Politecnico di Milano
Scuola di Ingegneria Industriale e dell'Informazione
Applied Statistics
July 12th, 2022

Problem n.4

The file `temperatures.txt` reports the measurements of daily maximum temperature y (in Celsius) recorded on June 25th, 2003 and June 25th, 2022 (variable `year` $= 2003, 2022$) within the city of Milan. The dataset also reports the UTM coordinates s_k, $k = 1, ..., 128$, of the measurement locations, and the indication of whether the measurement location is inside or outside a park or a garden (variable `park` $= 1, 0$, resp.). Consider for the temperatures $y_{ij}(s_k)$ the following model

$$y_{ij}(s_k) = \alpha_i + \beta_j + \delta(s_k),$$

with $\delta(s_k)$ a stationary residual with spherical model without nugget, $i = 1, 2$ the grouping induced by the variable `year` ($i = 1$ for 2003, $i = 2$ for 2022), and $j = 1, 2$ the grouping induced by the variable `park` ($j = 1$ for `park` $= 1$, $j = 2$ for `park` $= 0$).

a) Assuming $\alpha_1 \neq \alpha_2$, and $\beta_j = 0$, estimate the parameters α_i, $i = 1, 2$, of the model via generalized least squares. Report the model estimated for $\delta(s_k)$, and discuss the model assumptions.

b) Assuming $\alpha_i = 0$ for $i = 1, 2$, and $\beta_1 \neq \beta_2$, estimate the parameters β_j, $j = 1, 2$, of the model via generalized least squares. Report the model estimated for $\delta(s_k)$, and discuss the model assumptions.

c) How do these models compare with a stationary model? Which model, between those estimated at points (a) and (b), do you deem more appropriate to describe the data? Comment on your choice.

d) Use the model selected at point (c) to provide a point prediction $y^*(s_0)$ for the maximum temperature on June 25th, 2022 at location $s_0 = (513852.78, 5035411.95)$ (within Parco Sempione).

Upload your results here:
`https://forms.office.com/Pages/ResponsePage.aspx?id=K3EXCvNtXUKAjjCd8ope6-9AS0GWf2lHjvGX24HiqFVURThDQTJHNVhHMzJLR0NCUU9QTk9HMlBDUS4u`