Propose for graduate course.

SPATIAL ANALYSIS AND MODELLING OF ENVIRONMENTAL DATA.

Jose Camilo Fagua
Camilo.Fagua@nau.edu
Postdoctoral Research Associate
Global Earth Observation & Dynamics of Ecosystems Lab (GEODE)
School of Informatics, Computing & Cyber Systems-Northern Arizona University
Fagua ORCID , Fagua Google Scholar , Fagua Publons

Presentation

Spatial analysis is a type of statistical analysis that seeks to explain spatial data patterns in mathematical and geometric terms. Spatial modelling are the statistical techniques to approximate reality and make spatial predictions. Environmental data is any information that can be used to solve an environmental issue, including data on economics, society, biodiversity, climate, hydrology, biophysics, etc. The main objective of this course thus is to introduce grad students to the spatial analysis and modelling of environmental data using open-source software (R and Google Earth Engine-GEE). The course is divided in twelve modules where students develop spatial analyses on several environmental issues using real data (available for public use) and R codes mainly. Although the material of the course is in English (tutorials, books, data), the classes are going to be teach in Spanish.

Course Objectives

By the end of this course, the grand students will be able to:

1. Use R and GEE to summarize (spatial) data numerically and visually.

2. Explore (spatial) data patterns and relationships.

3. Create spatial models.

4. Conduct independent and reproducible research by applying techniques learned in class to novel datasets.

5. Effectively communicate results and implications.

The previous specific course objectives are also connected to broader objectives to help grad students in their thesis and future professional projects since this course only use powerful, free, and open resources (data and software), such as:

1. Develop specific skills, competencies, and points of view needed by professionals.

2. Learn how to find and use resources for answering questions or solving problems.

3. Learn to analyze and critically evaluate ideas, arguments, and points of view.

Prerequisites

Basic understanding of statistics, introductory programming (R, JavaScript, Python, etc.), and/or geospatial software (i.e. ArcGIS, QGIS) is helpful, but not required.

Course Materials

Students will use open resources (data and software), including three main textbooks:

1) An Introduction to Spatial Data Analysis and Visualization in R (Lansley and Cheshire, 2016).

2) R for Data Science (Grolemund and Wickham, 2017)

3) An Introduction to Statistical Learning with Applications in R (Witten et al., 2013).

4) Applied Spatial Data Analysis with R (Bivand et al., 2013).

Other books could be helpful.

5) Hands-On Machine Learning with R (Boehmke and Greenwell, 2019).

6) The Elements of Statistical Learning: Data Mining, Inference, and Prediction (Hastie et al., 2009).

7) Package 'raster.' (Hijmans et al., 2016)

8) The Art of R Programming. (Matloff, 2011)

9) A language and environment for statistical computing. (R Team, 2014)

Course Structure

The course is organized in twelve 5-hours modules and a final 4-hours session for the presentation of student projects. Each 5-hours module contains: a presentation given by the professor and a tutorial with background information required to complete the assignment for the module as well as additional resources if the students want to learn more about the topic. These additional resources are formed by published papers, books, and other R tutorials where the topics of the module are implemented. It is expected that the students spend additional hours on the 5-hours module, studying tutorial background information, additional resources, and developing their final projects.

Course Grading

Spatial analysis can only be learned by doing; therefore, the final grade (150 points) will come from the class-modules assignments (110 points total, 10 per assignment) and from the develop of a final project (40 points total). The points obtained by the student will be rescale from zero to five, the grading of Universidad Nacional de Colombia. Student must use data of their thesis or side projects for the final project. Students will start the final project after the Module 4. They will apply the tools and techniques learned in class to their own data and will present a 15-minute talk to the class as final project.

Course Policies

Outside materials: In the real world, researchers access online resources to solve complex problems; thus, students can open notes, book, computer, and any other academic resource.

Attendance: I would be not monitor attendance in class. However, it would be highly recommended for the students to attend classes since I would not assist student on topics reviewed during classes.

Academic Integrity

Students will be held accountable to the Honor Pledge which they have agreed to: I pledge, on my honor, to conduct myself with the foremost level of academic integrity. I would take appropriate actions in response to academic dishonesty, as defined the Universidad Nacional de Colombia. Acts of academic dishonesty include but are not limited to 1) cheating, 2) falsification, and 3) plagiarism.

Course Summary

Module 1: Introduction

- Introduction to R.
- Installing R and R-studio.
- Data exploration in R.
- Descriptive statistics and visualization in R.
- Using R-Markdown.

Module 2: Polygons in R.

- Polygon structure.
- Reading, creation, and projections of polygons.
- Point patterns.
- Polygon analysis (extract, overlay, proximity, statistics).
- Polygon visualization in R.

Module 3: Raster in R.

- Raster structure (spatial resolution, bands).
- Reading, creation, and projections of raster.
- Raster analysis (raster math).
- Raster visualization in R.

Module 4: Spatial data integration and spatial data availability.

- Polygon to raster and raster to polygon.
- Analysis integrating raster and polygon.

- Overview of available spatial data (social, economics, climate, land cover, topography, hydrology, conflicts).
- Using available spatial data (points, polygons, raster).

Module 5: Spatial Autocorrelation.

- Finding neighbors.
- Analyses for Global spatial autocorrelation.
- Analyses for Local spatial autocorrelation.
- Getis-Ord analysis.
- Hot Spot Spatial Analysis

Module 6: Spatial interpolation.

- Thiessen polygons analysis.
- Inverse Distance Weighting (IDW).
- Kriging analysis.
- Fixed Rank Kriging (FRK)

Module 7 and module 8: learning algorithms and spatial modelling.

- Regression vs classification.
- Training group and accuracy measures.
- Linear regression analysis.
- Tree analysis.
- Random Forest Analysis.
- Gradient Boosting Machine analysis.
- Support vector machine analysis.

Module 9: Remote sensing data and map modelling from scratch

- Remote sensors (spatial, temporal, and radiometric resolutions).
- Temporal and textural metrics for multispectral imagery.
- Smoothing algorithms (Asymmetric Gaussian function, Savitzky and Golay function)
- Map building from scratch.

Module 10: Remote sensing processing and map modelling in Google Earth Engine

- Introduction to GEE.
- Imagery processing
- Learning algorithms and spatial modelling.
- Modeling urban growth sustainability in GEE.

- Landsat-based forest height predictions using a GEDI asset on GEE.

Module 11: LiDAR in R.

- ALS-LiDAR (discreate form):
  Ground classification and digital terrain model.
  Height normalization and digital elevation model.
  Vertical and horizontal metrics.
- Spatial LiDAR (wave form):
  Processing of Geoscience Laser Altimeter System (GLAS)
  Processing of Global Ecosystem Dynamics Investigation (GEDI)

Module 12

- Final project: final time where students exchange the background of their projects with the professor before the presentation.

References

Bivand, R., Pebesma, E., Gómez-Rubio, V., 2013. Applied Spatial Data Analysis with R, 2nd ed. Springer, New York, US.

Boehmke, B., Greenwell, B.M., 2019. Hands-On Machine Learning with R.

Grolemund, G., Wickham, H., 2017. R for Data Science. New York, US.

Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edi. ed, Springer Series in Statistics. Springer Science+Business Media, New York.

Hijmans, R., Etten, J. van, Cheng, J., Mattiuzzi, M., Sumner, M., Greenberg, J., Perpinan, O., Bevan, A., Racine, E., Shortridge, A., 2016. Package 'raster.' r-project.org.

Lansley, G., Cheshire, J., 2016. An Introduction to Spatial Data Analysis and Visualisation in R. University College London, London, UK.

Matloff, N., 2011. The Art of R Programming: A Tour of Statistical Software Design, No Starch Press Series. No Starch Press.

Team, R.C., 2014. R: A language and environment for statistical computing.

Witten, J.G., Hastie, D., Tibshirani, R., 2013. An Introduction to Statistical Learning: with Applications in R. Springer, New York, US.