Week 6/7 Questions

Dr L Scott-Hayward

Introduction

We are going to use the same data as for the last four weeks of the 'Weekly Questions' to look at regression splines with model selection chosen using the spatially adaptive local smoothing algorithm (SALSA). There will also be some assessment of correlation and questions associated with GEEs.

2D Penalised regression spline using the `mgcv` library

```r
newdat <- read.table("../data/dataForWeeklyQuestions.csv", header = T)
newdat$response <- newdat$tobinsQ
```

```r
require(mgcv)
penreg2D <- gam(response ~ s(rd, ads) + as.factor(indclass), data = newdat)
summary(penreg2D)
```

```
Family: gaussian
Link function: identity

Formula:
response ~ s(rd, ads) + as.factor(indclass)

Parametric coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)            2.39349    1.07856   2.219  0.02649 *
as.factor(indclass)2   1.92353    1.12445   1.711  0.08717 .
as.factor(indclass)3  -1.60480    1.80770  -0.888  0.37469
as.factor(indclass)4  -1.83904    1.32307  -1.390  0.16456
as.factor(indclass)5   0.03692    1.72710   0.021  0.98295
as.factor(indclass)6  -0.40108    1.12433  -0.357  0.72130
as.factor(indclass)7   2.05538    1.14861   1.789  0.07357 .
as.factor(indclass)8   4.44531    1.35617   3.278  0.00105 **
as.factor(indclass)9   0.54122    1.11655   0.485  0.62788
as.factor(indclass)10 -0.24208    1.15218  -0.210  0.83359
as.factor(indclass)11 -0.01620    1.17912  -0.014  0.98904
as.factor(indclass)12  0.90374    1.09465   0.826  0.40904
as.factor(indclass)13  2.69947    1.09323   2.469  0.01355 *
as.factor(indclass)14  0.74328    1.13263   0.656  0.51168
as.factor(indclass)15  0.40993    1.22086   0.336  0.73705
as.factor(indclass)16 -1.58201    2.03339  -0.778  0.43657
as.factor(indclass)17 -0.12984    1.13479  -0.114  0.90891
as.factor(indclass)18 -1.27812    1.62764  -0.785  0.43232
as.factor(indclass)19 -0.87646    1.28861  -0.680  0.49641
as.factor(indclass)20 -0.63704    2.12276  -0.300  0.76411
as.factor(indclass)21  0.42278    1.10020   0.384  0.70078
as.factor(indclass)22 -0.14388    1.10857  -0.130  0.89674
as.factor(indclass)23  0.19374    1.12890   0.172  0.86374
```

```
as.factor(indclass)24  0.23033    1.75474   0.131  0.89557
as.factor(indclass)25  1.44507    1.84168   0.785  0.43268
as.factor(indclass)26  1.99296    1.29516   1.539  0.12388
as.factor(indclass)28 -0.51627    1.52488  -0.339  0.73494
as.factor(indclass)30  2.27336    2.37068   0.959  0.33760
as.factor(indclass)32  0.36045    1.14392   0.315  0.75269
as.factor(indclass)33 -0.41097    1.28199  -0.321  0.74854
as.factor(indclass)34  1.75604    1.12763   1.557  0.11943
as.factor(indclass)35  0.37248    1.09812   0.339  0.73446
as.factor(indclass)36  2.18453    1.08398   2.015  0.04389 *
as.factor(indclass)37  0.03178    1.08915   0.029  0.97672
as.factor(indclass)38  0.42224    1.10115   0.383  0.70139
as.factor(indclass)39  1.27813    1.18834   1.076  0.28215
as.factor(indclass)41  6.01025    1.89725   3.168  0.00154 **
as.factor(indclass)42  0.20230    1.11376   0.182  0.85587
as.factor(indclass)43 -0.02245    1.08702  -0.021  0.98352
as.factor(indclass)44  0.58689    1.09593   0.536  0.59230
as.factor(indclass)49 -1.31113    1.71651  -0.764  0.44498
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
            edf Ref.df     F p-value
s(rd,ads) 15.33   19.9 2.077 0.00329 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.0369   Deviance explained = 4.08%
GCV = 26.848  Scale est. = 26.736    n = 13525
```

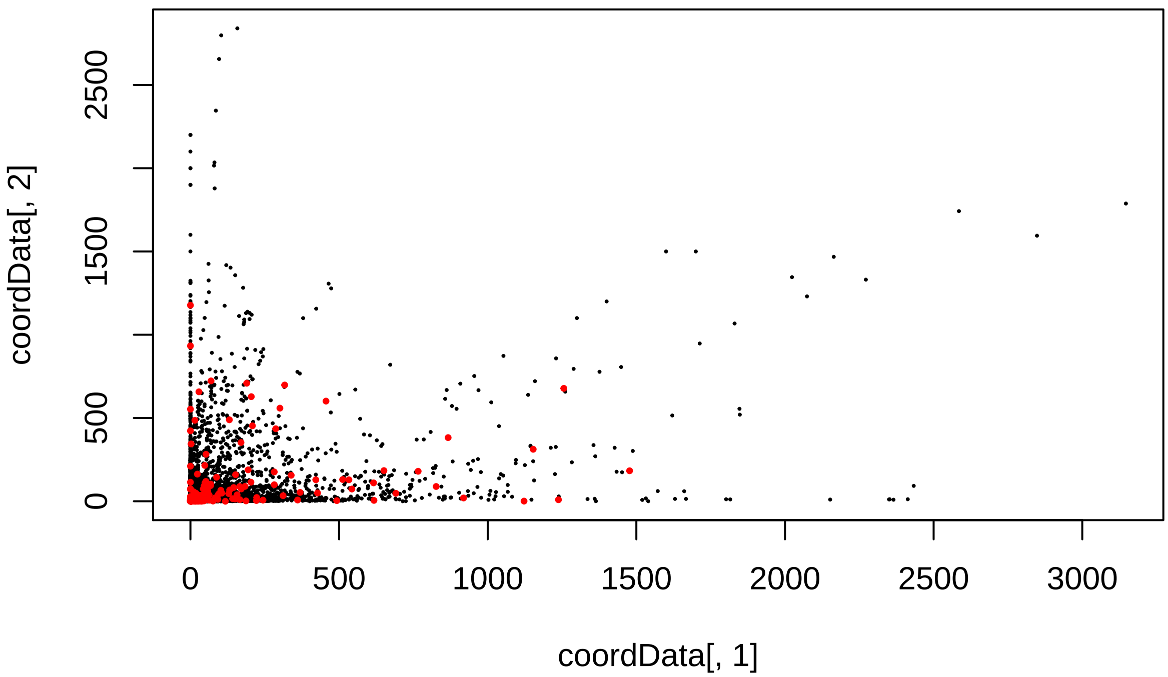
```r
require(boot)
set.seed(123)
cvpenreg2D <- cv.glm(penreg2D, K = 10, data = newdat)
cvpenreg2D$delta[2]
```

```
[1] 26.76536
```

CReSS with SALSA-based selection using the `MRSea` package.

```r
require(MRSea)
newdat$response <- newdat$tobinsQ
newdat$x.pos <- newdat$rd
newdat$y.pos <- newdat$ads
```

```r
knotgrid <- getKnotgrid(newdat[, c("x.pos", "y.pos")])
```




```
distMats <- makeDists(newdat[, c("x.pos", "y.pos")], na.omit(knotgrid))
```

Model 1

```
initialModel <- glm(response ~ as.factor(indclass), data = newdat)

# make parameter set for running salsa2d
salsa2dlist <- list(fitnessMeasure = "BIC", knotgrid = knotgrid, startKnots = 15,
    minKnots = 2, maxKnots = 20, gap = 0, interactionTerm = NULL)

salsa2dOutput <- runSALSA2D(initialModel, salsa2dlist, d2k = distMats$dataDist, k2k = distMats$knotDist
    splineParams = NULL, suppress.printout = TRUE)
```

```
summary(salsa2dOutput$bestModel)
```

```
Call:
gamMRSea(formula = response ~ as.factor(indclass) + LRF.g(radiusIndices,
    dists, radii, aR), data = newdat, splineParams = splineParams)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
 -6.074   -1.735   -0.881    0.440  188.673

Coefficients:
```

```
                                           Estimate Std. Error Robust S.E.
(Intercept)                               2.698e+00  1.093e+00   1.093e+00
as.factor(indclass)2                      2.116e+00  1.120e+00   1.120e+00
as.factor(indclass)3                     -1.944e+00  1.812e+00   1.812e+00
as.factor(indclass)4                     -1.899e+00  1.320e+00   1.320e+00
as.factor(indclass)5                      7.169e-01  1.740e+00   1.740e+00
as.factor(indclass)6                     -3.395e-01  1.123e+00   1.123e+00
as.factor(indclass)7                      2.030e+00  1.149e+00   1.149e+00
as.factor(indclass)8                      4.524e+00  1.354e+00   1.354e+00
as.factor(indclass)9                      6.190e-01  1.116e+00   1.116e+00
as.factor(indclass)10                    -2.583e-01  1.153e+00   1.153e+00
as.factor(indclass)11                     2.213e-02  1.179e+00   1.179e+00
as.factor(indclass)12                     9.073e-01  1.094e+00   1.094e+00
as.factor(indclass)13                     2.697e+00  1.093e+00   1.093e+00
as.factor(indclass)14                     7.505e-01  1.132e+00   1.132e+00
as.factor(indclass)15                     4.549e-01  1.221e+00   1.221e+00
as.factor(indclass)16                    -1.529e+00  2.033e+00   2.033e+00
as.factor(indclass)17                    -1.014e-01  1.135e+00   1.135e+00
as.factor(indclass)18                    -1.250e+00  1.627e+00   1.627e+00
as.factor(indclass)19                    -8.370e-01  1.288e+00   1.288e+00
as.factor(indclass)20                    -5.802e-01  2.122e+00   2.122e+00
as.factor(indclass)21                     4.501e-01  1.100e+00   1.100e+00
as.factor(indclass)22                    -1.340e-01  1.108e+00   1.108e+00
as.factor(indclass)23                     1.369e-01  1.126e+00   1.126e+00
as.factor(indclass)24                     1.475e-01  1.754e+00   1.754e+00
as.factor(indclass)25                     1.479e+00  1.841e+00   1.841e+00
as.factor(indclass)26                     2.001e+00  1.295e+00   1.295e+00
as.factor(indclass)28                    -5.160e-01  1.525e+00   1.525e+00
as.factor(indclass)30                     2.324e+00  2.370e+00   2.370e+00
as.factor(indclass)32                     3.875e-01  1.143e+00   1.143e+00
as.factor(indclass)33                    -4.512e-01  1.282e+00   1.282e+00
as.factor(indclass)34                     1.754e+00  1.127e+00   1.127e+00
as.factor(indclass)35                     3.480e-01  1.098e+00   1.098e+00
as.factor(indclass)36                     2.170e+00  1.084e+00   1.084e+00
as.factor(indclass)37                    -5.504e-03  1.089e+00   1.089e+00
as.factor(indclass)38                     4.019e-01  1.101e+00   1.101e+00
as.factor(indclass)39                     1.350e+00  1.184e+00   1.184e+00
as.factor(indclass)41                     6.011e+00  1.896e+00   1.896e+00
as.factor(indclass)42                     2.103e-01  1.114e+00   1.114e+00
as.factor(indclass)43                    -7.844e-02  1.087e+00   1.087e+00
as.factor(indclass)44                     5.706e-01  1.096e+00   1.096e+00
as.factor(indclass)49                    -1.257e+00  1.716e+00   1.716e+00
LRF.g(radiusIndices, dists, radii, aR)b1  2.383e+00  5.744e-01   5.744e-01
LRF.g(radiusIndices, dists, radii, aR)b2 -1.444e+03  3.984e+02   3.984e+02
LRF.g(radiusIndices, dists, radii, aR)b3  5.631e+03  1.548e+03   1.548e+03
LRF.g(radiusIndices, dists, radii, aR)b4 -4.503e+03  1.237e+03   1.237e+03
LRF.g(radiusIndices, dists, radii, aR)b5  3.144e+02  8.696e+01   8.696e+01
                                         t value Pr(>|t|)
(Intercept)                                2.468 0.013590 *
as.factor(indclass)2                       1.890 0.058825 .
as.factor(indclass)3                      -1.073 0.283401
as.factor(indclass)4                      -1.438 0.150397
as.factor(indclass)5                       0.412 0.680437
as.factor(indclass)6                      -0.302 0.762497
```

```
as.factor(indclass)7                         1.768 0.077154 .
as.factor(indclass)8                         3.340 0.000840 ***
as.factor(indclass)9                         0.555 0.579103
as.factor(indclass)10                       -0.224 0.822724
as.factor(indclass)11                        0.019 0.985023
as.factor(indclass)12                        0.829 0.407090
as.factor(indclass)13                        2.467 0.013619 *
as.factor(indclass)14                        0.663 0.507368
as.factor(indclass)15                        0.373 0.709427
as.factor(indclass)16                       -0.752 0.452027
as.factor(indclass)17                       -0.089 0.928824
as.factor(indclass)18                       -0.768 0.442538
as.factor(indclass)19                       -0.650 0.515795
as.factor(indclass)20                       -0.273 0.784600
as.factor(indclass)21                        0.409 0.682390
as.factor(indclass)22                       -0.121 0.903751
as.factor(indclass)23                        0.122 0.903246
as.factor(indclass)24                        0.084 0.932994
as.factor(indclass)25                        0.803 0.422012
as.factor(indclass)26                        1.545 0.122280
as.factor(indclass)28                       -0.338 0.735009
as.factor(indclass)30                        0.980 0.326903
as.factor(indclass)32                        0.339 0.734477
as.factor(indclass)33                       -0.352 0.724854
as.factor(indclass)34                        1.556 0.119721
as.factor(indclass)35                        0.317 0.751233
as.factor(indclass)36                        2.002 0.045281 *
as.factor(indclass)37                       -0.005 0.995966
as.factor(indclass)38                        0.365 0.715062
as.factor(indclass)39                        1.140 0.254356
as.factor(indclass)41                        3.170 0.001528 **
as.factor(indclass)42                        0.189 0.850245
as.factor(indclass)43                       -0.072 0.942490
as.factor(indclass)44                        0.521 0.602624
as.factor(indclass)49                       -0.732 0.464102
LRF.g(radiusIndices, dists, radii, aR)b1   4.148 3.38e-05 ***
LRF.g(radiusIndices, dists, radii, aR)b2  -3.625 0.000290 ***
LRF.g(radiusIndices, dists, radii, aR)b3   3.637 0.000277 ***
LRF.g(radiusIndices, dists, radii, aR)b4  -3.640 0.000274 ***
LRF.g(radiusIndices, dists, radii, aR)b5   3.615 0.000301 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 26.72598)

    Null deviance: 375420  on 13524  degrees of freedom
Residual deviance: 360239  on 13479  degrees of freedom
AIC:  82868

Max Panel Size = 1 (independence assumed); Number of panels = 13525
Number of Fisher Scoring iterations: 2
```

```
anova(salsa2dOutput$bestModel)
```

```
Analysis of 'Wald statistic' Table
```

```
Model: gaussian, link: identity
Response: response
Marginal Testing
Max Panel Size = 1 (independence assumed); Number of panels = 13525

                   Df     X2 P(>|Chi|)
as.factor(indclass) 40 503.48 < 2.2e-16 ***
s(x.pos, y.pos)      5  47.73  4.04e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
set.seed(123)
cv.gamMRSea(newdat, modelobject = salsa2dOutput$bestModel, K = 10)$delta[2]
```

```
[1] 26.75464
```

```r
BIC(salsa2dOutput$bestModel)
```

```
[1] 83221.51
```

```r
salsa2dOutput$bestModel$splineParams[[1]]$knotPos
```

```
[1] 283  65 287 192 181
```

Model 2:

```r
# make parameter set for running salsa2d
salsa2dlist <- list(fitnessMeasure = "BIC", knotgrid = knotgrid, startKnots = 10,
    minKnots = 2, maxKnots = 20, gap = 0, interactionTerm = NULL)

salsa2dOutputk10 <- runSALSA2D(initialModel, salsa2dlist, d2k = distMats$dataDist,
    k2k = distMats$knotDist, splineParams = NULL, suppress.printout = TRUE)
```

```r
set.seed(123)
cv.gamMRSea(newdat, salsa2dOutputk10$bestModel, K = 10)$delta[2]
```

```
[1] 26.78579
```

```r
BIC(salsa2dOutputk10$bestModel)
```

```
[1] 83228.71
```

```r
anova(salsa2dOutputk10$bestModel)
```

```
Analysis of 'Wald statistic' Table
Model: gaussian, link: identity
Response: response
Marginal Testing
Max Panel Size = 1 (independence assumed); Number of panels = 13525

                   Df     X2 P(>|Chi|)
as.factor(indclass) 40 484.80 < 2.2e-16 ***
s(x.pos, y.pos)      4  31.03 3.022e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
salsa2dOutputk10$bestModel$splineParams[[1]]$knotPos
```

```
[1] 120 202 207  65
```

Model 3:

```r
# make parameter set for running salsa2d
salsa2dlist <- list(fitnessMeasure = "BIC", knotgrid = knotgrid, startKnots = 5,
    minKnots = 2, maxKnots = 20, gap = 0, interactionTerm = NULL)

salsa2dOutputk5 <- runSALSA2D(initialModel, salsa2dlist, d2k = distMats$dataDist,
    k2k = distMats$knotDist, splineParams = NULL, suppress.printout = TRUE)
```

```r
set.seed(123)
cv.gamMRSea(newdat, salsa2dOutputk5$bestModel, K = 10)$delta[2]
```

```
[1] 26.80112
```

```r
BIC(salsa2dOutputk5$bestModel)
```

```
[1] 83219.31
```

```r
anova(salsa2dOutputk5$bestModel)
```

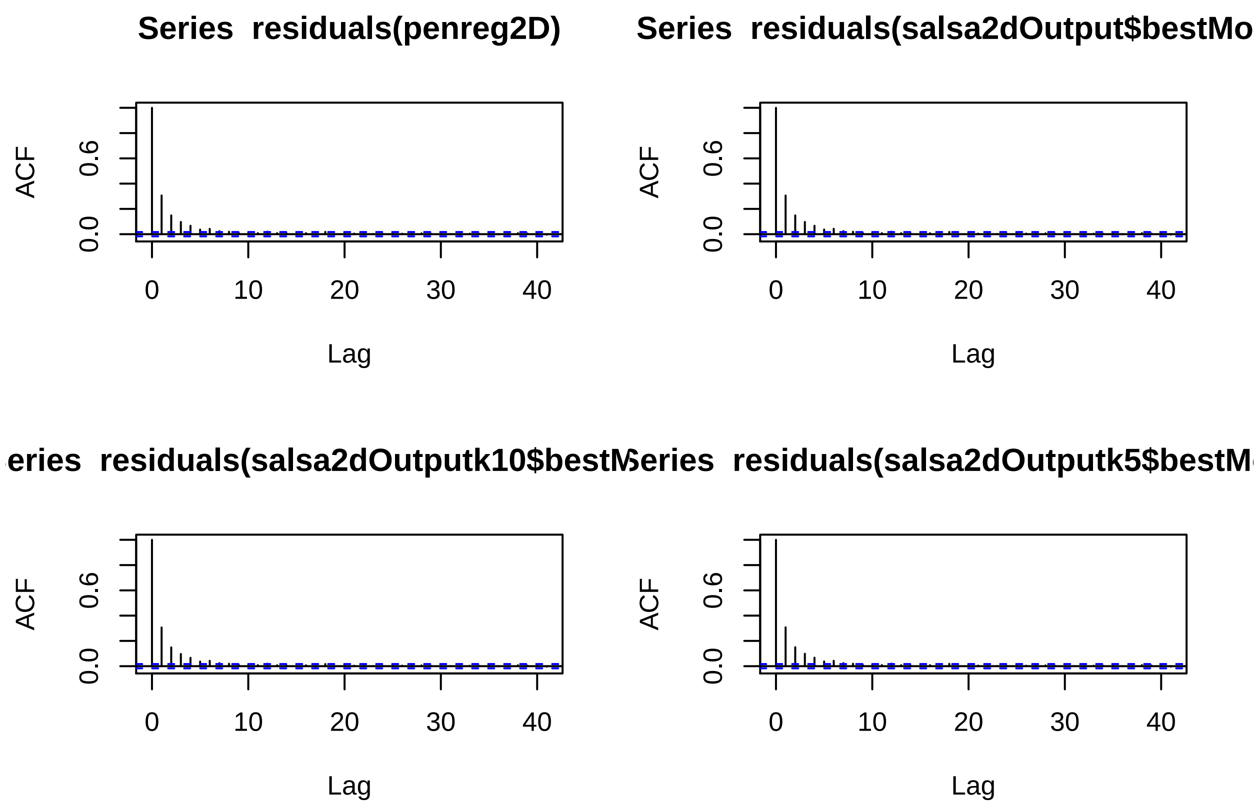
```
Analysis of 'Wald statistic' Table
Model: gaussian, link: identity
Response: response
Marginal Testing
Max Panel Size = 1 (independence assumed); Number of panels = 13525

                   Df     X2 P(>|Chi|)
as.factor(indclass) 40 514.63 < 2.2e-16 ***
s(x.pos, y.pos)      2  21.42 2.228e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
salsa2dOutputk5$bestModel$splineParams[[1]]$knotPos
```

```
[1]  55 270
```

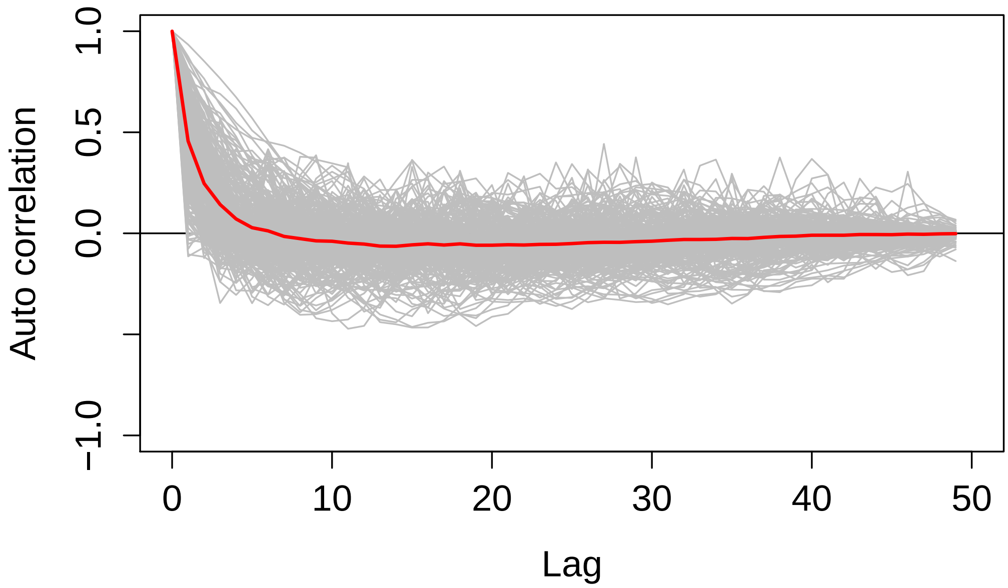
```r
par(mfrow = c(2, 2))
acf(residuals(penreg2D))
acf(residuals(salsa2dOutput$bestModel))
acf(residuals(salsa2dOutputk10$bestModel))
acf(residuals(salsa2dOutputk5$bestModel))
```




```r
newdat$panels <- rep(1:1353, each = 50)[1:13525]
```

Note: I have used an arbitrary choice of model for the correlation assessment and analysis that follows

```r
runACF(newdat$panels, salsa2dOutputk5$bestModel, suppress.printout = TRUE)
```




```
salsa.corrmodel <- make.gamMRSea(model = salsa2dOutputk5$bestModel, panelid = newdat$panels)
summary(salsa.corrmodel)
```

```
Call:
gamMRSea(formula = response ~ as.factor(indclass) + LRF.g(radiusIndices,
    dists, radii, aR), data = newdat, splineParams = splineParams)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
 -5.722   -1.743   -0.890    0.450  188.737

Coefficients:
                                        Estimate Std. Error Robust S.E.
(Intercept)                              3.18269    1.09526     0.73431
as.factor(indclass)2                     2.21831    1.11921     0.93548
as.factor(indclass)3                    -1.07429    1.79818     0.83258
as.factor(indclass)4                    -1.21959    1.31041     0.75082
as.factor(indclass)5                     0.13309    1.72161     1.07009
as.factor(indclass)6                    -0.31254    1.12357     0.71283
as.factor(indclass)7                     2.07864    1.14777     1.79244
as.factor(indclass)8                     4.34912    1.35529     2.82991
as.factor(indclass)9                     0.64261    1.11634     0.86922
as.factor(indclass)10                   -0.20241    1.15048     0.72914
as.factor(indclass)11                   -0.06280    1.17936     0.73756
as.factor(indclass)12                    0.88693    1.09535     0.74179
as.factor(indclass)13                    2.67230    1.09388     0.81808
as.factor(indclass)14                    0.66008    1.13251     0.82440
as.factor(indclass)15                    0.44975    1.22090     0.89423
as.factor(indclass)16                   -1.63906    2.03438     0.70845
```

```
as.factor(indclass)17                     -0.14808    1.13513     0.76895
as.factor(indclass)18                     -1.31956    1.62830     0.70094
as.factor(indclass)19                     -0.94858    1.28884     0.75910
as.factor(indclass)20                     -0.69596    2.12380     0.69547
as.factor(indclass)21                      0.42153    1.10080     0.73719
as.factor(indclass)22                     -0.13836    1.10919     0.72906
as.factor(indclass)23                      0.09676    1.12676     0.79094
as.factor(indclass)24                      0.23337    1.75426     0.69853
as.factor(indclass)25                      1.39822    1.84256     0.69547
as.factor(indclass)26                      1.97065    1.29571     1.13820
as.factor(indclass)28                     -0.53027    1.52578     0.82122
as.factor(indclass)30                      2.21798    2.37193     0.69547
as.factor(indclass)32                      0.47205    1.14385     0.82544
as.factor(indclass)33                     -0.39333    1.28144     0.96852
as.factor(indclass)34                      1.76023    1.12820     0.94766
as.factor(indclass)35                      0.34855    1.09830     0.74382
as.factor(indclass)36                      2.16701    1.08450     0.73768
as.factor(indclass)37                     -0.02555    1.08950     0.69810
as.factor(indclass)38                      0.39166    1.10156     0.75939
as.factor(indclass)39                      1.46034    1.18524     1.20425
as.factor(indclass)41                      6.00857    1.89766     0.94365
as.factor(indclass)42                      0.19140    1.11360     0.82260
as.factor(indclass)43                      0.04363    1.08508     0.70653
as.factor(indclass)44                      0.59300    1.09523     0.74184
as.factor(indclass)49                     -1.36884    1.71721     0.74317
LRF.g(radiusIndices, dists, radii, aR)b1 -0.86044    0.18646     0.22357
LRF.g(radiusIndices, dists, radii, aR)b2 -1.61488    0.68819     0.69848
                                          t value Pr(>|t|)
(Intercept)                                 4.334 1.47e-05 ***
as.factor(indclass)2                        2.371 0.017739 *
as.factor(indclass)3                       -1.290 0.196964
as.factor(indclass)4                       -1.624 0.104326
as.factor(indclass)5                        0.124 0.901024
as.factor(indclass)6                       -0.438 0.661063
as.factor(indclass)7                        1.160 0.246202
as.factor(indclass)8                        1.537 0.124356
as.factor(indclass)9                        0.739 0.459737
as.factor(indclass)10                      -0.278 0.781320
as.factor(indclass)11                      -0.085 0.932148
as.factor(indclass)12                       1.196 0.231851
as.factor(indclass)13                       3.267 0.001091 **
as.factor(indclass)14                       0.801 0.423333
as.factor(indclass)15                       0.503 0.615011
as.factor(indclass)16                      -2.314 0.020705 *
as.factor(indclass)17                      -0.193 0.847291
as.factor(indclass)18                      -1.883 0.059783 .
as.factor(indclass)19                      -1.250 0.211462
as.factor(indclass)20                      -1.001 0.316988
as.factor(indclass)21                       0.572 0.567461
as.factor(indclass)22                      -0.190 0.849482
as.factor(indclass)23                       0.122 0.902640
as.factor(indclass)24                       0.334 0.738319
as.factor(indclass)25                       2.010 0.044403 *
as.factor(indclass)26                       1.731 0.083409 .
```

```
as.factor(indclass)28                       -0.646 0.518478
as.factor(indclass)30                        3.189 0.001430 **
as.factor(indclass)32                        0.572 0.567409
as.factor(indclass)33                       -0.406 0.684661
as.factor(indclass)34                        1.857 0.063270 .
as.factor(indclass)35                        0.469 0.639368
as.factor(indclass)36                        2.938 0.003313 **
as.factor(indclass)37                       -0.037 0.970804
as.factor(indclass)38                        0.516 0.606032
as.factor(indclass)39                        1.213 0.225283
as.factor(indclass)41                        6.367 1.99e-10 ***
as.factor(indclass)42                        0.233 0.816019
as.factor(indclass)43                        0.062 0.950765
as.factor(indclass)44                        0.799 0.424092
as.factor(indclass)49                       -1.842 0.065515 .
LRF.g(radiusIndices, dists, radii, aR)b1  -3.849 0.000119 ***
LRF.g(radiusIndices, dists, radii, aR)b2  -2.312 0.020792 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 26.7721)

    Null deviance: 375420  on 13524  degrees of freedom
Residual deviance: 360941  on 13482  degrees of freedom
AIC:  82889

Max Panel Size = 50; Number of panels = 271
Number of Fisher Scoring iterations: 2
```

```
anova(salsa.corrmodel)
```

```
Analysis of 'Wald statistic' Table
Model: gaussian, link: identity
Response: response
Marginal Testing
Max Panel Size = 50; Number of panels = 271

                    Df         X2 P(>|Chi|)
as.factor(indclass) 40 3.4929e+10 < 2.2e-16 ***
s(x.pos, y.pos)      2 1.9000e+01 9.101e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Analysis of 'Wald statistic' Table
Model: gaussian, link: identity
Response: response
Marginal Testing
Max Panel Size = 50; Number of panels = 271

                    Df       X2 P(>|Chi|)
as.factor(indclass) 40 55906346 < 2.2e-16 ***
s(x.pos, y.pos)      3       13  0.004277 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Questions

1. Which of the following is **TRUE**?
   (a) CReSS includes a penalty term on the coefficients.
   (b) The thin plate spline smoother can have local smoothing parameters that vary across the surface.
   (c) The basis functions of the thin plate spline have decreasing values as the distance from knots increases.
   (d) The effective range of the bases in a CReSS-based smoother are specified by a range parameter that can vary across the knots.
   (e) The values in the columns of a CReSS basis function increase with distance from each knot.
2. Which of the following is **FALSE**?
   (a) The variance estimates for the `penreg2D` model and `gamMRSea` Model 1 (`salsa2dOutput`) are about the same.
   (b) The SALSA knot grid specifies the location of 300 candidate knots.
   (c) Model selection using BIC finds the same 'best' `gamMRSea` model as 10-fold CV (Models 1-3).
   (d) It is acceptable to compare the `mgcv` model with the `gamMRSea` models using cross-validation.
   (e) Knot locations are not available at every data point in the `gamMRSea` models.
3. TRUE or FALSE? Of all the models fitted here (`mgcv` and `MRSea`), the model with the largest degrees of freedom for the spatial term is the 'best' model (where best is determined by cross-validation score).
4. TRUE or FALSE? Changing the start location of the initial SALSA knots made a difference to the resulting model. I.e. the model did not converge on the same solution from each start point.
5. TRUE or FALSE? There is no overlap in the knot locations chosen by the three `gamMRSea` models.
6. When positive residual correlation is present in a model, which of the following is **FALSE**?
   (a) If this correlation is ignored, the standard errors and p-values will likely be too small.
   (b) If this correlation is accommodated using a GEE approach, the standard errors and p-values will likely be adjusted to be larger than reported assuming residual independence.
   (c) If this correlation is accommodated using a GEE approach, the nature of the adjustment to the standard errors and p-values (compared with those reported assuming residual independence) will depend, in part, on the nature of the correlation structure assumed if a model-based approach is taken.
   (d) If this correlation is accommodated using a GEE approach, the nature of the adjustment to the standard errors and p-values (compared with those reported assuming residual independence) will not depend on the nature of the correlation structure assumed if robust standard errors are used for inference.
   (e) If this correlation is accommodated using a GEE approach, the nature of the adjustment to the standard errors and p-values (compared with those reported assuming residual independence) will only depend on the correlation structure assumed if robust standard errors are used for inference, since they are robust to model mis-specification.
7. Which of the following is **FALSE**?
   (a) The ACF plots show positive correlation in model residuals, which necssarily means that all of the regression parameter robust standard errors will be larger when adjusted.
   (b) The mean residual correlation decays quite quickly to zero, after about a lag of 10, however, some panels decay even more quickly and others have a higher correlation for longer.
   (c) By using the panel feature to update a `gamMRSea` model, it is the equivalent of using a GEE (in SAS or R) with an independent working correlation matrix.
   (d) In this case, accounting for the correlation in our residuals will not change the regression estimates, and does not change the model selection choices for this model.
8. Which of the following about the correlation structures covered in class is **FALSE**?

(a) An AR(1) correlation structure decays with the distance between points (residuals in this case).
(b) A compound symmetry correlation structure has a 'common' correlation and does not depend on the distance between points (residuals in this case).
(c) When specifying a correlation structure the analyst must decide how to specify the 'panels', regardless of the correlation assumed within panels.
(d) The correlation structure assumed informs the content of the block-diagonal matrix.
(e) The size of the blocks in the block diagnonal matrix ($n_i \times n_i$) for all individuals is equal to the largest number of observations observed for any of the panels observed in the study.

9. Which of the following about robust (or empirical) standard errors is **FALSE**?

   (a) If the true correlation structure is known and this is specified in a model then the resulting standard errors will be closer to their true values than the values estimated using robust/empirical standard errors.
   (b) Empirical standard errors are based on the product of the sums of the within panel residuals.
   (c) Positive residual correlation observed within panels and accommodated using a GEE tends to result in deflated standard errors relative to standard errors obtained assuming residual independence because there is less variabiltiy in positive correlated residuals and so acknowledging this correlation as part of the model results in standard errors which are reduced in size.
   (d) If the analyst is not convinced that an appropriate correlation structure is available for the residual correlation, then a good solution is to use empirical standard errors when drawing any model conclusions.
   (e) Empirical standard errors are different to specifying an 'unstructured' correlation structure and the empirical approach doesn't require any correlation parameters (as part of a block diagonal matrix) to be formally estimated.

10. Which of the following about model selection and model assessment for GEEs is **FALSE**?

   (a) The QICu can be used to choose between models with different covariates but the same correlation structure.
   (b) The QIC(R) can be used to choose between models with the same covariates but with different correlation structures.
   (c) The QIC scores are a relative measure of fit and so, in practice, none of the available models may be a good description of the data.
   (d) Pearson residuals may assist in choosing between models since these can help diagnose model problems.
   (e) Cumulative residuals can help diagnose model problems and a large p-value for the test in the SAS output associated with the maximum of the absolute value covered in class suggests compelling evidence that there are problems with model fit.