Syllabus for Statistics 260: Spatial Statistics
University of California, Berkeley, Spring 2011

T/Th 3:30-5pm Lecture, 332 Evans

Instructor: Dr. Cari Kaufman
Office: 315 Evans Hall
Office Hours: TBA
e-mail: cgk@stat.berkeley.edu

About This Course. Spatial data arise in a wide range of scientific disciplines, including
the geosciences, environmental sciences, image processing, and public health. This course
will introduce you to methods and models that have been developed for spatial data. We
will cover three main types of spatial data: geostatistical (point-referenced) data, lattice or
areal data, and point process data. One unifying theme will be the specification and fitting
of probability models for such data, using stochastic processes and hierarchical models to
represent the complex dependencies that often arise. My goals are that by end of the
course, you will be able to

• Describe common methods and models for spatial data,
• Read and discuss new methods in the literature, based on an understanding of the
  basic spatial statistics approaches, principles, and main assumptions,
• Evaluate which methods to use for spatial datasets that may arise in your own
  research, and
• Implement methods using statistical software.

Course Structure. The structure of this course is designed to give you both an overview
of the main methods in spatial statistics, as well as to facilitate a deeper understanding of
methods that are of particular interest to you. To that end, the course will include

• Readings - from the textbook and occasional outside sources
• Lectures - overviewing main ideas, showing illustrative examples (3/4 of class time)
• Student presentation and discussion sessions - see below (1/4 of class time)
• Assignments - 4-5 problem sets
• Project - including small literature review and data analysis

Prerequisites. 210A, 200B, or instructor permission. Familiarity with topics from matrix
algebra and linear models in statistics is also assumed.

Textbook. The textbook for this course is
Statistics, CRC Press.
I will assign specific sections of the book to be read before class. My lectures will draw from additional sources and may not cover everything in the assigned reading. Homework questions will be based on both the book and the lectures.

Some additional references I recommend are

- “Statistics for Spatial Data” by Cressie
- “Environmental Statistics” by Smith (unpublished manuscript, available on bSpace)
- “Hierarchical Modeling and Analysis for Spatial Data” by Banerjee, Carlin, and Gelfand
- “Model-based Geostatistics” by Diggle and Ribeiro
- “Gaussian Markov Random Fields” by Rue and Held
- “Statistical Analysis of Spatial Point Processes” by Diggle
- “Applied Spatial Data Analysis with R” by Bivand, Pebesma, and Gomez-Rubio
- “Thematic Cartography and Geovisualization” by Slocum, McMaster, Kessler, and Howard

I will have these available during my office hours if you would like to review them.

**In-class Presentations.** Working in groups of two, you will choose a journal article to present to the class. These may be seminal works cited in the textbook, recent articles extending an idea we have discussed, or scientific applications. I will discuss your choice of article with you and make sure it serves the class as a whole. You will announce the article to the class at least one week prior to presenting it, and the rest of the class will read it. However, it is your job to deeply understand the article. This may involve reading additional sources, working through derivations, or implementing parts of an analysis yourself. You will have 40 minutes to present the article to the class, leaving time for discussion and questions. I expect you to spend approximately two weeks reading and preparing.

**Assignments.** I will post assignments to the class bSpace page. They will consist of both written problems and real data examples involving some computing. You will typically have two weeks to complete each assignment and turn in a hard copy at the beginning of class. As a general rule I will not accept late assignments. In special cases I may grant extensions, but the likelihood of this will increase if you talk to me in person well ahead of the due date.

**Computing.** Assignments will make use of the open source programming language and environment called R, and I will occasionally do worked examples in class that involve R programming. Previous knowledge of R is not required but will be useful. *If you would like an account for the department computer labs, please email me by Friday, January 21.*

**Final projects.** Again working in groups of two, you will carry out a novel analysis on a data set of your choosing. I expect you to choose an appropriate method for the analysis, as well as carrying out a small review of the relevant literature (statistical and subject-matter
appropriate). Additional details will be provided towards the end of the semester. You will have approximately three weeks to complete your project.

**Grading.** Your final grade will be a weighted average of your homework average (60%), in-class presentation (20%), and final project (20%). Pass/fail grading is also an option; please talk to me in person.

**Academic integrity.** Any work submitted by you and that bears your name is presumed to be your own original work that has not previously been submitted for credit in another course. While you are encouraged to work together on homework assignments, the code and writeup must be your own. For example, suggesting a function to another student is acceptable, whereas simply giving him or her your own code is not. If you are not clear about the expectations for completing an assignment, be sure to seek clarification from the instructor. Any evidence of cheating or plagiarism will be subject to disciplinary action.

**Students with disabilities.** If you need accommodations for any physical, psychological, or learning disability, please get in touch with me so that we can make the necessary arrangements.