Politecnico di Milano Scuola di Ingegneria Industriale e dell'Informazione

 $\begin{array}{c} \text{Applied Statistics} \\ \text{Exam 2025-06-12 - Part B - } \\ 2024/2025 \end{array}$

Problem 3: Spatial modelling of tree growth in a forest reserve

A study is conducted to analyze the impact of environmental conditions, seasonal variation, and geographic location on the growth rate of trees in a forest reserve between May and July 2023. The dataset treegrowth.txt reports the daily growth measurements of trees in different parts of the forest. It includes the UTM geographical coordinates s_i of the trees, a categorical variable season indicating if the measurement was taken in early or late summer (early for May–June and late for July), the recorded average daily soil moisture moisture_i (in percentage), and the observed growth rate $y(s_i)$ [mm/day]. Consider the following model:

$$y(s_i) = b_{0,i} + b_1 \text{moisture}_i + \delta(s_i) \tag{1}$$

where $\delta(s_i)$ represents stationary residuals with spherical variogram with nugget, and j = 0, 1 corresponds to the grouping induced by the variable season (j = 0 for early, j = 1 for late).

- a) Report a plot of the fitted variogram, *initialising* the variogram fit with the model vgm(0.1, "Sph", 1000, 0.1). Indicate the estimate of the range and the sill.
- b) Estimate the parameters $b_{0,0}$, $b_{0,1}$ and b_1 using the generalized least squares method.
- c) Using the model, provide an estimate of the total growth expected for a tree during July assuming a constant soil moisture of 25%.
- d) Due to their environment, trees can be categorized as growing in canopy-covered areas (canopy=TRUE) or in open areas (canopy=FALSE). Modify the model in Eq. (1), including this categorical effect, as follows:

$$y(s_i) = b_{0,i,k} + b_{1,k} \texttt{moisture}_i + \delta(s_i)$$

where $\delta(s_i)$ represents stationary residuals with exponential variogram without nugget, and k is the grouping induced by the variable canopy.

Initializing the variogram fit with a sill of 0.1 and a range of 3000, provide an estimate of the parameters $b_{0,0,0}$, $b_{0,0,1}$, $b_{1,0}$ and $b_{1,1}$.

How would you describe the effect of the canopy variable (during early summer)?

Upload your results here: https://forms.cloud.microsoft/e/v5vZu6i39g