Spatial Biostatistics

PubH 8472 **Spring 2008**

Instructor:	Dr. Brad Carlin
Office:	Mayo Building, Room A427; phone 624-6646
fax:	626-0660
e-mail:	brad@biostat.umn.edu
course website:	www.biostat.umn.edu/∼brad/ph8472.html

Class Meetings:	Tu Th 9:45 – 11:00 am, MoosT 2-580
Office Hours:	Instructor – Tu Th 3:45-5:15, or by appointment Shengde Liang (TA) – Tues 2:00-4:00 pm and Wed 1:00-2:00 pm; held in Mayo A452 (biostat TA room) (phone: 624-0131; e-mail: shengdel@biostat.umn.edu)
Prerequisites:	Stat 5101-02 and some experience with `S-plus` or `R`, or permission of instructor. PubH 7440 or 8431 (Bayes methods) and Stat 8311 (linear models) are also recommended.

I. Course Description

In recent years there has been enormous growth in interest in the analysis of spatial data, and the development of statistical methodologies that interact with geographical information systems (GISs). This course is about spatial data, spatial statistical models, and their proper fitting, summary, and interpretation. It is designed to introduce students to the nature of spatial data and the special analysis tools that help to analyze such data. The course covers a blend of theory, applications, and software related to the three major types of spatial data: geostatistical (point level), areal (block level), and spatial point process. Both classical (frequentist) and Bayesian statistical methods will be considered, though the latter will emerge as more powerful and, hence, generally preferred.

II. Learning Objectives

1. Understand the nature of spatial data; why such data merit special techniques for analysis and exploration. While most researchers agree that they have come across spatial data, many seem not to treat such data "spatially," either out of apprehension of the associated complexity, or failing to appreciate the need and benefits of doing so.

2. Learn the theoretical basics of spatial processes, and the way they are used in the estimation of important parameters and the testing of relevant hypotheses. Learn about the different types of spatial data and why they require different approaches.

3. Learn how to implement solutions and fit models in a classical as well as a Bayesian setup. Learn the functions in `geoR` and `spBayes` for classical analysis, and the `GeoBUGS` tools within the `WinBUGS` language for Bayesian implementation.

4. Understand the application of the tools in a variety of biostatistical settings, such as environmental monitoring, ecological studies, and spatial epidemiology.

III. Methods of Instruction and Work Expectations

Classroom experience will be primarily through traditional lecture, but will also include hands-on computer lab experience (using the `geoR`, `spBayes`, and `GeoBUGS` packages) in the SPH Computer Lab, Mayo C381. In addition, a permanent, named account on the Biostatistics Division UNIX workstation network will be issued to each biostatistics student registered for this class who does not already have one. Registered students from other divisions or departments will be issued temporary, class accounts, which will be closed one month following the end of the semester. The network offers `C`, `C++`, `Fortran`, and `Classic BUGS` for implementing numerically-intensive algorithms, `S-plus` for graphics and smaller programming jobs, LaTeX for numerical word processing, and several web browsers, email packages, and text editors. Several workstations and graphics X-terminals are specifically reserved for student use; these are located in Mayo A316.

Regarding work expectations, students are expected to attend the lecture and computer lab sessions, and complete all assigned homework, the midterm exam, and the final project.

IV. Evaluation and Grading

The grade for this course will be determined by several homework assignments, an in-class midterm exam, and a final project. The problems will include theoretical and applied questions, mostly from the text. Assignments will be given out as appropriate throughout the semester, and will generally be due a week after they are assigned. Students should try to do their own work on these problems; the TA and I are available for questions, of course. The midterm will be closed book and focus on theoretical and methodological issues. The final project involves preparing a short paper and giving a brief classroom presentation on some subtopic of interest to you. The attached pages (along with the reading and lecture material) should provide a variety of ideas for projects.

Grading Scheme:	Homework	40%
	Midterm exam	25%
	Final project	35%
		100%

For data analysis problems, your writeup must be a careful report of your models, methods, interpretations, and conclusions – as if you were making a final report to a supervisor who has statistical training, but doesn't want to get bogged down in the details. Include the relevant parts of your computer output as a technical appendix, or "cut and paste" them into your report, labeling all plots, variables, and so forth.

I take a very dim view of unexcused late assignments, especially in a class like this where most of the work is "take-home." As a general rule, *prior notification is essential* to my accepting a late paper of any kind. If illness or travel is going to cause you to miss a deadline, don't surprise me – call or send an e-mail message (I check my voice- and e-mail every day).

Grade changing option: Students may change grading options without written permission as specified by the University and without penalty during the initial registration period or during the first two weeks of the semester. *The grading option may not be changed after the second week of the term.*

Course withdrawal policy: Students may withdraw from a course through the second week of the semester without permission. After the second week, students will be required to obtain permission from their advisor and instructor (via email to the Student Services Center) and *a "W" will remain on their transcript*.

A/B/C and S/N cut-offs: Grading is "on the curve," so no absolute percentages will be applied. For S/N grading, work comparable or better to that required to obtain a "C" is necessary to receive an "S".

Incomplete Policy: An "I" is assigned at the discretion of the instructor when, due to extraordinary circumstances (e.g. hospitalization) a student is prevented from completing the work of the course on time. A written agreement between the instructor and the student for completing the work is required, and extension for completion shall not exceed one year, or else the "I" converts to an "F" or "N".

Scholastic Dishonesty: This is defined as "any act that violates the rights of another student in academic work, or that involves misrepresentation of your own work." This includes (but is not limited to) cheating, plagiarizing, depriving another student of necessary course materials, or interfering with another student's work. Scholastic dishonesty in any portion of the academic work for this course shall be grounds for awarding a grade of "F" or "N" for the entire course. For further information please consult the student conduct code at `www1.umn.edu/regents/policies/academic/StudentConduct.html`.

V. Course Text and Readings

Required: *Hierarchical Modeling and Analysis for Spatial Data*,
by Sudipto Banerjee, Bradley P. Carlin, and Alan E. Gelfand,
Chapman and Hall/CRC Press, 2004.
`WinBUGS` Users Manual, V.1.4.1, by D.J. Spiegelhalter et al.,
freely available from `www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml`

Helpful: Your favorite math stat and linear models books
Your favorite GIS reference manual
Statistics for Spatial Data, rev. ed., by Noel Cressie,
New York: John Wiley and Sons, 1993.
Bayes and Empirical Bayes Methods for Data Analysis, 2nd ed.,
by Bradley P. Carlin and Thomas A. Louis, Boca Raton, FL:
Chapman and Hall/CRC Press, 2000.

VI. Weekly Schedule (BCG = Banerjee, Carlin and Gelfand text; number gives chapter)

Week of	Topics	Text
Tue 1/22	Introduction, types of spatial data, fundamentals of cartography	BCG 1
Tue 1/29 Tue 2/5	Point-referenced data (geostatistics) basics: stationarity, variograms, EDA methods, classical estimation and prediction, kriging	BCG 2
Tue 2/12	Areal data (lattice) basics: Markov random fields, CAR/SAR models	BCG 3
Tue 2/19	Basics of Bayesian inference: point and interval estimation, testing, MCMC computation, basic modeling in `WinBUGS`	BCG 4
Tue 2/26 Tue 3/4	Hierarchical modeling for univariate data: spatial process models, Bayesian kriging, areal data models, disease mapping	BCG 5
	Note: **MIDTERM EXAM: THURS 3/6**	
Tue 3/11	Spatial misalignment: point and block level modeling	BCG 6
Tue 3/18	SPRING BREAK – no school!	
Tue 3/25 Tue 4/1	Multivariate spatial modeling: separable point level models, coregionalization models, areal data (MCAR) models	BCG 7
Tue 4/8 Tue 4/15	Spatio-temporal modeling: point and block level, with and without misalignment	BCG 8
Tue 4/22 Tue 4/29	Special methods: spatial survival (frailty and cure rate) models, spatially varying coefficients, spatial CDFs, wombling	BCG 9-10
Tue 5/6	Student project presentations (may carry over into finals week, depending on class size)	——

Final Project writeups due: **Thurs 5/15, 5:00 pm**

Disability Accommodation: It is University policy to provide, on a flexible and individualized basis, reasonable accommodations to students who have documented disability conditions (e.g., physical, learning, psychiatric, vision, hearing, or systemic) that affect their ability to participate in course activities or to meet course requirements. Students with disabilities are encouraged to contact Disability Services for a confidential discussion of their individual needs. Disability Services is located in Suite 180 McNamara Alumni Center, 200 Oak Street. For further information contact the University of Minnesota Disability Services website, `disserv3.stu.umn.edu/index2.html`, or call (612) 626-1333 (V/TTY).

VII. Final Project Information

The final project involves preparing a short (5-10 page) paper and giving a brief (15-20 minute) classroom presentation on some subtopic of interest to you. Once you have identified a topic of interest (see list below for suggestions), I will probably suggest a few papers for you to read as a starting point. This should in turn suggest several interesting project possibilities: extending an analytical result, simulating the performance of some procedure, undertaking a challenging data analysis, etc.

Project areas and presentation times will be allocated on a "first come, first served" basis, so it is best to let me know your selection early!

SIGN UP DEADLINE: Thursday April 10.

Topic Areas:

1. Methodology
   - Geostatistics fundamentals: variogram, semivariogram
   - Kriging (universal, indicator, block, etc.)
   - Lattice fundamentals: MRFs, CAR vs SAR, properties
   - Point process fundamentals: nearest neighbor methods, Ripley's K
   - Computing: Classical versus Bayes algorithms
   - GIS interfaces: Current statistical capabilities, combining with other statistical software
   - Statistical packages: `S+SpatialStats`, `GeoBUGS`, calling either from `R`
   - Spatial misalignment, modifiable areal unit problem (MAUP)
   - Confounding in aggregated data studies; the "ecological fallacy"
   - Spatio-temporal modeling
   - Multivariate modeling (e.g. MCAR)
   - spatial boundary analysis (wombling)
   - *Some other area of interest to you!*
2. Modeling and Applications
   - Disease mapping / cancer surveillance
   - Toxic exposure and/or related health outcome analysis
   - Forestry / environmental applications
   - Urban planning / Real estate valuation
   - Spatial longitudinal data analysis
   - Spatial survival analysis: frailty, cure rate models
   - Spatial factor analysis
   - *A geographically oriented model or application arising in your own work!*

Class presentation hints:

1. Time

   - Since you have only a few minutes to speak, think carefully about what to present and what *not* to present. An full slide typically takes at least 2 minutes to cover (assuming no one asks a question!).
   - Practice actually giving your talk at least once before your in-class presentation, to check its length and possibly uncover bugs in the material or your delivery.

2. Overheads

   - Use large type (In LaTeX, `\huge` or `\Huge` works best).
   - Avoid displaying huge tables with lots of digits; draw a figure instead if at all possible.
   - Don't put too much on each slide. That is, don't use them as notes for what to say; put your notes on a separate sheet.
   - Organize them so only seldom do you need to hide anything. If you do need to hide a portion of the slide until later, put the paper *underneath* the transparency so it won't fall off. After putting up each slide, check to make sure it is properly projected on the screen.

3. You

   - People can't see through your body; get out of the way.
   - Use a long pointer or your hand to point at the screen directly, or lay a pointer down on the transparency and step aside. If you are pointing at the transparency, touch your pen to it; small motions are greatly magnified on the screen.
   - Don't wave your hand over the projector – the magnification of motion will make the audience seasick!