Politecnico di Milano
Scuola di Ingegneria Industriale e dell'Informazione
Applied Statistics
June 28, 2024

Problem n.2

The file `whalesharks.txt` collects the number y of sightings of whale sharks during January 2024 at 64 observatory points in the Indian Ocean. The dataset also reports the UTM coordinates s_i, of those locations, and the logarithm of the Chlorophyll concentration $log[x(s_i)]$ [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) By using an exponential model without nugget for $\delta(s_i)$ (with initial parameters 4 and $50{\cdot}10^3$ for the sill and range, respectively), estimate via generalized least squares the parameters a_0, a_1 of the model. Briefly detail the implementation choices and procedure, reporting also the relevant R code. Report the model estimated for $\delta(s_i)$.

b) Construct now a spatially stationary model for a point prediction of the log-transformed Chlorophyll concentration $log[x(s_i)]$. Use a spherical model without nugget for the residual (with initial parameters 4 and $50{\cdot}10^3$ for the sill and range, respectively). Report the estimated model for the residual, and the point prediction $log[x(s_0)]$) at an observatory point $s_0 = (253844.8, 385997.7)$ located close to the island of Fenfushi (South Ari Atoll, Maldives).

c) Employing result in point b), provide a kriging prediction $log[y^*(s_0)]$ of the log-number of sightings at $s_0 = (253844.8, 385997.7)$, together with the kriging variance $\sigma^2(s_0)$ at this point prediction.

d) Would you deem the variance $\sigma^2(s_0)$ to be fully representative of the uncertainty associated with the prediction at point (c)?

Upload your solution **https://forms.office.com/e/3pe6NtjDNB**