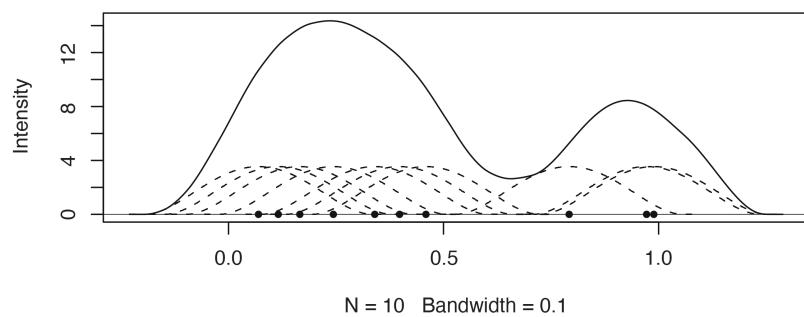


Estimating the intensity function

Estimating the intensity function

Kernel smoothing - choosing a bandwidth



ASDAR(1/2) p. (166/185)

Estimating the intensity function

Kernel smoothing - choosing a bandwidth

```
bw.diggle(pp0)
```

```
##      sigma  
## 0.05430528
```

```
pp0_density_diggle <- density(pp0, sigma = bw.diggle)  
bw.ppl(pp0)
```

```
##      sigma  
## 0.7071068
```

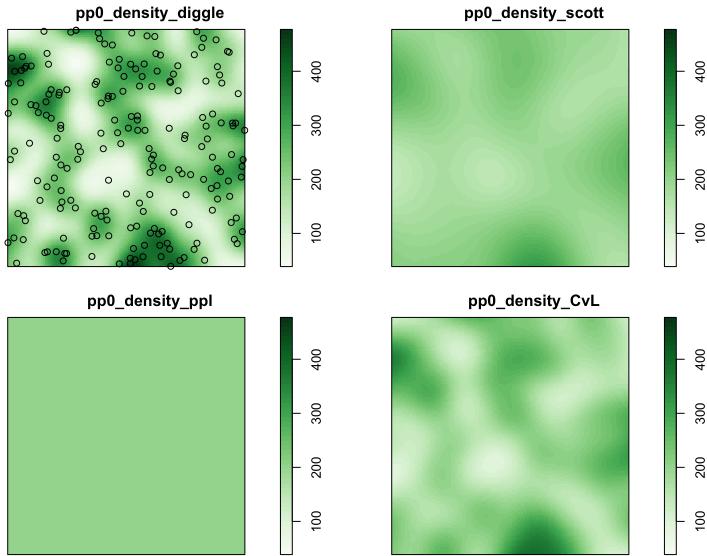
```
pp0_density_ppl <- density(pp0, sigma = bw.ppl)  
bw.scott(pp0)
```

```
## sigma.x  sigma.y  
## 0.1194348 0.1238363
```

```
pp0_density_scott <- density(pp0, sigma = bw.scott)  
bw.CvL(pp0)
```

```
##      sigma  
## 0.07660024
```

```
pp0_density_CvL <- density(pp0, sigma = bw.CvL)
```



7. Given what we know about the true intensity surface, which bandwidth is doing the best job?

```
colourmap_pp0 <- colourmap(col = colorRampPalette(RColorBrewer::brewer.pal(9, "Greens"))(5e1),  
                           range = range(pp0_density_diggle, pp0_density_ppl,  
                                         pp0_density_scott, pp0_density_CvL))  
layout(matrix(1:4, 2, 2))  
par(mar = c(1, 1, 1, 1))  
plot(pp0_density_diggle, col = colourmap_pp0)  
points(pp0)  
plot(pp0_density_ppl, col = colourmap_pp0)  
plot(pp0_density_scott, col = colourmap_pp0)  
plot(pp0_density_CvL, col = colourmap_pp0)
```

Estimating the intensity function

Kernel smoothing - choosing a bandwidth

```
bw.diggle(pp1)
```

```
##      sigma  
## 0.06066536
```

```
pp1_density_diggle <- density(pp1, sigma = bw.diggle)  
bw.ppl(pp1)
```

```
##      sigma  
## 0.1176992
```

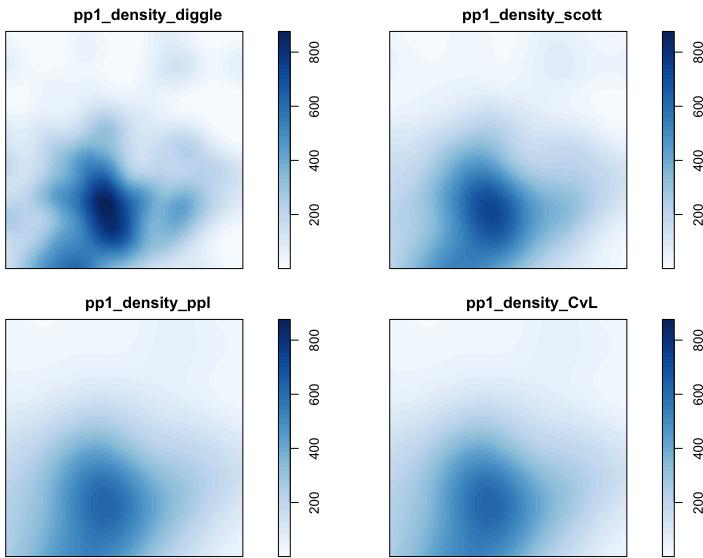
```
pp1_density_ppl <- density(pp1, sigma = bw.ppl)  
bw.scott(pp1)
```

```
##      sigma.x      sigma.y  
## 0.09089038 0.08294486
```

```
pp1_density_scott <- density(pp1, sigma = bw.scott)  
bw.CvL(pp1)
```

```
##      sigma  
## 0.1176992
```

```
pp1_density_CvL <- density(pp1, sigma = bw.CvL)
```



8. Given what we know about the true intensity surface, which bandwidth do you think is doing the best job?

```
colourmap_pp1 <- colourmap(col = colorRampPalette(RColorBrewer::brewer.pal(9, "Blues"))(5e1),  
                           range = range(pp1_density_diggle, pp1_density_ppl,  
                                         pp1_density_scott, pp1_density_CvL))  
layout(matrix(1:4, 2, 2))  
par(mar = c(1, 1, 1, 1))  
plot(pp1_density_diggle, col = colourmap_pp1)  
plot(pp1_density_ppl, col = colourmap_pp1)  
plot(pp1_density_scott, col = colourmap_pp1)  
plot(pp1_density_CvL, col = colourmap_pp1)
```

Estimating the intensity function

Kernel smoothing - choosing a bandwidth

```
bw.diggle(lightning_ppp)
```

```
##      sigma  
## 0.003498043
```

```
lightning_density_diggle <- density(lightning_ppp, sigma = bw.diggle)  
bw.ppl(lightning_ppp)
```

```
##      sigma  
## 0.02075112
```

```
lightning_density_ppl <- density(lightning_ppp, sigma = bw.ppl)  
bw.scott(lightning_ppp)
```

```
##      sigma.x      sigma.y  
## 0.09870609 0.05988603
```

```
lightning_density_scott <- density(lightning_ppp, sigma = bw.scott)  
bw.CvL(lightning_ppp)
```

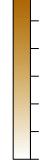
```
##      sigma  
## 0.7526689
```

```
lightning_density_CvL <- density(lightning_ppp, sigma = bw.CvL)
```

lightning_density_diggle



lightning_density_scott



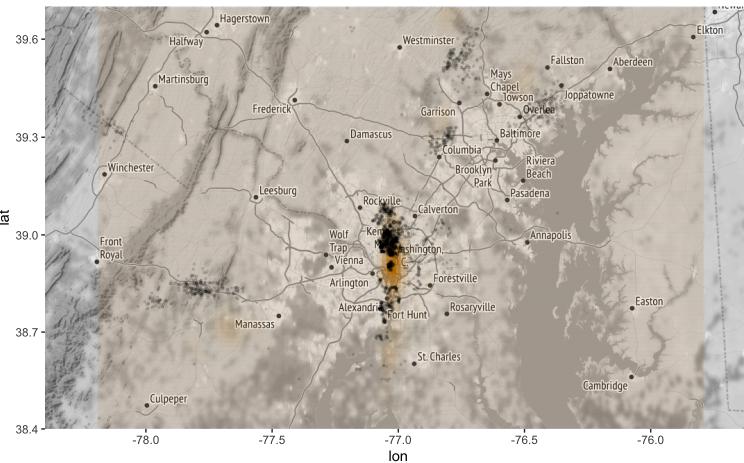
lightning_density_ppl



lightning_density_CvL



Estimating the intensity function



```
grid <- expand.grid(x = seq(-78.2, -75.8, 0.01875)[-1], y = seq(38.1, 39.9, 0.0140625)[-1])
lightning_density_df <- cbind(grid, v = c(t(lightning_density$v)))
ggmap(map) + geom_point(aes(x = x, y = y, alpha = v), shape = 15, size = 1.35,
                        data = lightning_density_df, col = "#bc7a00") +
  geom_point(aes(x = lon, y = lat), data = lightning, size = 0.5) +
  guides(alpha = F)
```

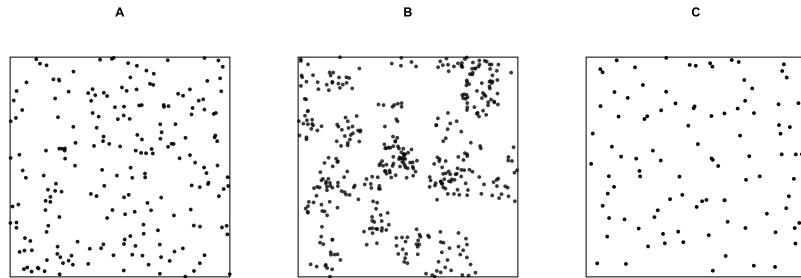
9. By modeling these data as an inhomogeneous Poisson process, we've made the implicit assumption that the lightning data are independent. Do you think that's reasonable?

2nd order properties

2nd order properties

The intensity surface captures first order properties like the mean and variance of \mathbf{s} , but it does not capture dependence across points.

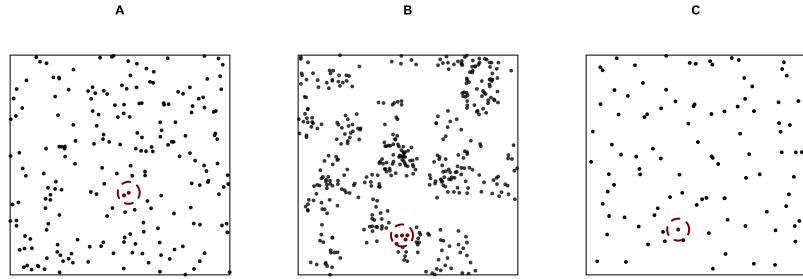
$K(r)$ is the standardized intensity of points within a disc of radius r centered on any realized point. For a CSR process $K(r) = \pi r^2$.



2nd order properties

The intensity surface captures first order properties like the mean and variance of \mathbf{s} , but it does not capture dependence across points.

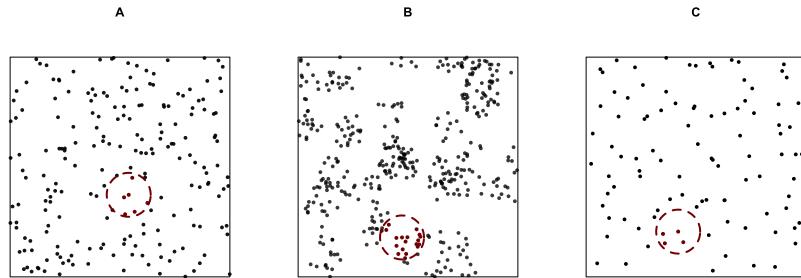
$K(r)$ is the expected density of points within a disc of radius r centered on any realized point. For a CSR process $K(r) = \pi r^2$.



2nd order properties

The intensity surface captures first order properties like the mean and variance of \mathbf{s} , but it does not capture dependence across points.

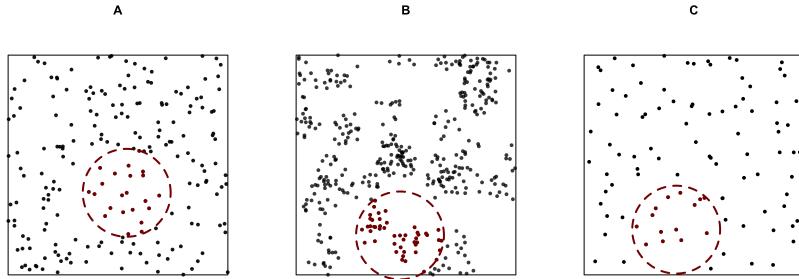
$K(r)$ is the expected density of points within a disc of radius r centered on any realized point. For a CSR process $K(r) = \pi r^2$.



2nd order properties

The intensity surface captures first order properties like the mean and variance of \mathbf{s} , but it does not capture dependence across points.

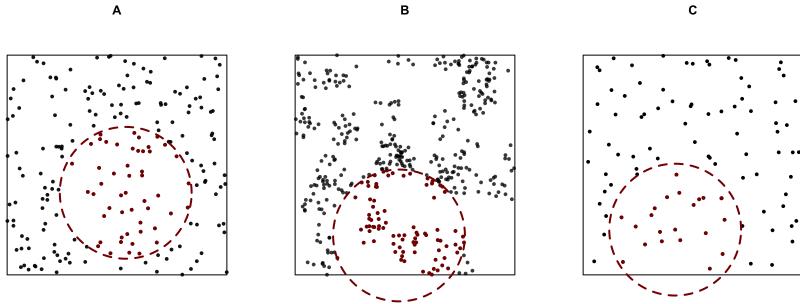
$K(r)$ is the expected density of points within a disc of radius r centered on any realized point. For a CSR process $K(r) = \pi r^2$.



2nd order properties

The intensity surface captures first order properties like the mean and variance of \mathbf{s} , but it does not capture dependence across points.

$K(r)$ is the expected density of points within a disc of radius r centered on any realized point. For a CSR process $K(r) = \pi r^2$.

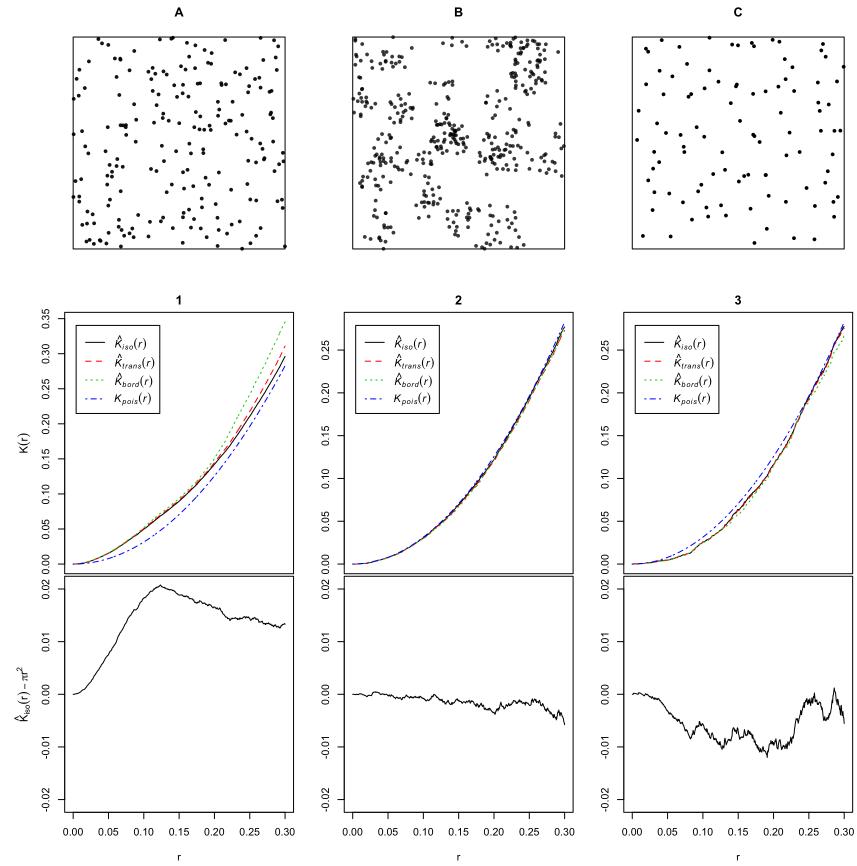


```
s <- 0.3
pts_in_circ_CSR <- which((ppCSR$x - ppCSR$x[14])^2 + (ppCSR$y - ppCSR$y[14])^2 < s^2)
pts_in_circ_Cl <- which((ppCl$x - ppCl$x[137])^2 + (ppCl$y - ppCl$y[137])^2 < s^2)
pts_in_circ_Reg <- which((ppReg$x - ppReg$x[15])^2 + (ppReg$y - ppReg$y[15])^2 < s^2)
##
layout(matrix(1:3, 1, 3))
par(mar = rep(2, 4))
plot(ppCSR, pch = 16, cex = 0.8, main = "A")
lines(ppCSR$x[14] + s * cos(thetas), ppCSR$y[14] + s * sin(thetas),
      col = "#880000", lty = 2, lwd = 2)
points(ppCSR[pts_in_circ_CSR], pch = 16, cex = 0.8, col = "#880000")
plot(ppCl, pch = 16, cex = 0.8, main = "B")
lines(ppCl$x[137] + s * cos(thetas), ppCl$y[137] + s * sin(thetas),
      col = "#880000", lty = 2, lwd = 2)
points(ppCl[pts_in_circ_Cl], pch = 16, cex = 0.8, col = "#880000")
plot(ppReg, pch = 16, cex = 0.8, main = "C")
lines(ppReg$x[15] + s * cos(thetas), ppReg$y[15] + s * sin(thetas),
      col = "#880000", lty = 2, lwd = 2)
points(ppReg[pts_in_circ_Reg], pch = 16, cex = 0.8, col = "#880000")
```

2nd order properties

The intensity surface captures first order properties like the mean and variance of \mathbf{s} , but it does not capture dependence across points.

10. All of the following processes are homogeneous Poisson processss. Match the estimate of $K(r)$ (1, 2, 3) with the correct process above (A, B, C). Which are CSR (independent)?



2nd order properties of lightning data

11. *How do you think $K(r)$ for the lightning will compare with CSR?*

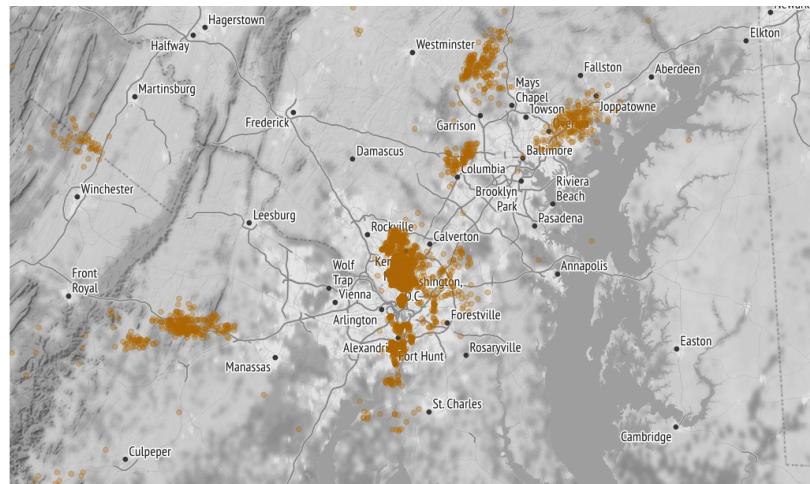
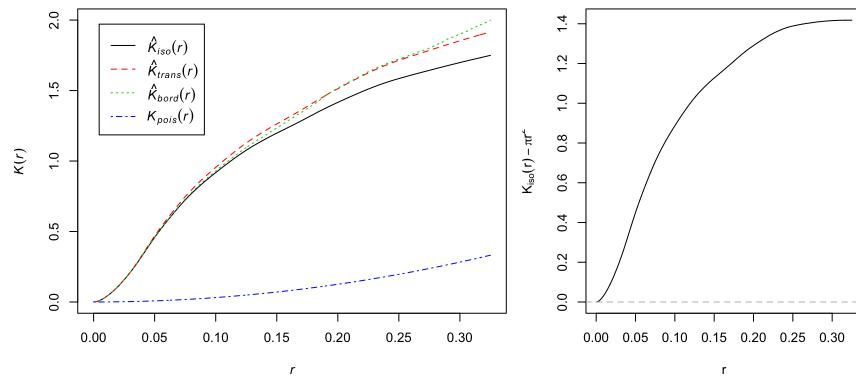
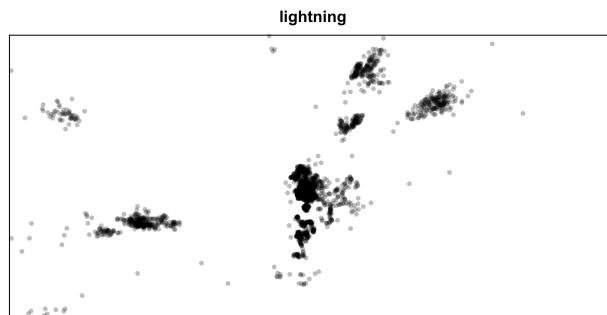


Figure 1: Sources of high power output associated with lightning strikes.

2nd order properties of lightning data



```
K_lightning <- Kest(lightning_ppp)
par(mar = c(4, 4, 1, 1))
layout(matrix(c(1, 2, 1, 3), 2, 2), widths = c(1, 0.7, 0.7))
plot(lightning_ppp, clipwin = lightning_ppp>window, main = "lightning",
  pch = 16, cex = 0.8)
plot(K_lightning, main = "")
plot(K_lightning$r, K_lightning$iso - K_lightning$theo, type = "l",
  xlab = expression(r), ylab = expression(hat(K)[iso](r) - pi*r^2))
abline(h = 0, lty = 2, col = "gray")
```

Asthma

"The data set is available from Prof. Peter J. Diggle's website..." –p. ASDAR(1/2) p. (159/177)

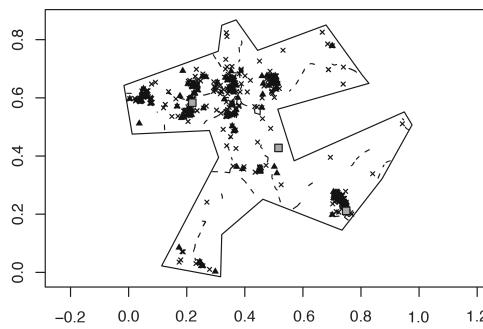
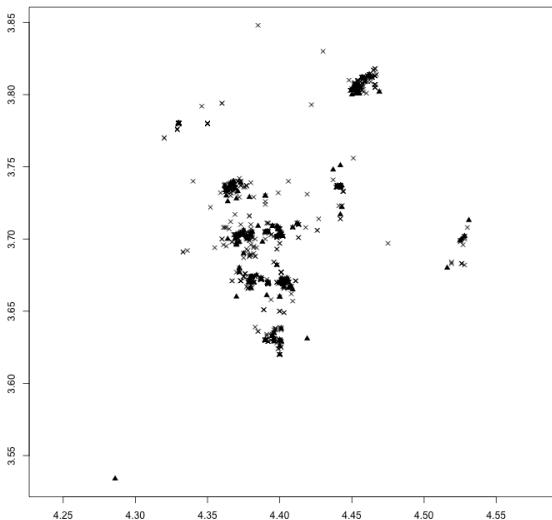


Fig. 7.2. Locations of the residence of asthmatic (cases, filled triangle) and non-asthmatic (controls, cross) in North Derbyshire, 1992 (Diggle and Rowlingson, 1994). The boundary has been taken to contain all points in the data set. The map shows the pollution sources (grey filled square) and the main roads (dashed lines)

```
library(sp)
asthma <- read.table(url("https://www.lancaster.ac.uk/staff/diggle/pointpatternbook/datasets/asthma.txt"), header = T)
coordinates(asthma) <- c('x', 'y')
plot(asthma, pch = 4 + 13 * asthma$Asthma, axes = T)
```

Asthma

"The data set is available from Prof. Peter J. Diggle's website..." –p. ASDAR(1/2) p. (159/177)

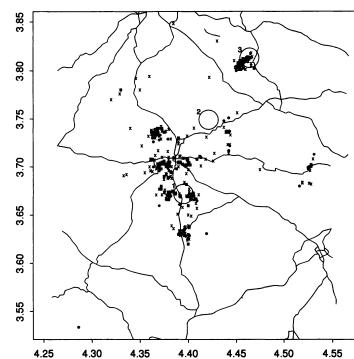
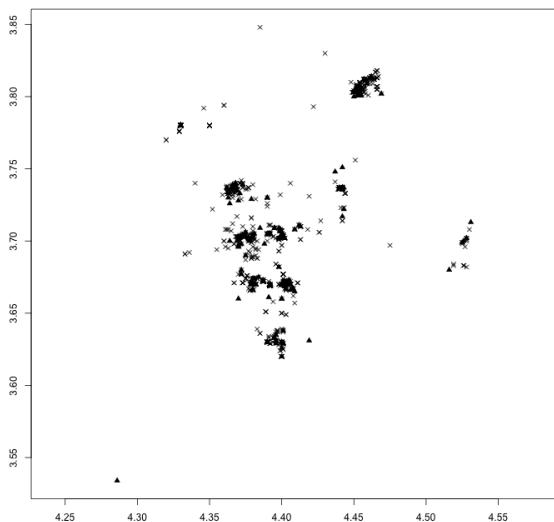


Fig. 3. Asthma in north Derbyshire: ●, cases; ×, controls; ~~, road network (the three point sources are labelled 1, 2 and 3; see text for details; the axis labels are Ordnance Survey eastings and northings, divided by 10000)

Asthma

"The data set is available from Prof. Peter J. Diggle's website..." –p. ASDAR(1/2) p. (159/177)

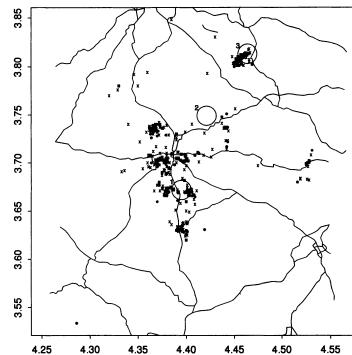
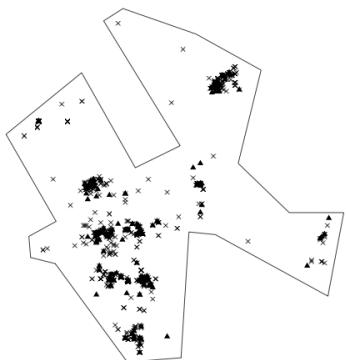


Fig. 3. Asthma in north Derbyshire: ●, cases; ×, controls; ~~, road network (the three point sources are labelled 1, 2 and 3; see text for details; the axis labels are Ordnance Survey eastings and northings, divided by 10000)



```
plot(asthma, pch = 4 + 13 * asthma$Asthma)
boundary <- data.frame(x = c(4.541321, 4.530354, 4.452367, 4.434089, 4.428606, 4.391440, 4.341480, 4.324421, 4.323202, 4.342090, 4.307361, 4.359758, 4.396924, 4.427997, 4.374990, 4.388394, 4.438964,
4.484050, 4.468209, 4.503546, 4.541321),
y = c(3.716894, 3.659013, 3.701662, 3.703490, 3.616364, 3.613318, 3.681556, 3.685821, 3.700444, 3.710801, 3.771119, 3.813768, 3.747967, 3.763199, 3.849715, 3.858245, 3.840576,
3.815596, 3.751013, 3.716894, 3.716894))
lines(boundary)
points(asthma[1252, ], pch = 1, cex = 5)
points(asthma[1252, ], pch = 4, cex = 5)
```

Asthma 7.5.3 Binary Regression Using Generalised Linear Models

Generalized linear models

$$\begin{aligned}z(\mathbf{s}_i) &\sim F_z \\ \mathbb{E}[z(\mathbf{s}_i)] &= \eta(\mathbf{s}_i) \\ g(\eta(\mathbf{s}_i)) &= \mathbf{x}'(\mathbf{s}_i)\boldsymbol{\beta}\end{aligned}$$

Asthma 7.5.3 Binary Regression Using Generalised Linear Models

Generalized linear models: cases and controls

$$\begin{aligned}z(\mathbf{s}_i) &\sim \text{Bern}(p(\mathbf{s}_i)) \\ \text{E}[z(\mathbf{s}_i)] &= p(\mathbf{s}_i) \\ \text{logit}(p(\mathbf{s}_i)) &= \mathbf{x}'(\mathbf{s}_i)\boldsymbol{\beta} \\ p(\mathbf{s}_i) &= \frac{\lambda_1(\mathbf{s}_i)}{\lambda_0(\mathbf{s}_i) + \lambda_1(\mathbf{s}_i)}\end{aligned}$$

12. Where is the point process part?

Asthma 7.5.3 Binary Regression Using Generalised Linear Models

Generalized linear models: cases and controls

$$\begin{aligned}z(\mathbf{s}_i) &\sim \text{Bern}(p(\mathbf{s}_i)) \\ \text{E}[z(\mathbf{s}_i)] &= p(\mathbf{s}_i) \\ \text{logit}(p(\mathbf{s}_i)) &= \mathbf{x}'(\mathbf{s}_i)\boldsymbol{\beta} \\ p(\mathbf{s}_i) &= \frac{\lambda_1(\mathbf{s}_i)}{\lambda_0(\mathbf{s}_i) + \lambda_1(\mathbf{s}_i)}\end{aligned}$$

```
head(asthma)
```

```
##      coordinates School Nsmokers Asthma HayFever Age Gender YearsHere
## 1 (4.396, 3.684)     1       2     0     0 8.9     1     8.9
## 2 (4.402, 3.672)     1       2     0     0 8.5     2     6.6
## 3 (4.402, 3.671)     1       2     0     1 8.5     2     8.5
## 4 (4.402, 3.672)     1       1     0     0 8.9     2     2.5
## 5 (4.402, 3.672)     1       1     0     0 6.9     1     1.9
## 6 (4.404, 3.671)     1       1     1     0 7.1     2     7.1
##   Income d2road d2source1 d2source2 d2source3
## 1      5 0.000089 0.000145 0.004850 0.021265
## 2      2 0.000089 0.000049 0.006290 0.023725
## 3      5 0.000083 0.000050 0.006445 0.024008
## 4      5 0.000089 0.000049 0.006290 0.023725
## 5      2 0.000089 0.000049 0.006290 0.023725
## 6      5 0.000121 0.000082 0.006373 0.023764
```

```
table(asthma$Asthma)
```

```
## 
##    0    1
## 1076 216
```

Asthma 7.5.3 Binary Regression Using Generalised Linear Models

```
asthma$smoking <- asthma$Nsmokers > 0
asthma_sub <- asthma[-1252, ] # omit far-flung point as per ASDAR
asthma_mf <- asthma_sub[asthma_sub$Gender == 1 | asthma_sub$Gender == 2, ]
asthma_glm <- glm(Asthma ~ I(sqrt(d2source1)) + I(sqrt(d2source2)) + I(sqrt(d2source3)) +
  I(sqrt(d2road)) + Gender + Age + HayFever + smoking,
  data = asthma_mf, family = binomial(link = "logit"))
summary(asthma_glm)
```

```
## 
## Call:
## glm(formula = Asthma ~ I(sqrt(d2source1)) + I(sqrt(d2source2)) +
##     I(sqrt(d2source3)) + I(sqrt(d2road)) + Gender + Age + HayFever +
##     smoking, family = binomial(link = "logit"), data = asthma_mf)
##
## Deviance Residuals:
##       Min      1Q   Median      3Q      Max
## -1.1611 -0.6096 -0.5333 -0.4492  2.3021
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.67741   0.60394 -1.122   0.2620
## I(sqrt(d2source1)) -1.70858   3.14550 -0.543   0.5870
## I(sqrt(d2source2)) -3.62168   4.91854 -0.736   0.4615
## I(sqrt(d2source3))  1.81377   3.40945  0.532   0.5947
## I(sqrt(d2road))    2.72382  16.36484  0.166   0.8678
## Gender        -0.37063   0.15527 -2.387   0.0170 *
## Age           -0.06695   0.03737 -1.792   0.0732 .
## HayFever      1.14370   0.18566  6.160 7.26e-10 ***
## smokingTRUE    0.23696   0.15679  1.511   0.1307
## ---
## Signif. codes:  0 '****' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 1159.9 on 1282 degrees of freedom
## Residual deviance: 1109.0 on 1274 degrees of freedom
## AIC: 1127
##
## Number of Fisher Scoring iterations: 4
```

13. Where is the spatial part?

Asthma 7.5.3 Binary Regression Using Generalised Additive Models

Generalized linear models: cases and controls

$$\begin{aligned} z(\mathbf{s}_i) &\sim \text{Bern}(p(\mathbf{s}_i)) \\ \text{E}[z(\mathbf{s}_i)] &= p(\mathbf{s}_i) \\ \text{logit}(p(\mathbf{s}_i)) &= \mathbf{x}'(\mathbf{s}_i)\boldsymbol{\beta} + \underbrace{\varepsilon(\mathbf{s}_i)}_{\text{small scale spatial dependence}} \\ p(\mathbf{s}_i) &= \frac{\lambda_1(\mathbf{s}_i)}{\lambda_0(\mathbf{s}_i) + \lambda_1(\mathbf{s}_i)} \end{aligned}$$

Asthma 7.5.3 Binary Regression Using Generalised Additive Models

Generalized linear models: cases and controls

$$\begin{aligned} z(\mathbf{s}_i) &\sim \text{Bern}(p(\mathbf{s}_i)) \\ \text{E}[z(\mathbf{s}_i)] &= p(\mathbf{s}_i) \\ \text{logit}(p(\mathbf{s}_i)) &= \mathbf{x}'(\mathbf{s}_i)\boldsymbol{\beta} + \underbrace{\mathbf{h}'(\mathbf{s}_i)\boldsymbol{\alpha}}_{\text{small scale spatial dependence}} \\ p(\mathbf{s}_i) &= \frac{\lambda_1(\mathbf{s}_i)}{\lambda_0(\mathbf{s}_i) + \lambda_1(\mathbf{s}_i)} \end{aligned}$$

Asthma 7.5.3 Binary Regression Using Generalised Additive Models

Generalized linear models: cases and controls

$$\begin{aligned} z(\mathbf{s}_i) &\sim \text{Bern}(p(\mathbf{s}_i)) \\ \text{E}[z(\mathbf{s}_i)] &= p(\mathbf{s}_i) \\ \text{logit}(\mathbf{p}) &= \mathbf{X}\boldsymbol{\beta} + \underbrace{\mathbf{H}\boldsymbol{\alpha}}_{\text{small scale spatial dependence}} \\ p(\mathbf{s}_i) &= \frac{\lambda_1(\mathbf{s}_i)}{\lambda_0(\mathbf{s}_i) + \lambda_1(\mathbf{s}_i)} \\ \boldsymbol{\alpha} &\sim N(\mathbf{0}, \sigma^2 \mathbf{I}) \\ \mathbf{H}\boldsymbol{\alpha} &\sim N(\mathbf{H}\mathbf{0}, \mathbf{H}\sigma^2 \mathbf{I}\mathbf{H}') \\ \mathbf{H}\boldsymbol{\alpha} &\sim N(\mathbf{0}, \sigma^2 \mathbf{H}\mathbf{H}') \end{aligned}$$

Asthma 7.5.3 Binary Regression Using Generalised Additive Models

Generalized linear models: cases and controls

$$z(\mathbf{s}_i) \sim \text{Bern}(p(\mathbf{s}_i))$$

$$\mathbb{E}[z(\mathbf{s}_i)] = p(\mathbf{s}_i)$$

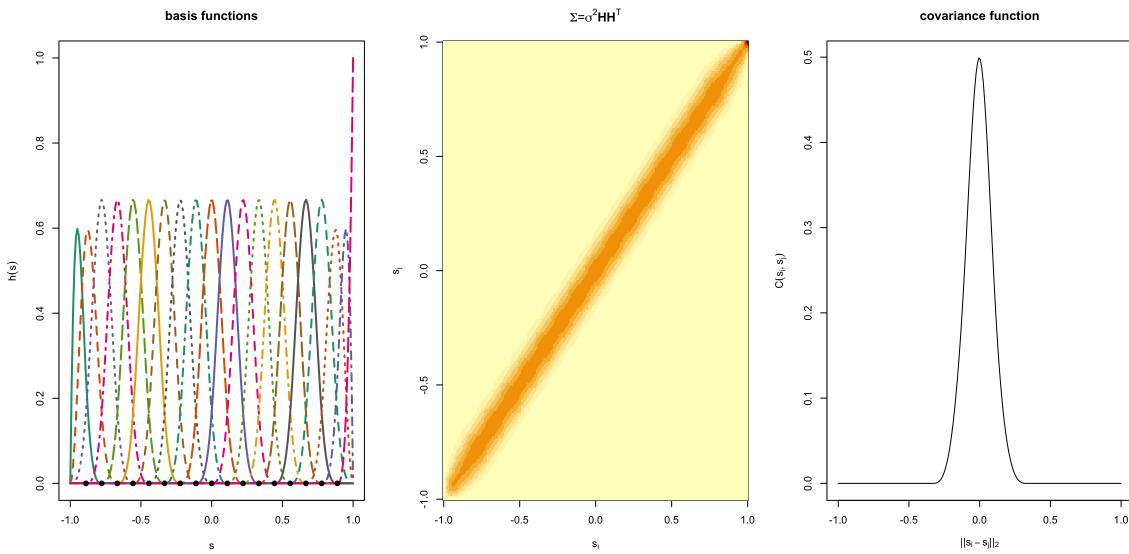
$$\text{logit}(\mathbf{p}) = \mathbf{X}\boldsymbol{\beta} + \underbrace{\mathbf{H}\boldsymbol{\alpha}}_{\text{small scale spatial dependence}}$$

$$p(\mathbf{s}_i) = \frac{\lambda_1(\mathbf{s}_i)}{\lambda_0(\mathbf{s}_i) + \lambda_1(\mathbf{s}_i)}$$

$$\boldsymbol{\alpha} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$$

$$\mathbf{H}\boldsymbol{\alpha} \sim N(\mathbf{H}\mathbf{0}, \mathbf{H}\sigma^2 \mathbf{I}\mathbf{H}')$$

$$\mathbf{H}\boldsymbol{\alpha} \sim N(\mathbf{0}, \sigma^2 \mathbf{H}\mathbf{H}')$$



```

library(splines)
H <- bs(x = seq(-1, 1, l = 2e2), df = 2e1)
layout(matrix(1:3, 1, 3))
matplot(seq(-1, 1, l = 2e2), H, type = "l", col = RColorBrewer::brewer.pal(8, "Dark2"), lwd = 2,
       xlab = "s", main = "basis functions", ylab = expression(h(s)))
points(attr(H, 'knots'), rep(0, 17), cex = 1.2, pch = 16)
Hht <- H %>% t(H)
image(seq(-1, 1, l = 2e2), seq(-1, 1, l = 2e2), Hht,
      main = expression(bold(Sigma) == sigma^2 * bold(H) * bold(H)^T),
      xlab = expression(s[i]), ylab = expression(s[j]))
plot(seq(-1, 1, l = 2e2), Hht[100, ], type = "l", xlab = expression("||" * s[i] - s[j] * ||^2), ylab = expression(C(s[i], s[j])), main = "covariance function")

```

Asthma 7.5.3 Binary Regression Using Generalised Additive Models

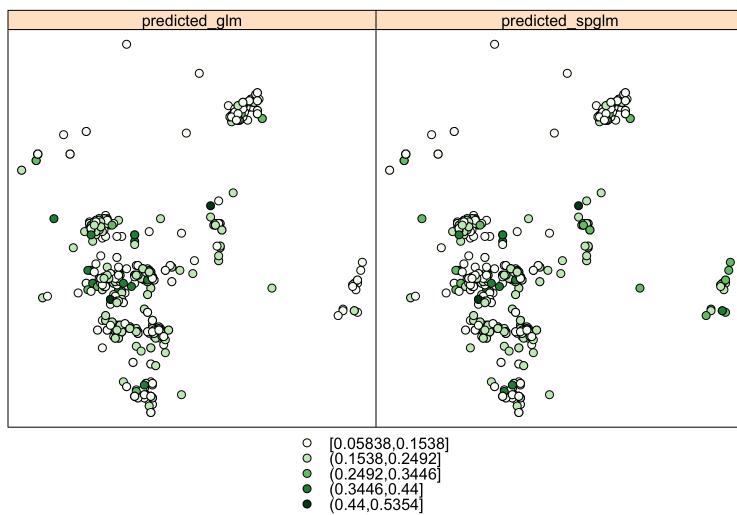
```
library(mgcv)
asthma_spglm <- gam(Asthma ~ I(sqrt(d2source1)) + I(sqrt(d2source2)) + I(sqrt(d2source3)) +
  I(sqrt(d2road)) + Gender + Age + HayFever + smoking + s(x, y),
  data = asthma_mf, family = binomial(link = "logit"))
summary(asthma_spglm)

##
## Family: binomial
## Link function: logit
##
## Formula:
## Asthma ~ I(sqrt(d2source1)) + I(sqrt(d2source2)) + I(sqrt(d2source3)) +
##   I(sqrt(d2road)) + Gender + Age + HayFever + smoking + s(x,
##   y)
##
## Parametric coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.68170  0.94734 -1.775  0.0759 .
## I(sqrt(d2source1)) 0.97439  6.07158  0.160  0.8725
## I(sqrt(d2source2)) -9.55514  5.77131 -1.656  0.0978 .
## I(sqrt(d2source3)) 11.22923  7.87250  1.426  0.1538
## I(sqrt(d2road)) 14.26143 17.10988  0.834  0.4046
## Gender      -0.34771  0.15620 -2.226  0.0260 *
## Age        -0.06795  0.03823 -1.777  0.0755 .
## HayFever    1.18775  0.18753  6.334 2.39e-10 ***
## smokingTRUE  0.16526  0.16103  1.026  0.3048
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##          edf Ref.df Chi.sq p-value
## s(x,y)  2  2.001  6.963  0.0308 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.0403  Deviance explained = 4.94%
## UBRE = -0.12344  Scale est. = 1      n = 1283
```

14. What differences do you see in the regression coefficient estimates?

Asthma 7.5.3 Binary Regression Using Generalised Additive Models

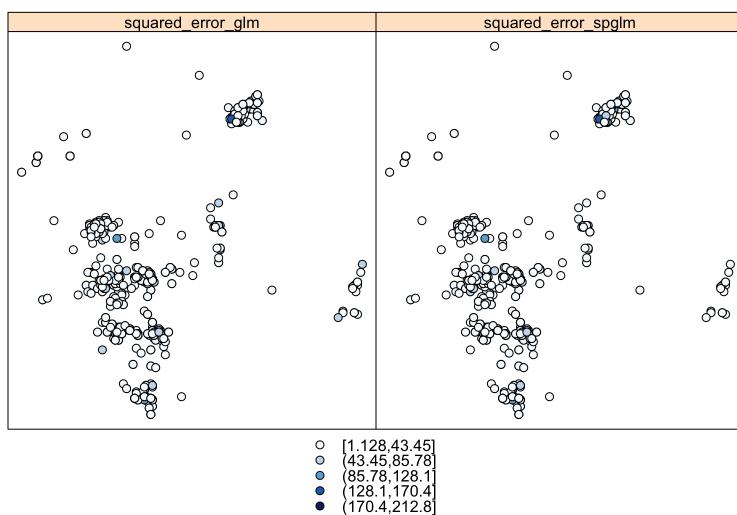
```
asthma_mf$predicted_glm <- asthma_glm$fitted.values  
asthma_mf$predicted_spglm <- asthma_spglm$fitted.values  
spplot(asthma_mf, c('predicted_glm', 'predicted_spglm'),  
       edge.col = "black", col.regions = RColorBrewer::brewer.pal(9, "Greens"))
```



15. What do the predicted values represent (e.g., what does a predicted value near 0 mean)?

16. Where do you see the biggest differences in the predictions between the models?

```
asthma_mf$squared_error_glm <- (asthma_glm$residuals)^2  
asthma_mf$squared_error_spglm <- (asthma_spglm$residuals)^2  
spplot(asthma_mf, c('squared_error_glm', 'squared_error_spglm'),  
       edge.col = "black", col.regions = RColorBrewer::brewer.pal(9, "Blues"))
```



17. What differences do you see in the squared errors of the two models?