Re-doing Lab 3

:)

25 April, 2023 11:04:24

Complete the following exercises before the submission deadline. In addition to the points detailed below, 5 points are assigned to the quality of the annotation, as well as to the 'cleanliness' of the code and resulting pdf document.

Exercise 1 – 1 point

We will again be working with the BC Parks dataset, which contains information on the locations of Provincial Parks in British Columbia. The parks belong to 5 different regions. There is also information on elevation (in m) and percent forest cover contained within the dataset.

- Import the BC park locations dataset and convert the data to a `ppp` object (for today you can exclude information on regions). – 1 point(s)

Note: You will need to load the `maptools` package and make use of the `as.owin()` function.

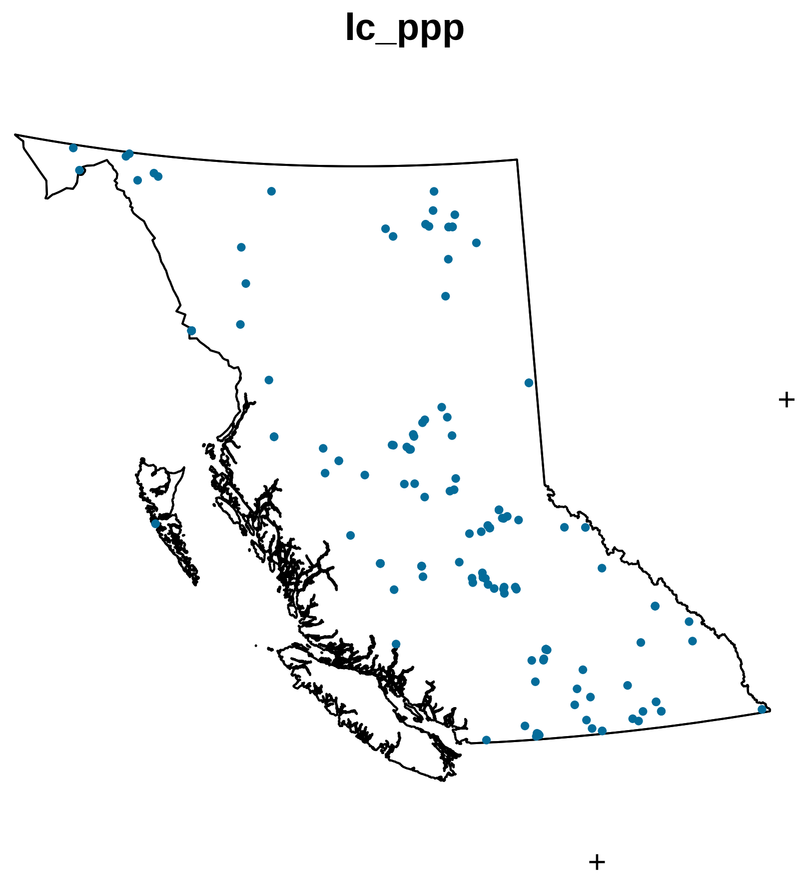
```
# 1
#quiet function
quiet <- function(x) {
  sink(tempfile())
  on.exit(sink())
  invisible(force(x))
}

#importing dataset
lc = read.csv("../datasets/processed/lc.csv")

#importing window
suppressMessages(library(spatstat))
suppressMessages(library(sf))
suppressMessages(library(maptools))
load("../datasets/raw/BC_Covariates.Rda")
bc_window_sf = st_as_sf(DATA$Window)
bc_window_owin = as.owin(bc_window_sf)

#converting to a ppp
lc_ppp = ppp(x = lc$decimalLongitude,
             y = lc$decimalLatitude,
             window = bc_window_owin,
             )
```

```
## Warning: 2 points were rejected as lying outside the specified window
```

```
## Warning: data contain duplicated points
```

```
plot(lc_ppp, pch = 16, cols = "#046C9A", cex = 0.6)
```

```
## Warning in plot.ppp(lc_ppp, pch = 16, cols = "#046C9A", cex = 0.6): 2 illegal
## points also plotted
```



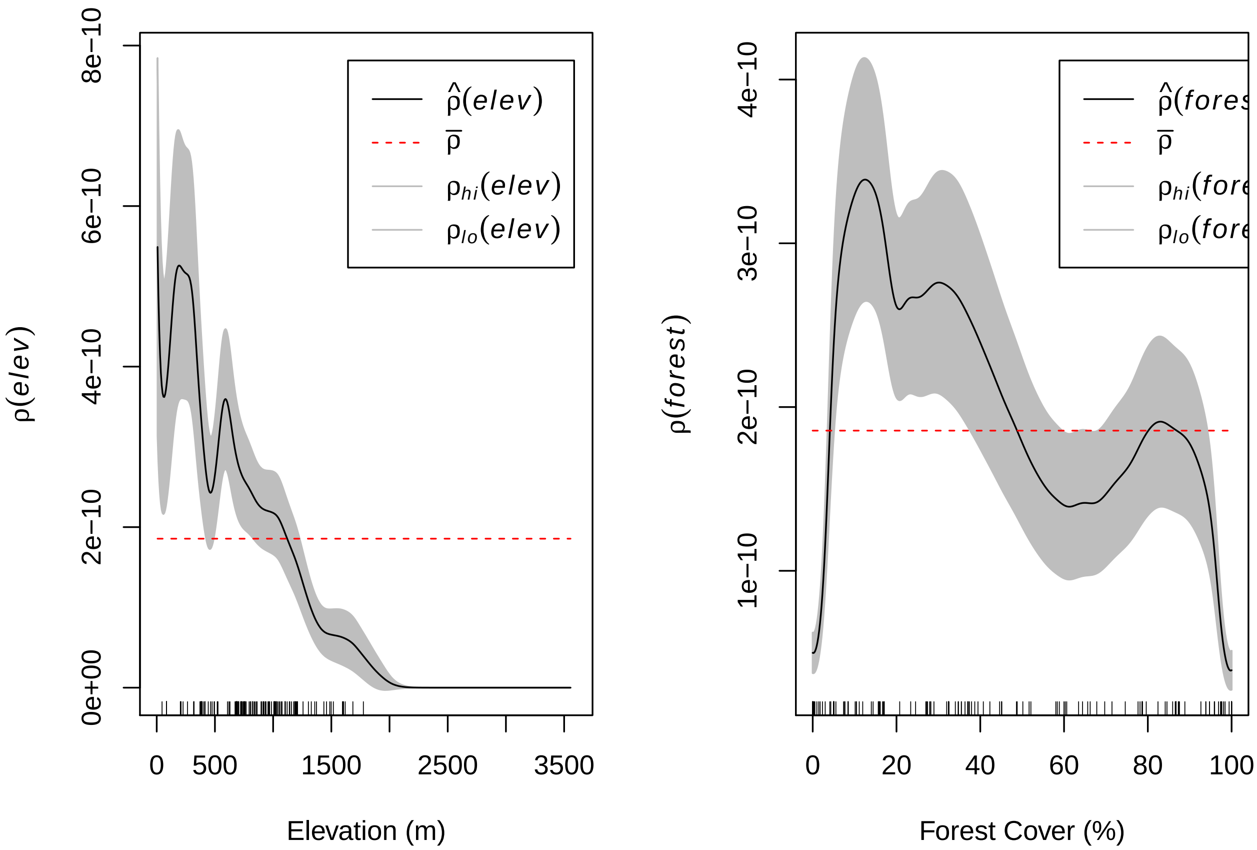
Exercise 2 – 4 points

- Estimate and plot ρ for the locations of parks as a function of both elevation and forest cover (be sure that the x-axis for elevation does not go below 0). – 2 point(s)
- Check for collinearity between elevation and forest cover (you will need to consider NA values). – 1 point(s)
- Based on these initial analyses, write down the expected form of the model. Provide justification for this starting point. – 1 point(s)

Note: Estimating rho can be slow ($\sim$ 1-2 min). Be sure to leave enough time for the document to knit.

```
# 1
elev = DATA$Elevation
rho_elev = rhohat(lc_ppp, elev)
forest = DATA$Forest
rho_forest = rhohat(lc_ppp, forest)
par(mfrow = c(1,2))
plot(rho_elev,
     main = "",
     xlab = "Elevation (m)",
     xlim=c(0, 3600))
plot(rho_forest,
```

```
     main = "",
     xlab = "Forest Cover (%)")
```



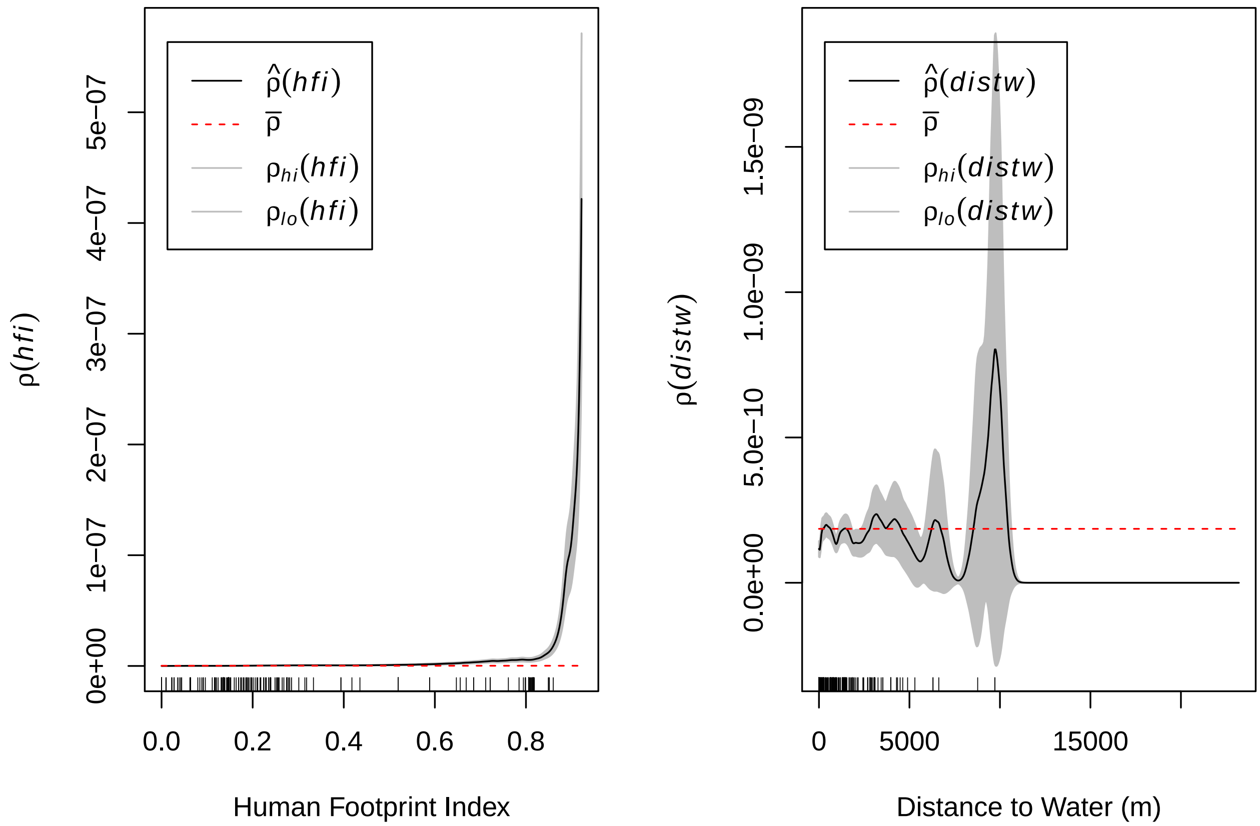

```
par(mfrow = c(1,1))
```

HFI is based on: population density, land transformation, human access, and power infrastructure.

```
# 1
hfi = DATA$HFI
rho_hfi = rhohat(lc_ppp, hfi)
```

```
## Warning: Values for 1 query point lying outside the pixel image domain were
## estimated by projection to the nearest pixel
```

```
distw = DATA$Dist_Water
rho_distw = rhohat(lc_ppp, distw)
par(mfrow = c(1,2))
plot(rho_hfi,
     main = "",
     xlab = "Human Footprint Index")
plot(rho_distw,
     main = "",
     xlab = "Distance to Water (m)")
```




```
par(mfrow = c(1,1))

#correlation
cor.im(elev, forest, hfi, distw,  use = 'complete.obs')
```

```
##              ..1         ..2         ..3         ..4
## ..1  1.00000000 -0.26225376 -0.26625626 -0.03493453
## ..2 -0.26225376  1.00000000  0.06618592  0.04818598
## ..3 -0.26625626  0.06618592  1.00000000  0.13246899
## ..4 -0.03493453  0.04818598  0.13246899  1.00000000
```

Exercise 3 – 4 points

- Fit the model you have defined in exercise 2 and inspect the model output. – 1 point(s)
- Fit a null, intercept only model. – 1 point(s)
- Use AIC and a likelihood ratio test to determine if the model you defined is a better fit than the intercept only model. – 1 point(s)
- Write down the equation for the selected model. – 0.5 point(s)
- Use this equation to estimate the intensity of parks at 500m elevation and 50% forest cover.

```
# 0
#mean centering and scaling the elevation and distance to water variables
mu <- mean(DATA$Elevation)
stdev <- sd(DATA$Elevation)
Elevation_scaled <- eval.im((elev - mu)/stdev, DATA)
mu <- mean(DATA$Dist_Water)
stdev <- sd(DATA$Dist_Water)
Dist_Water_scaled <- eval.im((distw - mu)/stdev, DATA)
```

```r
# 1
fit1 = ppm(lc_ppp ~ Elevation_scaled + I(Elevation_scaled^2) + forest + I(forest^2) + hfi + I(hfi^2) + I
```

```
## Warning: Values of the covariate 'hfi' were NA or undefined at 0.44% (3 out of
## 689) of the quadrature points. Occurred while executing: ppm.ppp(Q = lc_ppp,
## trend = ~Elevation_scaled + I(Elevation_scaled^2) +

## Warning: glm.fit: algorithm did not converge
```

```r
fit1
```

```
## Nonstationary Poisson process
## Fitted to point pattern dataset 'lc_ppp'
##
## Log intensity:  ~Elevation_scaled + I(Elevation_scaled^2) + forest +
## I(forest^2) + hfi + I(hfi^2) + Dist_Water_scaled + I(Dist_Water_scaled^2)
##
## Fitted trend coefficients:
##           (Intercept)      Elevation_scaled  I(Elevation_scaled^2)
##         -2.345722e+01         -5.896126e-01          -2.159342e-01
##                forest           I(forest^2)                    hfi
##         -7.122462e-03          6.673426e-05           8.483340e+00
##              I(hfi^2)     Dist_Water_scaled I(Dist_Water_scaled^2)
##         -5.814717e+00         -2.233870e-01          -2.987220e-03
##
##                             Estimate         S.E.       CI95.lo       CI95.hi
## (Intercept)            -2.345722e+01 2.395698e-01 -2.392677e+01 -22.987674680
## Elevation_scaled       -5.896126e-01 1.371295e-01 -8.583814e-01  -0.320843734
## I(Elevation_scaled^2)  -2.159342e-01 8.596659e-02 -3.844256e-01  -0.047442748
## forest                 -7.122462e-03 8.619863e-03 -2.401708e-02   0.009772159
## I(forest^2)             6.673426e-05 8.647235e-05 -1.027484e-04   0.000236217
## hfi                     8.483340e+00 1.175667e+00  6.179076e+00  10.787604237
## I(hfi^2)               -5.814717e+00 1.345916e+00 -8.452663e+00  -3.176769595
## Dist_Water_scaled      -2.233870e-01 1.193615e-01 -4.573313e-01   0.010557262
## I(Dist_Water_scaled^2) -2.987220e-03 4.558796e-02 -9.233798e-02   0.086363542
##                        Ztest         Zval
## (Intercept)              *** -97.91393254
## Elevation_scaled         ***  -4.29967786
## I(Elevation_scaled^2)      *  -2.51183826
## forest                        -0.82628484
## I(forest^2)                    0.77174104
## hfi                      ***   7.21577068
## I(hfi^2)                 ***  -4.32026696
## Dist_Water_scaled             -1.87151621
## I(Dist_Water_scaled^2)        -0.06552651
## Problem:
##  Values of the covariate 'hfi' were NA or undefined at 0.44% (3 out of 689) of
## the quadrature points
##
## *** Fitting algorithm for 'glm' did not converge ***
```

```r
# 1.5
fit2 = ppm(lc_ppp ~ Elevation_scaled + I(Elevation_scaled^2) + hfi + I(hfi^2))
```

```
## Warning: Values of the covariate 'hfi' were NA or undefined at 0.44% (3 out of
## 689) of the quadrature points. Occurred while executing: ppm.ppp(Q = lc_ppp,
```

```
## trend = ~Elevation_scaled + I(Elevation_scaled^2) +

## Warning: glm.fit: algorithm did not converge
```

```r
fit2
```

```
## Nonstationary Poisson process
## Fitted to point pattern dataset 'lc_ppp'
##
## Log intensity:  ~Elevation_scaled + I(Elevation_scaled^2) + hfi + I(hfi^2)
##
## Fitted trend coefficients:
##           (Intercept)      Elevation_scaled I(Elevation_scaled^2)
##           -23.5142168            -0.5496413            -0.1806727
##                   hfi              I(hfi^2)
##             7.9345208            -5.1926304
##
##                          Estimate       S.E.      CI95.lo     CI95.hi  Ztest
## (Intercept)           -23.5142168 0.17827617 -23.8636317 -23.1648020   ***
## Elevation_scaled       -0.5496413 0.13384529  -0.8119732  -0.2873093   ***
## I(Elevation_scaled^2)  -0.1806727 0.08343033  -0.3441932  -0.0171523     *
## hfi                     7.9345208 1.15645659   5.6679075  10.2011340   ***
## I(hfi^2)               -5.1926304 1.33813978  -7.8153361  -2.5699246   ***
##                             Zval
## (Intercept)           -131.897701
## Elevation_scaled        -4.106542
## I(Elevation_scaled^2)   -2.165552
## hfi                      6.861062
## I(hfi^2)                -3.880484
## Problem:
##  Values of the covariate 'hfi' were NA or undefined at 0.44% (3 out of 689) of
## the quadrature points
##
## *** Fitting algorithm for 'glm' did not converge ***
```

```r
# 2
fit_intercept = ppm(lc_ppp ~ 1)
fit_intercept
```

```
## Stationary Poisson process
## Fitted to point pattern dataset 'lc_ppp'
## Intensity: 1.856026e-10
##              Estimate       S.E.   CI95.lo   CI95.hi Ztest      Zval
## log(lambda) -22.40741 0.07537784 -22.55515 -22.25968   *** -297.2679
```

```r
# 3
AIC(fit1); AIC(fit2); AIC(fit_intercept)
```

```
## [1] 7963.644

## [1] 7965.547

## [1] 8241.409
```

```r
AIC(fit2) - AIC(fit1) #delta AIC
```

```
## [1] 1.903073
```

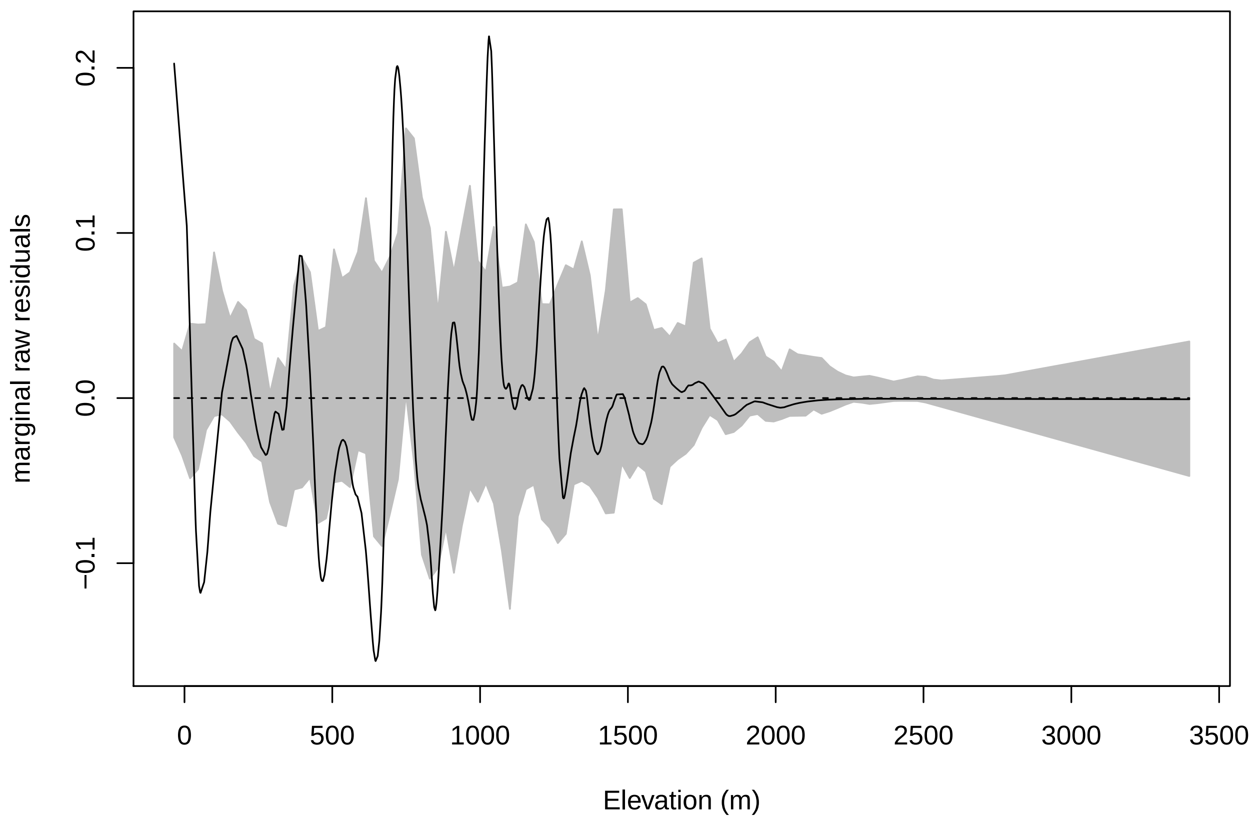
```r
anova(fit2, fit1, test = "LRT") #LRT
```

```
## Analysis of Deviance Table
##
## Model 1: ~Elevation_scaled + I(Elevation_scaled^2) + hfi + I(hfi^2)   Poisson
## Model 2: ~Elevation_scaled + I(Elevation_scaled^2) + forest + I(forest^2) + hfi + I(hfi^2) + Dist_Wa
##   Npar Df Deviance Pr(>Chi)
## 1    5
## 2    9  4   9.9031  0.04209 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
#the positive delta AIC and the small p-value from the LRT both point
#to the more complex model being more parsimonious (better fit and
#worth it)
```

```r
#lurking variable plot
lurking(fit2, elev, type = "raw", cumulative = F, envelope = T, xlab = "Elevation (m)" )
```

```
## Generating 39 simulated patterns ...1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 1
## Processing.. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24,
## Done.
```




```r
#GAM
```

Exercise 4 – 4 points

- Visualise the fitted model. Note: log scale the estimated intensity when plotting, ignore the standard error. You can use the `n` argument to adjust the resolution – 1 point(s)

- Plot the effects of the individual coefficients. Note: use the median value(s) of the other coefficients. – 2 point(s)
- Visually, do you think the model predictions are a good match to the data? – 1 point(s)

```
# 1
# plot(fit1,
#      se = FALSE,
#      superimpose = FALSE,
#      log = TRUE, n=80)
#
# plot(parks_ppp,
#      pch = 16,
#      cex = 0.6,
#      cols = "white",
#      add = TRUE)
# plot(parks_ppp,
#      pch = 16,
#      cex = 0.5,
#      cols = "black",
#      add = TRUE)
#
# # 2
# elev_effect = effectfun(fit1, "elev", forest = median(forest), se.fit = T)
# forest_effect = effectfun(fit1, 'forest', elev = median(elev), se.fit = T)
# par(mfrow = c(1,2))
# plot(elev_effect,
#      legend = FALSE,
#      main = "Elevational effect at median \n forest coverage")
# plot(forest_effect,
#      legend = FALSE,
#      main = "Effect of forest coverage at \n median elevation")
# par(mfrow = c(1,1))

# 3
```

```
#diagnose.ppm(fit1)
```

The model predictions are mediocre. It does alright on the island, but misses a lot of the density on the lower part of BC, and overpredicts in the top right corner. There is definitely room for improvement.

Exercise 5 – 1 point

- Test whether the observed data deviate significantly from the model predictions. – 1 point(s)

```
# 1
#quadrat.test(fit1, nx = 2, ny = 4)
#The small p value tells us that there's a significant deviation
#from our model's predictions.
```

Exercise 5 – 2 points

- Calculate and plot the model residuals. – 1 point(s)
- Based on the residuals, do you think the model performing well? – 1 point(s)

```
# 1
#plot(residuals(fit1), cols = "transparent")
```

If the model is performing well, there should be no trend in the residuals. Although it is difficult to tell based on this plot, there is a trend in the residuals. There are larger residuals nearer the coast than there is inland. Perhaps 'distance to the coast' should be included as a covariate to account for this.

Exercise 6 – 3 points

- Calculate the partial residuals as a function of both elevation and forest cover. – 1 point(s)
- Do you think that the terms are accurately capturing trends in the data? – 1 point(s)
- Do you have enough information to further refine the model and improve it's accuracy? – 1 point(s)

```
# 1
# par_res_elev = suppressWarnings(parres(fit1, "elev"))
# par_res_forest = suppressWarnings(parres(fit1, "forest"))
# par(mfrow = c(1,2))
# plot(par_res_elev,
#      legend = FALSE,
#      lwd = 2,
#      main = "",
#      xlab = "Elevation (m)")
# plot(par_res_forest,
#      legend = FALSE,
#      lwd = 2,
#      main = "",
#      xlab = "Forest Coverage (%)")
# par(mfrow = c(1,1))

# 2
#The terms actually seem to be doing a pretty good job. However,
#there are certainly some bumps and curves that are not being modelled.
#That being said, there is certainly room for improvement.

# 3
#No, I don't think so. The only thing I could do would be to increase
#the order of the polynomials, but as mentioned in lab, this is unstable.
#I think another relevant covariate would be needed in order to improve
#modelling accuracy.
```

Exercise 7 – 1 points

- Based on these analyses, what have you learned about the spatial distribution of parks in BC? – 1 point(s)