Universidad Adolfo Ibáñez
Facultad de Ciencias

MB/mb

Prueba N°2

1. El archivo raindata.txt contiene información sobre la cantidad de lluvia (tercera columna) mensual en 500 localizaciones espaciales (primera y segunda columnas) en una cierta región de España en el mes de febrero 2010. El archivo alt.txt contiene contiene las alturas asociadas a la localizaciónes espaciales (secunda columna).

   Para analizar la cantidad de lluvia (Y) se considera el siguiente modelo (campo aleatorio) multiplicativo

   $$Y(\mathbf{s}) = \mu(\mathbf{s})U(\mathbf{s}), \quad \mathbf{s} \in \mathbb{R}^2 \tag{1}$$

   donde U es un campo aleatorio positivo de media 1, es decir $E(U(\mathbf{s})) = 1$, y la media $E(Y(\mathbf{s})) = \mu(\mathbf{s})$ se especifica con un modelo de regresión $\mu(\boldsymbol{s}) = e^{X(\boldsymbol{s})^T\boldsymbol{\beta}}$ donde $X(\boldsymbol{s}) = (1, X_1(\boldsymbol{s}))^T$ y $\boldsymbol{\beta} = (\beta_0, \beta_1)^T$ son los parámetros de regresión. Acá $X_1(\boldsymbol{s})$ es la altura asociada a la localización espacial $\boldsymbol{s}$

   Para el campo aleatorio U se especifica los siguiente dos modelos:

   a)
   $$U(\boldsymbol{s}) = \nu(\kappa)\left(\frac{1}{2}\sum_{k=1}^{2} Z_k(\boldsymbol{s})^2\right)^{1/\kappa}, \tag{2}$$

   donde Z_1, Z_2 son dos campos aleatorios independiente Gaussianos estándar con correlación $\rho(\boldsymbol{h})$ y donde $\nu(\kappa) = \Gamma^{-1}(1 + 1/\kappa)$ y $\kappa > 0$ es un parámetro de forma.
   La distribución marginal de U es $U(\boldsymbol{s}) \sim Weibull(\kappa, \nu(\kappa))$.

   b)
   $$U(\boldsymbol{s}) = exp\{\sigma Z(\boldsymbol{s}) - \sigma^2/2\} \tag{3}$$

   donde Z es un campo aleatorio Gaussiano estándar con correlación $\rho(\boldsymbol{h})$ y $\sigma > 0$ es un parámetro de forma.
   La distribución marginal de U es $U(\boldsymbol{s}) \sim LogGaussian(1, \sigma^2)$.

   En ambos casos se especifica un modelo de Matern para la función de correlación del campo aleatorio Gaussiano latente, es decir

   $$\rho(\boldsymbol{h}) = \frac{2^{1-\nu}}{\Gamma(\nu)}\left(\frac{||\boldsymbol{h}||}{\alpha}\right)^{\nu}\mathcal{K}_\nu\left(\frac{||\boldsymbol{h}||}{\alpha}\right), \qquad ||\boldsymbol{h}|| \geq 0.$$

   Utilizando el software R conteste a las siguientes preguntas:

   a) Realizar un análisis exploratorio de los datos. ¿Por qué el modelo Gaussiano no es un modelo adecuado?.
   b) Explore la dependencia en los datos graficando el semivariograma isotrópico y el semivariograma en diferentes direcciones. ¿Es razonable la hipótesis de isotropía?,
   c) ¿Por qué el campo aleatorio Y no es débilmente estacionario (independiente de la elección de U)?.

d) Asumiendo el parámetro de nugget fijo e igual a cero, ocupe el método de máxima verosimilitud compuesta condicional para estimar (en ambos modelos) los parámetros de regresión β_0, β_1, de dependencia espacial α y de forma σ (para el caso log-Gaussiano) y κ (para el caso Weibull) (se asume que el parámetro de suavizamiento del modelo Matern $\nu = 0{,}5$). Utilize una vecindad de orden 4 para los pesos de la verosimilitud compuesta.

e) Elija el mejor modelo utilizando el criterio de información de verosimilitud compuesta. Utilice también cross-validacion para averiguar cuáles de los dos modelos tiene mejor capacidad predictiva segùn el criterio del root mean squared error.

f) Utilizando el modelo elegido en la letra e), calcule los residuos y compare el semivariograma estimado versus el semivariograma empírico de los residuos.

g) Utilizando el modelo elegido en la letra e), calcule el predictor lineal óptimo y realize un mapa utilizando las localizaciones espaciales en el archivo coordstopred.txt (las altura asociadas están en el archivo alt2.txt).