Stat 605 – Spatial Statistics – Spring 2022
Homework 6. Due: Friday, February 25, midnight.

This homework assignment is about analyzing the Wolfcamp aquifer data. This data set is provided with the geoR package.

1. Load the data into memory, `data(wolfcamp)`; add a border to the data set, plot it, discuss briefly. What are these data?

2. Act as a frequentist; fit a spatial model to these data, and obtain a smoothed map. Did you include a trend term or terms? Did you need to log-transform the data? Explain your reasoning. It might be helpful to show results from fitting one or two multiple linear regression models.

   Be sure to discuss how you selected a type of semivariogram and how you estimated its parameters. Be sure to state your final fitted model. Discuss briefly.

3. Fit a Bayesian model to these data by modifying the R code, `Bayesian_kriging.r`, which is posted on Canvas. Be sure to include diagnostic plots shown in class (traceplots and density plots for the model parameters) and a final smoothed map.

4. State the statistical model you are using. This means: State the likelihood (e.g.: $Y_i = \beta_0 + \beta_1 X_{i,1} + \cdots + \epsilon_i$ where $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \Sigma)$) and state the prior distribution for the various parameters in the model (e.g. for the β or β's, τ^2, σ^2,...) Page 118 of the lecture notes shows an example of what I mean by this. Also, be sure to state the posterior distribution, for example, as shown on page 119 of the class notes.

5. Summarize the (marginal) posterior distributions of the model parameters in a table; one row for each parameter and the following columns for each:

   - posterior mean
   - posterior standard deviation
   - posterior median
   - 95% credible set
   - Monte Carlo error
   - Effective sample size

   Please include not too many digits and not too few. (Generally speaking, 3 digits seems about right.) Do not simply copy-and-paste R output; I want you to convince me that you've looked at the numbers, however briefly.

6. Provide a summary of your results; specifically,

   - How many iterations did you use for your MCMC, and how much thinning (if any) did you do? What about burn-in?
   - Discuss the convergence of the MCMC for the Bayesian model. Do the various plots (trace plots) indicate any issues with convergence? What do the Monte Carlo errors suggest about convergence or lack of convergence?

7. Discuss (briefly) whether your results for the frequentist and Bayesian approaches are similar; are there any major differences? (A table plus discussion is the best way to do this.)