥ 77.5	C+	≥ 72.5	C	≥ 70.0	C-
≥ 67.5	D+	≥ 62.5	D	≥ 60.0	D-
< 60.0	F				

Project

- A small project in which the student applies **spatial analysis** to a **problem of interest** relevant to student's field of study;
  - Usually part of students' graduate research or senior thesis.
  - May work in groups if independent contributions can be evaluated.
- **Spatial** or **Spatio-temporal** analysis must be prominent;
- As much as possible, presented as **reproducible research**
- Prefer use of **own** data, **similar** obtained from colleague/advisor, or **publically-available** data; little credit for extensive data manipulation (not the main purpose of this course);
- Should be able to complete during **allocated lab. time and related self-study time**; graded as such.

Project milestones

1. **brief proposal**: 07-March (week 7 Tuesday)
   - instructor checks (1) within scope of course, (2) feasible (available data sources)
2. **full proposal** literature review / detailed project plan: 21 March (week 9 Tuesday)
   - Graded as **mid-term**: background, objectives, proposed methods, expected results ...
3. in-lab presentation of preliminary results 2/3 May (week 14 Tu/W)
   - because of CU schedule this is the latest possible
4. **report**: ≈19-May (on scheduled exam date, TBD, 16–19-May)

Lectures

- Tuesday, Thursday 0805–0920.
- Bradfield 105
  - lecture PDF on Canvas (with other module materials)
  - recorded lecture on Canvas (after live lecture)
- Content is equivalent to a 50-minute lecture, this gives more time for review, repetition and **interaction** – stay awake!
- Start with Q&A of previous lectures/labs; questions and discussion of the key points in the assigned papers, as motivation for the lecture.
- Some in-class quizzes on Thursdays
- On some Thursdays discuss a paper which uses techniques covered during the week.

Computer laboratory

- Tuesday (Emerson 133) **or** Wednesday (Mann B30B) 1330-1630
- brief orientation; most time is self-paced with instructor available to answer questions; a few breaks to review key points
- use your own computer
  - any standard Mac OS/X, Windows, Linux system is suitable
  - most labs. use the **R Project for Statistical Computing**[1] in the **RStudio** environment[2]
  - one lab with GeoDA[3] spatial analysis program
  - one lab with Google Earth Engine[4]
- discussion forum for asynchronous problem-solving and suggestions
- due following Tuesday morning, after grading in-class discussion of key points

[1] https://www.r-project.org/
[2] https://www.rstudio.com
[3] https://spatial.uchicago.edu/geoda
[4] https://earthengine.google.com/

Pre-class

Please complete these tasks during winter break.

R Environment

- Install R[5] to your computer system
- Install R Studio[6] to your computer system
- Follow the R Markdown tutorial from R Studio[7] through the "Inline Code" section
- Download and review the R Markdown Cheatsheet[8]

Thought exercise See pre-class reading

[5] `https://www.r-project.org`
[6] `https://www.rstudio.com`
[7] `https://rmarkdown.rstudio.com`
[8] `https://github.com/rstudio/cheatsheets/raw/main/rmarkdown-2.0.pdf`

Lecture schedule I

- Week 1 (week of 23-Jan) Naïve analysis; Spatial concepts; georeferencing, Coördinate Reference Systems and transformation
- Week 2 (week of 30-Jan) Spatial prediction from points: Universal model of spatial variation; feature-space regression, trend surfaces
- Week 3 (week of 06-Feb) Local spatial dependence, variogram analysis, Ordinary Kriging (OK), regression kriging
- Week 4 (week of 13-Feb) spatially-explicit data-driven methods ("machine learning")
- Week 5 (week of 20-Feb) model tuning, model evaluation,
- Week 6 (week of 27-Feb) Tu *February mini-break*
  Th Remote sensing as a data source; Principal Components Analysis
- Week 7 (week of 06-Mar) Areal Data Spatial Analysis
- Week 8 (week of 13-Mar) Point Pattern analysis
- Week 9 (week of 20-Mar) Time series analysis; Spatio-temporal analysis
- Week 10 (week of 27-Mar) Spatial sampling; clustering

Lecture schedule II

Spring Break (week of 03-Apr)

Week 11 (week of 10-Apr) Big data, open data

Week 12 (week of 17-Apr) Geographically-weighted methods; Geospatial simulation

Week 13 (week of 24-Apr) Uncertainty, data quality, metadata

Week 14 (week of 01-May) Pattern analysis

Week 15 (week of 08-May) (Tu only; last day of classes) Summary: newest developments in spatial analysis

Lab schedule

- **Week 1** R spatial: using R with spatial data; creating a literate data analysis with R Markdown
- **Week 2** Trend surfaces, spatial regression
- **Week 3** Point geostatistics
- **Week 4** Data-driven methods: classification & regression trees; random forests
- **Week 5** Big data, open data, remote sensing, principal components analysis (Google Earth Engine)
- **Week 6** (no lab, *February mini-break*)
- **Week 7** Areal Data Spatial Analysis (GeoDa)
- **Week 8** Point-pattern analysis
- **Week 9** Time series analysis; Spatio-temporal analysis
- **Week 10** Spatial sampling; clustering
- **Weeks 11–14** Individual project work (instructor available for consultation)
- **Week 15** Student project (ongoing) presentations/discussion

Accessibility statement

- **Lectures** are presented “live”, these follow structured PDFs (see below) but there is also interaction. If there are hearing or vision problems, these are common to all lectures and Cornell should make the appropriate accomodations. Lectures are recorded with Zoom, automatically subtitled and posted on Canvas under the appropriate module.
- **Computer labs** use the student’s own computer, for any accessibility issues there please consult Cornell IT.
- **Documents** are almost all PDF created with LaTeX, which does not (yet) provide a mechanism for making them compliant with accessibility standards. However, they can be automatically converted in Canvas to various accessible formats (HTML, audio …), see `Alternative formats` button when the PDF is displayed within Canvas.

  Cornell Center for Teaching Innovation is aware of this issue, which especially affects STEM fields, and is working on accessibility support in LaTeX).
- **Any other issues**, please discuss with the instructor.

Diversity statement

A diverse class with diverse experiences and objectives! We try to be useful to everyone's intellectual journey.

- Students come from all over the University in many disciplines, which have **different ways of thinking about and describing space and spatial analysis**.
- Students are interested in **different application areas** and orientations: some more social, some more natural resources.
- Students are studying for PhD (at **different points in their projects**), MSc, MPS and even senior BSc; this leads to different ways of thinking about research.
- Students have **different levels of previous knowledge** of (spatial) statistics, computer programming (especially R), GIS, remote sensing, empirical-statistical models, etc.
- Student research may **more or less emphasize concepts vs. computation** and may think more **laterally** or **linearly**.
- Students have **different life experiences**, e.g., previous study, fieldwork, employment, organizations

Key texts/references I

Concepts of spatial modelling

- O'Sullivan, D., & Unwin, D. (2010). Geographic information analysis (2nd ed). Wiley.
  http://library.books24x7.com/toc.aspx?site=KD7O8&bookid=35218

Ecological modelling

- Legendre, P. (2012). Numerical Ecology (3rd ed.). San Diego: Elsevier Science & Technology Books.
  http://www.sciencedirect.com/science/bookseries/01678892/24
- Borcard, D. (2011). Numerical ecology with R. New York: Springer.
  http://link.springer.com/book/10.1007%2F978-1-4419-7976-6

Statistics review

- Electronic Statistics Textbook https://docs.tibco.com/data-science/textbook
- Dalgaard, P. (2008). Introductory Statistics with R (2nd ed.). Springer.
  http://link.springer.com/book/10.1007%2F978-0-387-79054-1

Key texts/references II

Spatial analysis in R

- Bivand, R., Pebesma, E., & Gómez-Rubio, V. (2013). Applied spatial data analysis with R (2nd ed). New York: Springer.
  http://link.springer.com/book/10.1007%2F978-1-4614-7618-4.
  Datasets etc. at http://www.asdar-book.org/
- Pebesma, E. (2018). Simple Features for R: Standardized support for spatial vector data. The R Journal, 10(1), 439–446.
  https://doi.org/10.32614/RJ-2018-009
- Hijmans, R. J. (2019–2021). Spatial Data Science with R — R Spatial.
  https://rspatial.org/

General GIS

- Burrough, P. A., McDonnell, R., & Lloyd, C. D. (2015). Principles of geographical information systems (3rd edition). Oxford University Press.

Key texts/references III

Geostatistics

- Webster, R., & Oliver, M. A. (2008). Geostatistics for environmental scientists. John Wiley & Sons Ltd. http://onlinelibrary.wiley.com/book/10.1002/9780470517277
- Goovaerts, P. (1997). Geostatistics for natural resources evaluation. New York; Oxford: Oxford University Press.
- Diggle, P. J., & Ribeiro Jr., P. J. (2007). Model-based geostatistics. Springer. http://link.springer.com/book/10.1007%2F978-0-387-48536-2
- Krivoruchko, K. (2011). Spatial statistical data analysis for GIS users. DVD containing book + data. Redlands: ESRI.

Key texts/references IV

Statistical modelling

- Fox, J. (2008). Applied regression analysis and generalized linear models (2nd ed.). Los Angeles: Sage.
- Fox, J., & Weisberg, S. (2011). An R companion to applied regression (2nd ed.). Thousand Oaks, Calif.: SAGE Publications.
- Venables, W., & Ripley, B. (2002). Modern Applied Statistics with S. Fourth Edition. Springer.
  `http://link.springer.com/book/10.1007%2F978-0-387-21706-2`
- Hosmer, D. W., & Lemeshow, S. (2013). Applied logistic regression (3rd ed.). Wiley-Interscience Publication.
  `http://onlinelibrary.wiley.com/book/10.1002/9781118548387`

Key texts/references V

Machine learning

- Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). The elements of statistical learning data mining, inference, and prediction (2nd ed). New York: Springer. http://link.springer.com/book/10.1007%2F978-0-387-84858-7
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning: with applications in R. New York: Springer. http://link.springer.com/book/10.1007%2F978-1-4614-7138-7 (A simplified version of Hastie *et al.* 2009).

Land surface modelling

- Hengl, T., & Reuter, H. I. (Eds.). (2009). Geomorphometry: concept, software, applications. Developments in soil science 33. http://www.sciencedirect.com/science/bookseries/01662481/33

Key texts/references VI

Spatial sampling

- Brus, D. J. (2022). Spatial sampling with R. https://dickbrus.github.io/SpatialSamplingwithR/
- de Gruijter, J., Brus, D. J., Bierkens, M. F. P., & Knotters, M. (2006). Sampling for Natural Resource Monitoring. Springer. http://link.springer.com/book/10.1007%2F3-540-33161-1

Time series, spatio-temporal

- Shumway, R. H., & Stoffer, D. S. (2017). Time series analysis and its applications: with R examples (4th ed). New York: Springer. https://link.springer.com/book/10.1007%2F978-3-319-52452-8
- Cowpertwait, P. S. P., & Metcalfe, A. V. (2009). Introductory Time Series with R. Springer. https://doi.org/10.1007/978-0-387-88698-5
- Christakos, G. (2012). Modern spatiotemporal geostatistics. Mineola, N.Y: Dover Publications. http://app.knovel.com/hotlink/toc/id:kpMSG00001/modern-spatiotemporal-geostatistics

Key texts/references VII

Data science

- Wickham, H., & Grolemund, G. (2016). R for Data Science. O'Reilly Media `https://r4ds.had.co.nz`
- Ahmed, Zia (2020). Geospatial Data Science in R. `https://zia207.github.io/geospatial-r-github.io/index.html`

Graphics

- Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis (2nd ed.). Springer International Publishing. `https://doi.org/10.1007/978-3-319-24277-4`