TASK 2

Task 2: Intensity and Randomness

The nest data from islet "nucli 84" is stored in nucli84.txt. Additionally, the coordinates of the islet are in poly84.txt.

1) Build a ppp object using the "nucli 84" data.

```r
rm(list = ls())
nucli84 <- read.delim("T2/nucli84.txt")

min.X = min(nucli84$X)
min.Y = min(nucli84$Y)
max.X = max(nucli84$X)
max.Y = max(nucli84$Y)

nucli84$X = nucli84$X - min.X
nucli84$Y = nucli84$Y - min.Y

n84 = ppp(x = nucli84$X, y = nucli84$Y, range(nucli84$X), range(nucli84$Y))

poligon = read.delim("T2/poly84.txt")

poligon$X = poligon$X - min.X
poligon$Y = poligon$Y - min.Y
pol.illa <- list(x = poligon$X, y = poligon$Y)

min.pX = min(poligon$X)
min.pY = min(poligon$Y)
max.pX = max(poligon$X)
max.pY = max(poligon$Y)

n84p = ppp(nucli84$X, nucli84$Y, poly = pol.illa, range(poligon$X), range(poligon$Y))

islet_window = owin(poly = pol.illa)
```
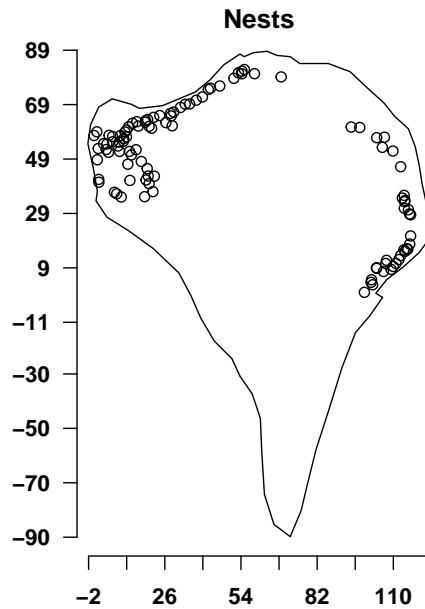
```r
par(mfrow = c(1, 1), font = 2, font.axis = 2, font.lab = 4, las = 1, mar = c(2, 0,
    1, 0))
plot(n84p, main = "Nests")
axis(1, at = c(round(seq(min.pX, max.pX, length = 10), digits = 0)))  #, pos=c(0,0))
axis(2, at = c(round(seq(min.pY, max.pY, length = 10), digits = 0)), pos = c(min.pX -
    4, min.pY - 50))
```

Nests

2) Draw a plot with the intensity of the point process computed by the non-parametric approach. Briefly comment the results.

```
par(mfrow = c(1, 1), font = 2, font.axis = 2, font.lab = 4, las = 1, mar = c(0, 0,
    1, 0))
plot(density(n84p, dimax.Yx = c(256, 256), sigma = 15), main = "Intensity of the point process")
```

Intensity of the point process
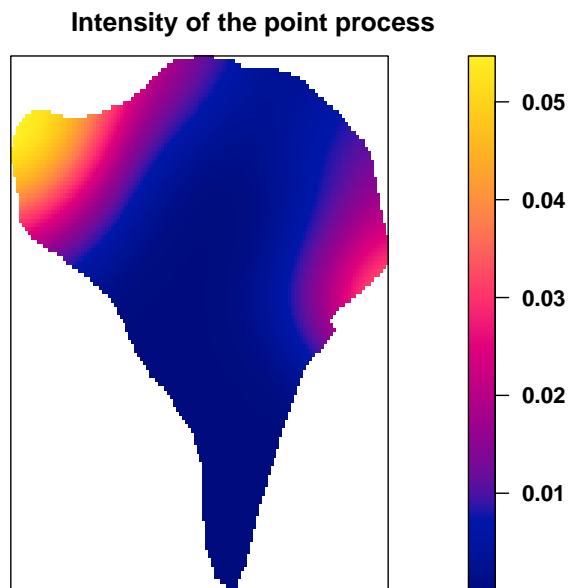
In the density plot, we can see that the nests are concentrated in two main areas: one in the northwest corner and one in the eastern part, both located on the seaside. No nests were observed in the central or southern areas of the islet.

This may suggest that a geographical feature is influencing the location of the nests.

2

We also try to fit a homogeneous Poisson process:

```
modelPois = ppm(n84, ~1)
modelPois
```

```
## Stationary Poisson process
## Intensity: 0.01091393
##             Estimate        S.E.   CI95.lo   CI95.hi Ztest      Zval
## log(lambda) -4.517715 0.09805807 -4.709905 -4.325525   *** -46.07183
```

So the intensity has point estimate `r exp(modelPois$coef)` and 95% confidence interval
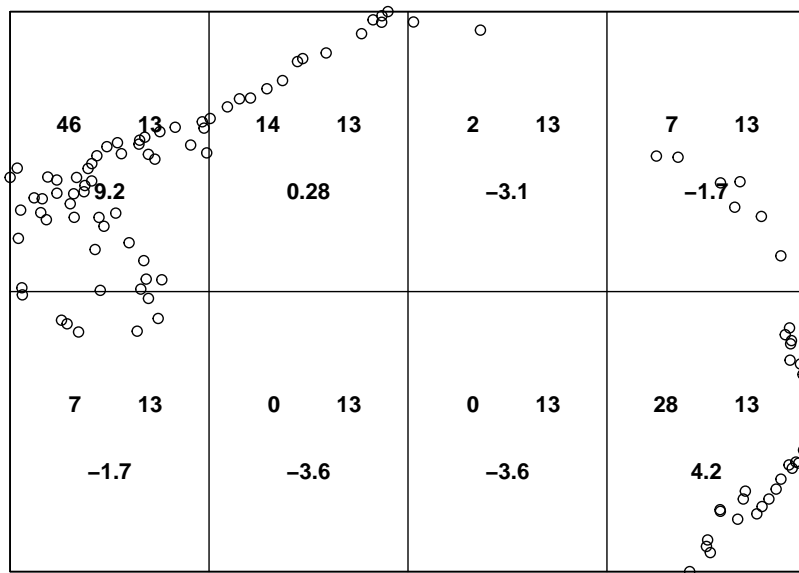
```
##       2.5 %     97.5 %
##  0.00900563 0.01322661
```

3) Assess the Completely Spatial Randomness hypothesis

- via Chi-square test:

We divide the region in 8 subareas with equal areas and under CSR we would expect more or less same number of nests in each subregion.

```
M <- quadrat.test(n84, nx = 4, ny = 2)
par(mfrow = c(1, 1), font = 2, font.axis = 2, font.lab = 4, las = 1, mar = c(0, 0,
    0, 0))
plot(n84, main = "")
plot(M, add = TRUE)
```


Nonetheless we notice that some subregions have more nests than other. Indeed after performing a Chi-square test we get:

```
M
```

```
##
##  Chi-squared test of CSR using quadrat counts
##
## data:  n84
## X2 = 142, df = 7, p-value < 2.2e-16
## alternative hypothesis: two.sided
##
## Quadrats: 4 by 2 grid of tiles
```

The p-value is extremely low ($p < 2.2e - 16$), which indicates that we can reject the null hypothesis of complete spatial randomness (CRS).

- via Kolmogorov-Smirnov test:

```
KS = cdf.test(n84, covariate = "x")
KS
```

```
##
##  Spatial Kolmogorov-Smirnov test of CSR in two dimensions
##
## data:  covariate 'x' evaluated at points of 'n84'
##      and transformed to uniform distribution under CSR
## D = 0.26906, p-value = 5.772e-07
## alternative hypothesis: two-sided
```

```
plot(KS)
```


Spatial Kolmogorov–Smirnov test of CSR in two dimensions
based on distribution of x coordinate
p–value= 5.772e–07
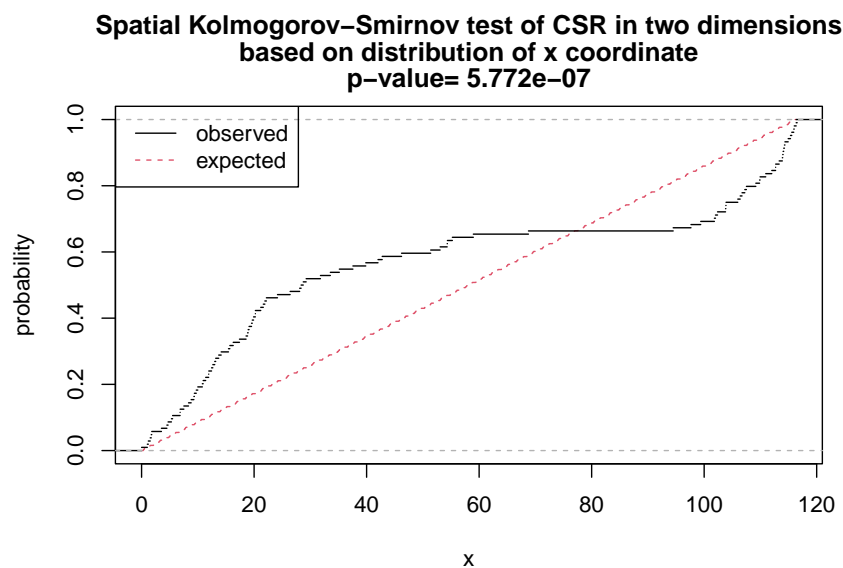
```
KS = cdf.test(n84, covariate = "y")
KS
```

```
##
##  Spatial Kolmogorov-Smirnov test of CSR in two dimensions
##
## data:  covariate 'y' evaluated at points of 'n84'
##       and transformed to uniform distribution under CSR
## D = 0.18466, p-value = 0.001662
## alternative hypothesis: two-sided
```

```
plot(KS)
```

Spatial Kolmogorov–Smirnov test of CSR in two dimensions
based on distribution of y coordinate
p–value= 0.001662
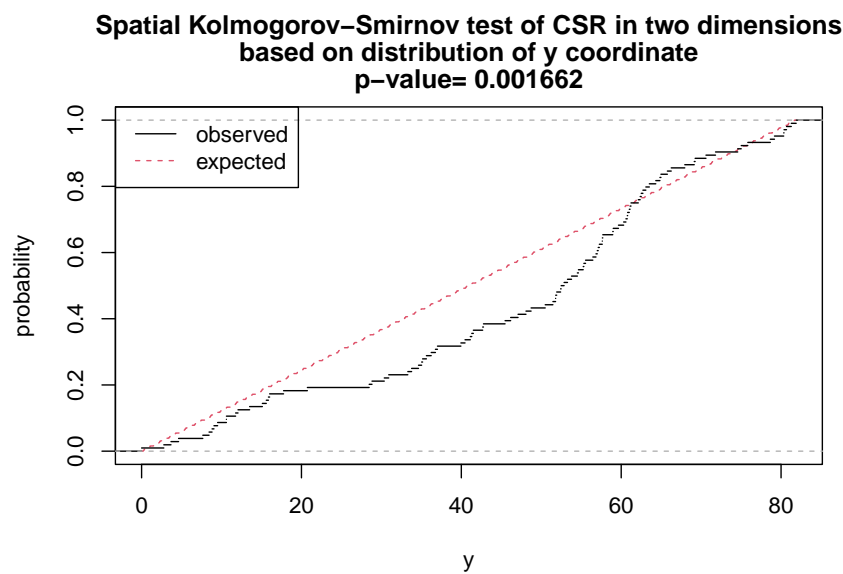
The observed distribution of Z at data points differs significantly from the expected distribution, allowing us to reject the null hypothesis of CSR and conclude that there is a dependence between the intensity of the points and both the Cartesian coordinates.

**4) Assess the relation between the intensity of the point process and the covariates height and vegetation.**

```
grid <- read.delim("T2/grid.txt")
grid.veg = read.delim("T2/grid_veg.txt")
veg = as.matrix(read.delim("T2/veg.txt", header = FALSE))
height = as.matrix(read.delim("T2/height.txt", header = FALSE))
```

First we visualize the values of Height and Vegetation in the islet:

```
par(mfrow = c(1, 2), font = 2, font.axis = 2, font.lab = 4, las = 1, mar = c(0, 0,
    1, 2))
```

5

```
Height = im(mat = height, xcol = grid$x, yrow = grid$y)
plot(Height, main = "Height", clipwin = islet_window)
plot(n84p, add = T, cex = 0.5, pch = 16)

Vegetation = im(mat = veg, xcol = grid.veg$x, yrow = grid.veg$y)
plot(Vegetation, main = "Vegetation", clipwin = islet_window)
plot(n84p, add = T, cex = 0.5, pch = 16)
```
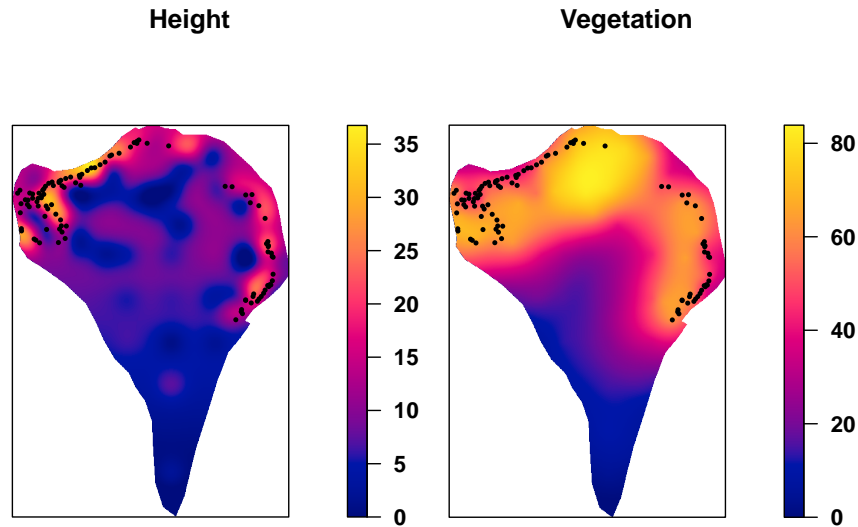
Height   **Vegetation**

The nests seem to be concentrated in areas with higher vegetation and elevation.

Now we categorize the covariates using the suggested intervals:

- Height. [0,10], (10,20], (20,40]
- Vegetation. [0,20], (20,50], (50,100]

```
# height
brks.h <- c(0, 10, 20, 40)
Hcut <- cut(Height, breaks = brks.h, labels = c("0-10", "10-20", "20-40"))
H <- tess(image = Hcut)

# vegetation
brks.v <- c(0, 20, 50, 100)
Vcut <- cut(Vegetation, breaks = brks.v, labels = c("0-20", "20-50", "50-100"))
V <- tess(image = Vcut)
```

```
par(mfrow = c(1, 2), font = 2, font.axis = 2, font.lab = 4, las = 1, mar = c(0, 0,
    1, 3.5))
plot(H, main = "Height")
plot(V, main = "Vegetation")
```

Height   **Vegetation**
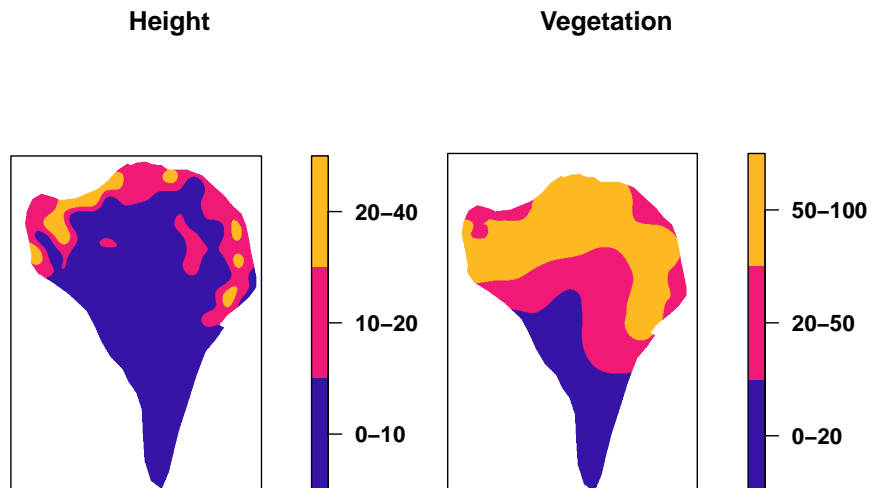
And to assess quantitatively this relation we use:
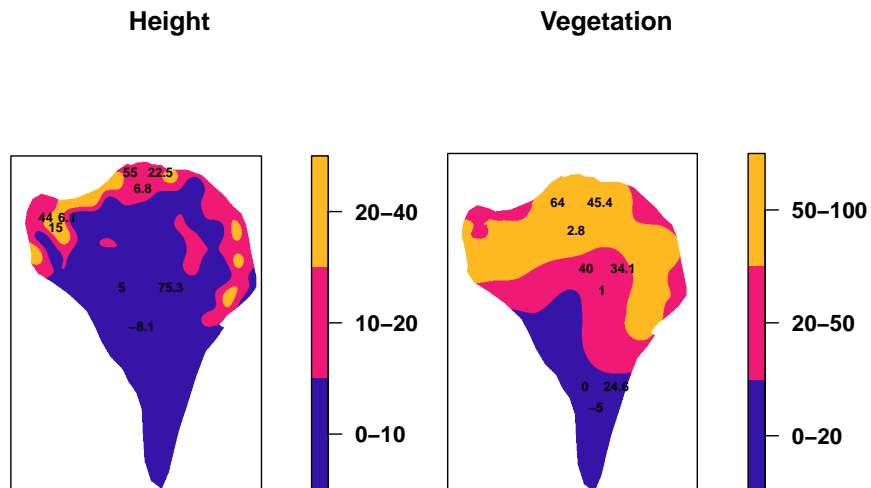
- a Chi-squared test:

```
M.h = quadrat.test(n84, tess = H)
M.h
```

```
##
##  Chi-squared test of CSR using quadrat counts
##
## data:  n84
## X2 = 345.67, df = 2, p-value < 2.2e-16
## alternative hypothesis: two.sided
##
## Quadrats: 3 tiles (levels of a pixel image)
```

```
M.v = quadrat.test(n84p, tess = V)
M.v
```

```
##
##  Chi-squared test of CSR using quadrat counts
##
## data:  n84p
## X2 = 33.266, df = 2, p-value = 1.195e-07
## alternative hypothesis: two.sided
##
## Quadrats: 3 tiles (levels of a pixel image)
```

```
par(mfrow = c(1, 2), font = 2, font.axis = 2, font.lab = 4, las = 1, mar = c(0, 0,
    1, 3.5))
plot(M.h, valuesAreColours = FALSE, main = "Height", cex = 0.6)
plot(M.v, valuesAreColours = FALSE, main = "Vegetation", cex = 0.6)
```

**Height**   **Vegetation**

The Chi-squared tests on the two covariates Height and Vegetation have p-values: $p_1 < 1 \cdot 10^{-16}$ and $p_2 = 1.195 \cdot 10^{-7}$.

- a Kolmogorov-Smirnov test:

```
KS <- cdf.test(n84, Height)
KS
```

```
##
##  Spatial Kolmogorov-Smirnov test of CSR in two dimensions
##
## data:  covariate 'Height' evaluated at points of 'n84'
##      and transformed to uniform distribution under CSR
## D = 0.72281, p-value < 2.2e-16
## alternative hypothesis: two-sided
```
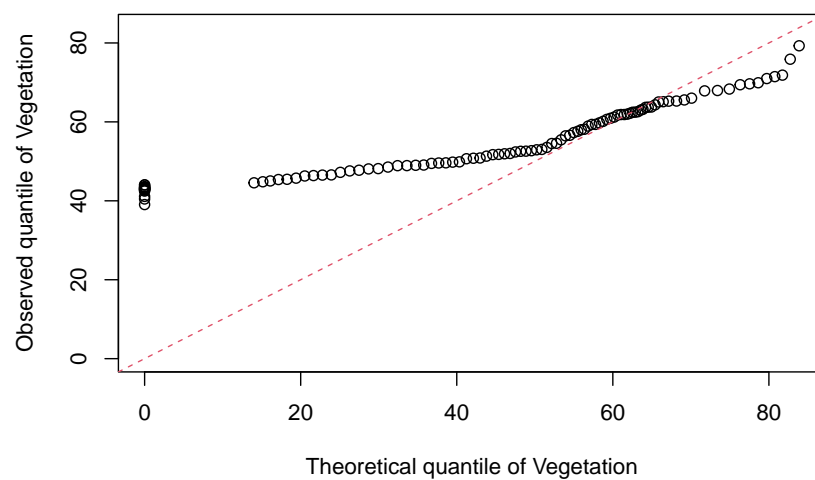
```
plot(KS, style = "QQ")
```
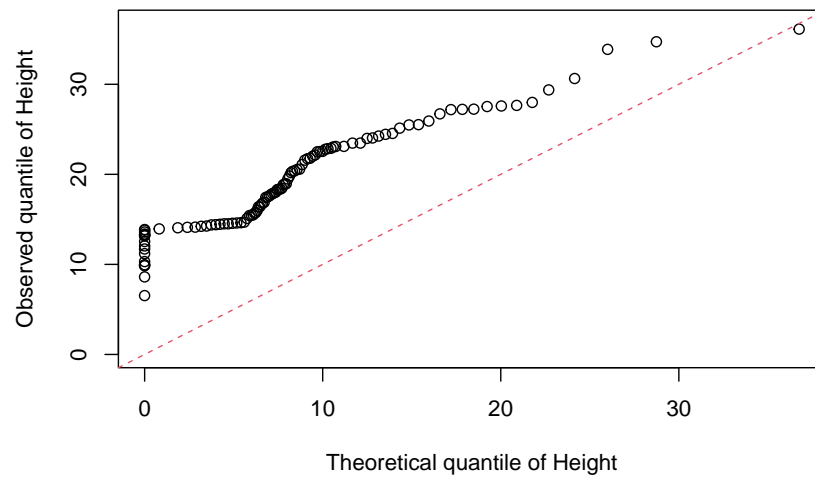
```
KS <- cdf.test(n84, Vegetation)
KS
```

```
##
##  Spatial Kolmogorov-Smirnov test of CSR in two dimensions
##
## data:  covariate 'Vegetation' evaluated at points of 'n84'
##       and transformed to uniform distribution under CSR
## D = 0.38481, p-value = 8.404e-14
## alternative hypothesis: two-sided
```

```
plot(KS, style = "QQ")
```

Also the Kolmogorov test has very little p-values and the plots suggest a general difference of the observed quantiles from the theoretical ones.

So, in the light of the above considerations, we can state there is in fact a relation between the covariates and the point process.

5) Fit an inhomogeneous Poisson model to data

We'll fit the model:
$$\lambda \sim x + y + \text{Height} + \text{Vegetation}$$

```
model = ppm(n84, ~x + y + Height + Vegetation)
model
```

```
## Nonstationary Poisson process
##
## Log intensity:  ~x + y + Height + Vegetation
##
## Fitted trend coefficients:
##  (Intercept)            x            y        Height    Vegetation
## -6.773462579 -0.001683221 -0.014827821   0.162420734   0.016598983
##
##                 Estimate         S.E.      CI95.lo       CI95.hi Ztest
## (Intercept) -6.773462579 0.486190000 -7.726377469 -5.820547689   ***
## x           -0.001683221 0.002823778 -0.007217724  0.003851283
## y           -0.014827821 0.005412996 -0.025437099 -0.004218544    **
## Height       0.162420734 0.012634209  0.137658140  0.187183328   ***
## Vegetation   0.016598983 0.007100232  0.002682784  0.030515183     *
##                   Zval
## (Intercept) -13.9317192
## x            -0.5960882
## y            -2.7393001
## Height       12.8556318
## Vegetation    2.3378085
```

We notice that we can remove the x variable as it's not significant, since it has the Z statistic very close to 0.

```
model1 = ppm(n84, ~y + Height + Vegetation)
model1
```

```
## Nonstationary Poisson process
##
## Log intensity:  ~y + Height + Vegetation
##
## Fitted trend coefficients:
## (Intercept)           y       Height   Vegetation
## -6.91203934 -0.01415209   0.16458833   0.01620508
##
##                Estimate        S.E.       CI95.lo       CI95.hi Ztest        Zval
## (Intercept) -6.91203934 0.433891191 -7.762450448 -6.061628232   *** -15.930352
## y           -0.01415209 0.005238198 -0.024418771 -0.003885411    **  -2.701710
## Height       0.16458833 0.012134934  0.140804296  0.188372364   ***  13.563183
## Vegetation   0.01620508 0.007127567  0.002235302  0.030174850     *   2.273578
```

Now all the covariates are significant (they have the Z statistic far enough from zero and their confidence intervals don't contain 0), so we can keep the model:

$$\lambda \sim y + \text{Height} + Vegetation$$

whose fitted parameters have point estimates:

```
chosen.model = model1

coef(chosen.model)
```

```
## (Intercept)            y       Height  Vegetation
## -6.91203934 -0.01415209   0.16458833  0.01620508
```
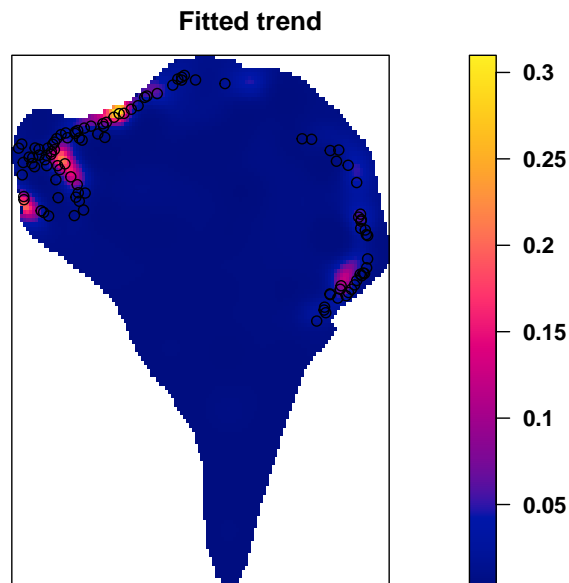
and 95% confidence interval:

```
confint(chosen.model)
```

```
##                    2.5 %        97.5 %
## (Intercept) -7.762450448 -6.061628232
## y           -0.024418771 -0.003885411
## Height       0.140804296  0.188372364
## Vegetation   0.002235302  0.030174850
```

The trend fitted by the model is:

```
par(mfrow = c(1, 1), font = 2, font.axis = 2, font.lab = 4, las = 1, mar = c(0, 0,
    1, 0))
plot(chosen.model, se = F, locations = islet_window)
```


Fitted trend

Now we assess the goodness of fit of the chosen model:

- via Chi-squared test

11

```
M <- quadrat.test(chosen.model, nx = 4, ny = 2)
M
```

```
##
##  Chi-squared test of fitted Poisson model 'chosen.model' using quadrat
##  counts
##
## data:  data from chosen.model
## X2 = 35.127, df = 4, p-value = 8.749e-07
## alternative hypothesis: two.sided
##
## Quadrats: 4 by 2 grid of tiles
```

```
par(mfrow = c(1, 1), font = 2, font.axis = 2, font.lab = 4, las = 1, mar = c(0, 0,
    0, 0))
plot(n84, main = "")
plot(M, add = TRUE, cex = 1.5, col = "red")
```
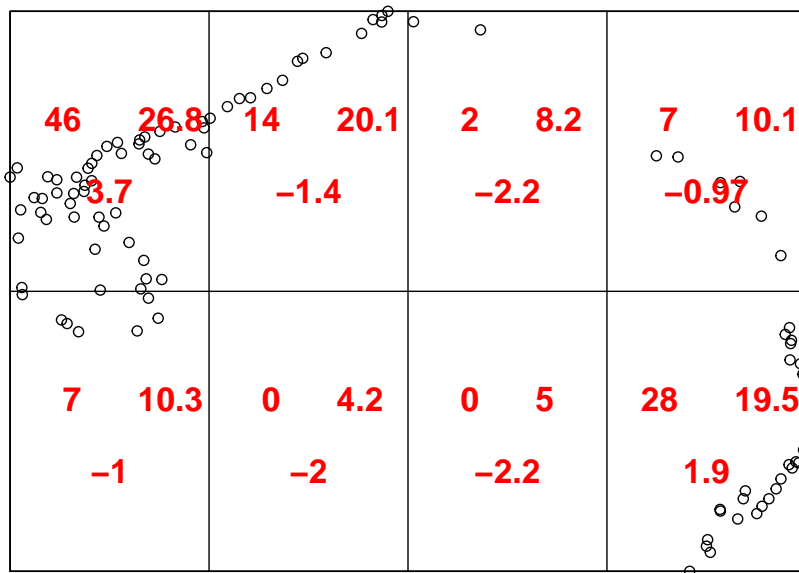

The p-value is very low, even though in almost all the subregions the standardized residuals have absolute value around 2 or smaller.

- via Kolmogorov-Smirnov test, to check the fitting of the model by each covariate separately.

```
KS1 = cdf.test(chosen.model, "y")
KS1
```

```
##
##  Spatial Kolmogorov-Smirnov test of inhomogeneous Poisson process in two
##  dimensions
##
## data:  covariate 'y' evaluated at points of 'n84'
```

```
##        and transformed to uniform distribution under 'chosen.model'
## D = 0.10476, p-value = 0.2038
## alternative hypothesis: two-sided
```

```
KS2 = cdf.test(chosen.model, Vegetation)
KS2
```

```
##
##  Spatial Kolmogorov-Smirnov test of inhomogeneous Poisson process in two
##  dimensions
##
## data:  covariate 'Vegetation' evaluated at points of 'n84'
##        and transformed to uniform distribution under 'chosen.model'
## D = 0.13628, p-value = 0.042
## alternative hypothesis: two-sided
```

```
KS3 = cdf.test(chosen.model, Height)
KS3
```

```
##
##  Spatial Kolmogorov-Smirnov test of inhomogeneous Poisson process in two
##  dimensions
##
## data:  covariate 'Height' evaluated at points of 'n84'
##        and transformed to uniform distribution under 'chosen.model'
## D = 0.20645, p-value = 0.0002825
## alternative hypothesis: two-sided
```

The y variable seems to fits quite well the data, while Vegetation and Height have very little significance.

- looking at the Smoothed Pearson residuals

```
band = 40
par(mfrow = c(1, 1), font = 2, font.axis = 2, font.lab = 4, las = 1, mar = c(0, 0,
    1, 0))
smooth = diagnose.ppm(chosen.model, which = "smooth", type = "pearson", sigma = band)
```

Smoothed Pearson residuals
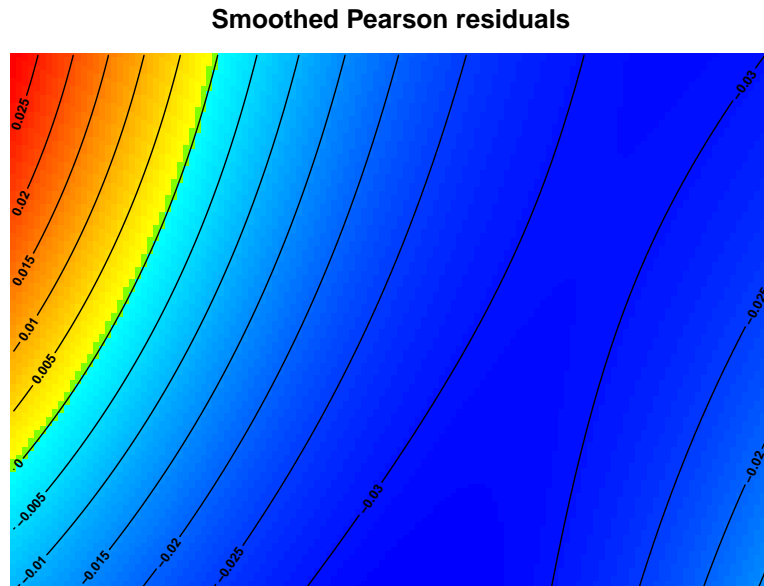

```
smooth
```

```
## Model diagnostics (Pearson residuals)
## Diagnostics available:
##   smoothed residual field
## range of smoothed field =  [-0.03264, 0.02842]
## Null standard deviation of smoothed Pearson residual field: 0.007052
```

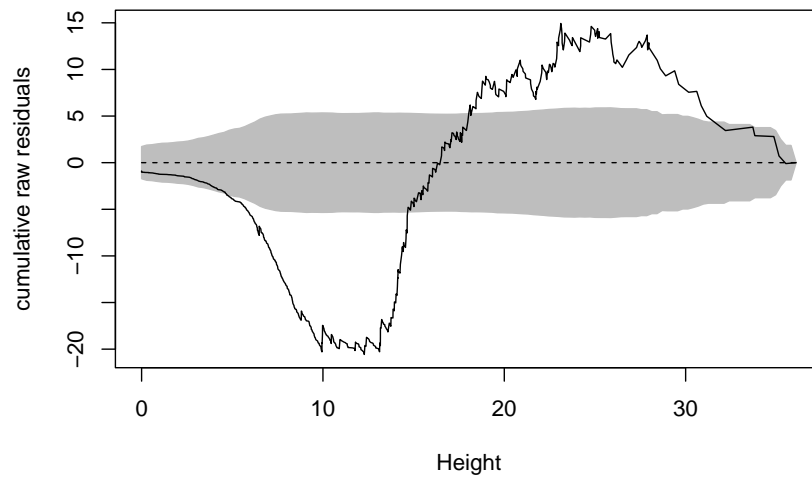The standard deviation of the smoothed residuals under the null hypothesis of correct fitting is

```
smooth$smooth$sdp
```

```
## [1] 0.00705237
```

so residuals whose modulus is larger than this value can be considered as extreme. This means that almost all subregions are not well fitted.

- through a Lurking plot

```
lurking(chosen.model, Height, type = "raw")
```

The residuals lie outside the envelope. The fit is not good:

- When the height is lower than 10 the residuals are consistently negative. Therefore, there are less points than expected with low height.
- From $height = 10$ to $height = 25$ the residuals are growing, so there are more points than expected.

After conducting these analyses, we can conclude that the model we have built, although it had the best fit from a statistical standpoint, does not actually fit the data well. Despite this, it was the most significant model based on the data we have.