Lab 3. Kriging

Math 586. Due April 10

From `http://infohost.nmt.edu/~olegm/586/Matlab/`, download the files `forLab3.m, krige.m, /data/Sev-cov.csv, /data/janNNM.csv`.

Kriging weights and dependence on variogram

We will continue exploring NNM precipitation (compare Lecture 10, pp.4-7). In particular, how does the variogram model affect the results?
In Lecture 10, we used the exponential model with variance ("sill") $\sigma^2 = 0.1144$ and scale $\ell = 10.445$. What happens when we vary σ^2 and ℓ?

In the code file `forLab3.m` the matrix **C** is computed row by row according to the distances and the covariance function as:

```
dist = zeros(n,n);   % Euclidean distances between points
C = zeros(n,n);
for i = 1:n,
  dist(:,i) = sqrt( (x(i) - x).^2 + (y(i) - y).^2);
  C(:,i) = sill*exp( -dist(:,i)/scale);
end;
```

Just before that, the values for `sill` and `scale` are assigned.

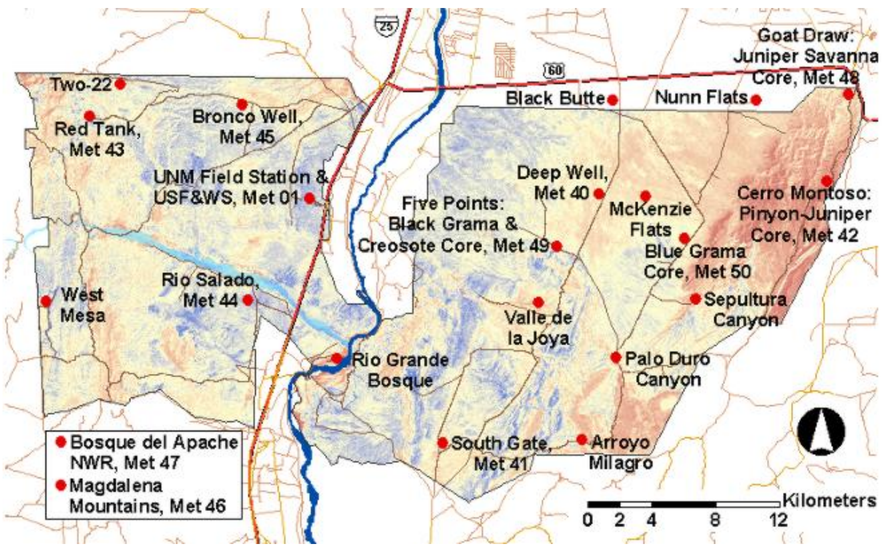
Change the value of the scale to 50, re-compute the kriging map and kriging weights for the point $(470, 4010)$ - example used in Lecture 10, p.6.
⋆ *Question:* In what way did these change, compared to the Lecture 10 output?

Now, return the scale to 10.445, and change the value of sill to 0.4.
⋆ *Question:* How did the kriging map and kriging weights change?

Now, return the sill value to 0.1144 and modify the code (including the function `krige.m`) to include the nugget of size 0.05.
⋆ *Question:* What is the effect of introducing this nugget value?

Ordinary Kriging (OK) and estimation of the mean: Spatial averaging

Prediction at "far away" locations should not depend on the data, except through the mean. (Can you see the forest for the trees?) To illustrate this point, consider the Sevilleta temperature data. [We have said that the kriging weights do not depend on the data. They do, indirectly, through the estimated variogram or covariance.]



If we were to define the average temperature of the area, which stations will matter more?

We will not fit the variogram, but use the matrix $\mathbf{C}$ = covariances obtained from the data for the year 2003 for the above 10 stations. Look at the data in `Sev-cov.csv`. There's a great deal of correlation!

The stations contributing to matrix $\mathbf{C}$ are listed in their number order: 1, 40, ..., 50. Stations Met47 and Met46 are not in the region and are thus omitted.

Station Met41 (third row of $\mathbf{C}$) is somewhat less correlated with the others, because it's farther apart.

Now, calculate the OK weights for a faraway point:

```
lastrow = [ones(1,n) 0];
Caug = [C  -ones(n,1); lastrow];       % building the OK matrix
Cinv = Caug^(-1);
b = zeros(n,1);                         % this simulates the "faraway point"
baug = [b ; 1];
```

```
lamOK = Cinv * baug
```

The results are very unexpected and challenge our common-sense perception of "the mean"!

Accidentally, the last entry in `lamOK` is the Lagrange multiplier μ, here it has the meaning of uncertainty (variance) in estimating the mean. Compare this with the value of $C[1, 1]$: we might as well use Station 1 and discard all the rest!

Exercises:

1. Answer the three questions in the text.

2. (Based on log precipitation Northern NM model)

- (a) Compare the kriging st.dev. maps for the original case in Lecture 10 (with scale = 10.445), and when the scale parameter changes to 50. Explain what you see.
- (b) Change the scale parameter to 5 and plot all the weights from both SK and OK (use `plot(lamSK)` and `hold on` commands). What are the main differences between SK and OK weights?
- (c) What is the uncertainty in estimating the mean m for ordinary kriging (original parameters)?

3. For the Sevilleta example:

- (a) Compute the correlation coefficient between first and second stations on the Sevilleta list (based on the matrix $\mathbf{C}$).
- (b) Verify that OK weights add up to 1.