Spatial Data Analysis

Lab Exercise 3.

1.

I'm not sure what is wrong with my answer, since I got a result that tent covariance is indeed nonnegative definite. I tried to specify all steps quite clearly.

Tent covariance function:

$r = |h| = |s_i - s_j|$

$$C(r, \theta) = \begin{cases} \theta_1(1 - r/\theta_2) & \text{for } 0 \leq r \leq \theta_2 \\ 0 & \text{for } \theta_2 < r \end{cases}$$

Sites $s_1, ..., s_{48}$, $s_i \in \mathbb{R}^2$

The grid of sites is defined as follows:

$$G_{6\times 8} = \begin{pmatrix} s_1 & s_2 & s_3 & s_4 & s_5 & s_6 & s_7 & s_8 \\ & & & \vdots & & & & \\ s_{41} & s_{42} & s_{43} & s_{44} & s_{45} & s_{46} & s_{47} & s_{48} \end{pmatrix}$$

By the given spacing, the distance between horizontally and vertically adjacent sites is $r = \dfrac{\theta_2}{\sqrt{2}}$, and the distance between diagonally adjacent sites is $r = \theta_2$. Therefore, given a site s_i, only first horizontal and vertical neighbours are correlated since for adjacent sites, it holds that

$$\theta_1(1 - r/\theta_2) = \theta_1(1 - \frac{\theta_2/\sqrt{2}}{\theta_2} = \theta_1(1 - \frac{1}{\sqrt{2}}) = c > 0$$

and for diagonally adjacent sites

$$\theta_1(1 - r/\theta_2) = \theta_1(1 - \frac{\theta_2}{\theta_2} = \theta_1(1 - 1) = 0$$

Hence, for each site s_i we have to consider only the 3x3 neighbourhood, say $\begin{pmatrix} & \cdot & \\ \cdot & s_i & \cdot \\ & \cdot & \end{pmatrix}$ since elsewhere the covariance between sites is 0.

Moreover, by the given definition of parameters $a_i = (-1)^{k+1}$, we can see that they are mapped to the grid $G_{6\times 8}$ as follows:

$$a = \begin{pmatrix} \mathbf{1} \\ \mathbf{-1} \\ \mathbf{1} \\ \mathbf{-1} \\ \mathbf{1} \\ \mathbf{-1} \end{pmatrix}$$

By these information, we can calculate the sum $\displaystyle\sum_i^n \sum_j^n a_i a_j C(s_i - s_j)$. Lets first focus on sites that are so called inner sites, that is where site has all 8 adjacent neighbours. Here we have to consider the neighbourhood $\begin{pmatrix} & \circ & \\ \bullet & s_i & \bullet \\ & \circ & \end{pmatrix}$, where $|s_i - \bullet| = |s_i - \circ|$. Because $\bullet$'s are on the same row that s_i and $\circ$'s on the different row, it also holds that $a_i a_j C(s_i - \bullet) = -a_i a_j C(s_i - \circ)$, and hence for all inner sites, their respective contribution to the sum is 0.

Lets now consider all edge sites on the grid by examining the horizontal and vertical edges separately. For the sites on horizontal edge, the neighbourhood is $\begin{pmatrix} \bullet & s_i & \bullet \\ & \circ & \end{pmatrix}$ or $\begin{pmatrix} & \circ & \\ \bullet & s_i & \bullet \end{pmatrix}$. In both cases, this neighbourhood's contribution to the sum is positive, since we have two sites at the same row that where s_i is, compared to the one site at the different row. To be more exact, for all edge sites the respective contribution to the sum is $a_i a_j C(s_i - \bullet) = a_i a_j c = c > 0$, since $a_i a_j = 1$.

For the sites on vertical edge, the neighbourhood is $\begin{pmatrix} \circ & & \\ s_i & \bullet & \\ \circ & & \end{pmatrix}$ or $\begin{pmatrix} & & \circ \\ & \bullet & s_i \\ & & \circ \end{pmatrix}$. Based on the same comment given above, for all vertical edge sites, their contribution to the sum is negative, say $a_i a_j C(s_i - \circ) = a_i a_j c = -c$.

For all corner sites, their respective contribution to the sum is 0. Therefore, for 6×8 grid, the overall sum can be written as

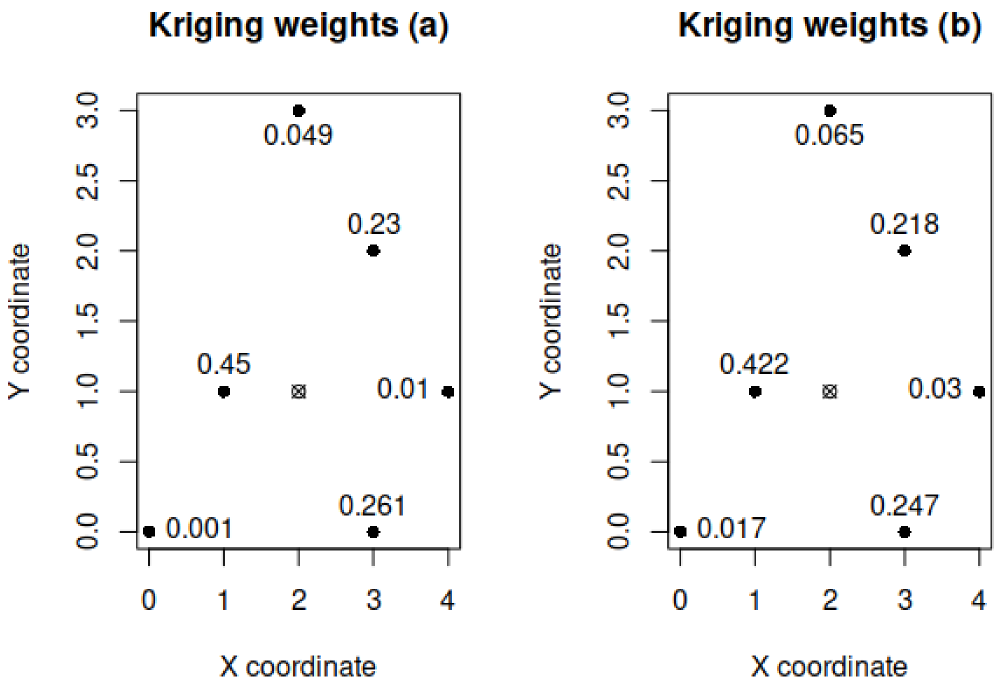
$$\begin{aligned} \sum_i^n \sum_j^n a_i a_j C(s_i - s_j) &= 2 * 6 * c + 2 * 4 * (-c) \\ &= 4c \\ &> 0 \end{aligned}$$

because the gird has more columns than rows.

Therefore tent covariance function is nonnegative definite?

2.

i)



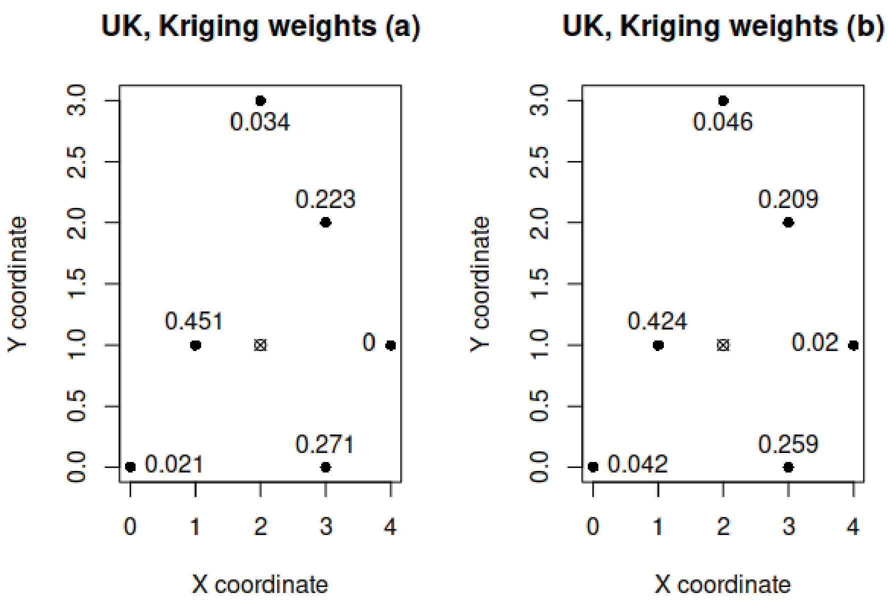
Prediction variances:

a) 0.254

b) 0.609

Based on the weight plots, it seems that in the case of semivariogram used in b, the weights are generally more similar in terms of magnitude. For example, in scenario a, the weights in locations (0,0), (4,1) and (2,3) are relatively small compared to other locations, but in scenario b, these weights are a bit larger. This is mostly due to the differences in magnitude of spatial dependence between models (semivariogram a causes a larger spatial dependence), since I noticed that increase of the nugget effect doesn't seem to cause any effects to weights.

ii)



3.

```
i)
# e1
> soil.kre1$predict
    data
4.471681
> soil.kre1$krige.var
[1] 0.01992799

# e2
> soil.kre2$predict
   data
4.69337
> soil.kre2$krige.var
[1] 0.0199287

# e3
> soil.kre3$predict
   data
4.49676
> soil.kre3$krige.var
[1] 0.01420468

ii)
> mean(cross.wnls$error^2)
[1] 0.03312797
> mean(cross.wnls$std)
[1] 0.000231071
> sqrt(mean(cross.wnls$std^2))
[1] 0.9744802
>
> mean (cross.reml$error^2)
[1] 0.03223426
> mean(cross.reml$std)
[1] 8.207015e-06
> sqrt(mean(cross.reml$std^2))
[1] 0.9998676

Based on the results, it seems that REML-fitting is better.
```