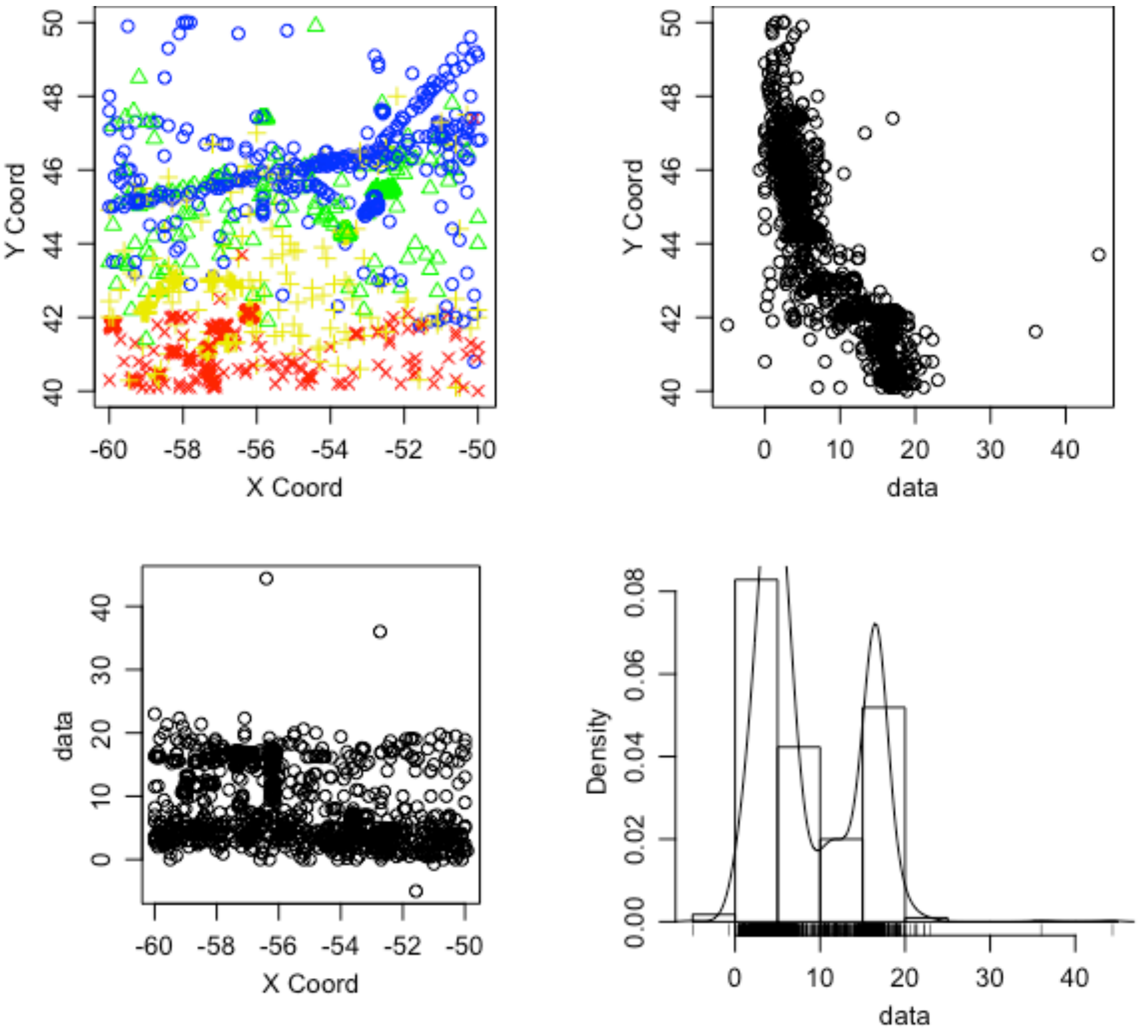
Daanish Ahsan
Candidate number: 124070
Statistical Modelling Space and Time

Statistical Modelling in Space and Time Assignment 1

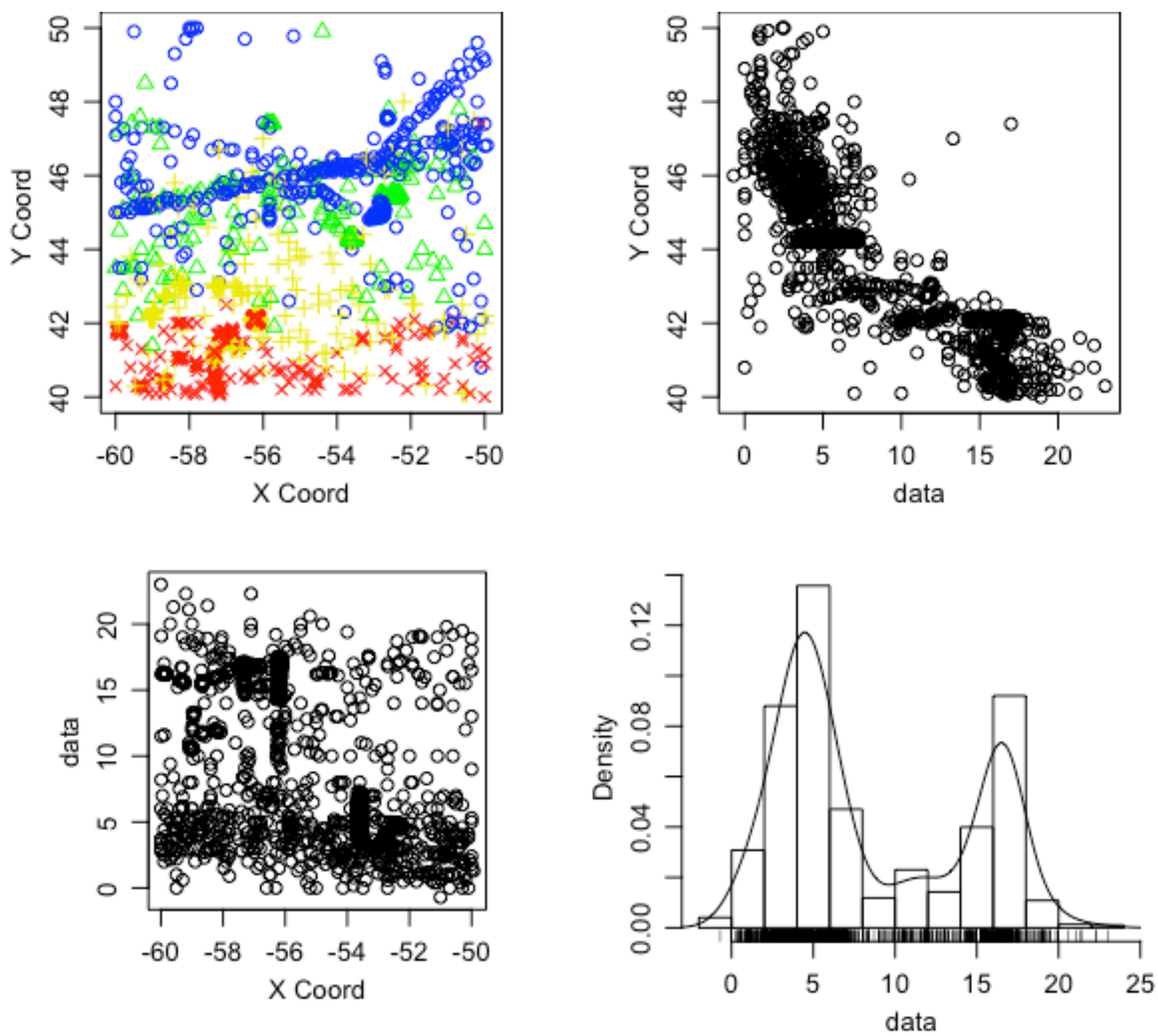
Plot of all data



1)

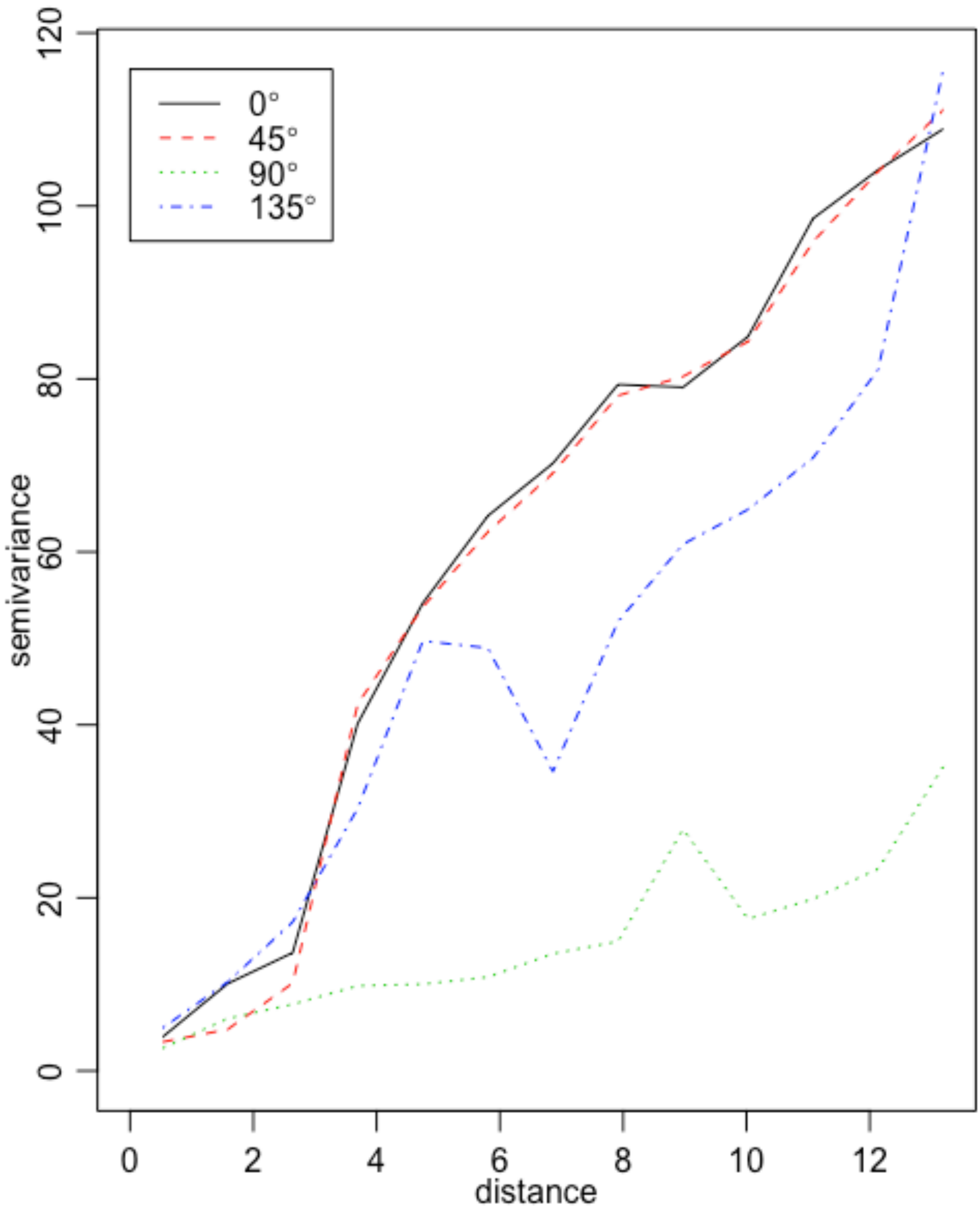
The label ‘data’ on the x axes of the plots is the sea surface temperature. The first 4 plots are of all the data in the Grandbanks dataset, and shows that there are a couple of outliers with some very high temperatures, and one with a very low temperature. The second group of plots show the data with the outliers removed. It is clear that the outliers were vastly different in terms of temperature compared to the other datapoints, as the scale for the x axis now ends at 25 on the histogram, rather than 40.

Plot of data with outliers removed

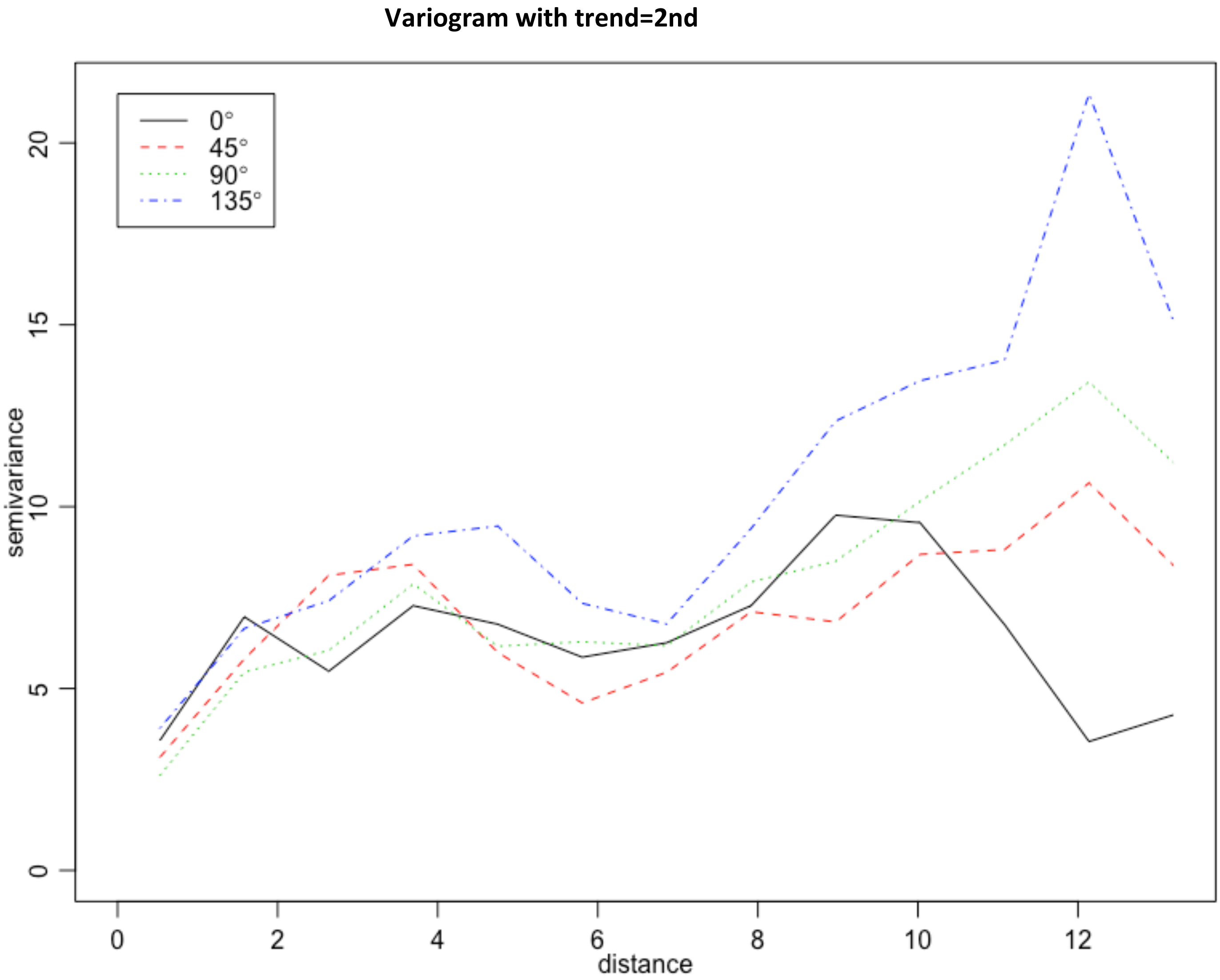


Daanish Ahsan
Candidate number: 124070
Statistical Modelling Space and Time
2)

Variogram with no trend



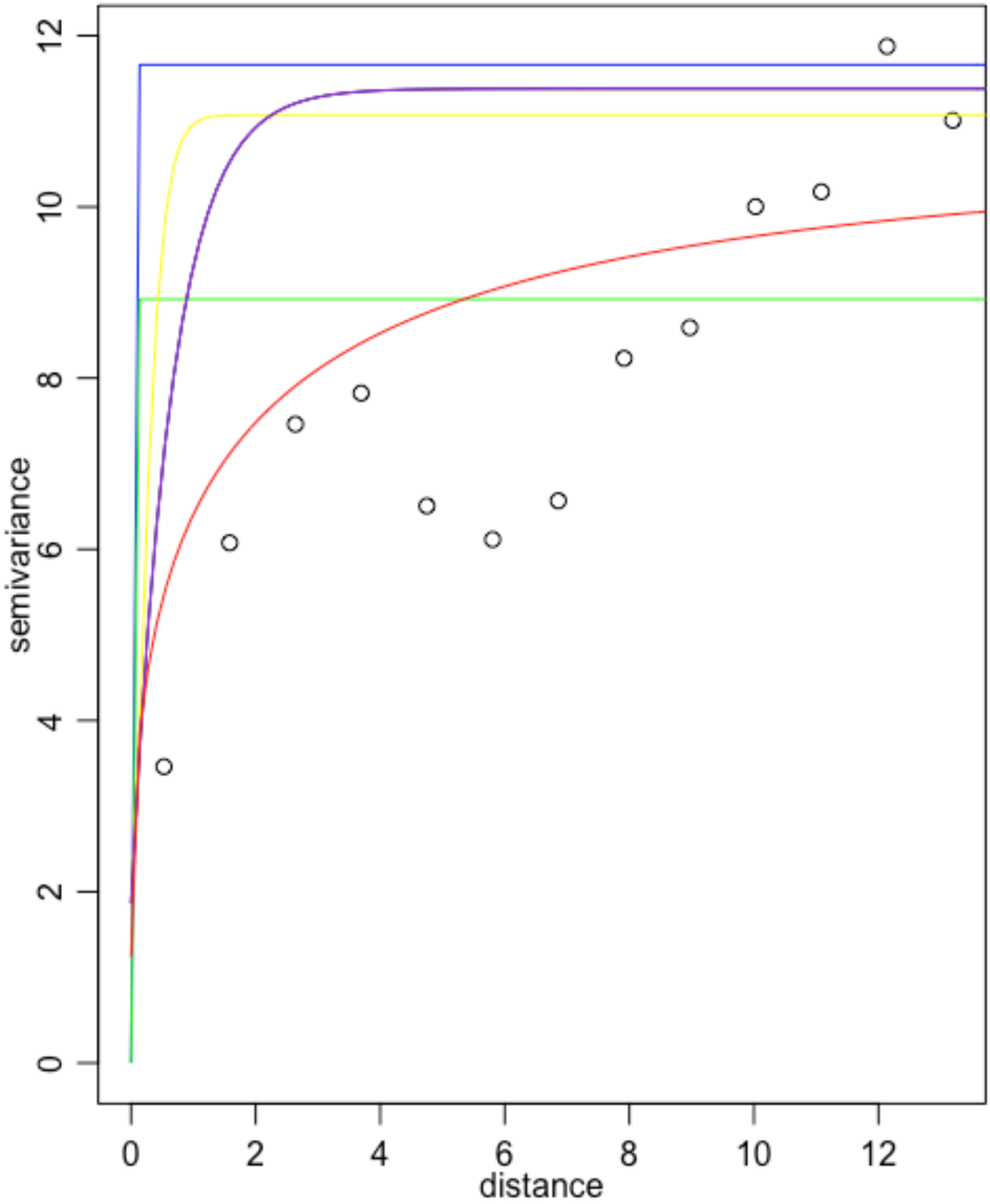
Daanish Ahsan
Candidate number: 124070
Statistical Modelling Space and Time



Spatial dependence is the concept that observations closer together in space are more correlated than those further apart. Semi-variance is a measure of spatial dependence between two observations as a function of the distance between them. A variogram is a graph of how semivariance changes as the distance between observations change. The 2 plots above are directional variograms of the data. A process is isotropic when covariance depends only on distance not direction. Thus the first variogram is not isotropic as the lines do not converge. However, the second plot, with trend=2^{nd} added, seems more isotropic than the first as lines converge much more.

Daanish Ahsan
Candidate number: 124070
Statistical Modelling Space and Time
3)

Different Spatial Models



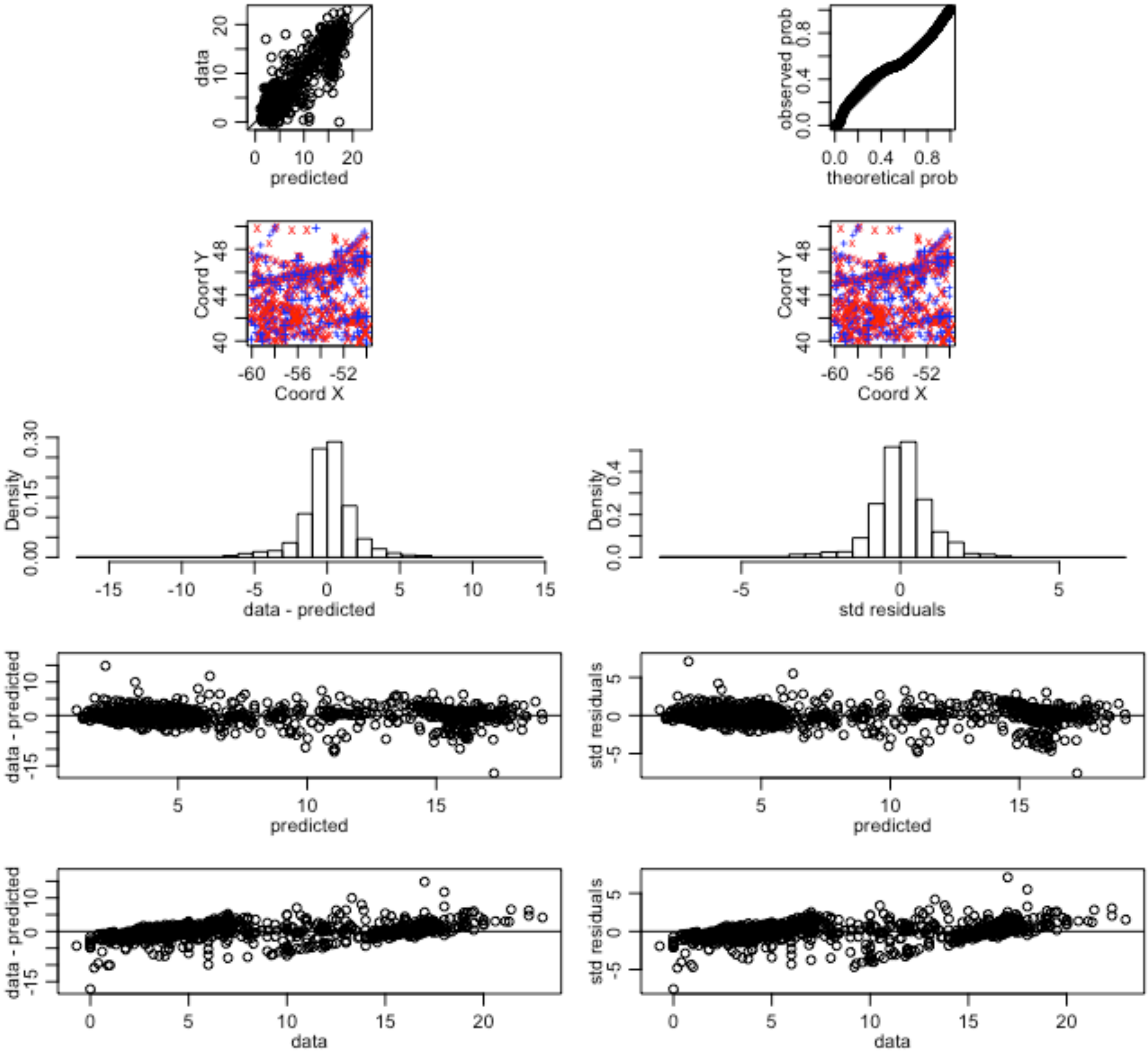
Estimation method: maximum likelihood

Parameters of the mean component (trend):

beta0	beta1	beta2	beta3	beta4	beta5
367.0118	**-3.7780**	**-19.3260**	**-0.0099**	**0.2319**	**0.0566**

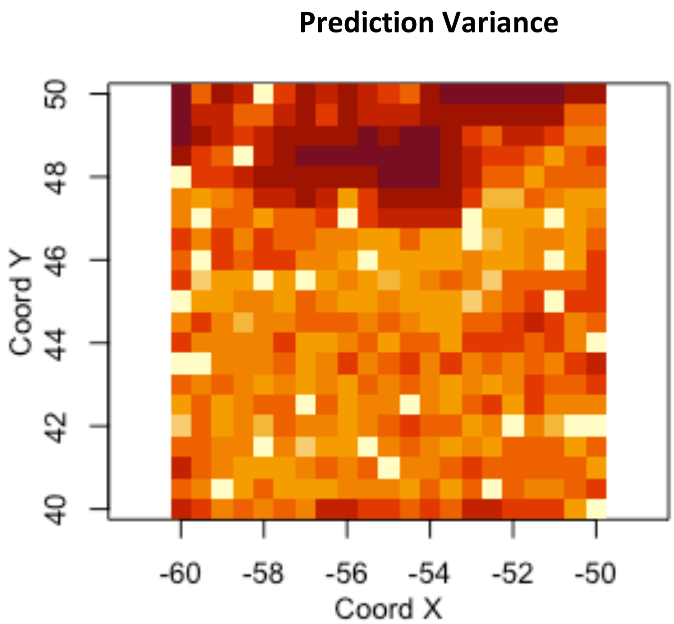
The plot to the left shows how various spatial models fit to the variogram. Spherical (blue), pure.nugget (green) , matern (yellow), exponential (purple) and powered exponential (red) models were fitted, with a variety of kappa values and nugget adjustments tried. The powered exponential was found to be the best fit on the variogram, although it still isn't a particularly strong fit. This model also had the lowest AIC of all the models, at 6207, indicating it is a relatively higher quality model compared to the others. Shown above are the parameter estimates that were obtained after fitting the models using the likfit function.

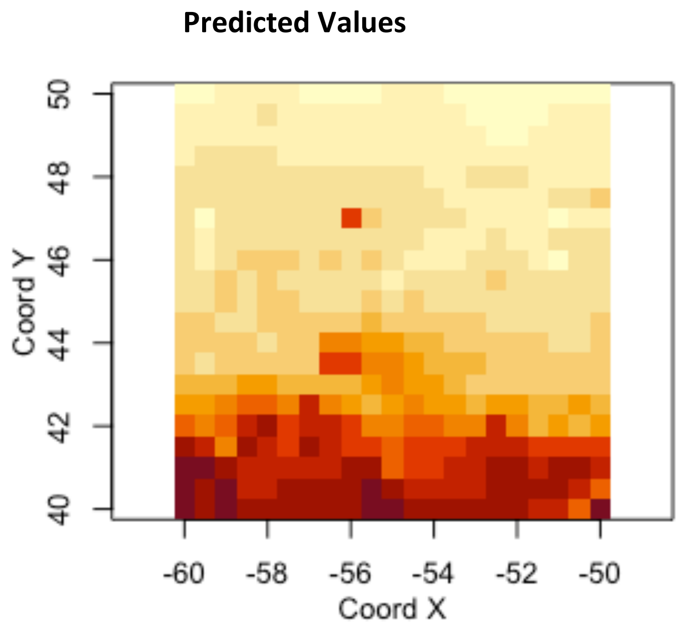
Cross validation by kriging



Daanish Ahsan
Candidate number: 124070
Statistical Modelling Space and Time

Spatial prediction-conventional kriging

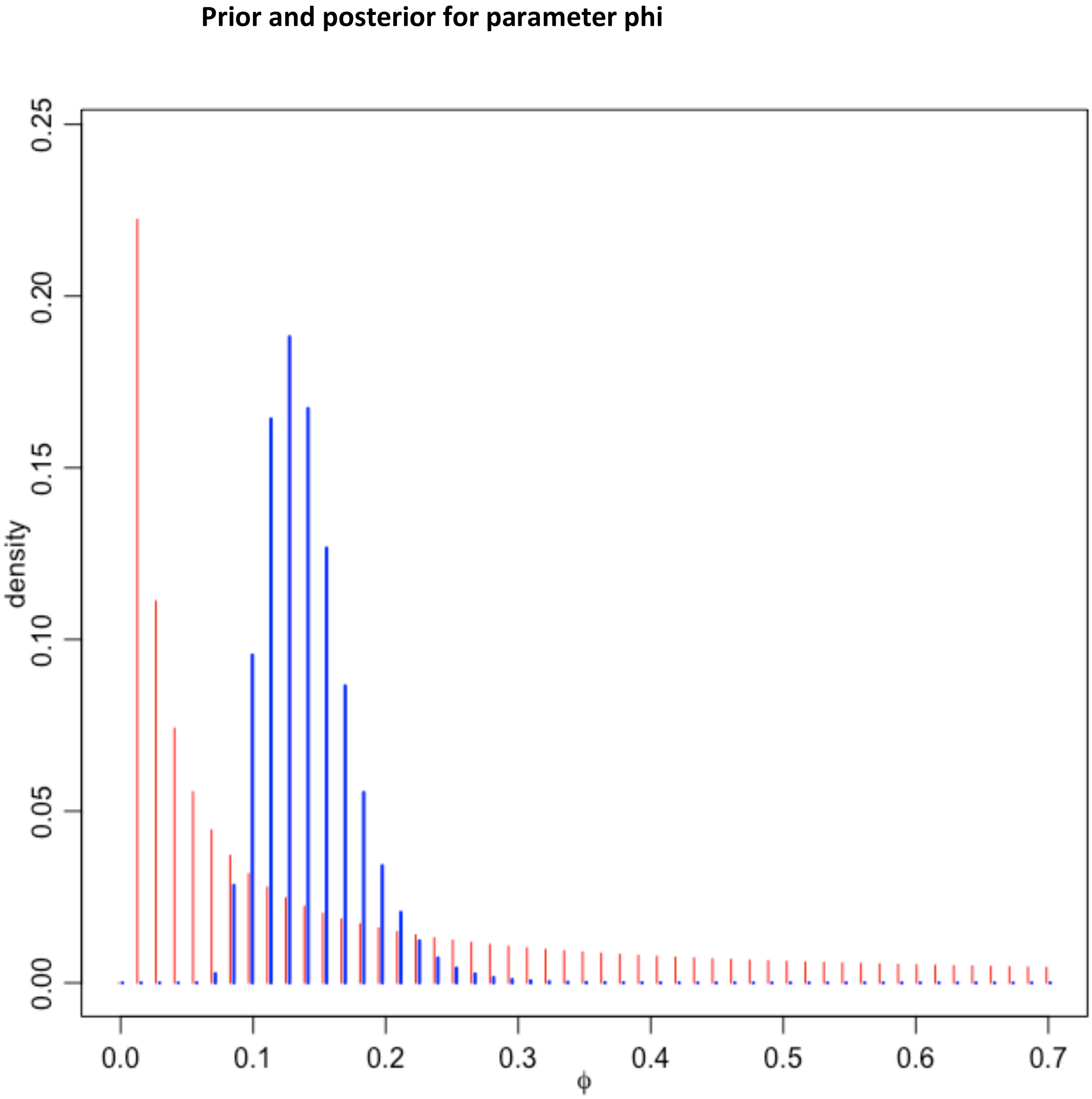




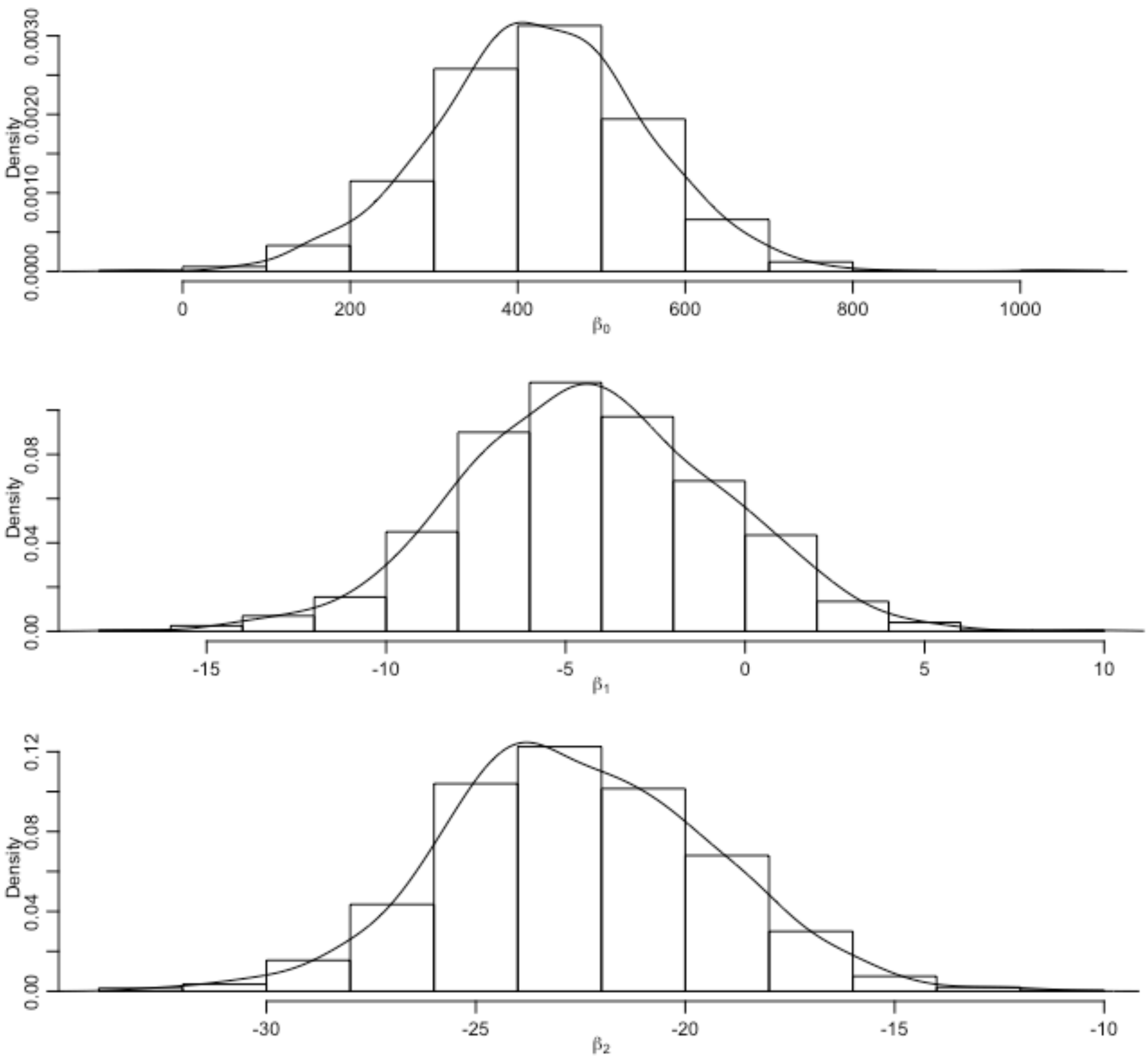
The plots on the left page show model validation by comparing observed and predicted values by kriging; the upper right plot shows there is some underestimation. The xvalid function was used for this. The above plots are variance and values predictions of sea surface temperature using spatial prediction kriging and the krige.conv function which performs spatial prediction for fixed covariance parameters. The predicted values plot shows that the hotter regions are Southern, and colder Northern as expected. But there is a clear outlier, with a hot temperature fairly North. The variance prediction shows that the northern regions have much higher temperature variation in general.

Daanish Ahsan
Candidate number: 124070
Statistical Modelling Space and Time

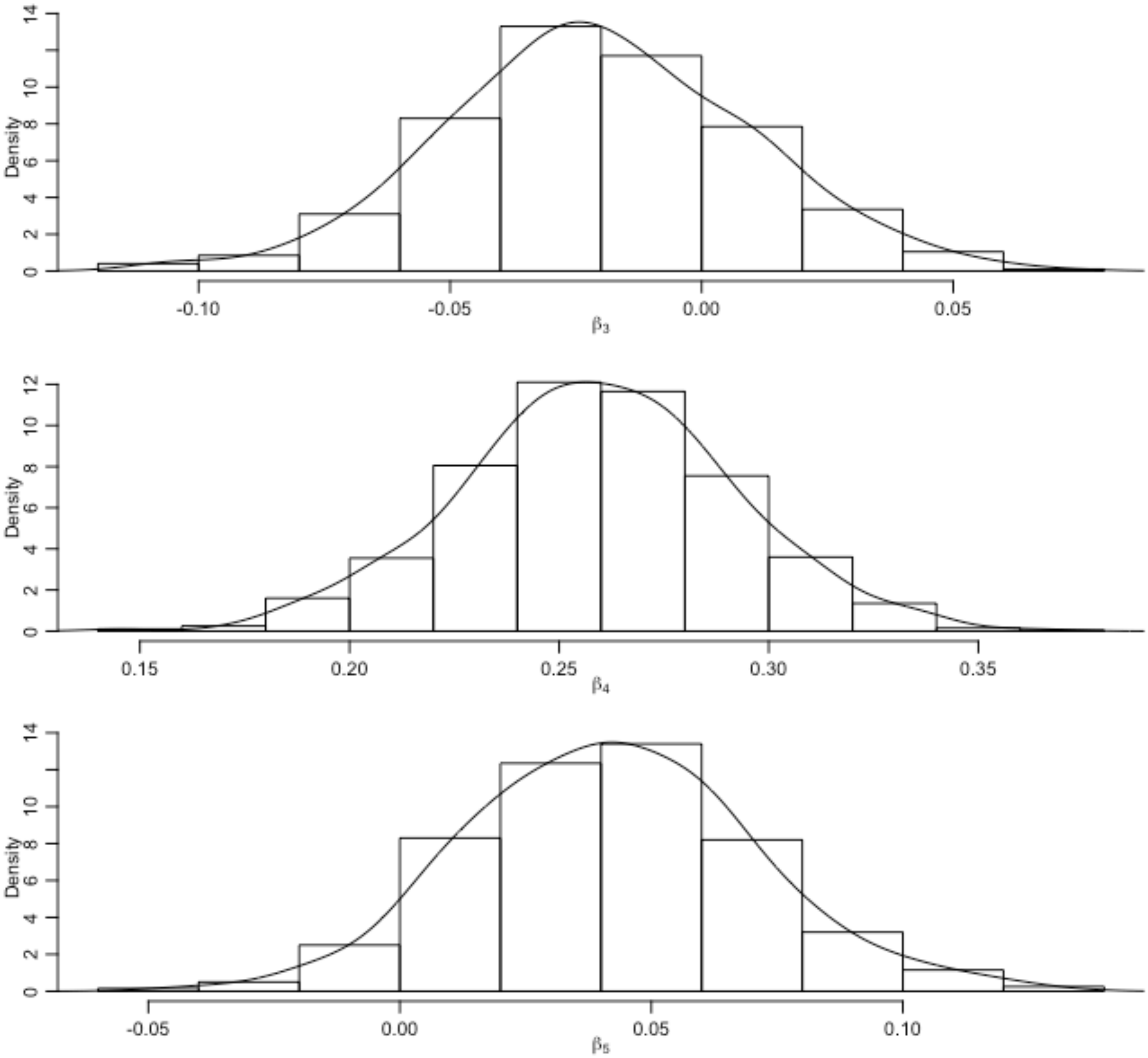
4)



Histograms with samples from the posterior

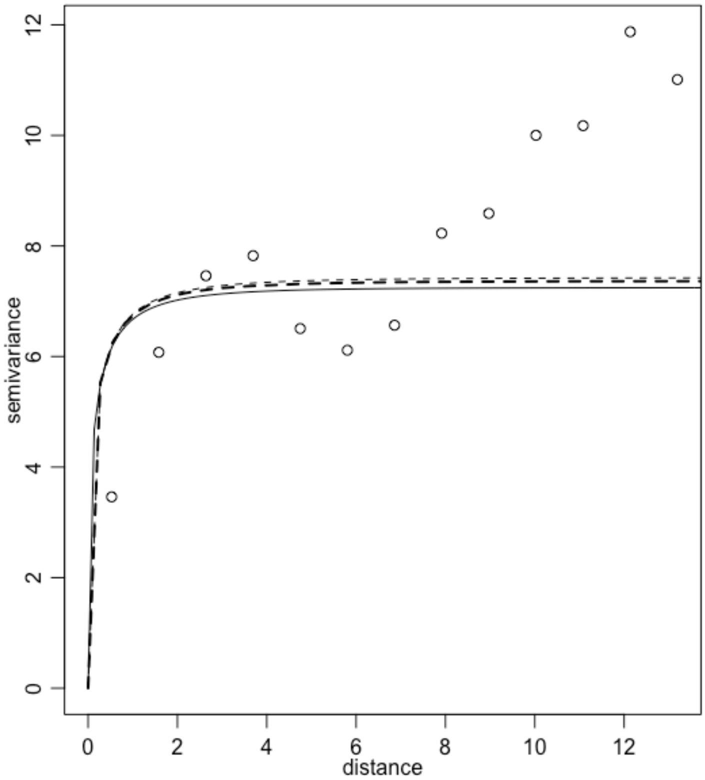


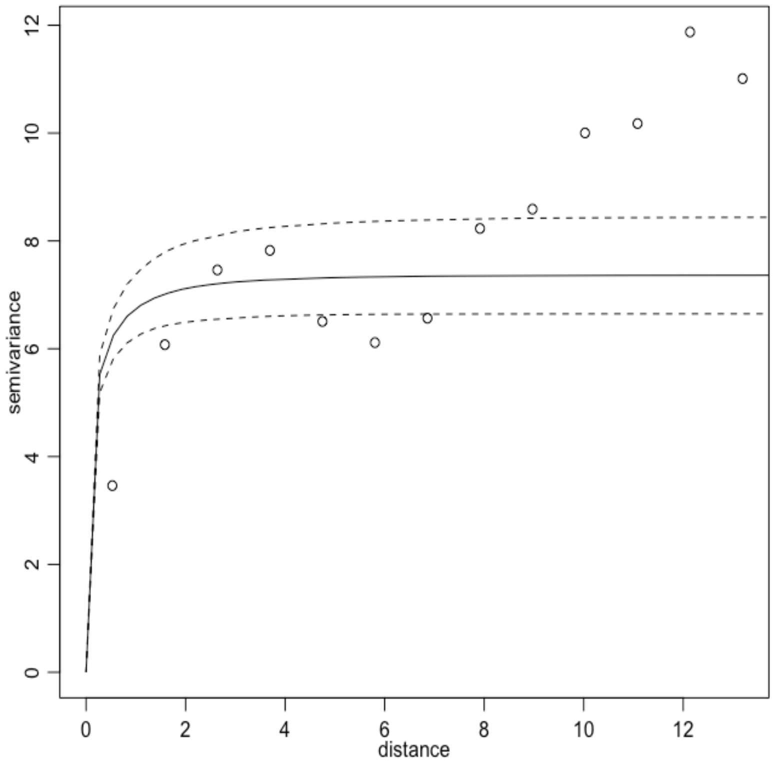
More histograms with samples from the posterior



Daanish Ahsan
Candidate number: 124070
Statistical Modelling Space and Time

Empirical Variograms

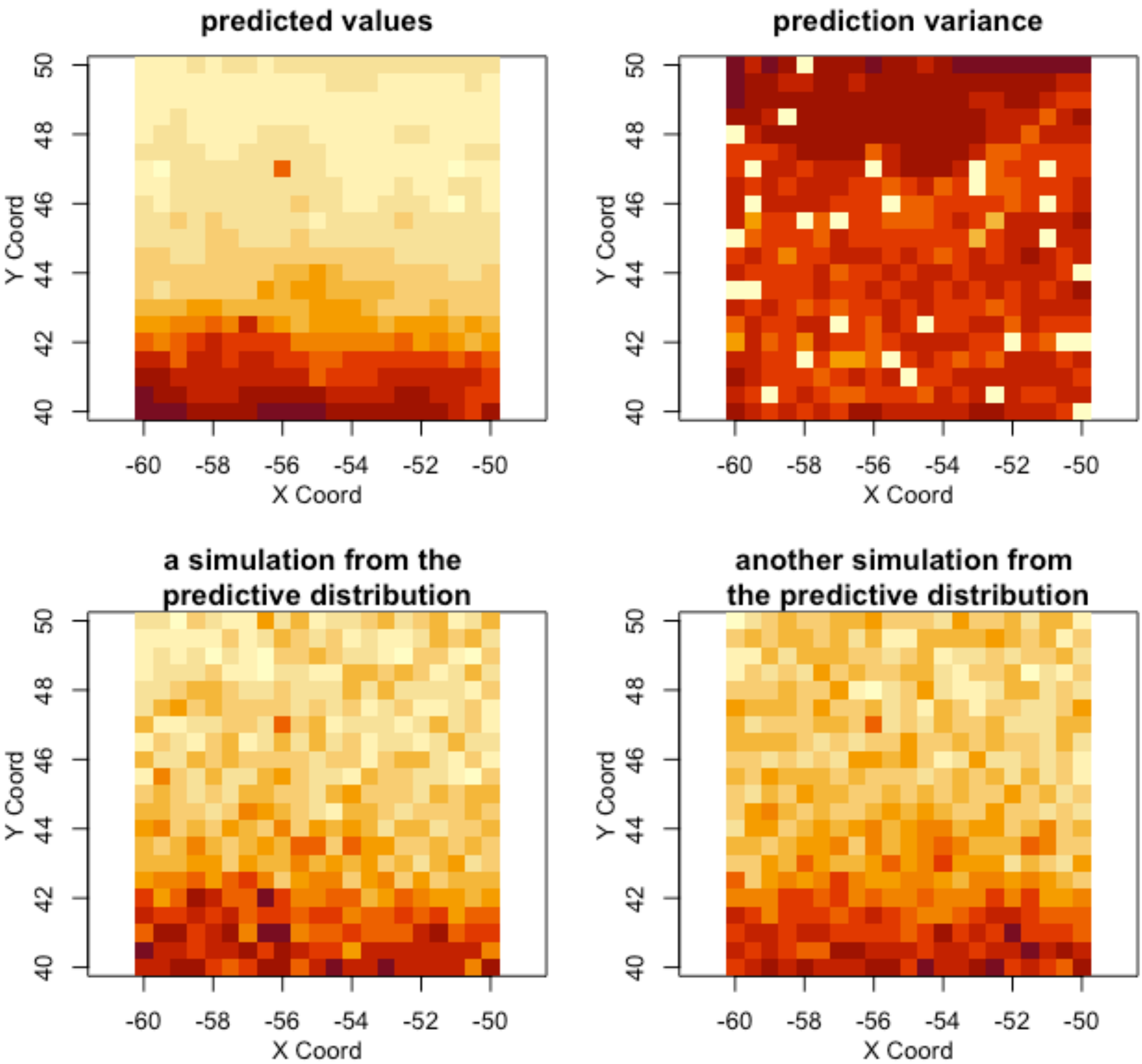




For the Bayesian model, the krige.bayes function was used with a powered exponential as it was the best fitting model previously; a reciprocal phi prior was also used. The prior and posterior plot for phi shows the prior in red and the posterior in blue. The histograms show density for each parameter, using samples from the posterior, and look broadly very similar. The above empirical variogram on the left shows lines which are the summaries of the posterior of the binned variogram, and a line that is the summary of the posterior of the parameters. The variogram on the right shows lines with median and quantile estimates. The predicted values plot on page 13 looks identical to the ordinary kriging plot on page 8, but the variance prediction plots differ, with the Bayesian one looking much redder in general, implying higher temperature variation, whereas the non-Bayesian is more orange. Lastly, the bottom two plots show simulations from the predictive distributions, and broadly reflect that sea surface temperatures are higher in the South compare to the North, as seen in the predicted values plot.

Daanish Ahsan
Candidate number: 124070
Statistical Modelling Space and Time

Prediction Results



Daanish Ahsan
Candidate number: 124070
Statistical Modelling Space and Time

5) In terms of the two methods of estimation, the ordinary powered exponential model seems to fit the variogram better than the Bayesian model. The predicted values plot for both models seems identical, but they differ quite significantly in the prediction variance plot.

Daanish Ahsan
Candidate number: 124070
Statistical Modelling Space and Time

Code Appendix

```
library(tidyverse)
library(geoR)
library(fit.models)
library(MASS)
set.seed(362)

#  Read data

#1 Plot the data and check that that it looks reasonable. Are there bad
 data that need to
#be removed? yes, outliers have been removed
data <- read.csv('Spacetimedata.csv')
gdata <- as.geodata(data,coords.col=2:3,data.col=6)
dup <-dup.coords(gdata)
gdata2 <- jitterDupCoords(gdata,max=0.1,min=0.05)
par(mfrow=c(1,1))
plot(gdata2)

#removing bad data
data <- read.csv('Spacetimedata.csv')
data <- data[-c(366, 664, 678),]
gdata <- as.geodata(data,coords.col=2:3,data.col=6)
dup <-dup.coords(gdata)
gdata2 <- jitterDupCoords(gdata,max=0.1,min=0.05)

par(mfrow=c(1,1))
plot(gdata2)

#2  Check for isotropy (use the function variog4 in geoR)? Do you need a
 trend in the
#model?

f <- variog4(gdata2, coords = gdata2$coords, data = gdata2$data,
             uvec = "default", breaks = "default", trend = "2nd", lambda =
 1,
             option = c("bin", "cloud", "smooth"),
             estimator.type = c("classical", "modulus"),
             pairs.min = 2,
             bin.cloud = FALSE, direction = c(0, pi/4, pi/2, 3*pi/4),
 tolerance = pi/8,
             unit.angle = c("radians", "degrees"))

plot(f)



#3 Decide what spatial model you want to fit. You may want to try several
```

```
 and see
# which one fits best. Estimate the parameters of your chosen model by
 Maximum
# Likelihood and plot the expected value and variance for the estimate on
 the required
# grid.

ml <- likfit(gdata2, ini=c(0.5, 0.5), fix.nug = TRUE,
 cov.model='spherical',kappa=2, trend = '1st')
ml
summary(ml)
ml2 <- likfit(gdata2, ini=c(0.5, 0.5), fix.nug = TRUE,
 cov.model='pure.nugget', kappa=2, trend = '1st')
summary(ml2)
m1 <- likfit(gdata2, ini=c(0.5, 0.5), nugget = 0, cov.model='matern',
 kappa=3/2, trend = '2nd')
m2 <- likfit(gdata2, ini=c(0.5, 0.5), nugget = 3, cov.model='matern',
 kappa=3/2, trend = '2nd')
m3 <- likfit(gdata2, ini=c(0.5, 0.5), nugget = 3, cov.model='exponential',
 kappa=2, trend = '2nd')
m4 <- likfit(gdata2, ini=c(0.5, 0.5), nugget = 8, cov.model='exponential',
 kappa=2, trend = '2nd')
m5 <- likfit(gdata2, ini=c(0.5, 0.5), fix.nugget = FALSE,
 cov.model='powered.exponential', kappa=0.45, trend = '2nd') #this one is
 best on variogram
plot(variog(gdata2, trend='2nd'))
lines(ml, col='blue')
lines(ml2, col='green')
lines(m2, col='yellow')
lines(m3, col='brown')
lines(m4, col='purple')
lines(m5, col='red') #i like the black one

xvR.ml <- xvalid(gdata2, model = m5)
par(mfcol = c(5, 2), mar = c(3, 3, 1, 0.5), mgp = c(1.5, 0.7, 0))
plot(xvR.ml)


#  ordinary kriging- do this to chosen model?

pred.grid <- pred_grid(gdata2$coords, by=0.5)
kc <- krige.conv(gdata2, loc = pred.grid, krige = krige.control(obj.m =
 m5))

image(kc, val=kc$krige.var, loc = pred.grid, xlab = "Coord X", ylab =
 "Coord Y")
image(kc, loc = pred.grid, xlab = "Coord X", ylab = "Coord Y")


#bayesian
ex.grid <- as.matrix(expand.grid(seq(0,1,l=21), seq(0,1,l=21)))
```

```
#
# computing posterior and predictive distributions
# (warning: the next command can be time demanding)
ex.bayes <- krige.bayes(gdata2, loc=pred.grid,
                        model = model.control(trend.d='2nd', trend.l='2nd',
 cov.m="powered.exponential", kappa=0.45),
                        prior = prior.control(phi.discrete=seq(0, 0.7,
 l=51),
                                              phi.prior="reciprocal"))



# Plotting empirical variograms and some Bayesian estimates:
# Empirical variogram
plot(variog(gdata2, max.dist = 1), ylim=c(0, 15))
# Since ex.data is a simulated data we can plot the line with the "true"
 model
lines.variomodel(gdata2, lwd=2)
# adding lines with summaries of the posterior of the binned variogram
lines(ex.bayes, summ = mean, lwd=1, lty=2)
lines(ex.bayes, summ = median, lwd=2, lty=2)
# adding line with summary of the posterior of the parameters
lines(ex.bayes, summary = "mode", post = "parameters")

# Plotting again the empirical variogram
plot(variog(gdata2, max.dist=1), ylim=c(0, 15))
# and adding lines with median and quantiles estimates
my.summary <- function(x){quantile(x, prob = c(0.05, 0.5, 0.95))}
lines(ex.bayes, summ = my.summary, ty="l", lty=c(2,1,2), col=1)

# Plotting some prediction results
op <- par(no.readonly = TRUE)
par(mfrow=c(2,2), mar=c(4,4,2.5,0.5), mgp = c(2,1,0))
image(ex.bayes, main="predicted values")
image(ex.bayes, val="variance", main="prediction variance")
image(ex.bayes, val= "simulation", number.col=1,
      main="a simulation from the \npredictive distribution")
image(ex.bayes, val= "simulation", number.col=2,
      main="another simulation from \nthe predictive distribution")
```