Politecnico di Milano
Scuola di Ingegneria Industriale e dell'Informazione
Applied Statistics
February 9th, 2022

Problem n.4

The file `walesharks.txt` collects the number y of sightings of wale sharks during January 2022 at 64 observatory points in the Indian Ocean. The dataset also reports the UTM coordinates s_i of those locations, and the Chlorophyll concentration $x(s_i)$ (mg/m^3) measured at the same locations. Consider for the variable $y(s_i)$, $i = 1, ...64$, the following model

$$log[y(s_i)] = a_0 + a_1 \cdot log[x(s_i)] + \delta(s_i),$$

with $\delta(s_i)$ a stationary residual.

a) Estimate via generalized least squares the parameters a_0, a_1 of the model. Report the model estimated for $\delta(s_i)$, and discuss the model assumptions (use an exponential model without nugget for $\delta(s_i)$).

b) Provide a kriging prediction $log[y^*(s_0)]$ of the log-number of sightings at an observatory point located close to the island of Fenfushi (South Ari Atoll, Maldives) $s_0 = (253844.8, 385997.7)$. For this purpose, use a point prediction of the log-transformed Chlorophyll concentration $log[x(s_0)]$ obtained through a spatially stationary model (use a spherical model without nugget for $log[x(s)]$; report the estimated model, and the point prediction $log[x(s_i)]^*$).

c) Report the kriging variance $\sigma^2(s_0)$ of the point prediction at point (b). Would you deem the variance $\sigma^2(s_0)$ to be fully representative of the uncertainty associated with the prediction $y^*(s_0)$?

Upload your results here:
`https://forms.office.com/Pages/ResponsePage.aspx?id=K3EXCvNtXUKAjjCd8ope6-9AS0GWf2lHjvGX24HiqFVUNldUTTFCNDZMSVBFR0IySTNFSDZUOEpEMy4u`