Politecnico di Milano
Scuola di Ingegneria Industriale e dell'Informazione
Applied Statistics
Exam 2024-07-08 - Part B - 2023/2024

Problem 3: Spatial Modelling of uranium concentration

A study is conducted to analyze the spatial distribution of uranium concentration in the Tim Mersoi Basin in northern Niger. The dataset `uranium.txt` reports the uranium concentration at various locations within the study area. It includes the UTM geographical coordinates s of the sampling locations, a categorical variable `rock_type` indicating the type of rock (granite, sandstone, or shale), the depth of the sample `depth` (in meters), and the measured uranium concentration $y(s)$ [ppm]. Consider the following model:

$$y(s) = b_{0,j} + b_1 \texttt{depth} + \delta(s)$$

where $\delta(s)$ represents a stationary residual with a spherical variogram with a nugget effect and $j = 0, 1, 2$ is the grouping induced by the variable `rock_type` ($j = 0$ for granite, $j = 1$ for sandstone, $j = 2$ for shale).

a) Report a plot of the fitted variogram, *initialising* the variogram fit with the model `vgm(200000, "Sph", 2000, 100000)`. Indicate the estimate of the range and the sill.

b) Provide an estimate of the mean uranium concentration at the surface of a sandstone rock type area.

c) Independently of the position, by which quantity the uranium concentration increases when the sample is taken 1m lower?

d) Consider a new location with coordinates $s_0 = (687000, 2234000)$, which is in a sandstone rock type area. Which depth must we reach to find an uranium concentration of at least 3000 ppm at that location?

e) Consider now the model update:

$$y(s) = b_{0,j} + b_{1,j} \texttt{depth} + \delta(s) \tag{1}$$

Indicate the estimate of the sill, fitting the variogram with the same initialisation as in a). Should this model be preferred to the first one? Justify your answer

Upload your results here: `https://forms.office.com/e/sPn3UhfHMF`