Cornell University
PLSCS/NTRES 6200
Spatial Modelling and Analysis for agronomic, natural resources, environmental and regional studies

Spring semester 2021
Course information

D G Rossiter
d.g.rossiter@cornell.edu
http://www.css.cornell.edu/faculty/dgr2/

November 30, 2020

Orientation to course

1. for whom?
2. objectives
3. instructor
4. method
5. assignments & grading
6. project
7. schedule
8. 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 inferential statistics (e.g., STSCI 3200/BTRY 3020)
  - Helpful but not necessary: remote sensing, computer programming, GIS theory and practice (e.g., PLSCS 4200)

- Especially useful for students starting a **graduate research project** that uses spatial information in agronomy, soil science, natural resource management, regional science, hydrology, ecology . . .
  - "spatial": information is from **known locations** with **known spatial extent**

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 the interactive **GeoDA**[1] spatial analysis program

[1] https://spatial.uchicago.edu/geoda

Instructor

David G. Rossiter

- Adjunct Associate Professor CALS (10th 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) (6th 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) (8th year – postponed due to COVID-19)

- Visiting Professor (讲座教授), Nanjing Normal University, School of Geography (南京师范大学地理学学院) (6th year – postponed due to COVID-19)

- 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**
  - Each lecture period begins with $\approx$ 20 minutes discussion of an **assigned pre-class reading**, and with Q&A.
- 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 1-10: nine (**9**) set **computer lab. exercises**, small hand-in assignment; due before Tues. of following week
- Weeks 11-13 three (**3**) short **question sets** based on reading; due before before Tues. of following week
- Week 14: 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
- **Three** 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:

$\geq$ 97.5	A+	$\geq$ 92.5	A	$\geq$ 90.0	A-
$\geq$ 87.5	B+	$\geq$ 82.5	B	$\geq$ 80.0	B-
$\geq$ 77.5	C+	$\geq$ 72.5	C	$\geq$ 70.0	C-
$\geq$ 67.5	D+	$\geq$ 62.5	D	$\geq$ 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 pairs if independent contributions can be evaluated.
- **Spatial** or **Spatio-temporal** analysis must be prominent;
- 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;
- Milestones: (1) **brief proposal**: 16-March (week 6 Tuesday) ; (2) **proposal presentation** (literature review / detailed project plan): 6 April (week 9 Tuesday); (3) presentation of preliminary results (11 or 12 May, week 14 Tu/Wed); (4) **report**: $\approx$24-May (on scheduled exam date, TBD)

Week schedule

Lectures: Tuesday, Thursday 0805–0920. In person Plant Science 404, also streaming on-line; lectures recorded and subtitled on Canvas

- Content is equivalent to a 50-minute lecture, this gives more time for review and interaction.
- Start with questions and discussion of the key points in the assigned papers, as motivation for the lecture.
- On some Thursdays discuss a paper which uses techniques covered during the week.

...

. . .

Computer laboratory: Tuesday **or** Wednesday 1330-1420 **or** 1440-1530, Mann Library 409 (first 50 minutes) and on-line, with on-line consultation for remainder of week.

- you should select one labs, because of room occupation restrictions (8 maximum), but if there is space you can come to an additional lab.
- 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** in the **RStudio** environment, one lab with GeoDA.
- discussion forum for asynchronous problem-solving and suggestions

Schedule – lectures

Week 1 Naïve analysis; The R environment; Literate data analysis; Spatial concepts; Coördinate Reference Systems

Week 2 The Universal Model of Spatial Variation; Local spatial dependence, Ordinary Kriging (OK) ;

Week 3 Model-based methods: Feature-space regression, trend surfaces

Week 4 Data-driven methods (“machine learning”)

Week 5 Remote sensing as a data source; PCA (Th only)

Week 6 Big data, open data

Week 7 Area-Based Spatial Data Analysis

...

...

Week 8 Time series analysis; Spatio-temporal analysis

Week 9 Point Pattern analysis

Week 10 Spatial sampling; clustering

Week 11 Uncertainty, data quality, metadata;

Week 12 Geospatial simulation

Week 13 Bayesian methods for spatial analysis

Week 14 Summary: newest developments in spatial analysis

Schedule – labs

Week 1 The R environment; Literate Data Analysis

Week 2 Geostatistics with R/gstat

Week 3 Trend surfaces, spatial regression

Week 4 Data-driven methods: classification & regression trees; random forests; optimization

Week 5 (no lab)

Week 6 Google Earth Engine

Week 7 Area spatial data analysis: GeoDa

Week 8 Time series analysis; Spatio-temporal analysis

...

. . .

Week 9 Point-pattern analysis

Week 10 Spatial sampling; clustering

Weeks 11–14 Individual project work (instructor available for consultation)

Week 15 Student project (ongoing) presentations/discussion

Key texts/references

Cornell has access to most of the e-books.

Spatial modelling in general

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

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). **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.

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.

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`

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

Spatial sampling

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.proxy.library.cornell.edu/book/10.1007/978-3-319-52452-8

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

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