Patrick Liu

CS&SS554

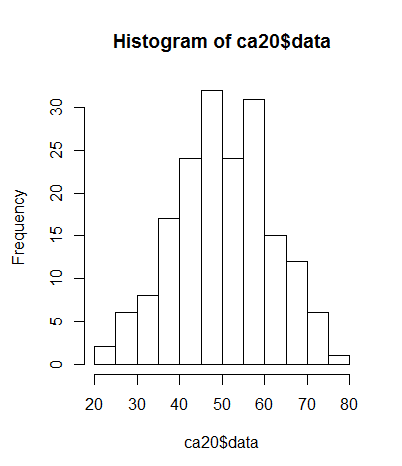
Homework 5

3/05/2017

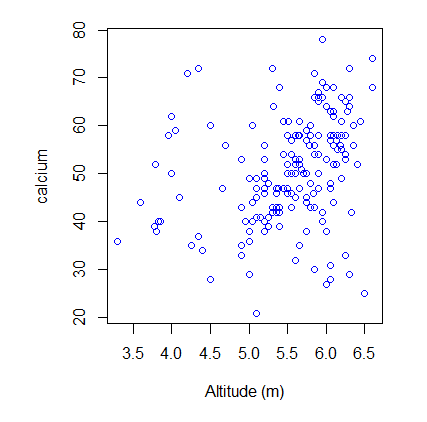
We are exploring a dataset of the calcium content measured in soil samples (mmolc/dm3) taken from the 0-20 cm layer in 178 locations encompassed in our study area. The whole study area is further divided into 3 subregions: (1) an area that is typically flooded during the rain season—these represent natural content of calcium in the region, (2) an area that had received fertilisers a while ago, and (3) an area that recently received fertilisers. Included in this dataset are the calcium levels, the altitude at each point, and a factor variable for the sub area a point belongs to.

We plan to model the data using two approaches, through the variogram and through an ordinary kriging model. First, we prepare some exploratory plots to better understand the distribution of the data, and how it relates to our covariates.

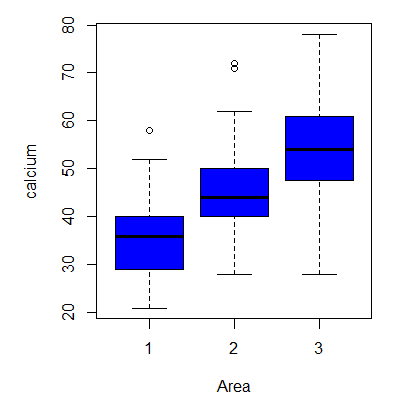
The distribution of calcium looks relatively symmetrically distributed from a concentration of 20-80 mmol/dm3.



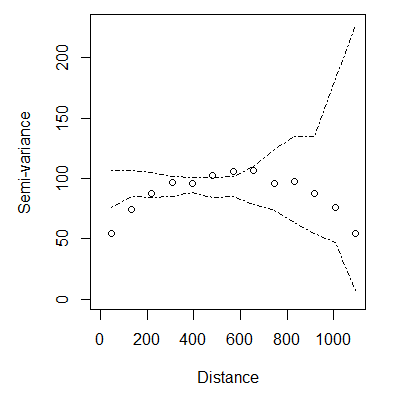
Concentration of calcium seems to increase with altitutde.



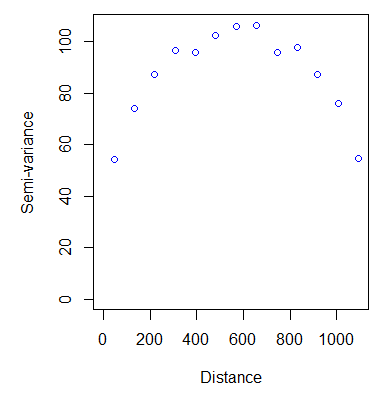
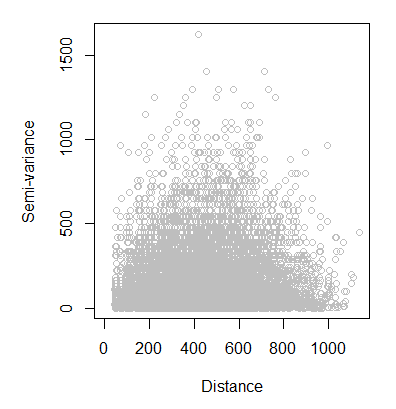
The Mean concentration of calcium is increasing over the categories of designated sub areas.



We fit and plot Monte Carlo intervals of no spatial dependence, which shows that there is clear spacial dependence in the data.



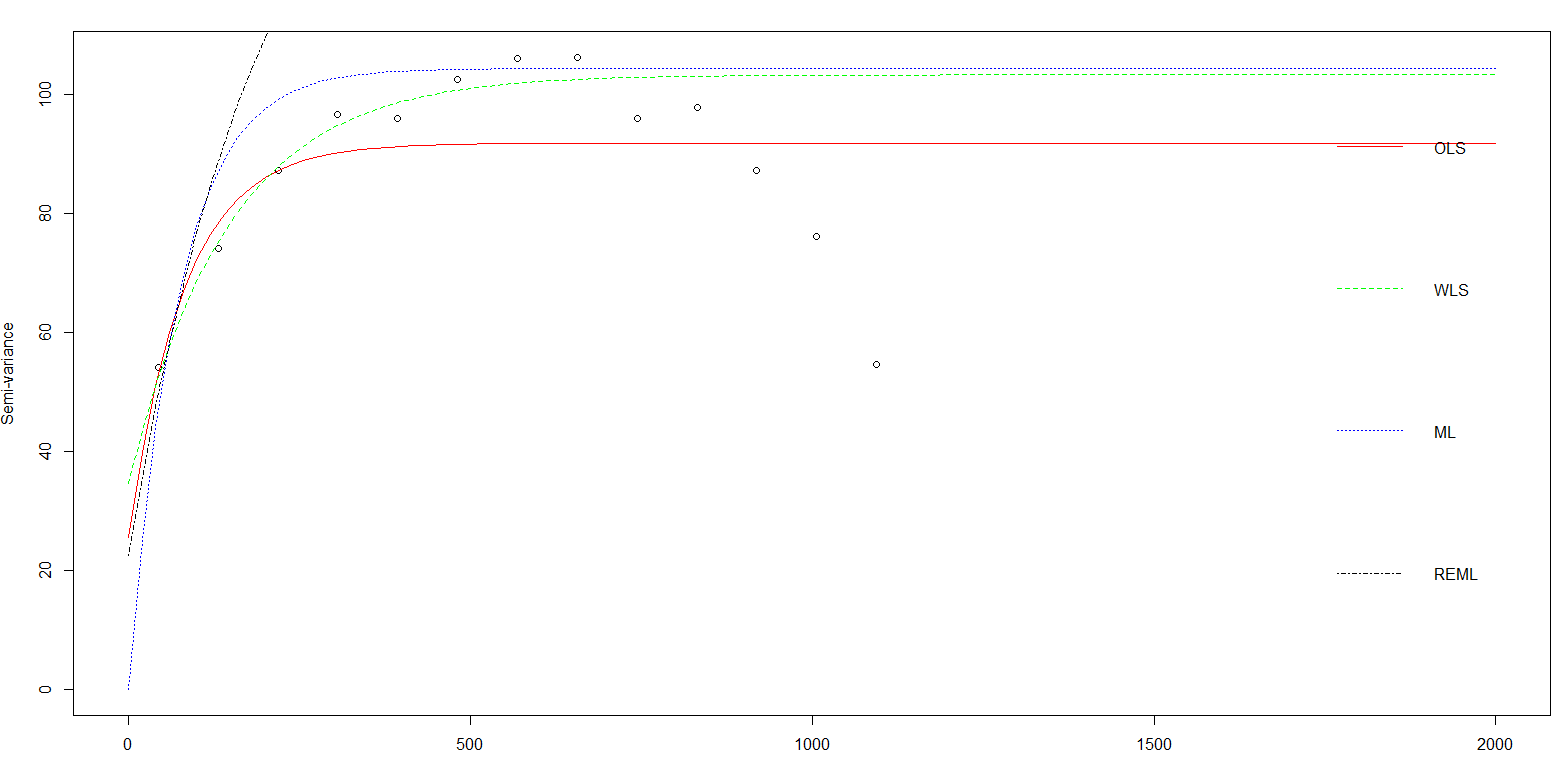
We first attempt to model the data using the variogram approach. We first get a sense of the emperical variogram by plotting the cloud and binned variograms.



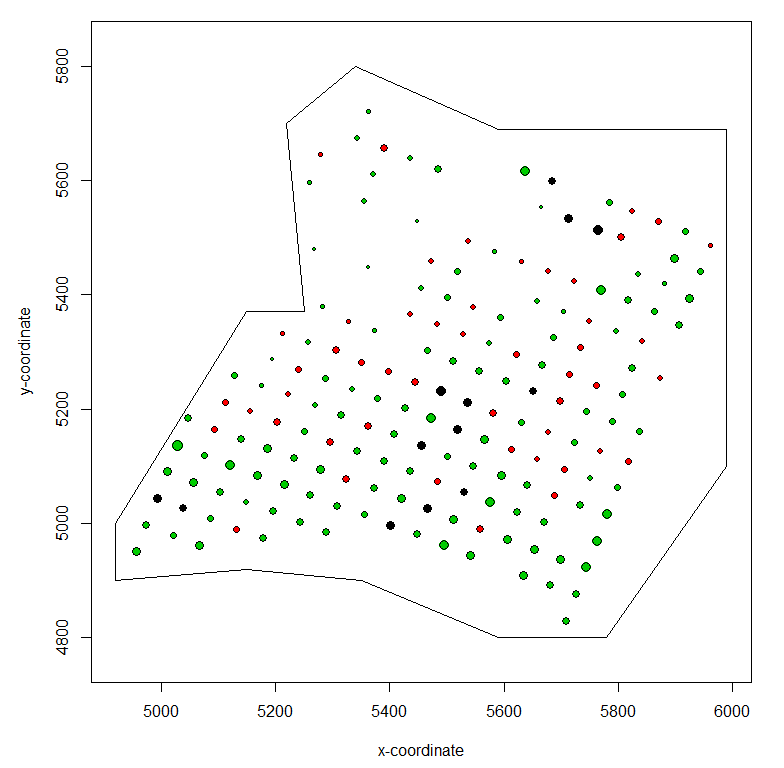
We next fit a series of models (1) OLS, (2) Weighted LS, (3) ML, (4) Restricted ML to estimate the variogram.

|  |  |  |  |
| --- | --- | --- | --- |
| Model |  |  |  |
| OLS | 25.5 | 66.3 | 81.9 |
| WLS | 34.7 | 68.8 | 147.0 |
| ML | 0.00 | 104.4 | 73.0 |
| REML | 22.6 | 137.5 | 199.99 |

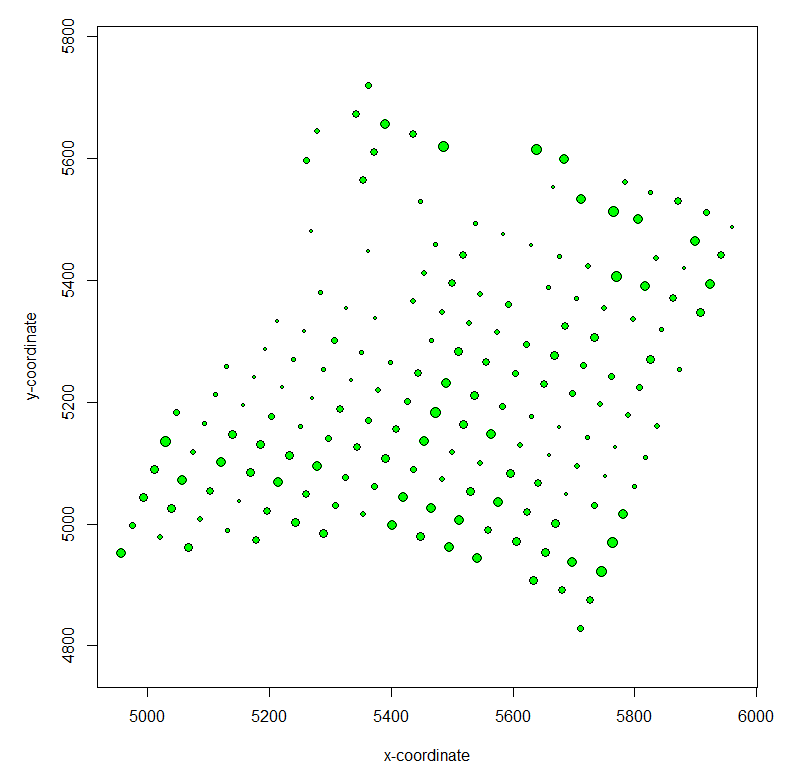
Plotting the four models against the empirical variogram. Weighted least squares seems to fit the data the best, however none of the four models, because of the nature of the variogram approach is able to capture the dip at the end.



We plot a spatial map of the points in the datset, each color green, red, and black, corresponds to areas 1, 2, and 3, respectively. Surprised to see that the “areas” aren’t distinct and points form each area are scattered throughout the study region.



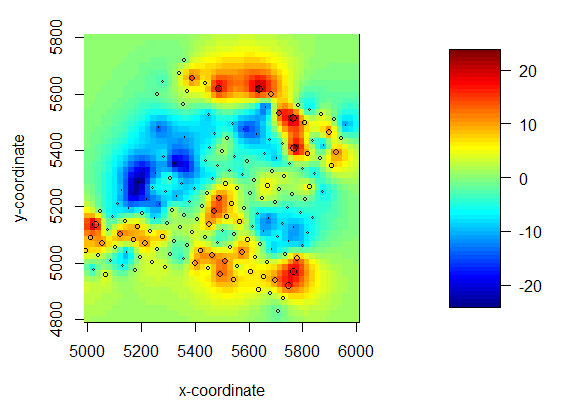
We next carry out a kriging model to estimate the surface of calcium in the study region. We first plot and examine the detrended points (removing altitutde and area trends).



We first fit a ML omdel on the detrended data.

|  |  |  |  |
| --- | --- | --- | --- |
| Model |  |  |  |
| ML | 0.00 | 101.6 | 69.5 |

Using this, we fit an ordinary kriging model to obtain spatial predictions of calcium content in the soil. We plot the predictions of the detrended concentrations with the data superimposed. We see higher concentrations in the most northern and southern regions and areas of low concentrations in the middle belt, particularly on the west-most region.



We plot the standard deviation of the estimates which, as expected, is smallest close to the datapoints.