# Day 8, PM Lab
# Intro to Spatial Statistics
# Bayesian Geostatistics

## Overview

In this lab, we are going to 1.) simulate geostatistical data, 2.) examine spatial dependence in errors, 3.) fit a Bayesian geostatistical model to simulated data using a.) JAGS and b.) spBayes, 4.) obtain spatial predictions (using spBayes), 5.) compare spatial predictions to "true" simulated process, 6.) create Bayesian model for Iowa temperature data and obtain predictions in a rectangular region encompassing the data.

## Objectives

The objectives of this lab are to:

1. Gain familiarity with spatial processes and models.
2. Obtain experience simulating Gaussian spatial processes.
3. Learn how to manipulate spatial data in R.
4. Use variograms to assess spatial structure in data and/or residuals.
5. See how to fit Bayesian geostatistical models and obtain spatial predictions.

## Procedure

1. Open the MCMC algorithm file called 'lab_8_script.R' in a text editor (or R). Examine the different sections of this script which are denoted by the comments '####'.

2. In R, change directory to the lab folder.

3. In this lab, we are first going to

4. Following the R script:

   (a) Load packages: geoR, maps, mvtnorm, rjags, spBayes, rgdal, and maptools. Download them first if you don't have them installed.

   (b) Simulate data based on a continuous spatial process encompassing the state of Utah:

      i. Convert Utah map to UTM coordinates.
      ii. Find the boundaries of Utah.
      iii. Specify points locations that span the spatial domain.
      iv. Make a map to check these locations. Note that even though these are truly point locations, we are going to visualize them in a regular grid for convenience. They are NOT areal data in this case, as a different simulation and modeling procedure would be used in that case.
      v. Create Euclidean distance matrix that specifies the distance between all points in full domain.
      vi. Specify an exponential covariance model: $\Sigma_{i,j} = \sigma^2 \exp(-d_{i,j}\phi)$. Note that in this specification the range parameter $\phi$ is on top in the exponential (preferred by the spBayes package).
      vii. Simulate the spatially correlated error process from a multivariate normal: $\boldsymbol{\varepsilon} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma})$.
      viii. Specify how much actual "data" you want keep. Here we take a subset (perhaps randomly) of the simulated process and act like we don't have the rest, that way we can assess our predictive ability at those "missing" locations.
      ix. Make map of these correlated errors.

   (c) Create an empirical semi-variogram to check the spatial structure evident in the correlated errors and compare with "known" simulated semi-variogram.

   (d) Create a couple of spatially-structure covariates. Here we just make up some interesting functions to construct $\mathbf{X}$.

   (e) Choose a known set of values for the regression coefficients $\boldsymbol{\beta}$.

   (f) Create observed data vector $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$.

   (g) Make figure containing maps of $\mathbf{y}$, $\boldsymbol{\epsilon}$, and the covariates.

   (h) Use JAGS to fit model:

      i. Set up variables so that we can use JAGS to fit geostatistical Bayesian model.
      ii. View the JAGS model specification 'lab_8_jags_model.txt' file to see how this spatial model might be handled.
      iii. Fit model using JAGS. Note that we only use 1000 iterations to get a feel for how this model is fit. JAGS is too slow for this problem to obtain a large number of MCMC samples and it is WAY too slow to produce spatial predictions.
      iv. View MCMC trace plots based on JAGS output.

(i) Use spBayes to fit model

  i. Set up variables so that we can use spBayes to fit geostatistical Bayesian model.
  ii. Use 'spLM' to first fit the model.
  iii. Use 'spRecover' to obtain MCMC samples for $\boldsymbol{\beta}$ so that we can assess trace plots.
  iv. Check trace plots of model parameters for convergence.
  v. Use 'spPredict' to obtain predictions and prediction standard deviations at all locations in the spatial domain. Note that these quantities represent the mean and standard deviation of the spatial posterior predictive distribution.
  vi. Make figure that shows maps of the 'true' process, data, predictions, and the prediction standard deviation.

# Challenge

1. Read in the Iowa temperature data in file 'iowa_temperature.txt'.

2. Create map of these data using circles with size corresponding to temperature with Iowa state boundary overlaid.

3. Using the spatial coordinates as covariates, create a Bayesian predictive surface map over the region encompassing the data. Also construct a map illustrating the uncertainty in predictions.