

# Stat 534 Homework 6

Kenny Flagg

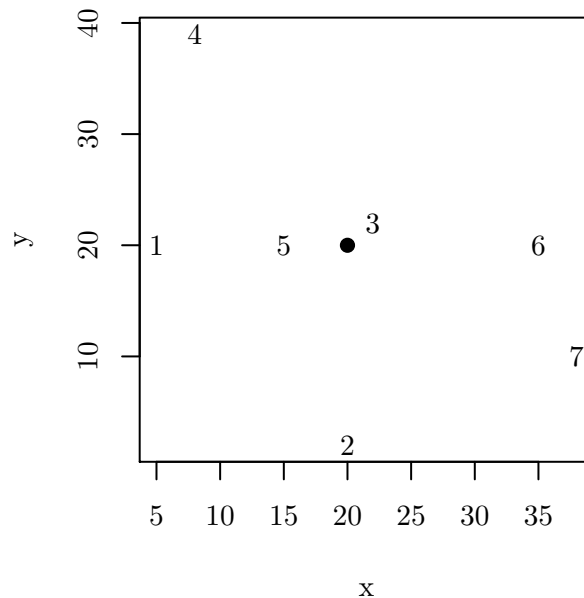
February 24, 2017

1. Refer to the 7 point ordinary kriging example we discussed in class. Repeat the example with the following changes. We move point  $s_3$  to  $(22, 22)$  and point  $s_5$  to  $(15, 20)$ . Compare and contrast the results to the example being sure to address the following. You need to be careful with the nugget effect and range parameters.

Recall that Model D is a pure nugget effect model. You do not need the above function to fit that one. A table showing predictions, kriging variances, and weights will help.

```
prob1 <- data.frame(x = c( 5, 20, 22,  8, 15, 35, 38),
                    y = c( 20,  2, 22, 39, 20, 20, 10),
                    z = c(100, 70, 60, 90, 50, 80, 40))
plot(y ~ x, data = prob1, pch = as.character(1:7),
     main = expression('Locations with Relocated '*bold(s)[3]*' and '*bold(s)[5]'))
points(20, 20, pch = 19)
```

Locations with Relocated  $s_3$  and  $s_5$



```

# Eighth row/column are for s_0.
dists <- as.matrix(dist(rbind(probl[,1:2], c(20, 20))))

# A function to automate all this.
probl_function <- function(model, eff_range, sill, nugget){
  if(model == 'exponential')
    eff_range <- eff_range / 3 # Adjust exponential range by 3.
  else if(model == 'gaussian')
    eff_range <- eff_range / sqrt(3) # Adjust gaussian range by sqrt(3).
  if(model == 'nugget')
    Sigma <- diag(rep(nugget, 8))
  else
    Sigma <- cov.spatial(dists, cov.model = model,
                        cov.pars = c(sill - nugget, eff_range)) +
              diag(rep(nugget, 8))
  sig <- Sigma[8, 1:7]
  SigmaStar <- Sigma
  SigmaStar[8, 1:7] <- 1
  SigmaStar[1:7, 8] <- 1
  SigmaStar[8, 8] <- 0
  sigStar <- c(sig, 1)
  lambdaStar <- solve(SigmaStar) %*% sigStar
  p_ok <- t(lambdaStar[-8]) %*% probl$z
  if(model == 'nugget')
    sig2_ok <- nugget - t(lambdaStar) %*% sigStar
  else
    sig2_ok <- sill - t(lambdaStar) %*% sigStar
  return(c(p_ok, sig2_ok, lambdaStar[-8]))
}

probl_table <- cbind(data.frame(Model = c(LETTERS[1:6], 'Distance')),
                    as.data.frame(rbind(
                      probl_function('exponential', 20, 10, 0),
                      probl_function('exponential', 10, 10, 0),
                      probl_function('exponential', 20, 10, 5),
                      probl_function('nugget', 0, 0, 10),
                      probl_function('exponential', 20, 20, 0),
                      probl_function('gaussian', 20, 10, 0),
                      c(NA, NA, dists[8, 1:7])
                    )))
colnames(probl_table)[-1] <- c('\(p_{ok}\)', '\(sigma^2_{ok}\)',
                             paste0('\(lambda_{', 1:7, '}\)'))
print(xtable(probl_table, align = 'ccrrrrrrrr', hline.after = c(0, 6))

```

Model	$p_{ok}$	$\sigma_{ok}^2$	$\lambda_1$	$\lambda_2$	$\lambda_3$	$\lambda_4$	$\lambda_5$	$\lambda_6$	$\lambda_7$
A	58.83	5.02	0.02	0.04	0.57	0.02	0.30	0.03	0.03
B	62.59	8.11	0.06	0.06	0.46	0.06	0.23	0.06	0.06
C	64.51	8.91	0.08	0.09	0.35	0.08	0.24	0.09	0.08
D	70.00	11.43	0.14	0.14	0.14	0.14	0.14	0.14	0.14
E	58.83	10.04	0.02	0.04	0.57	0.02	0.30	0.03	0.03
F	51.45	0.36	-0.09	0.02	0.68	-0.01	0.41	-0.03	0.01
Distance			15.00	18.00	2.83	22.47	5.00	15.00	20.59

- (a) *Do the same conclusions regarding the effects of the sill, range, and nugget effect still hold?*

Yes, the sill still just rescales the variance, and decreasing the range or increasing the nugget still reduces the variability of the weights.

It's interesting that the weights for  $\mathbf{s}_3$  and  $\mathbf{s}_5$  are closer to the weights of the other points in Model B (effective range of 10) than they are in Model A (effective range of 20). The range doesn't act like a magic threshold that separates highly correlated points from points with low correlation; it simply adjusts how quickly the correlation decays over distance. Even though  $\mathbf{s}_3$  and  $\mathbf{s}_5$  are the closest points to  $\mathbf{s}_0$ , under Model B they are still less correlated with  $\mathbf{s}_0$  than they are under Model A, so under Model B they get weights that are closer to the weights of the farthest points.

- (b) *Which points are screening?*

$\mathbf{s}_1$  is screened by  $\mathbf{s}_5$  and  $\mathbf{s}_6$  is screened by  $\mathbf{s}_3$ . The screening is apparent from the table because, for most models,  $\lambda_1$  and  $\lambda_6$  are smaller than  $\lambda_2$  even though  $\mathbf{s}_2$  is a farther from  $\mathbf{s}_0$ . It is especially obvious for the gaussian model (F) where  $\lambda_1$  and  $\lambda_6$  are negative.

- (c) *What happens to the kriging variance?*

For all of the models except the pure nugget model, the kriging variance is lower. This makes sense because moving some correlated observations closer to the location of the prediction provides more information about  $Z(\mathbf{s}_0)$  and should reduce the amount of uncertainty in the prediction.

- (d) *In which models does the predicted value change and in which does it stay the same?*

The predicted value stays the same for the pure nugget model, and it changes for all the other models. For the models with spatial dependence, the predicted value is pulled towards the values  $Z(\mathbf{s}_3) = 60$  and  $Z(\mathbf{s}_5) = 50$  because those nearby points have the largest weights.

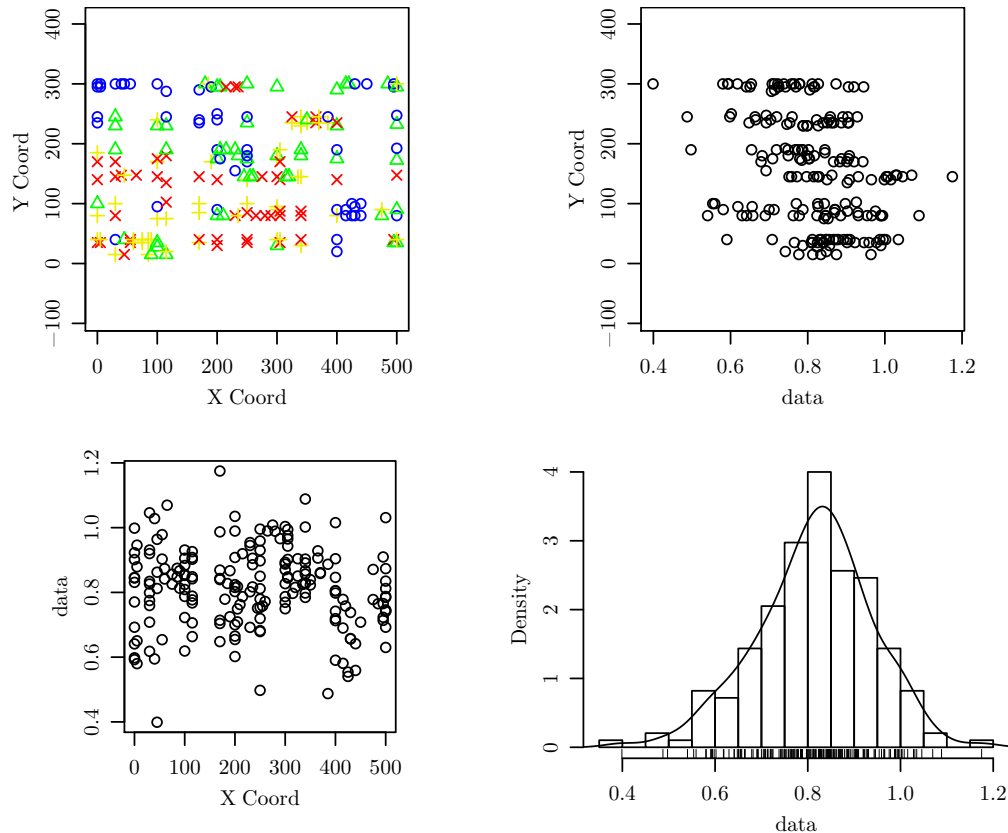
2. We are going to use ordinary kriging to predict values of total carbon over a grid. You should still have the carbon/nitrogen data set from the last homework but let me know if you need to have it emailed to you. Choose your grid to be the same as in the CN ratio example I worked in class. Set up R code is below.

```
CN.dat <- read.table('../hw05/CN.dat', header = TRUE)
pred.grid <- expand.grid(seq(-50, 550, 1 = 100), seq(-15, 330, 1 = 100))
TC.geodat <- as.geodata(CN.dat, coords.col = 1:2, data.col = 4)
```

We will work with total carbon (column labeled TC in the data set) in this problem. You can use either the *ksline* or *krige.conv* functions. Be a bit patient; it should take only a minute or so. Don't worry about anisotropy.

(a) Plot the data using

```
plot(TC.geodat, breaks = 20)
```



Do you see any evidence of trend? Outliers? Other potential anomalies?

There appears to be a slight trend of decreasing total carbon to the north. There is a possible low outlier in the northwest corner, and a possible high outlier near (200, 150), but neither of these look very extreme in the histogram.

```
control <- krige.control(cov.model = 'exponential',
                        cov.pars = c(0.01418, 154.62223/3),
                        nugget = 0.00241)
TC_pred <- krige.conv(TC.geodat, locations = pred.grid, krige = control)
par(mfrow = c(2, 1), mar = c(4, 4, 2, 1))
contour(TC_pred, main = 'Kriging Predictions')
contour(TC_pred, val = sqrt(TC_pred$krige.var), main = 'Kriging Standard Errors')
```



- (c) *Evaluate the predictions using cross validation ( $CV_2$ ). You can use the R code I provided in class notes.*

```
CV2_term <- function(i, geodata = TC.geodat, ...){
  pred <- krige.conv(coords = geodata$coords[-i,],
    data = geodata$data[-i],
    locations = rbind(geodata$coords[i,]), ...)
  return((geodata$data[i] - pred$predict) / sqrt(pred$krige.var))
}
CV2 <- sqrt(mean(sapply(seq_along(TC.geodat$data), CV2_term, krige = control)^2))

CV2

[1] 1.07633
```

- (d) *Briefly discuss your results.*

The plots of the observed values in part (a) shows a random-looking arrangement of colors and shapes, indicating that some large total carbon values occur small values, but some banding is visible, suggesting that there is a trend. Unsurprisingly, the kriging prediction surface is noisy because of the variation between nearby measurements. The predictions tend to be higher in the southern half of the site than in the northern half, so the carbon values may be non-stationary and it would be worthwhile to fit a linear trend model. However, the  $CV_2$  value is close to 1, meaning that the model accurately predicts the observed values when one at a time is omitted. The standard error surface shows small dips at the observed locations and larger standard errors in the vertical gaps where no measurements were recorded.