Stat 5810 Applied Spatial Statistics	Fall Semester 2000

Final (Wednesday 12/6/2000 — 5pm–7pm)

100 Points

Work on the following questions using S+SpatialStats and ArcView/XGobi. Put your name on any document you produce, i.e., in any textfile or Web page you create and any sheet of paper you use. Also, put your initials electronically in any graphic you produce (e.g., as part of the title).

As in the homework assignments, your solutions should consist of the S–Plus commands you have typed, any S–Plus functions you have written, and the output (text and graphics) produced in S–Plus or ArcView/XGobi.

Only those documents that are printed and turned in to me at the end of the exam will count towards your grade. Files (text or graphics) that are not printed cannot be turned in at a later time (nor will you be allowed to recreate text or graphics that are unintenionally being deleted during the exam).

If you print something during the exam, do not go to the printer. I will distribute all printouts to the corresponding owner.

When you explain your answers, always use the appropriate vocabulary in your answers to obtain full credit.

Start with the questions that are easiest for you. However, try to do at least some work on each question to obtain some partial credit.

You can use all your books, notes, copies, and the "public" Web. You are not allowed to access the files or "private" Web pages from anybody else in this class. Also, you are not allowed to communicate verbally or electronically with anybody else besides me.

Good luck!

P.S.: If you want, you can pick your graded final and the solutions during my office hours next semester.

(i) (30 Points) In Section 2.3.3, we introduced the nearest neighbor event–event distance W and the empirical cdf $\hat{G}(w)$ of W as

$$\hat{G}(w) = \frac{\#(w_i \leq w)}{n}.$$

This version of $\hat{G}(w)$ does not contain any edge correction.

(a) (15 Points) Bailey/Gatrell, page 102, indicate the following possibility to correct for edge effects:

$$\hat{G}(w) = \frac{\#(b_i > w \geq w_i)}{\#(b_i > w)},$$

where b_i is the distance from event i to the nearest point on the boundary of $\mathcal{R}$.

Implement this edge corrected version of $\hat{G}(w)$ in S–Plus. You may assume that $\mathcal{R}$ has a rectangular boundary (but not necessarily the unit square). Some initial sketches might help in determining which distances to calculate. You do not have to optimize your function with respect to speed.

(b) (5 Points) Simulate a CSR process that consists of 100 points in the rectangular region with lower left corner $(0, 0)$ and upper right corner $(2, 1)$ and plot these point locations.

(c) (5 Points) Test your function developed in (a) with the data created in (b). Plot your $\hat{G}(w)$ [with edge correction], the original $\hat{G}(w)$ [already available in S+SpatialStats], and the theoretical distribution function $G(w)$ in one plot against w (where w varies in the interval 0 to 0.5 in 20 steps of width 0.025). Explicitly state which value you use for λ in the theoretical $G(w)$.

(d) (5 Points) Comment on your results obtained in (c).

(ii) (25 Points) In this and the next question you have to work with the following data set:

`/home/symanzik/splus/final/final.txt`

Copy the data set into your directory and make the required modifications to read this data set into S–Plus.

In this question, we assume that this data set represents a spatial point pattern, with spatial locations s_i represented by easting and northing.

(a) (15 Points) Plot the spatial point pattern and create the following plots in order to assess whether the point pattern might represent a homogeneous Poisson process: an intensity (image) plot, a contour plot, and a plot of Lhat with corresponding simulation envelopes (based on 40 simulation runs).

(b) (10 Points) Precisely explain why the plots from (a) suggest that the data is or is not from a CSR process.

If the point pattern indeed represents a CSR process, you are done. Otherwise, indicate in which way the pattern departs from CSR and suggest a possible alternative model (you do not formally have to verify that your suggested model indeed fully describes the given point pattern).

(iii) (25 Points) In this question, we use the same data set as in question (ii). However, in this question we are interested in the variable XXXXX measured at each point location. We assume that this is a spatially continuous data set now.

(a) (12 Points) First use ArcView/XGobi to determine whether there are any global outliers or any spatial outliers in this data set. Also, do some preliminary assessment using ArcView/XGobi whether there is some spatial dependence in this data. Use a cutoff distance of 2000. To support your answers, you are allowed to produce at most 6 graphics. Comment your results.

The .apr file for this question is available at

`/home/symanzik/splus/final/final.apr`

(b) (13 Points) Using S–Plus, verify your answer from (a) about the spatial dependence, i.e., calculate a robust variogram, fit a theoretical variogram model, and clearly indicate the values for the range, sill, and nugget effect (if any). You may assume that the data already meets the necessary preconditions, i.e., that there exists no large scale trend in the data. Comment your results.

(iv) (20 Points) We already discussed this example from Mark Monmonier's book "How to Lie with Maps (Second Edition)", University of Chicago Press, Chicago, IL, 1996, in class. The four choropleth maps on the next page show the effect of selecting different class widths and starting points when visualizing data in a geographic context using choropleth maps.

(a) (5 Points) This first part aims at the potential misuse of standard choropleth maps. Answer the following parts, assuming that someone wants to support a particular (political) opinion. Shortly explain why you would select a particular choropleth map in each situation.

i. If you were the governor of New York, which graphic would you use to demonstrate how advanced your state is?

ii. If you were the governor of Connecticut, which graphic would you use to demonstrate how advanced your state is?

iii. If you were the governor of New Jersey, which graphic would you use to demonstrate how advanced your state is?

iv. If you were the governor of Virginia, which graphic would you use to demonstrate how far behind your state is and desperately needs federal funding?

v. If you were a historician who wants to show how few telephones were around in 1960, which graphic would you use?

(b) (3 Points) Now find the "true" interval for the following 5 states. Do this by calculating the intersection of the class intervals used within the 4 maps below, e.g., if we know that a value is ≤ 25 and ≥ 20, it obviously falls in the interval $[20, 25]$:

Maine

New York

Connecticut

New Jersey

Virginia

(c) (5 Points) Explain, by drawing a crude sketch, how a micromap display of the same data might look like. You can assume that the "exact" value (and not only the approximate interval) is available.

(d) (7 Points) Assuming that the data itself is not manipulated, can you manipulate micromaps as easily as choropleth maps to express different political opinions? Justify your answer.