Spatial autocorrelation

Eddie Wu

2024-03-28

Introduction

This .Rmd file is to test the spatial autocorrelation in our black carp data file, and try to find the best distance that reduces the spatial autocorrelation when doing subsampling. In this file, we:

1. Show the distribution of our data points on a map.
2. Use moran's I test to test the global spatial autocorrelation in our dataset.
3. Use the correlog plots to examine how spatial autocorrelation changes with distances in all 6 temperatures and 3 different models for each temperature (18 in total).

- *Annual average air. / Cold quarter air. / Warm quarter air. / GDD 0. / Annual average water. / Cold quarter water.*
- *Simple regression model. / Additive model. / Interaction model.*

4. Subsample at 250 km and 550 km, and check the local moran's I after sub-sampling to see if spatial autocorrelation has been reduced.
5. Conduct latitudinal stratification analysis.
6. Remove the South Ukarine data point and conduct the same spatial autocorrelation analyses.
7. Look at the Chinese dataset - 17 data points.

```
library(gstat)
library(ggplot2)
library(dplyr)
library(spdep)
library(sp)
library(nlme)
library(ape)
library(MuMIn)
library(raster)
library(ncf)
library(knitr)
library(rnaturalearth)
library(sf)
```

```
## Import location data
location <- read.csv("location_no_temps.csv")
location <- unique(location)

location.new <- location %>% # remove duplicating locations
  filter(Location != "South Ukraine") # remove South Ukraine

# Clean the data
```

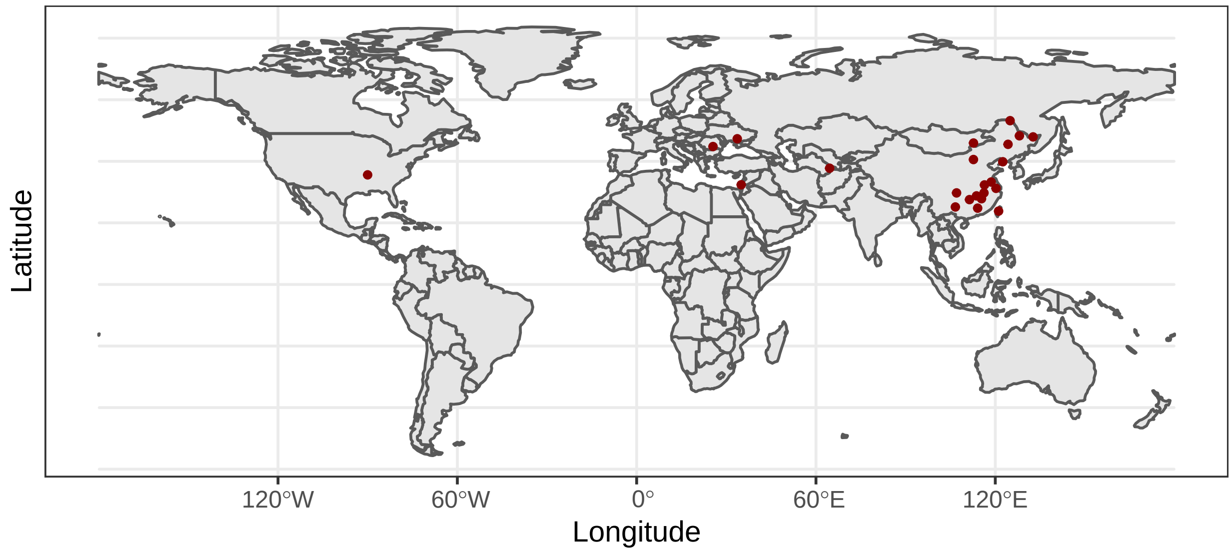
```r
carp <- read.csv("eddie_carp_new.csv")
carp.r <- carp %>%
  filter(!row_number() == 5) %>%
  filter(sex != "male")


## Download one file to get the spatial points
# (this defines what projection to use when converting to spatial objects)
tmin.1979 <- brick("cpc/tmin.1979.nc", varname = "tmin")
```

```
## Loading required namespace: ncdf4
```

```r
tmin.1979<-rotate(tmin.1979)


## Define a function to plot the correlogram for local moran's I

plot_correlog <- function(model){
  # Define margins
  par(mfrow=c(2,1),mar=c(4,4,2,2))

  # Use the coorelog function to develop the relationship
  test <- correlog(coords$long, coords$lat, model$residuals,
                   increment=50, resamp=500, latlon=T)

  # Plot with the entire distance range
  plot(test, main=paste(model$call))
  abline(h=0)

  # Reduce the distance range to 2500 km
  plot(test, main="", xlim=c(0,2500))
  abline(h=0)
}
```

Data distribution

```r
## Get the world data
world <- ne_countries(scale = "small", returnclass = "sf")

## Plots (global plot)
world %>%
  filter(admin != "Antarctica") %>%
  # remove antarctica
  ggplot()+
  geom_sf()+
  geom_point(aes(x=Longitude, y=Latitude), data = location,
             size = 1, color = "darkred")+
  theme_bw()+
  xlab("Longitude") + ylab("Latitude")
```



Section 1: Spatial autocorrelation examination for entire dataset

Model definition

We first need to define all 18 models (3 models for each temperature).

```
# Define the models for AnnualTemp
annual.lin <- lm(log(AAM)~AnnualTemp, data = carp.r)
annual.add <- lm(log(AAM)~AnnualTemp+condition, data = carp.r)
annual.int <- lm(log(AAM)~AnnualTemp*condition, data = carp.r)

# Define the models for ColdTemp
cold.lin <- lm(log(AAM)~ColdTemp, data = carp.r)
cold.add <- lm(log(AAM)~ColdTemp+condition, data = carp.r)
cold.int <- lm(log(AAM)~ColdTemp*condition, data = carp.r)

# Define the model for WarmTemp
warm.lin <- lm(log(AAM)~WarmTemp, data = carp.r)
warm.add <- lm(log(AAM)~WarmTemp+condition, data = carp.r)
warm.int <- lm(log(AAM)~WarmTemp*condition, data = carp.r)

# Define the models for GDD0
gdd.lin <- lm(log(AAM)~average_gdd_0, data = carp.r)
gdd.add <- lm(log(AAM)~average_gdd_0+condition, data = carp.r)
gdd.int <- lm(log(AAM)~average_gdd_0*condition, data = carp.r)

# Define the models for WaterTemp
```

```
water.lin <- lm(log(AAM)~WaterTemp, data = carp.r)
water.add <- lm(log(AAM)~WaterTemp+condition, data = carp.r)
water.int <- lm(log(AAM)~WaterTemp*condition, data = carp.r)

# Define the models for WaterCold
waterC.lin <- lm(log(AAM)~WaterCold, data = carp.r)
waterC.add <- lm(log(AAM)~WaterCold+condition, data = carp.r)
waterC.int <- lm(log(AAM)~WaterCold*condition, data = carp.r)
```

Global Moran.I

```
## Make spatial dataframe
coords <- data.frame("long"=location[,3],"lat"=location[,2])
df <- data.frame(a = 1:nrow(location[3]))
spatial.data <- SpatialPointsDataFrame(coords,df,proj4string = tmin.1979@crs)

# Get a distance matrix from all points
dists <- spDists(spatial.data, longlat = TRUE)


## Global Moran.I
# Annual Temp
Moran.annual.lin <- Moran.I(annual.lin$residuals, dists)
Moran.annual.add <- Moran.I(annual.add$residuals, dists)
Moran.annual.int <- Moran.I(annual.int$residuals, dists)

global.moran.annual <- data.frame(
  Model = c("simple linear", "linear additive", "interaction"),
  Observed = c(Moran.annual.lin$observed, Moran.annual.add$observed,
               Moran.annual.int$observed),
  Expected = c(Moran.annual.lin$expected, Moran.annual.add$expected,
               Moran.annual.int$expected),
  sd = c(Moran.annual.lin$sd, Moran.annual.add$sd, Moran.annual.int$sd),
  p.value = c(Moran.annual.lin$p.value, Moran.annual.add$p.value,
              Moran.annual.int$p.value)
)

kable(global.moran.annual)
```

Model	Observed	Expected	sd	p.value
simple linear	-0.0276451	-0.0454545	0.0317115	0.5743823
linear additive	-0.0171313	-0.0454545	0.0317411	0.3722215
interaction	-0.0154402	-0.0454545	0.0318226	0.3455901

```
# Cold Temp
Moran.cold.lin <- Moran.I(cold.lin$residuals, dists)
Moran.cold.add <- Moran.I(cold.add$residuals, dists)
Moran.cold.int <- Moran.I(cold.int$residuals, dists)

global.moran.cold <- data.frame(
  Model = c("simple linear", "linear additive", "interaction"),
  Observed = c(Moran.cold.lin$observed, Moran.cold.add$observed,
```

```r
                Moran.cold.int$observed),
  Expected = c(Moran.cold.lin$expected, Moran.cold.add$expected,
                Moran.cold.int$expected),
  sd = c(Moran.cold.lin$sd, Moran.cold.add$sd, Moran.cold.int$sd),
  p.value = c(Moran.cold.lin$p.value, Moran.cold.add$p.value,
               Moran.cold.int$p.value)
)

kable(global.moran.cold)
```

Model	Observed	Expected	sd	p.value
simple linear	-0.0328926	-0.0454545	0.0318228	0.6930295
linear additive	-0.0201217	-0.0454545	0.0318073	0.4257723
interaction	-0.0160914	-0.0454545	0.0319118	0.3575019

```r
# Warm Temp
Moran.warm.lin <- Moran.I(warm.lin$residuals, dists)
Moran.warm.add <- Moran.I(warm.add$residuals, dists)
Moran.warm.int <- Moran.I(warm.int$residuals, dists)

global.moran.warm <- data.frame(
  Model = c("simple linear", "linear additive", "interaction"),
  Observed = c(Moran.warm.lin$observed, Moran.warm.add$observed,
                Moran.warm.int$observed),
  Expected = c(Moran.warm.lin$expected, Moran.warm.add$expected,
                Moran.warm.int$expected),
  sd = c(Moran.warm.lin$sd, Moran.warm.add$sd, Moran.warm.int$sd),
  p.value = c(Moran.warm.lin$p.value, Moran.warm.add$p.value,
               Moran.warm.int$p.value)
)

kable(global.moran.warm)
```

Model	Observed	Expected	sd	p.value
simple linear	-0.0408894	-0.0454545	0.0315415	0.8849196
linear additive	-0.0335627	-0.0454545	0.0315865	0.7065566
interaction	-0.0363573	-0.0454545	0.0315894	0.7733589

```r
# GDD0
Moran.gdd.lin <- Moran.I(gdd.lin$residuals, dists)
Moran.gdd.add <- Moran.I(gdd.add$residuals, dists)
Moran.gdd.int <- Moran.I(gdd.int$residuals, dists)

global.moran.gdd <- data.frame(
  Model = c("simple linear", "linear additive", "interaction"),
  Observed = c(Moran.gdd.lin$observed, Moran.gdd.add$observed,
                Moran.gdd.int$observed),
  Expected = c(Moran.gdd.lin$expected, Moran.gdd.add$expected,
                Moran.warm.int$expected),
  sd = c(Moran.gdd.lin$sd, Moran.gdd.add$sd, Moran.gdd.int$sd),
  p.value = c(Moran.gdd.lin$p.value, Moran.gdd.add$p.value,
```

```r
              Moran.gdd.int$p.value)
)

kable(global.moran.gdd)
```

Model	Observed	Expected	sd	p.value
simple linear	-0.0247012	-0.0454545	0.0317304	0.5130789
linear additive	-0.0154093	-0.0454545	0.0318183	0.3450276
interaction	-0.0136061	-0.0454545	0.0318872	0.3178994

```r
# Annual water temp
Moran.water.lin <- Moran.I(water.lin$residuals, dists)
Moran.water.add <- Moran.I(water.add$residuals, dists)
Moran.water.int <- Moran.I(water.int$residuals, dists)

global.moran.water <- data.frame(
  Model = c("simple linear", "linear additive", "interaction"),
  Observed = c(Moran.water.lin$observed, Moran.water.add$observed,
               Moran.water.int$observed),
  Expected = c(Moran.water.lin$expected, Moran.water.add$expected,
               Moran.water.int$expected),
  sd = c(Moran.water.lin$sd, Moran.water.add$sd, Moran.water.int$sd),
  p.value = c(Moran.water.lin$p.value, Moran.water.add$p.value,
              Moran.water.int$p.value)
)

kable(global.moran.water)
```

Model	Observed	Expected	sd	p.value
simple linear	-0.0408571	-0.0454545	0.0316097	0.8843611
linear additive	-0.0261522	-0.0454545	0.0316502	0.5419511
interaction	-0.0213996	-0.0454545	0.0317488	0.4486524

```r
# Cold water temp
Moran.waterC.lin <- Moran.I(waterC.lin$residuals, dists)
Moran.waterC.add <- Moran.I(waterC.add$residuals, dists)
Moran.waterC.int <- Moran.I(waterC.int$residuals, dists)

global.moran.waterC <- data.frame(
  Model = c("simple linear", "linear additive", "interaction"),
  Observed = c(Moran.waterC.lin$observed, Moran.waterC.add$observed,
               Moran.waterC.int$observed),
  Expected = c(Moran.waterC.lin$expected, Moran.waterC.add$expected,
               Moran.waterC.int$expected),
  sd = c(Moran.waterC.lin$sd, Moran.waterC.add$sd, Moran.waterC.int$sd),
  p.value = c(Moran.waterC.lin$p.value, Moran.waterC.add$p.value,
              Moran.waterC.int$p.value)
)

kable(global.moran.waterC)
```

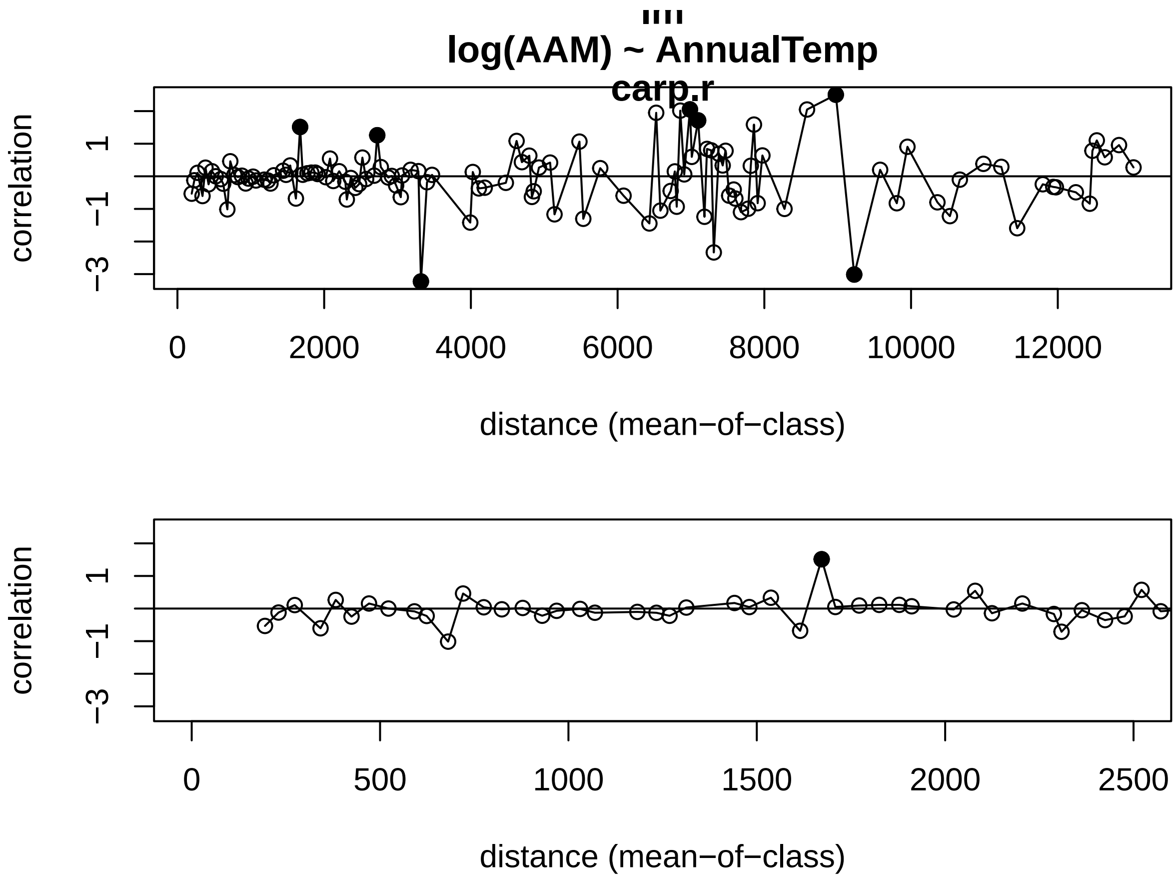
Model	Observed	Expected	sd	p.value
simple linear	-0.0282095	-0.0454545	0.0318057	0.5876811
linear additive	-0.0190927	-0.0454545	0.0319068	0.4086825
interaction	-0.0211682	-0.0454545	0.0320296	0.4483028

There is no global spatial autocorrelation on the entire dataset for all three temperature metrics.

Local spatial autocorrelation

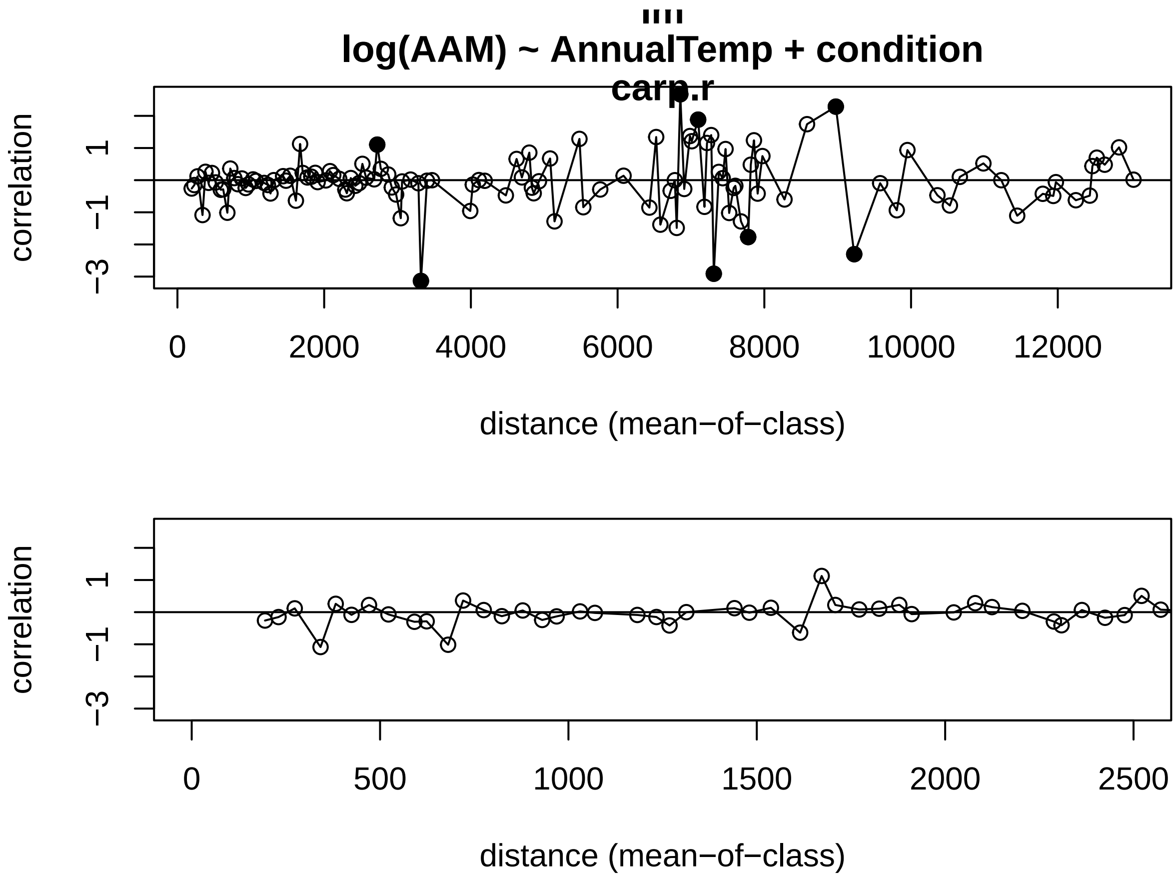
```
## Annual temperature
plot_correlog(annual.lin)
```

```
## 50   of   500 100   of   500 150   of   500 200   of   500 250   of   500 300   of   500 350   of   500 400   of   5
```



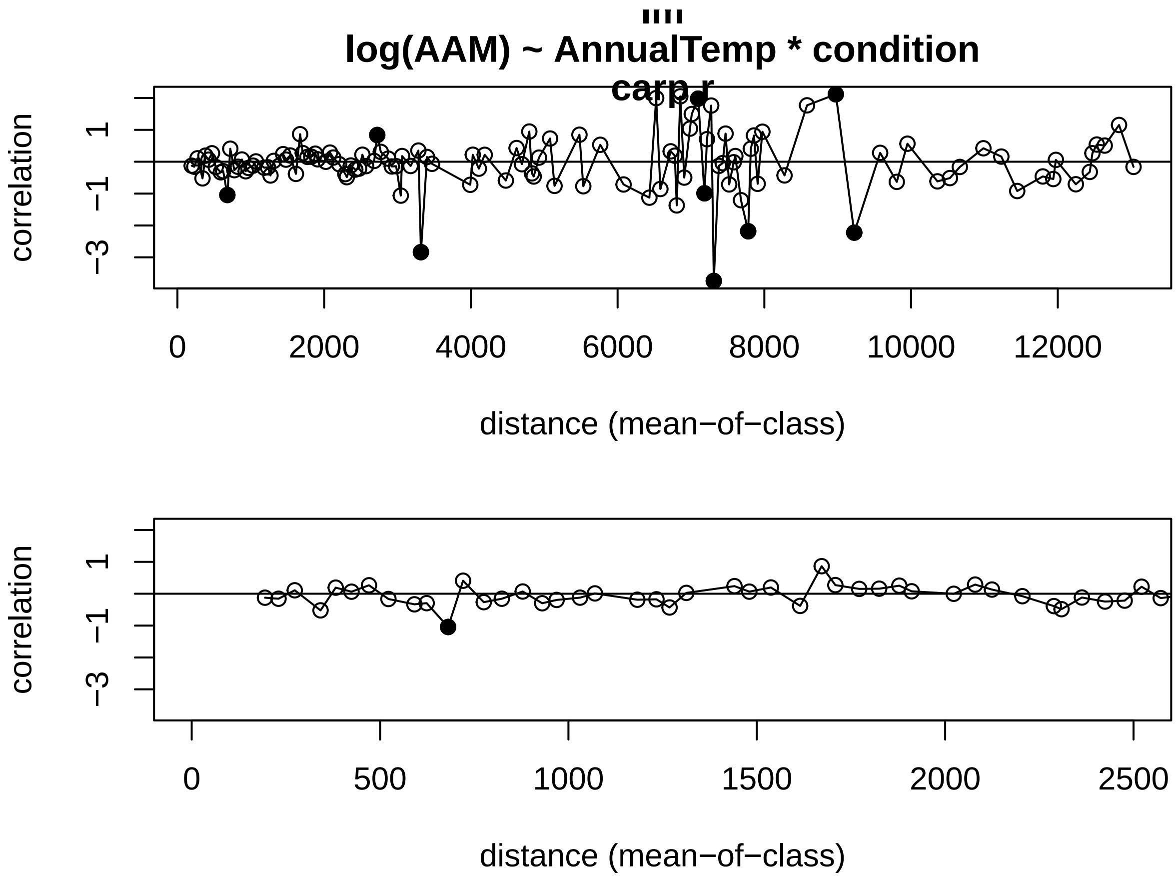

```
plot_correlog(annual.add)
```

```
## 50   of   500 100   of   500 150   of   500 200   of   500 250   of   500 300   of   500 350   of   500 400   of   5
```



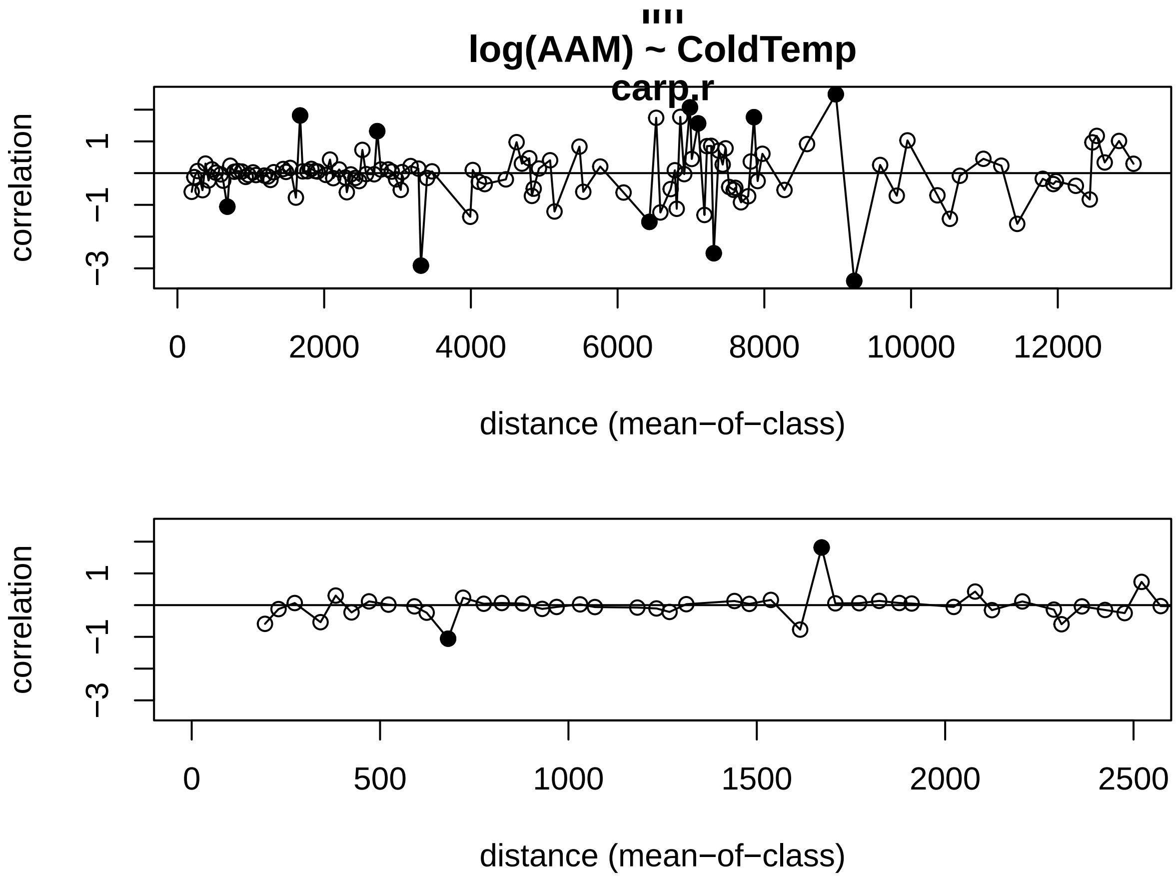

```
plot_correlog(annual.int)
```

```
## 50   of   500 100   of   500 150   of   500 200   of   500 250   of   500 300   of   500 350   of   500 400   of   5
```



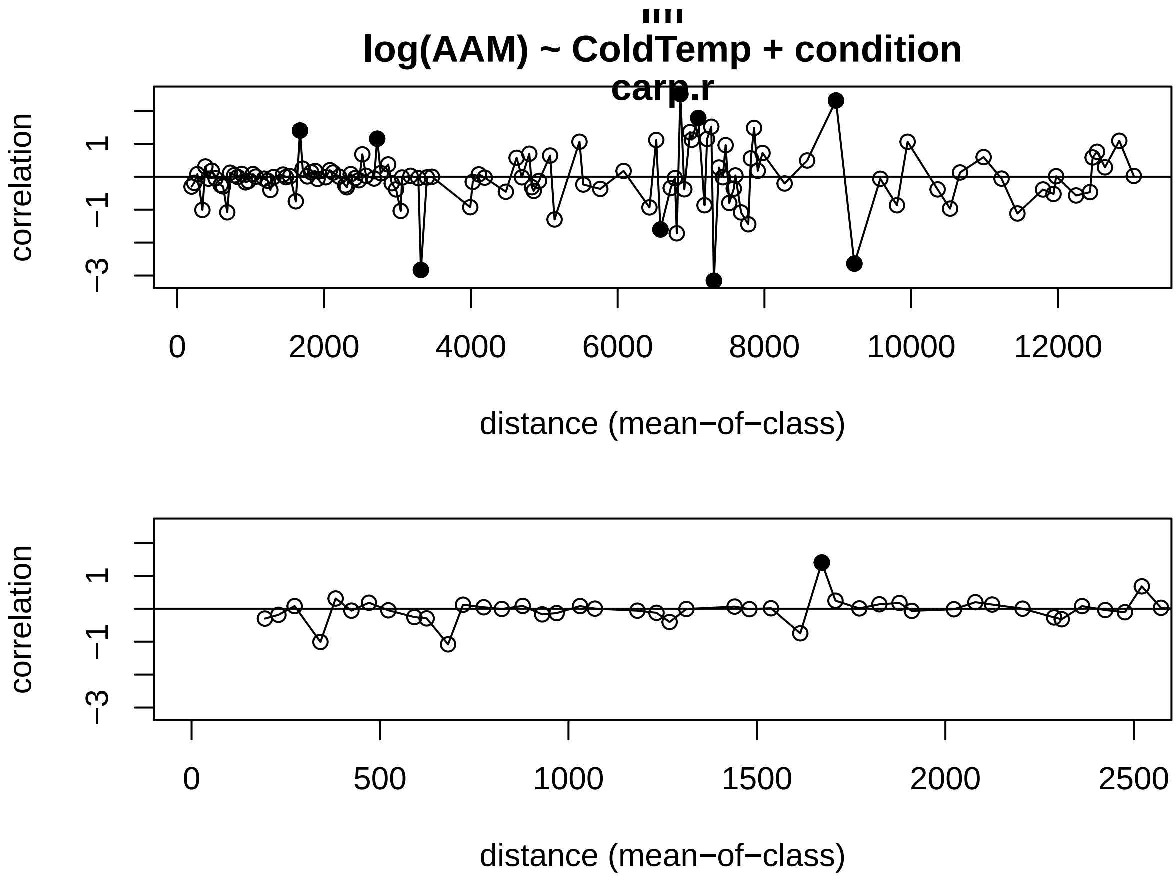

```r
## Cold temperature
plot_correlog(cold.lin)
```

```
## 50   of   500 100   of   500 150   of   500 200   of   500 250   of   500 300   of   500 350   of   500 400   of   5
```



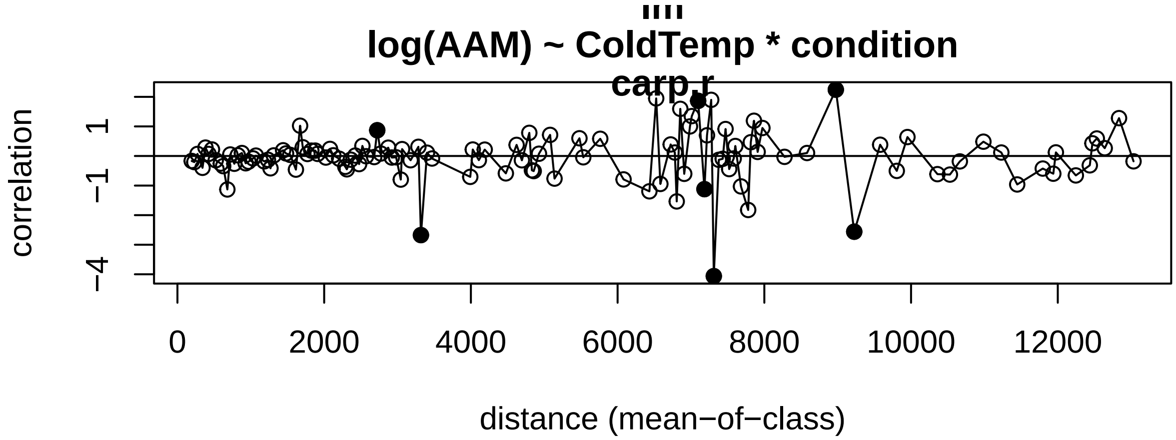

```
plot_correlog(cold.add)
```

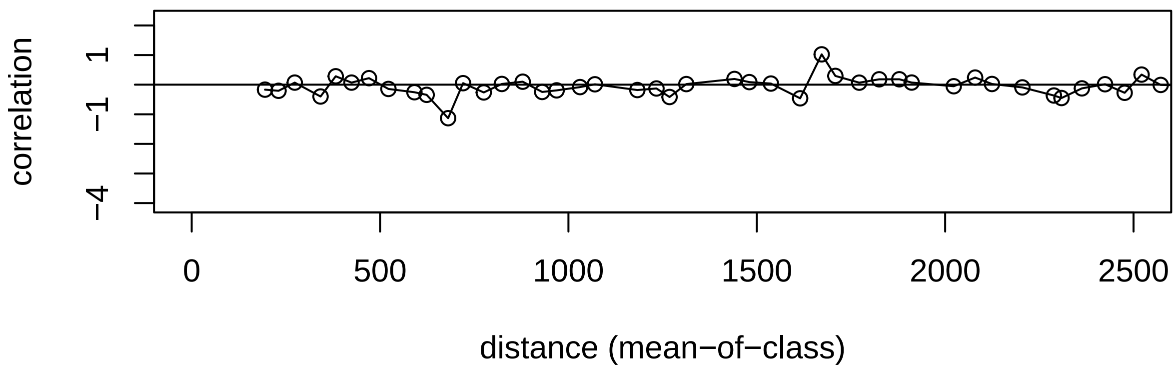
```
## 50  of  500 100  of  500 150  of  500 200  of  500 250  of  500 300  of  500 350  of  500 400  of  5
```




```
plot_correlog(cold.int)
```

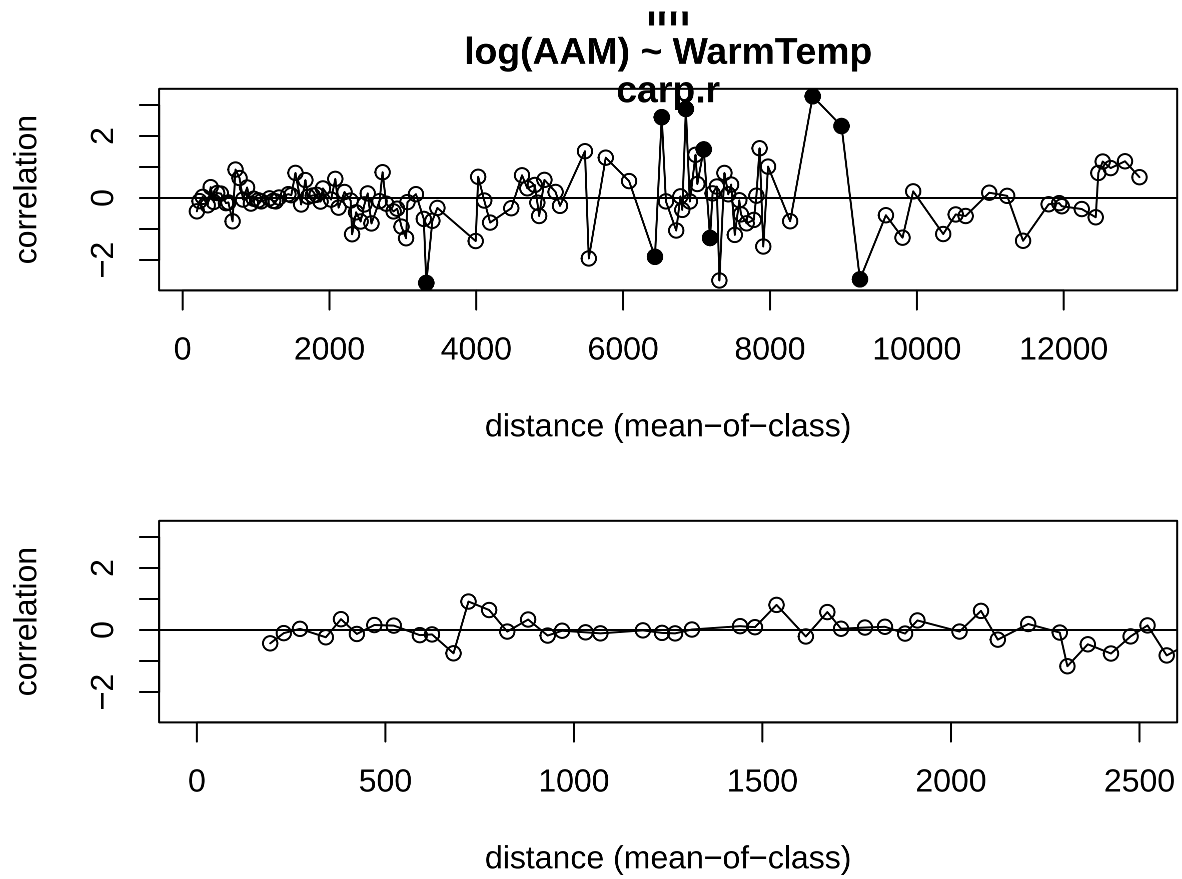
```
## 50  of  500 100  of  500 150  of  500 200  of  500 250  of  500 300  of  500 350  of  500 400  of  5
```





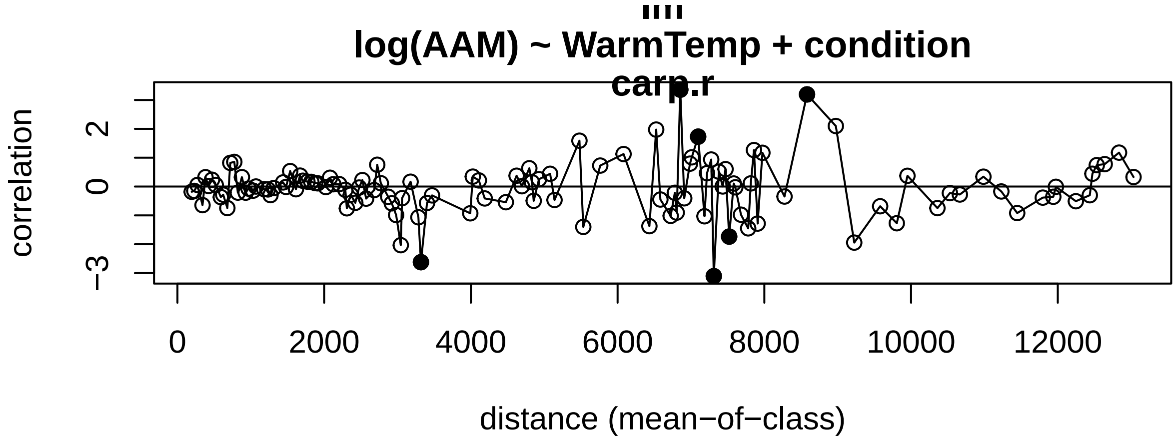

```r
## Warm temperature
plot_correlog(warm.lin)
```

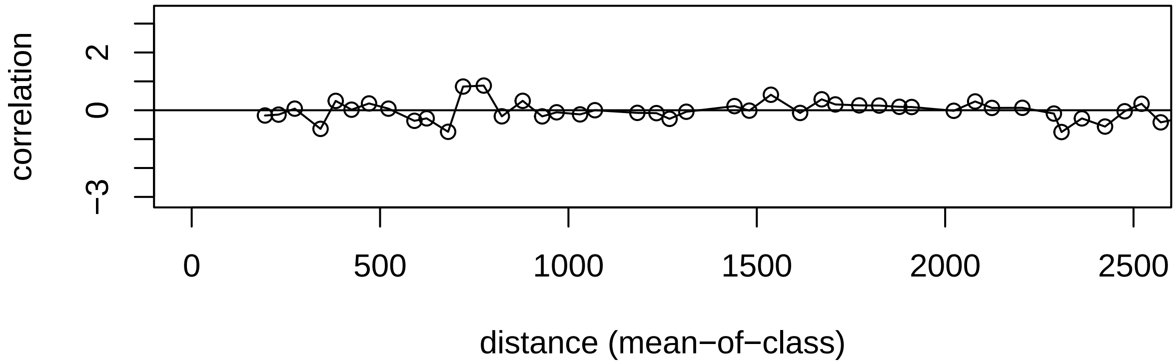
```
## 50   of   500 100   of   500 150   of   500 200   of   500 250   of   500 300   of   500 350   of   500 400   of   5
```




```
plot_correlog(warm.add)
```

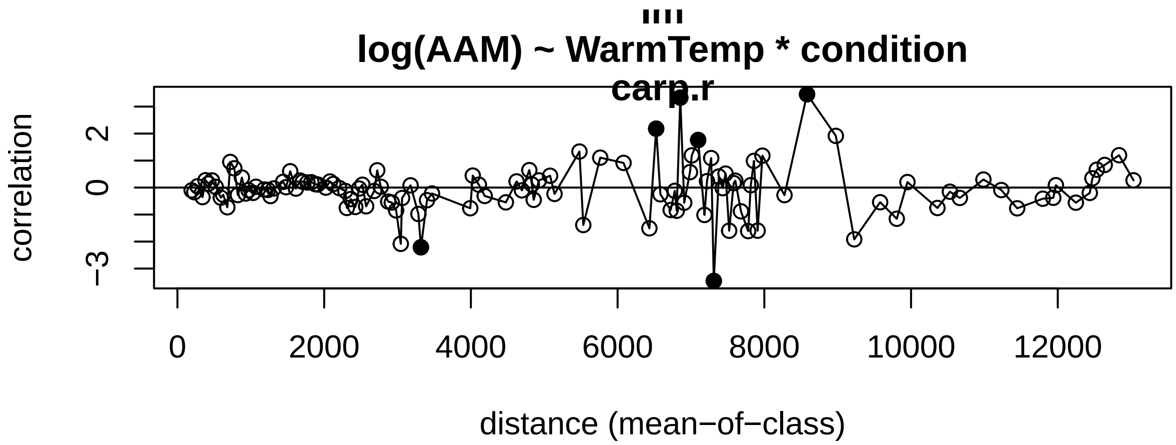
```
## 50   of   500 100   of   500 150   of   500 200   of   500 250   of   500 300   of   500 350   of   500 400   of   5
```

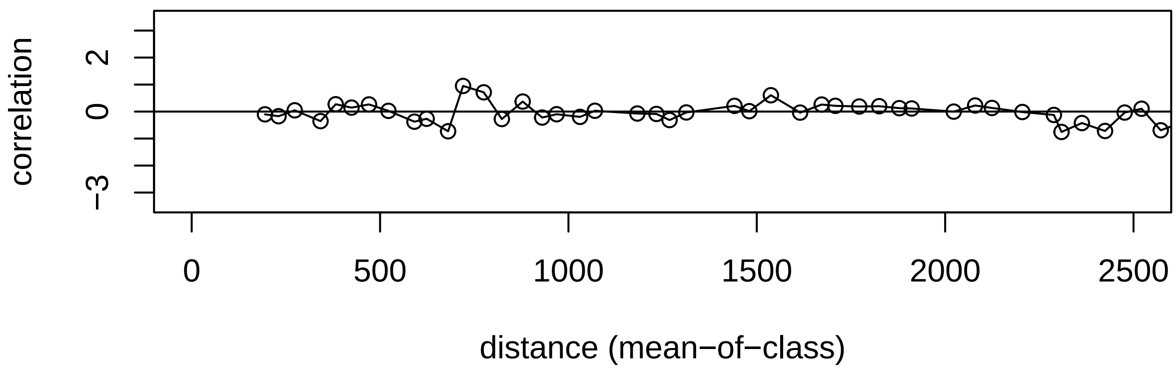





```
plot_correlog(warm.int)
```

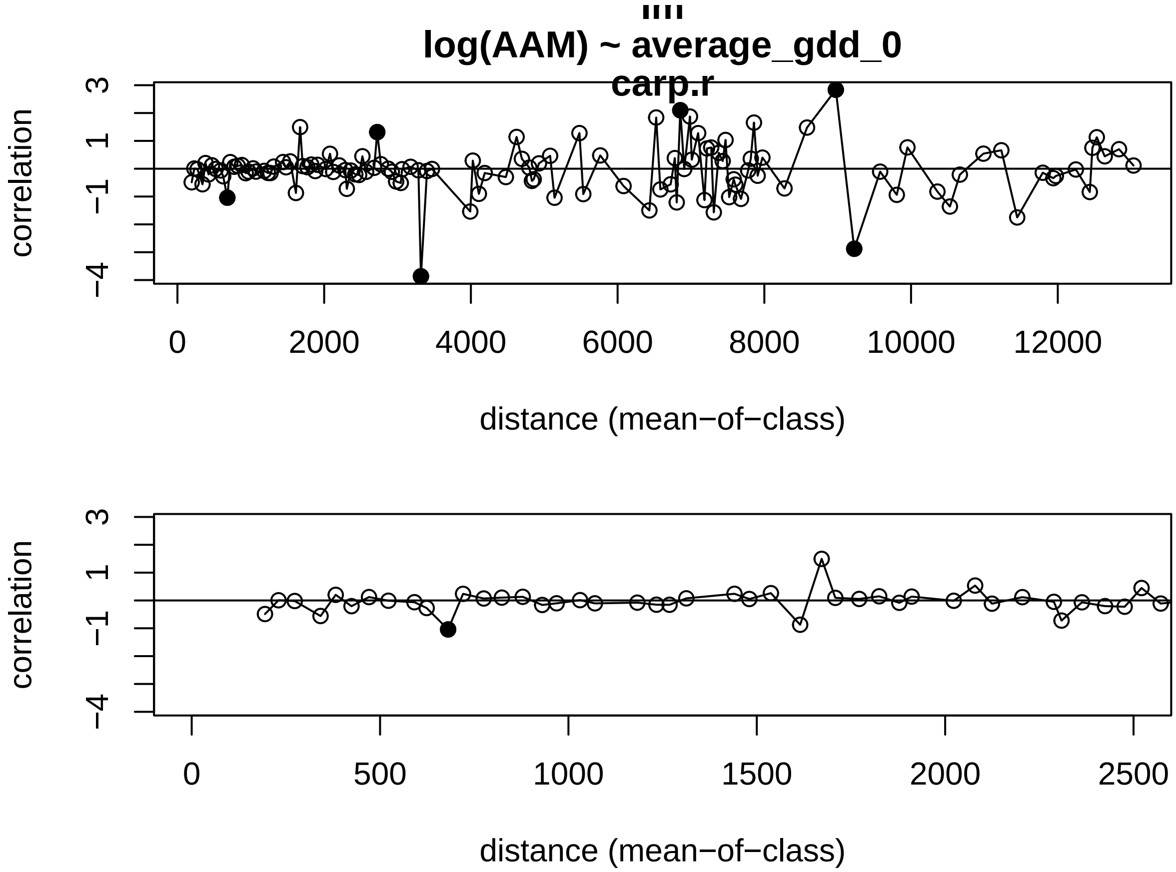
```
## 50   of   500 100   of   500 150   of   500 200   of   500 250   of   500 300   of   500 350   of   500 400   of   5(
```





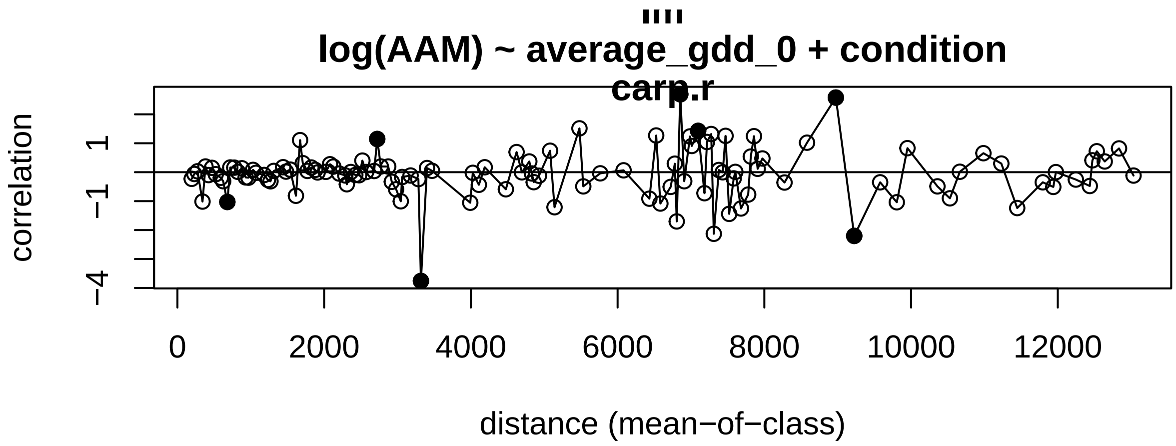

```
## GDD0
plot_correlog(gdd.lin)
```

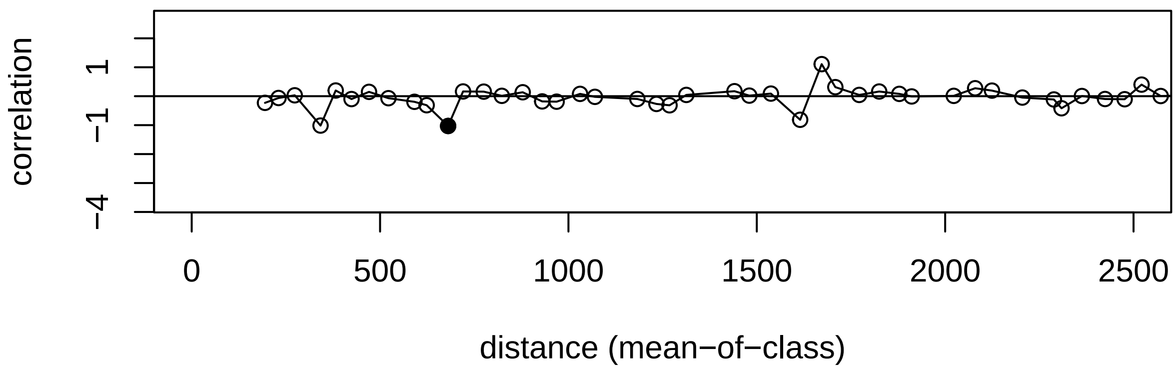
```
## 50   of   500 100   of   500 150   of   500 200   of   500 250   of   500 300   of   500 350   of   500 400   of   5
```




```
plot_correlog(gdd.add)
```

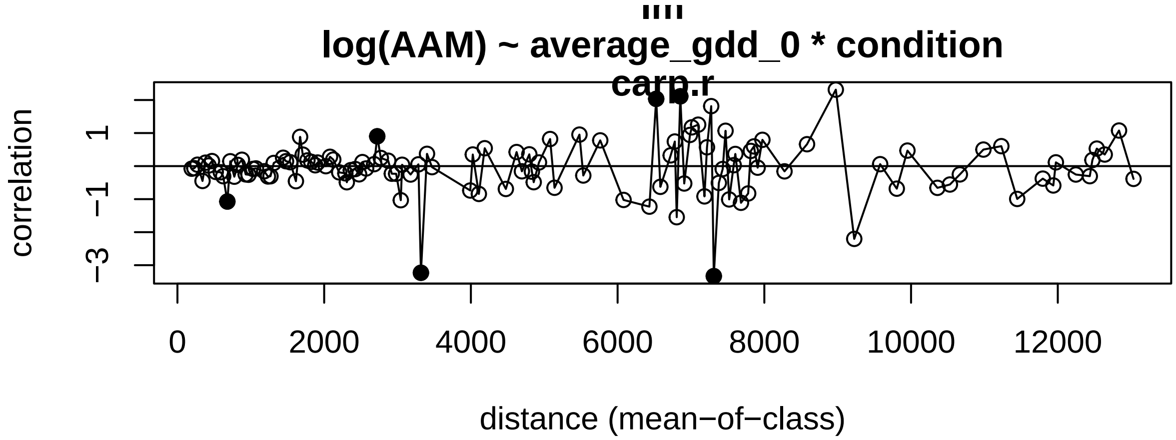
```
## 50  of  500 100  of  500 150  of  500 200  of  500 250  of  500 300  of  500 350  of  500 400  of  5
```

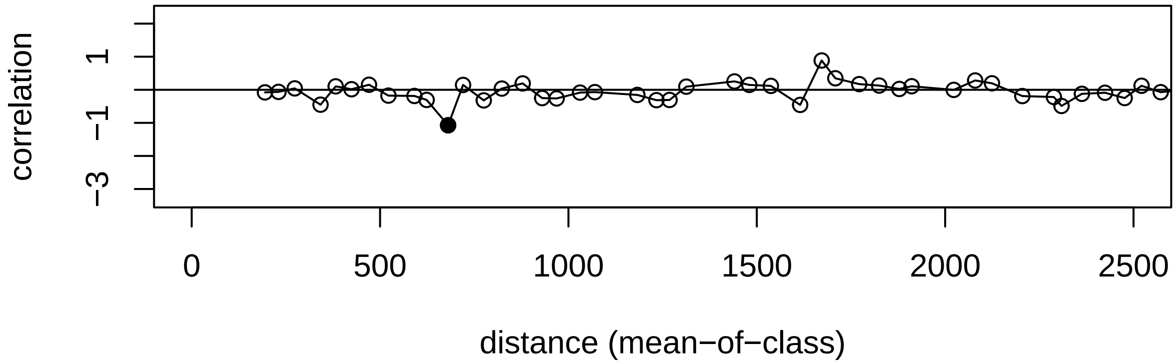





```
plot_correlog(gdd.int)
```

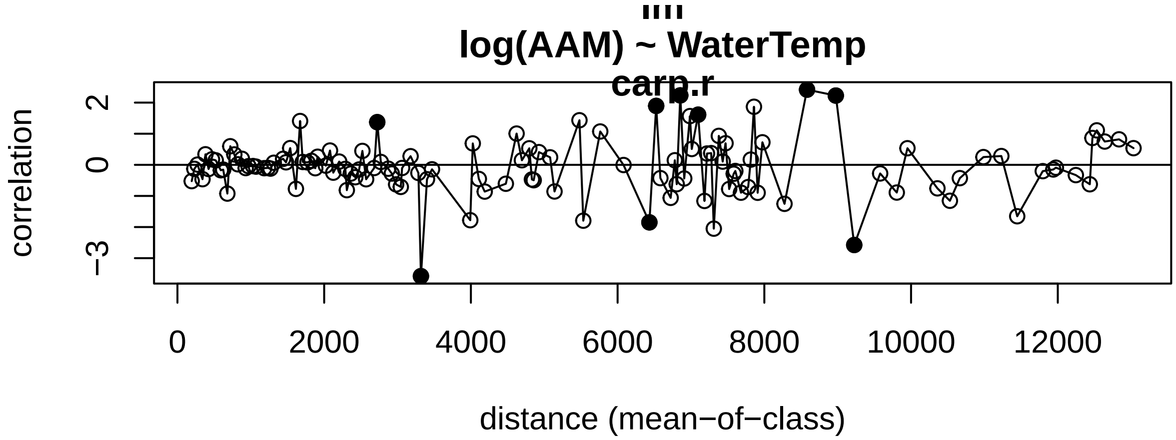
```
## 50  of  500 100  of  500 150  of  500 200  of  500 250  of  500 300  of  500 350  of  500 400  of  5
```

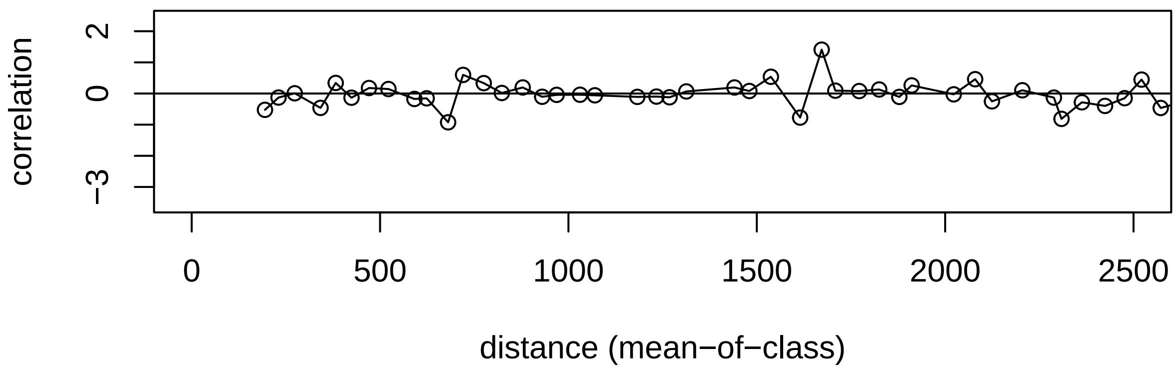





```r
## Annual water
plot_correlog(water.lin)
```

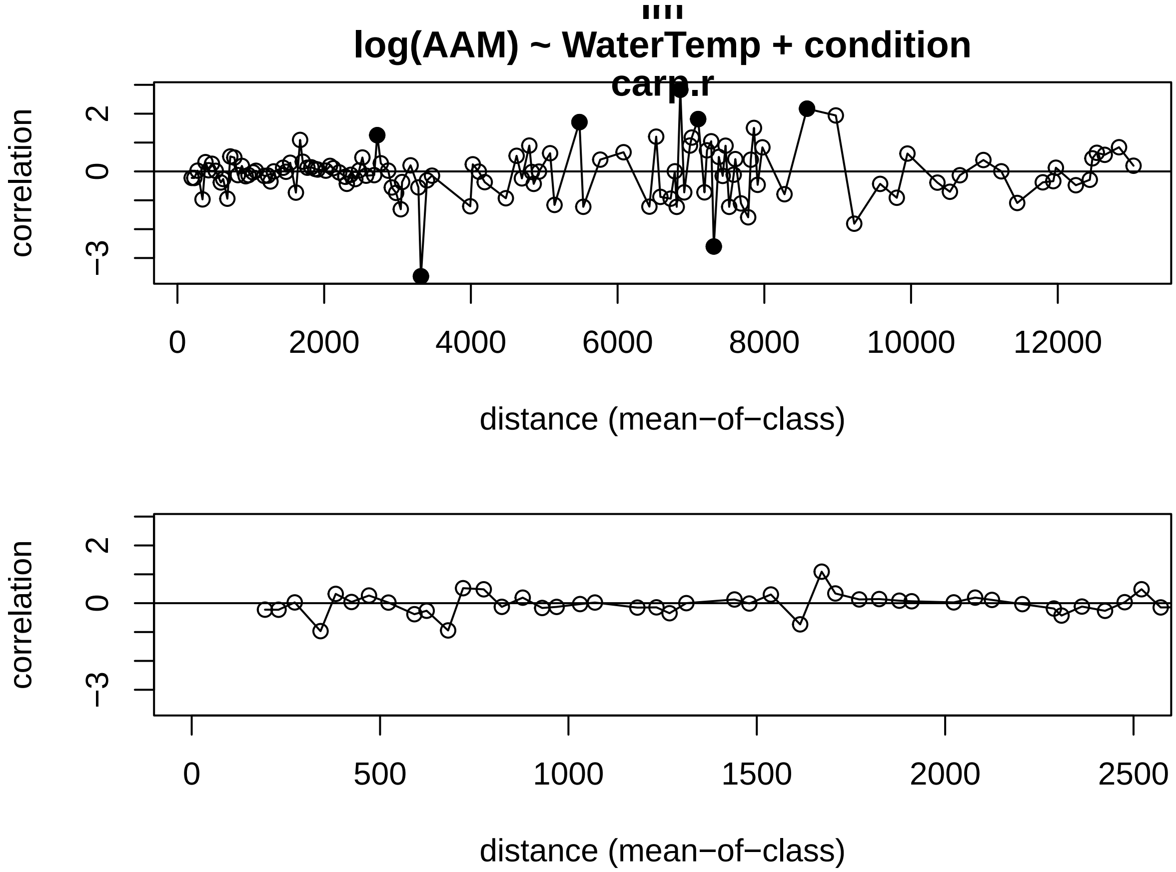
```
## 50   of   500 100   of   500 150   of   500 200   of   500 250   of   500 300   of   500 350   of   500 400   of   5
```





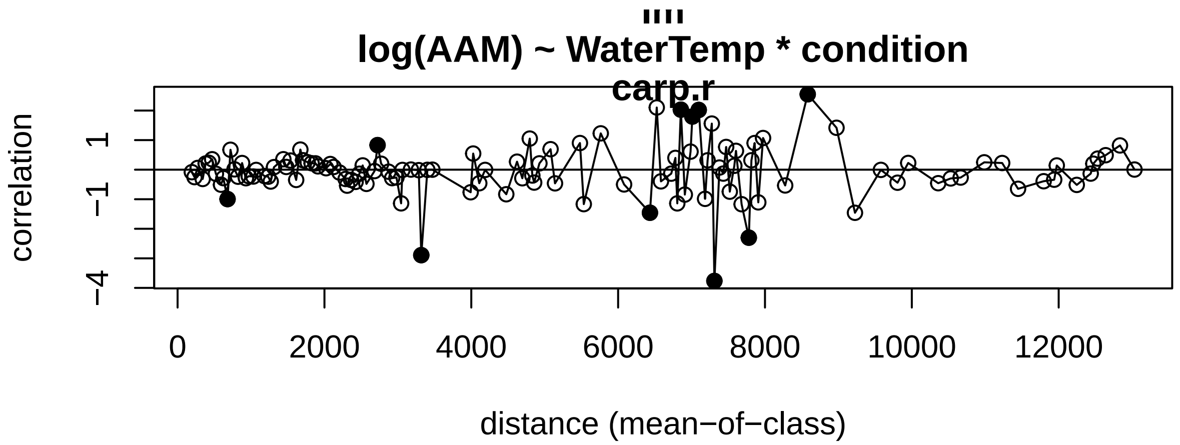

```
plot_correlog(water.add)
```

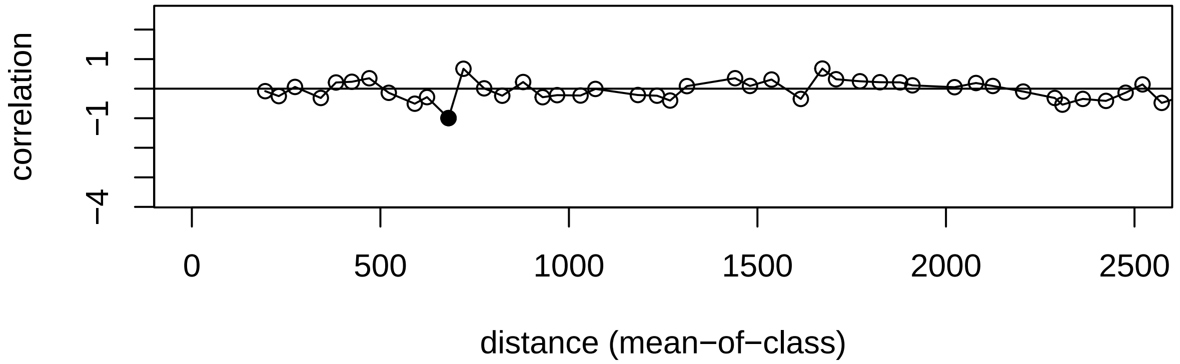
```
## 50   of   500 100   of   500 150   of   500 200   of   500 250   of   500 300   of   500 350   of   500 400   of   50
```




```
plot_correlog(water.int)
```

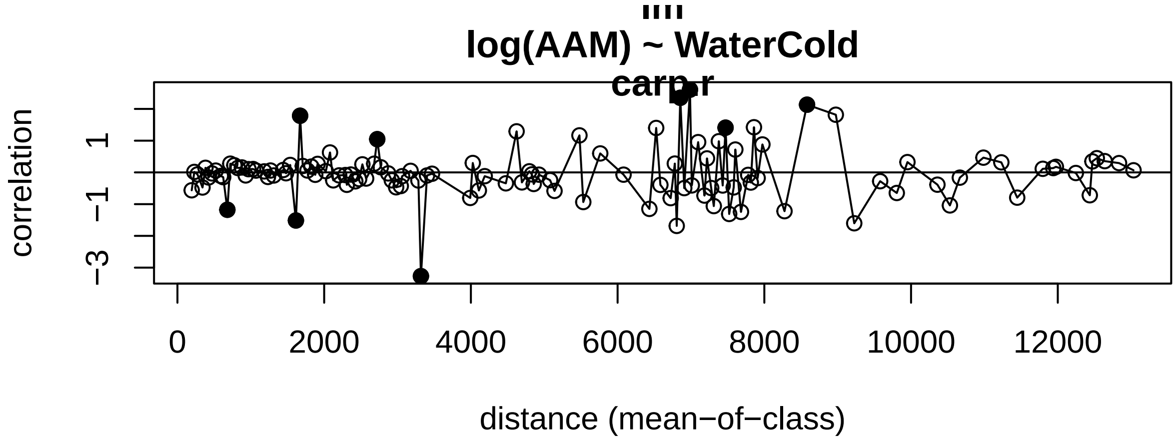
```
## 50   of   500 100   of   500 150   of   500 200   of   500 250   of   500 300   of   500 350   of   500 400   of   5
```

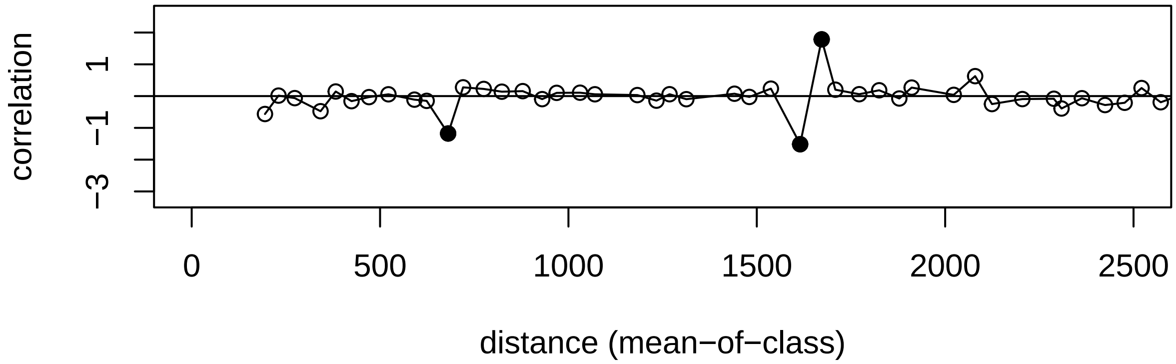





```
## Cold water
plot_correlog(waterC.lin)
```

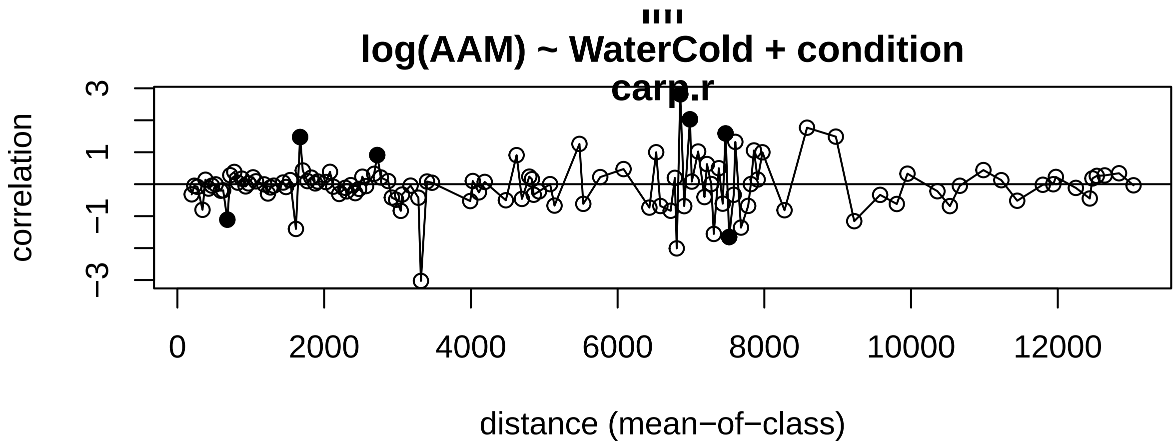
```
## 50   of   500 100   of   500 150   of   500 200   of   500 250   of   500 300   of   500 350   of   500 400   of   5(
```

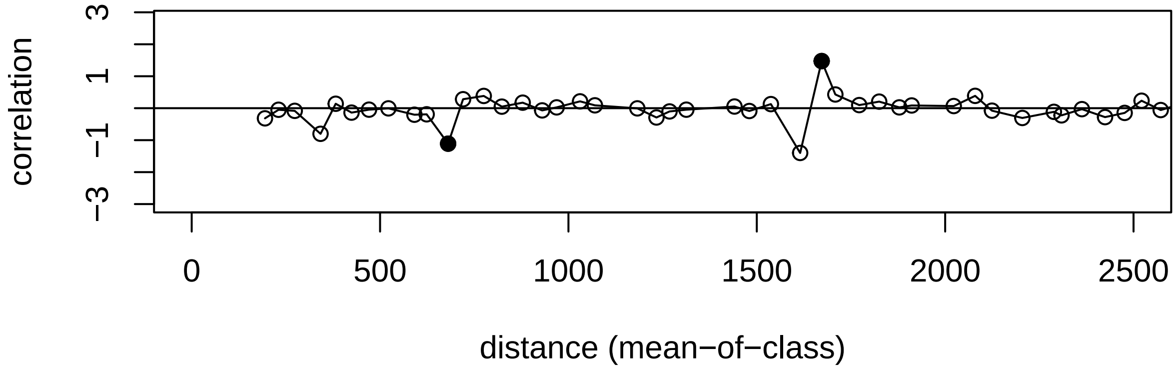





```r
plot_correlog(waterC.add)
```

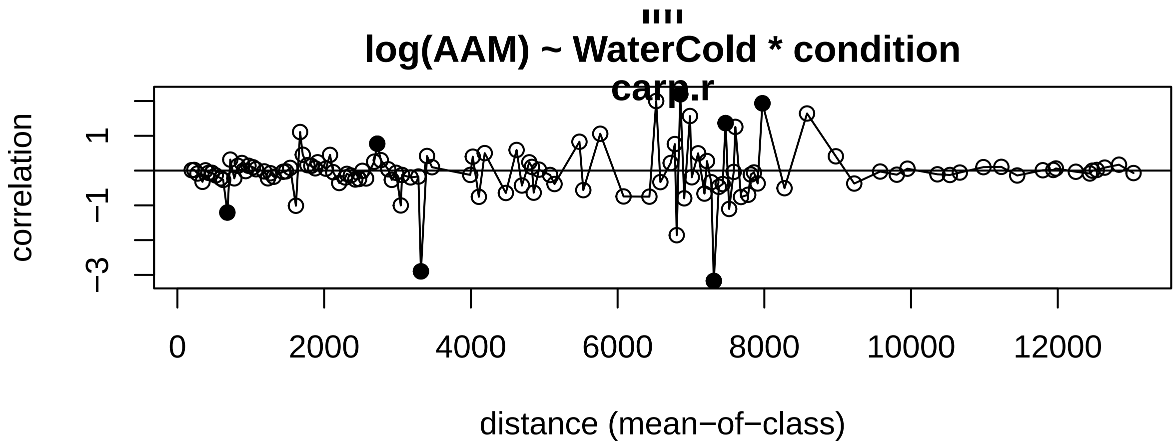
```
## 50   of   500 100   of   500 150   of   500 200   of   500 250   of   500 300   of   500 350   of   500 400   of   5
```

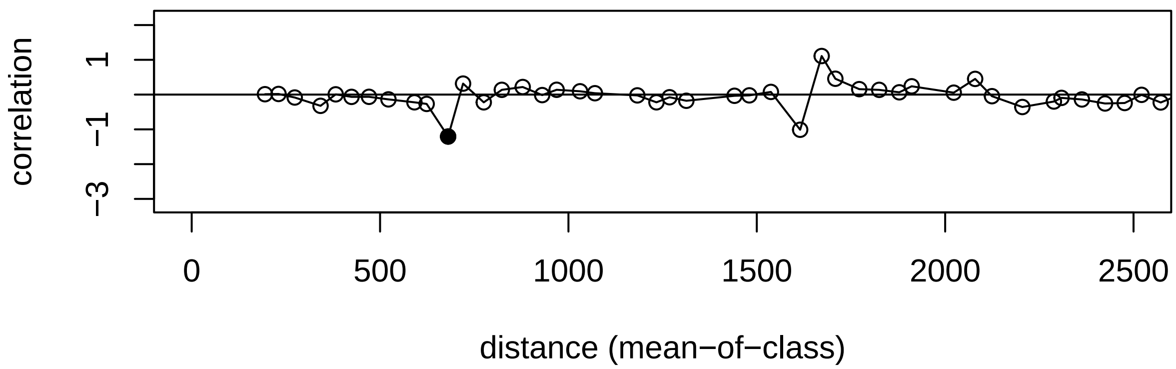





```
plot_correlog(waterC.int)
```

```
## 50   of   500 100   of   500 150   of   500 200   of   500 250   of   500 300   of   500 350   of   500 400   of   50
```





In general, the correlog plots suggest that for annual average, cold, gdd, water cold, we have a significant negative spatial autocorrelation at 550 km.

Howeve, warm air temperature, and annual water temperature does not have this negative spatial autocorrelation at 550 km.

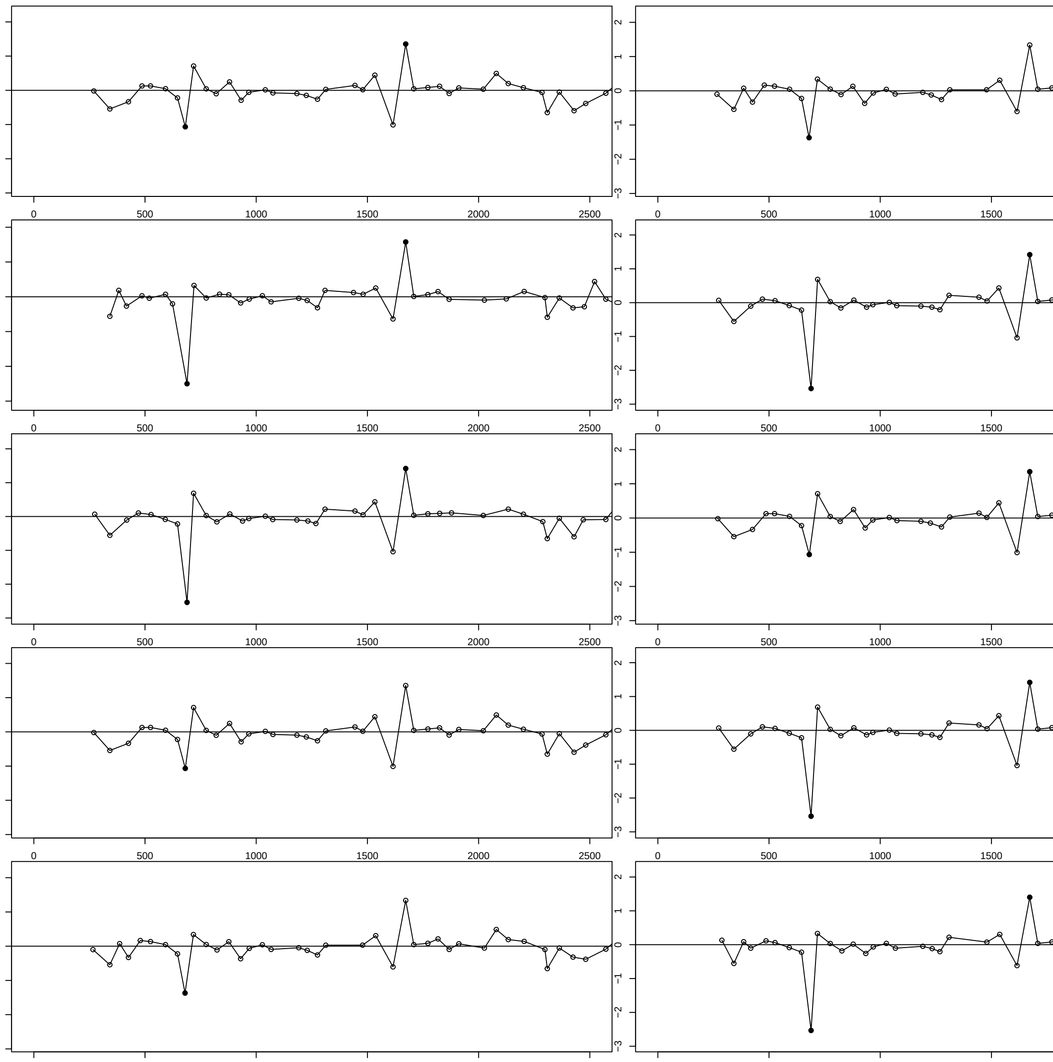
Section 2: Subsampling

Subsampling at 250km

Here, we try to sub-sample at 250km and check the local moran's I (correlog plot) after sub-sampling.

```
# Store the r2 values
r2.250.raw <- matrix(NA,10, 2)
colnames(r2.250.raw) <- c("annual","cold")

table(carp.r$spatial.code.250)

par(mfrow = c(5, 2))
par(mar=c(1,1,1,1))

## Check the local moran's results after sub-sampling
for (i in 1:10){
  sub <- carp.r %>% group_by(spatial.code.250) %>% sample_n(size=1)
  reg.sub.annual <- lm(log(sub$AAM)~sub$AnnualTemp)
  test <- correlog(sub$longitude, sub$latitude, reg.sub.annual$residuals,
```

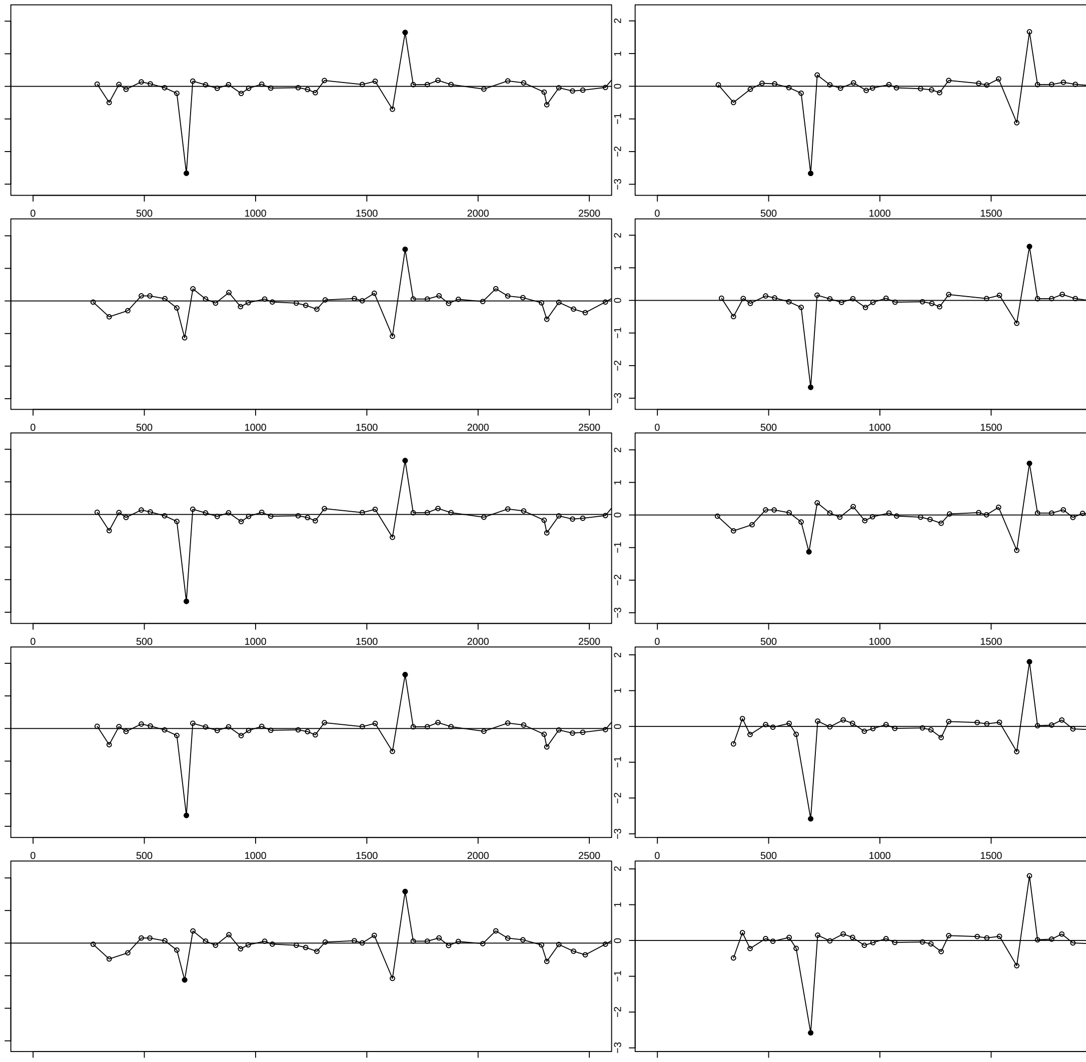
```r
                     increment=50, resamp=500, latlon=T)
  plot(test, main="", xlim=c(0,2500))
  abline(h=0)
  r2.250.raw[i,1] <- summary(reg.sub.annual)$adj.r.squared #get the r2
}
```



Local moran's I - annual 250

```r
table(carp.r$spatial.code.250)

par(mfrow = c(5, 2))
par(mar=c(1,1,1,1))

## Check the local moran's results after sub-sampling
for (i in 1:10){
  sub <- carp.r %>% group_by(spatial.code.250) %>% sample_n(size=1)
  reg.sub.cold <- lm(log(sub$AAM)~sub$ColdTemp)
  test <- correlog(sub$longitude, sub$latitude, reg.sub.cold$residuals,
                     increment=50, resamp=500, latlon=T)
  plot(test, main="", xlim=c(0,2500))
```

```
  abline(h=0)
  r2.250.raw[i,2] <- summary(reg.sub.cold)$adj.r.squared #get the r2
}
```



Local moran's I - cold 250

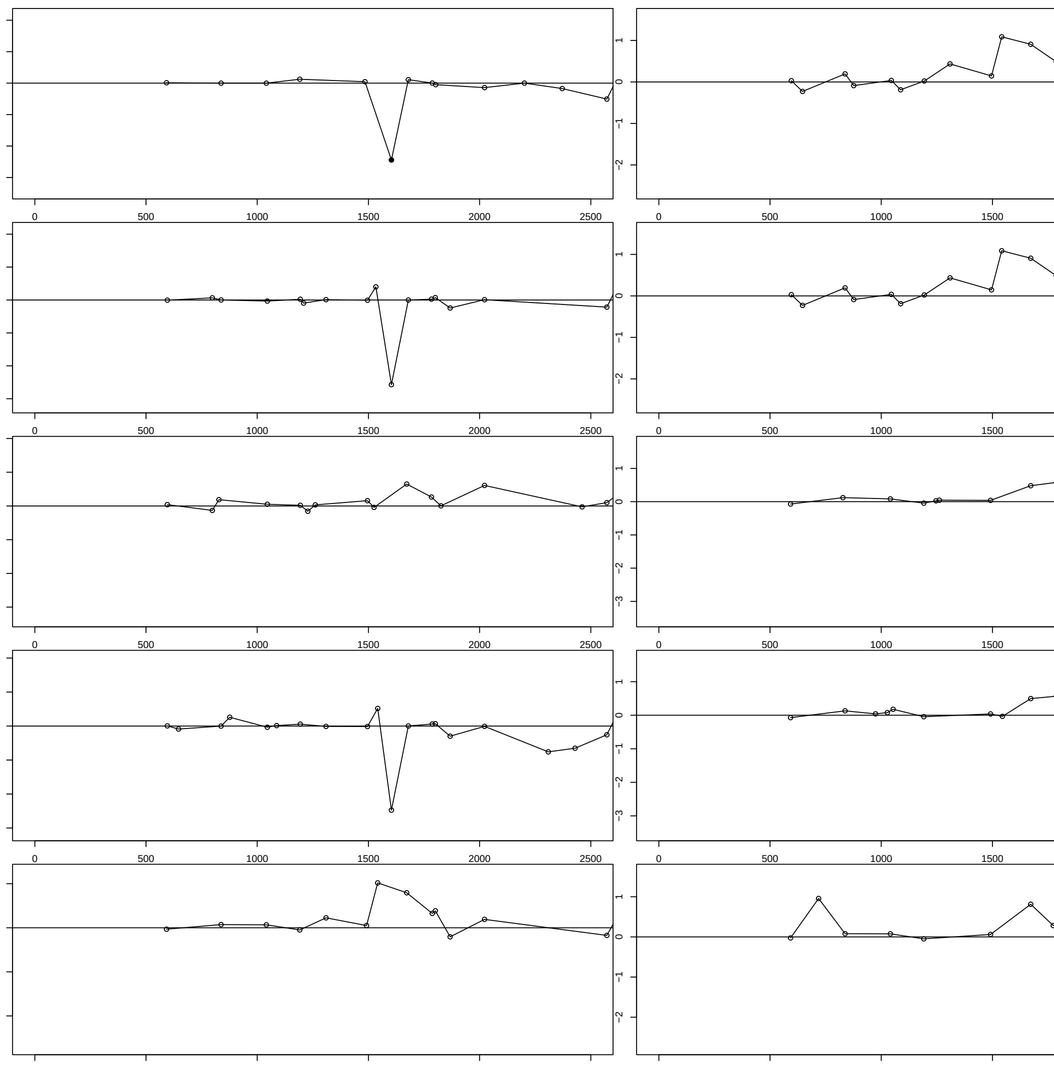
Subsampling at 250 km does not reduce the spatial autocorrelation in our dataset.

Subsampling at 550km

Now, we try to subsample at a larger distance - 550km.

```
table(carp.r$spatial.code.550)

# Store the r2 values
r2.550.raw <- matrix(NA,10, 2)
colnames(r2.550.raw) <- c("annual","cold")

par(mfrow = c(5, 2))
par(mar=c(1,1,1,1))
```

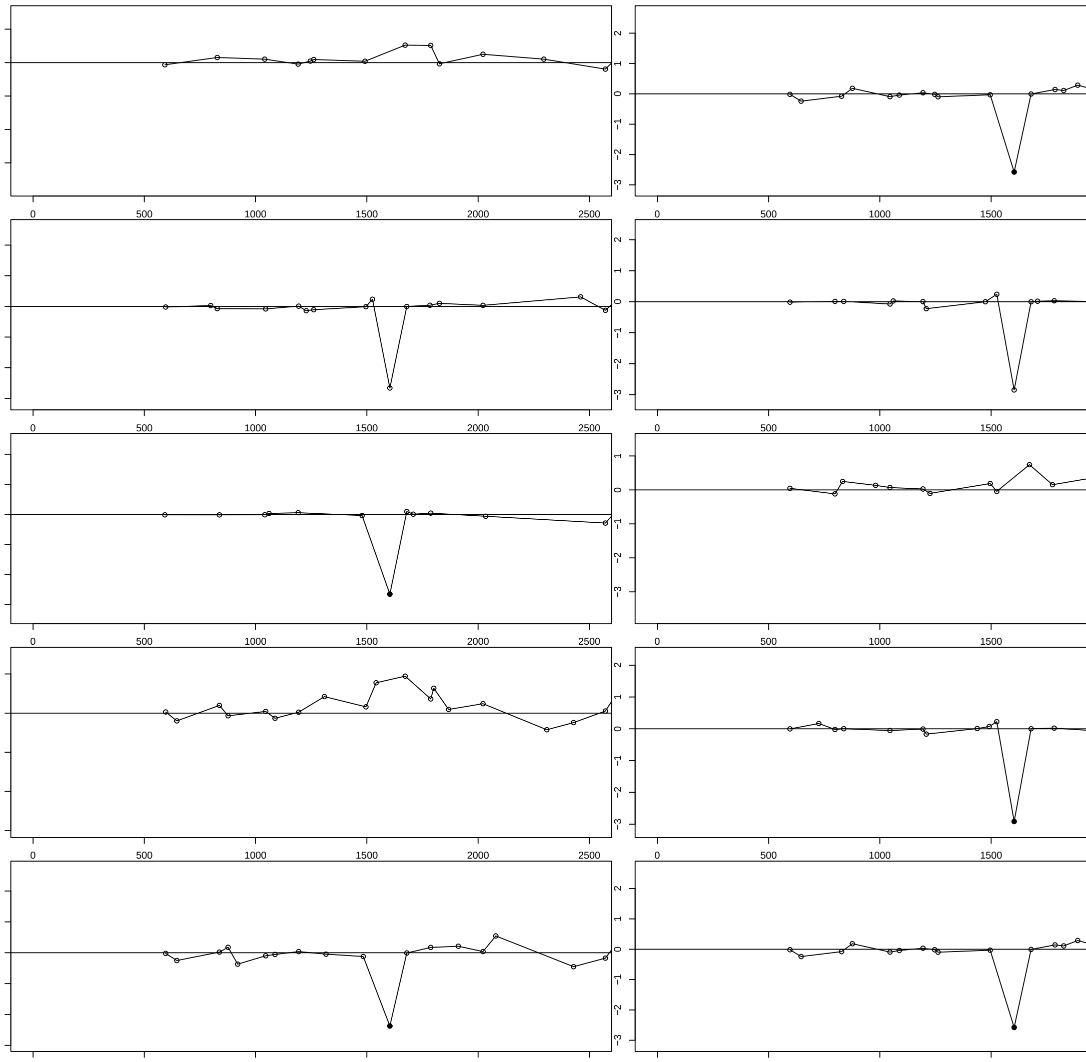
```r
## Check the local moran's results after sub-sampling
for (i in 1:10){
  sub <- carp.r %>% group_by(spatial.code.550) %>% sample_n(size=1)
  reg.sub.annual <- lm(log(sub$AAM)~sub$AnnualTemp)
  test <- correlog(sub$longitude, sub$latitude, reg.sub.annual$residuals,
                   increment=50, resamp=500, latlon=T)
  plot(test, main="", xlim=c(0,2500))
  abline(h=0)
  r2.550.raw[i,1] <- summary(reg.sub.annual)$adj.r.squared #get the r2
}
```



Local moran's I - annual 550

```r
table(carp.r$spatial.code.550)

par(mfrow = c(5,2))
par(mar=c(1,1,1,1))

## Check the local moran's results after sub-sampling
for (i in 1:10){
```

```
  sub <- carp.r %>% group_by(spatial.code.550) %>% sample_n(size=1)
  reg.sub.cold <- lm(log(sub$AAM)~sub$ColdTemp)
  test <- correlog(sub$longitude, sub$latitude, reg.sub.cold$residuals,
                   increment=50, resamp=500, latlon=T)
  plot(test, main="", xlim=c(0,2500))
  abline(h=0)
  r2.550.raw[i,2] <- summary(reg.sub.cold)$adj.r.squared #get the r2
}
```



Local moran's I - cold 550

We have seen that subsampling at 550 km would largely reduce the spatial autocorrelation in our dataset at low distances, even though we are only left with 12 data points.

Compare the r2 values

```
r2 <- data.frame(
  Time = c("Original", "Sub 250","Sub 550"),
  Annual = c(summary(annual.lin)$adj.r.squared,
             mean(unique(r2.250.raw[,1])),
             mean(unique(r2.550.raw[,1]))),
```

```
  Cold = c(summary(cold.lin)$adj.r.squared,
           mean(unique(r2.250.raw[,2])),
           mean(unique(r2.550.raw[,2])))
)
kable(r2)
```

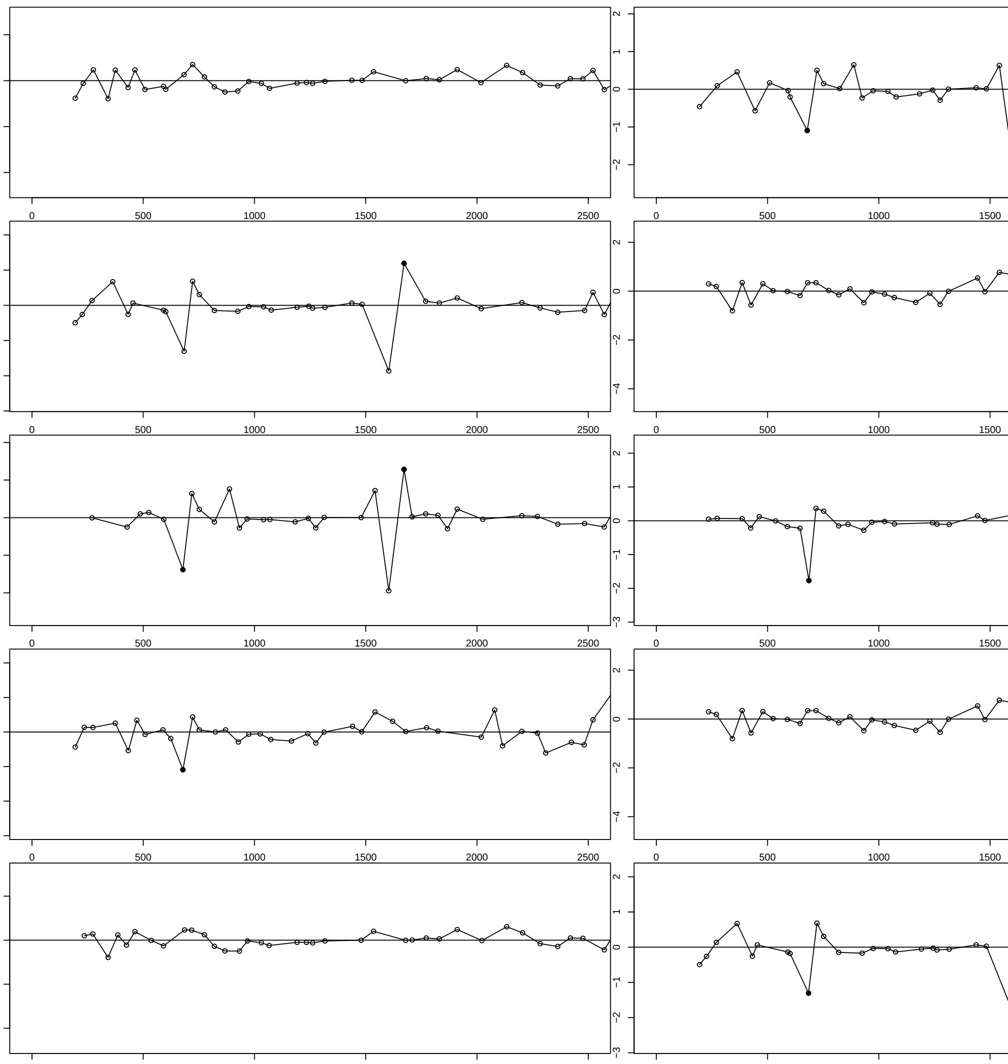
Time	Annual	Cold
Original	0.2533656	0.2751848
Sub 250	0.2430559	0.2691274
Sub 550	0.1036730	0.1703567

However, subsampling at 550 km would reduce our R2 value as well, suggesting that the model would not fit for the sub-dataset.

Latitudinal stratification

Now we would like to explore if stratified sub-sampling using latitude would reduce our spatial autocorrelation at low distances.

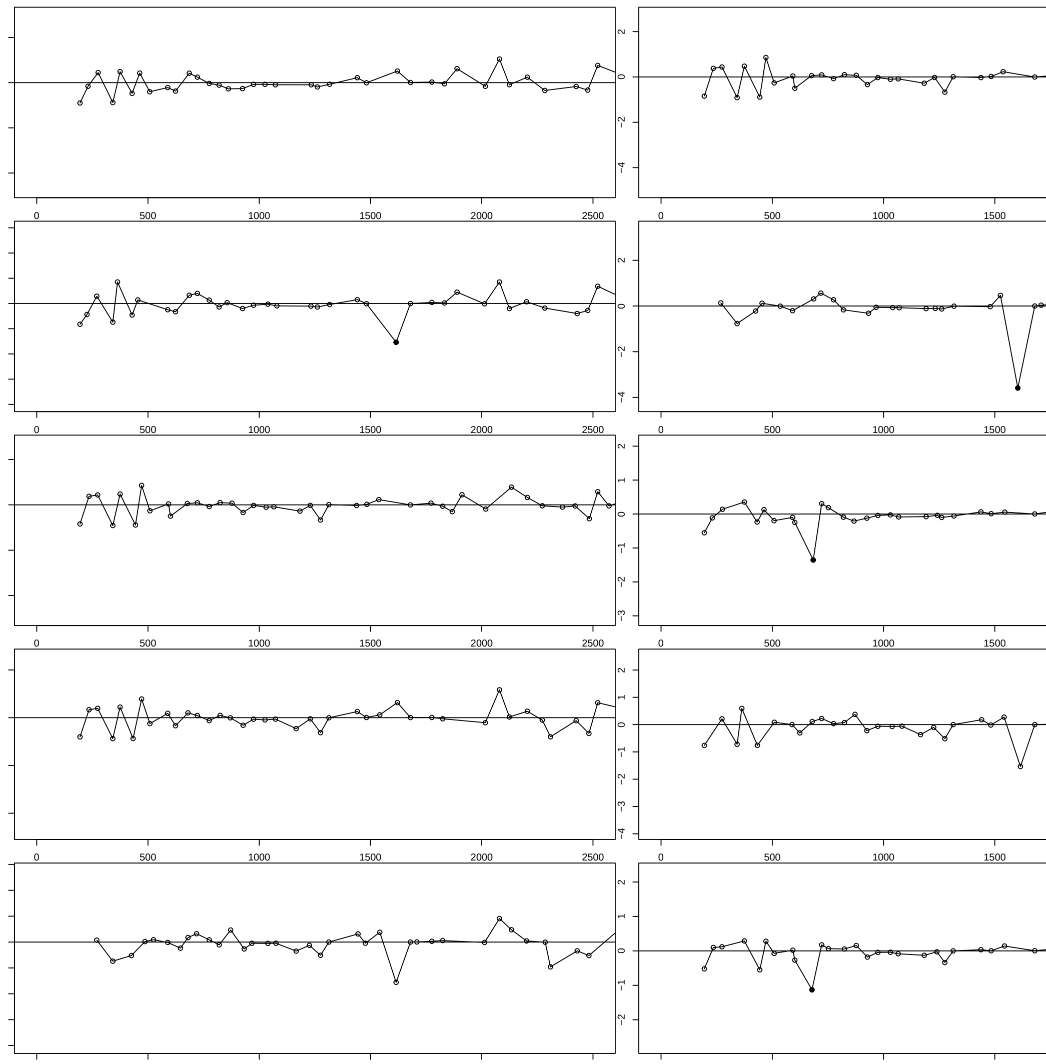
```
table(carp.r$lat.one)

par(mfrow = c(5,2))
par(mar=c(1,1,1,1))

for (i in 1:10){
  sub <- carp.r %>% group_by(lat.one) %>% sample_n(size=1)
  reg.sub.annual <- lm(log(sub$AAM)~sub$AnnualTemp)
  test <- correlog(sub$longitude, sub$latitude, reg.sub.annual$residuals,
                   increment=50, resamp=500, latlon=T)
  plot(test, main="", xlim=c(0,2500))
  abline(h=0)
}
```



one degree latitude bin - annual

```
dev.off()
```

```
table(carp.r$lat.one)

par(mfrow = c(5,2))
par(mar=c(1,1,1,1))

for (i in 1:10){
  sub <- carp.r %>% group_by(lat.one) %>% sample_n(size=1)
  reg.sub.cold <- lm(log(sub$AAM)~sub$ColdTemp)
  test <- correlog(sub$longitude, sub$latitude, reg.sub.cold$residuals,
                   increment=50, resamp=500, latlon=T)
  plot(test, main="", xlim=c(0,2500))
  abline(h=0)
}
```



one degree latitude bin - cold

```
dev.off()
```

- For some subsamples, the negative autocorrelation at 550 km is reduced.
- It seems that the South Ukarine data point is the main cause that we have negative spatial autocorrelation at around 550 km. Therefore, we ought to remove this datapoint from our analyses.

Section 3: Spatial autocorrelation without SU

Now we would like to re-examine the global and local spatial autocorrelation with the South Ukraine data point removed.

Model construction

Again, we first need to define all 18 models (3 models for each temperature). This time we remove the SU data point.

```
# Remove the SU data point
carp.new <- carp.r %>%
  filter(!row_number() == 20) # remove South Ukraine

# Define the models for AnnualTemp
```

```r
annual.lin <- lm(log(AAM)~AnnualTemp, data = carp.new)
annual.add <- lm(log(AAM)~AnnualTemp+condition, data = carp.new)
annual.int <- lm(log(AAM)~AnnualTemp*condition, data = carp.new)

# Define the models for ColdTemp
cold.lin <- lm(log(AAM)~ColdTemp, data = carp.new)
cold.add <- lm(log(AAM)~ColdTemp+condition, data = carp.new)
cold.int <- lm(log(AAM)~ColdTemp*condition, data = carp.new)

# Define the model for WarmTemp
warm.lin <- lm(log(AAM)~WarmTemp, data = carp.new)
warm.add <- lm(log(AAM)~WarmTemp+condition, data = carp.new)
warm.int <- lm(log(AAM)~WarmTemp*condition, data = carp.new)

# Define the models for GDD0
gdd.lin <- lm(log(AAM)~average_gdd_0, data = carp.new)
gdd.add <- lm(log(AAM)~average_gdd_0+condition, data = carp.new)
gdd.int <- lm(log(AAM)~average_gdd_0*condition, data = carp.new)

# Define the models for WaterTemp
water.lin <- lm(log(AAM)~WaterTemp, data = carp.new)
water.add <- lm(log(AAM)~WaterTemp+condition, data = carp.new)
water.int <- lm(log(AAM)~WaterTemp*condition, data = carp.new)

# Define the models for WaterCold
waterC.lin <- lm(log(AAM)~WaterCold, data = carp.new)
waterC.add <- lm(log(AAM)~WaterCold+condition, data = carp.new)
waterC.int <- lm(log(AAM)~WaterCold*condition, data = carp.new)

# Define the models for days_below5
below5.lin <- lm(log(AAM)~below5, data = carp.new)
below5.add <- lm(log(AAM)~below5+condition, data = carp.new)
below5.int <- lm(log(AAM)~below5*condition, data = carp.new)

# Define the models for days_below5_3d
below5.3d.lin <- lm(log(AAM)~below5_3days, data = carp.new)
below5.3d.add <- lm(log(AAM)~below5_3days+condition, data = carp.new)
below5.3d.int <- lm(log(AAM)~below5_3days*condition, data = carp.new)
```

Global Moran.I

```r
## Make spatial dataframe
coords <- data.frame("long"=location.new[,3],"lat"=location.new[,2])
df <- data.frame(a = 1:nrow(location.new[3]))
spatial.data <- SpatialPointsDataFrame(coords,df,proj4string = tmin.1979@crs)

# Get a distance matrix from all points
dists <- spDists(spatial.data, longlat = TRUE)


## Global Moran.I
# Annual Temp
Moran.annual.lin <- Moran.I(annual.lin$residuals, dists)
```

```r
Moran.annual.add <- Moran.I(annual.add$residuals, dists)
Moran.annual.int <- Moran.I(annual.int$residuals, dists)

global.moran.annual <- data.frame(
  Model = c("simple linear", "linear additive", "interaction"),
  Observed = c(Moran.annual.lin$observed, Moran.annual.add$observed,
               Moran.annual.int$observed),
  Expected = c(Moran.annual.lin$expected, Moran.annual.add$expected,
               Moran.annual.int$expected),
  sd = c(Moran.annual.lin$sd, Moran.annual.add$sd, Moran.annual.int$sd),
  p.value = c(Moran.annual.lin$p.value, Moran.annual.add$p.value,
              Moran.annual.int$p.value)
)

kable(global.moran.annual)
```

Model	Observed	Expected	sd	p.value
simple linear	-0.0327532	-0.047619	0.0317899	0.6400513
linear additive	-0.0346096	-0.047619	0.0320210	0.6845372
interaction	-0.0452347	-0.047619	0.0321136	0.9408124

```r
# Cold Temp
Moran.cold.lin <- Moran.I(cold.lin$residuals, dists)
Moran.cold.add <- Moran.I(cold.add$residuals, dists)
Moran.cold.int <- Moran.I(cold.int$residuals, dists)

global.moran.cold <- data.frame(
  Model = c("simple linear", "linear additive", "interaction"),
  Observed = c(Moran.cold.lin$observed, Moran.cold.add$observed,
               Moran.cold.int$observed),
  Expected = c(Moran.cold.lin$expected, Moran.cold.add$expected,
               Moran.cold.int$expected),
  sd = c(Moran.cold.lin$sd, Moran.cold.add$sd, Moran.cold.int$sd),
  p.value = c(Moran.cold.lin$p.value, Moran.cold.add$p.value,
              Moran.cold.int$p.value)
)

kable(global.moran.cold)
```

Model	Observed	Expected	sd	p.value
simple linear	-0.0284140	-0.047619	0.0318113	0.5460304
linear additive	-0.0295230	-0.047619	0.0319752	0.5714354
interaction	-0.0409897	-0.047619	0.0321866	0.8368166

```r
# Warm Temp
Moran.warm.lin <- Moran.I(warm.lin$residuals, dists)
Moran.warm.add <- Moran.I(warm.add$residuals, dists)
Moran.warm.int <- Moran.I(warm.int$residuals, dists)

global.moran.warm <- data.frame(
  Model = c("simple linear", "linear additive", "interaction"),
```

```
  Observed = c(Moran.warm.lin$observed, Moran.warm.add$observed,
               Moran.warm.int$observed),
  Expected = c(Moran.warm.lin$expected, Moran.warm.add$expected,
               Moran.warm.int$expected),
  sd = c(Moran.warm.lin$sd, Moran.warm.add$sd, Moran.warm.int$sd),
  p.value = c(Moran.warm.lin$p.value, Moran.warm.add$p.value,
              Moran.warm.int$p.value)
)

kable(global.moran.warm)
```

Model	Observed	Expected	sd	p.value
simple linear	-0.0707320	-0.047619	0.0314208	0.4619778
linear additive	-0.0729461	-0.047619	0.0316630	0.4237719
interaction	-0.0841580	-0.047619	0.0315116	0.2462355

```
# GDD0
Moran.gdd.lin <- Moran.I(gdd.lin$residuals, dists)
Moran.gdd.add <- Moran.I(gdd.add$residuals, dists)
Moran.gdd.int <- Moran.I(gdd.int$residuals, dists)

global.moran.gdd <- data.frame(
  Model = c("simple linear", "linear additive", "interaction"),
  Observed = c(Moran.gdd.lin$observed, Moran.gdd.add$observed,
               Moran.gdd.int$observed),
  Expected = c(Moran.gdd.lin$expected, Moran.gdd.add$expected,
               Moran.warm.int$expected),
  sd = c(Moran.gdd.lin$sd, Moran.gdd.add$sd, Moran.gdd.int$sd),
  p.value = c(Moran.gdd.lin$p.value, Moran.gdd.add$p.value,
              Moran.gdd.int$p.value)
)

kable(global.moran.gdd)
```

Model	Observed	Expected	sd	p.value
simple linear	-0.0325368	-0.047619	0.0321680	0.6391696
linear additive	-0.0345263	-0.047619	0.0324863	0.6869312
interaction	-0.0448821	-0.047619	0.0323991	0.9326769

```
# Annual water temp
Moran.water.lin <- Moran.I(water.lin$residuals, dists)
Moran.water.add <- Moran.I(water.add$residuals, dists)
Moran.water.int <- Moran.I(water.int$residuals, dists)

global.moran.water <- data.frame(
  Model = c("simple linear", "linear additive", "interaction"),
  Observed = c(Moran.water.lin$observed, Moran.water.add$observed,
               Moran.water.int$observed),
  Expected = c(Moran.water.lin$expected, Moran.water.add$expected,
               Moran.water.int$expected),
  sd = c(Moran.water.lin$sd, Moran.water.add$sd, Moran.water.int$sd),
```

```
  p.value = c(Moran.water.lin$p.value, Moran.water.add$p.value,
              Moran.water.int$p.value)
)

kable(global.moran.water)
```

Model	Observed	Expected	sd	p.value
simple linear	-0.0393254	-0.047619	0.0316407	0.7932289
linear additive	-0.0380113	-0.047619	0.0318923	0.7632192
interaction	-0.0470504	-0.047619	0.0319071	0.9857805

```
# Cold water temp
Moran.waterC.lin <- Moran.I(waterC.lin$residuals, dists)
Moran.waterC.add <- Moran.I(waterC.add$residuals, dists)
Moran.waterC.int <- Moran.I(waterC.int$residuals, dists)

global.moran.waterC <- data.frame(
  Model = c("simple linear", "linear additive", "interaction"),
  Observed = c(Moran.waterC.lin$observed, Moran.waterC.add$observed,
               Moran.waterC.int$observed),
  Expected = c(Moran.waterC.lin$expected, Moran.waterC.add$expected,
               Moran.waterC.int$expected),
  sd = c(Moran.waterC.lin$sd, Moran.waterC.add$sd, Moran.waterC.int$sd),
  p.value = c(Moran.waterC.lin$p.value, Moran.waterC.add$p.value,
              Moran.waterC.int$p.value)
)

kable(global.moran.waterC)
```

Model	Observed	Expected	sd	p.value
simple linear	-0.0265963	-0.047619	0.0325361	0.5181925
linear additive	-0.0273127	-0.047619	0.0328074	0.5359456
interaction	-0.0297237	-0.047619	0.0328997	0.5864851

```
# days below 5 for 5
Moran.below5.lin <- Moran.I(below5.lin$residuals, dists)
Moran.below5.add <- Moran.I(below5.add$residuals, dists)
Moran.below5.int <- Moran.I(below5.int$residuals, dists)

global.moran.below5 <- data.frame(
  Model = c("simple linear", "linear additive", "interaction"),
  Observed = c(Moran.below5.lin$observed, Moran.below5.add$observed,
               Moran.below5.int$observed),
  Expected = c(Moran.below5.lin$expected, Moran.below5.add$expected,
               Moran.below5.int$expected),
  sd = c(Moran.below5.lin$sd, Moran.below5.add$sd, Moran.below5.int$sd),
  p.value = c(Moran.below5.lin$p.value, Moran.below5.add$p.value,
              Moran.below5.int$p.value)
)

kable(global.moran.below5)
```

Model	Observed	Expected	sd	p.value
simple linear	-0.0229027	-0.047619	0.0316404	0.4347061
linear additive	-0.0231094	-0.047619	0.0317310	0.4398668
interaction	-0.0291273	-0.047619	0.0319756	0.5630571

```
# days below 5 for 3
Moran.below53.lin <- Moran.I(below5.3d.lin$residuals, dists)
Moran.below53.add <- Moran.I(below5.3d.add$residuals, dists)
Moran.below53.int <- Moran.I(below5.3d.int$residuals, dists)

global.moran.below53 <- data.frame(
  Model = c("simple linear", "linear additive", "interaction"),
  Observed = c(Moran.below53.lin$observed, Moran.below53.add$observed,
             Moran.below53.int$observed),
  Expected = c(Moran.below53.lin$expected, Moran.below53.add$expected,
             Moran.below53.int$expected),
  sd = c(Moran.below53.lin$sd, Moran.below53.add$sd, Moran.below53.int$sd),
  p.value = c(Moran.below53.lin$p.value, Moran.below53.add$p.value,
            Moran.below53.int$p.value)
)

kable(global.moran.below53)
```

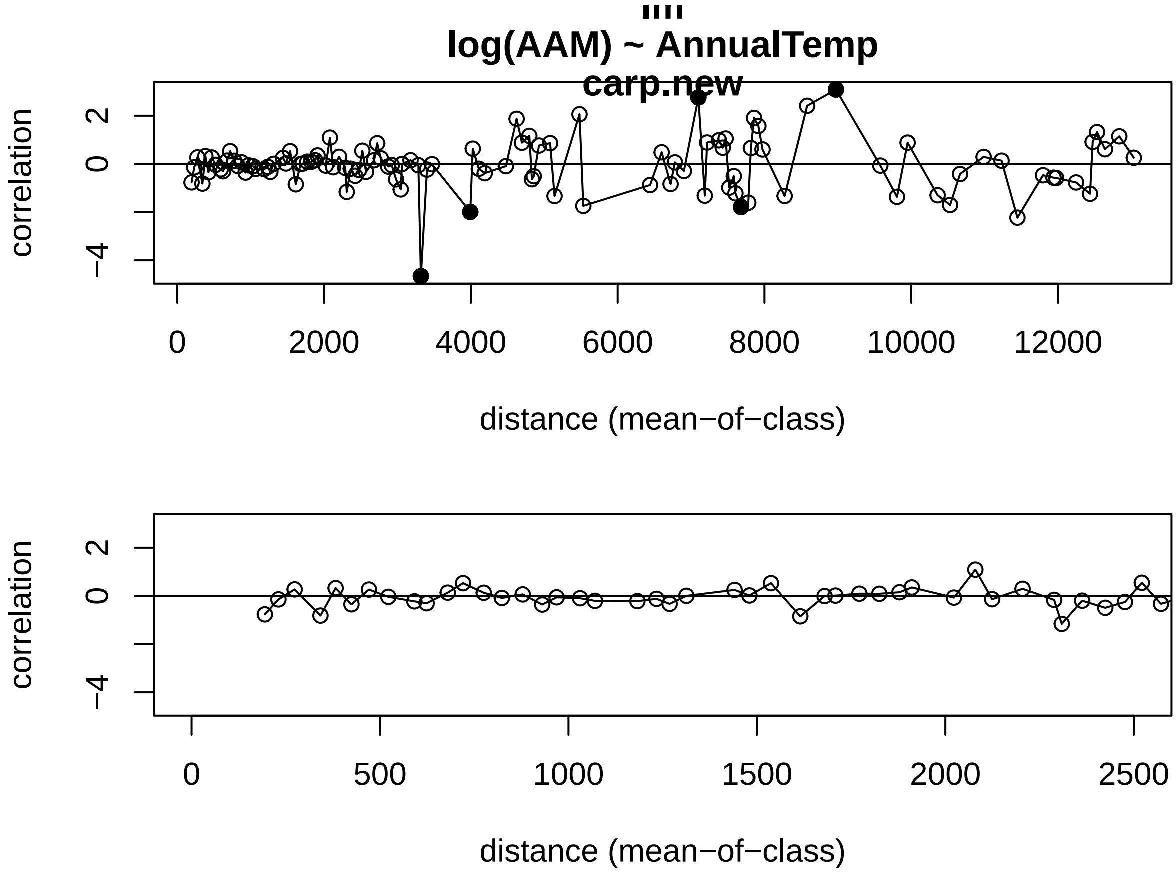
Model	Observed	Expected	sd	p.value
simple linear	-0.0235594	-0.047619	0.0315933	0.4463319
linear additive	-0.0243676	-0.047619	0.0317099	0.4634029
interaction	-0.0302327	-0.047619	0.0318756	0.5854495

There is no global spatial autocorrelation on the entire dataset for all six temperature metrics.

Local spatial autocorrelation

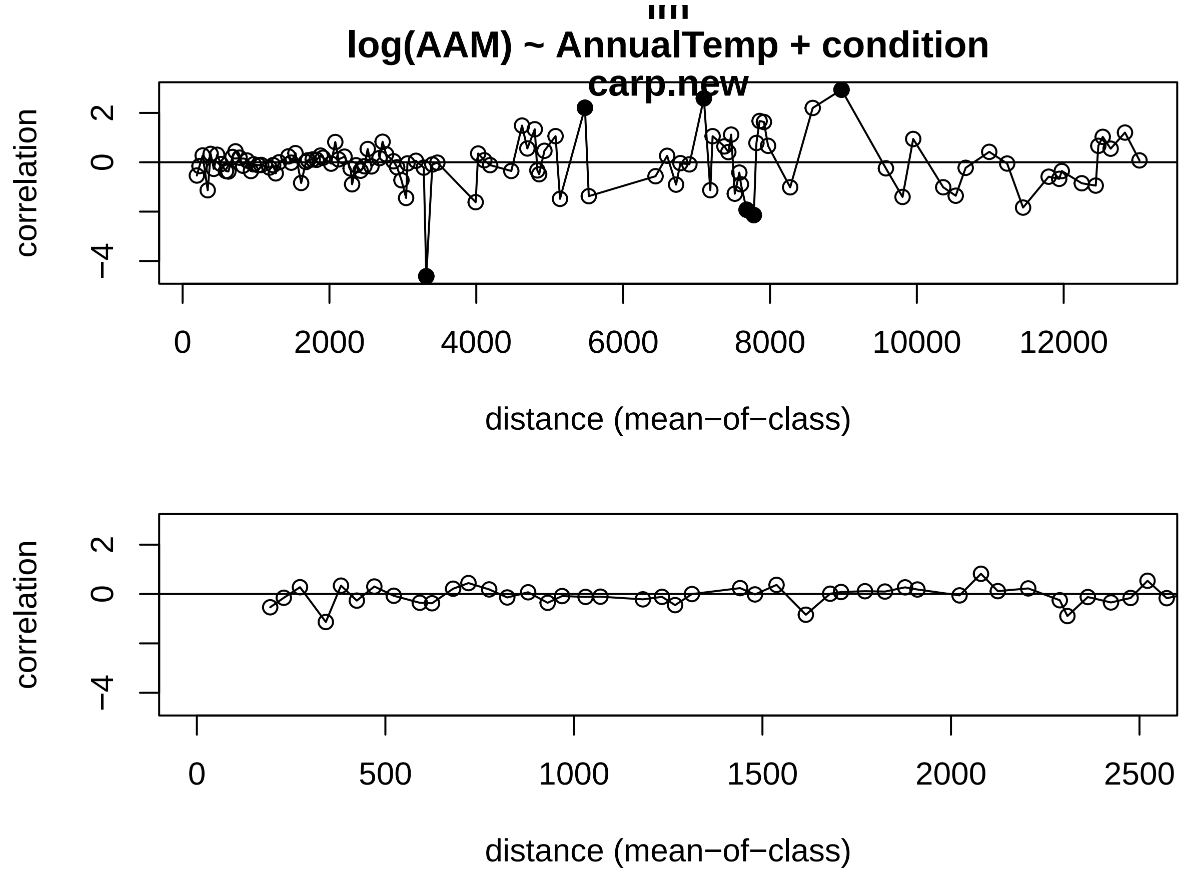
```
## Annual temperature
plot_correlog(annual.lin)
```

```
## 50   of   500 100   of   500 150   of   500 200   of   500 250   of   500 300   of   500 350   of   500 400   of   5(
```



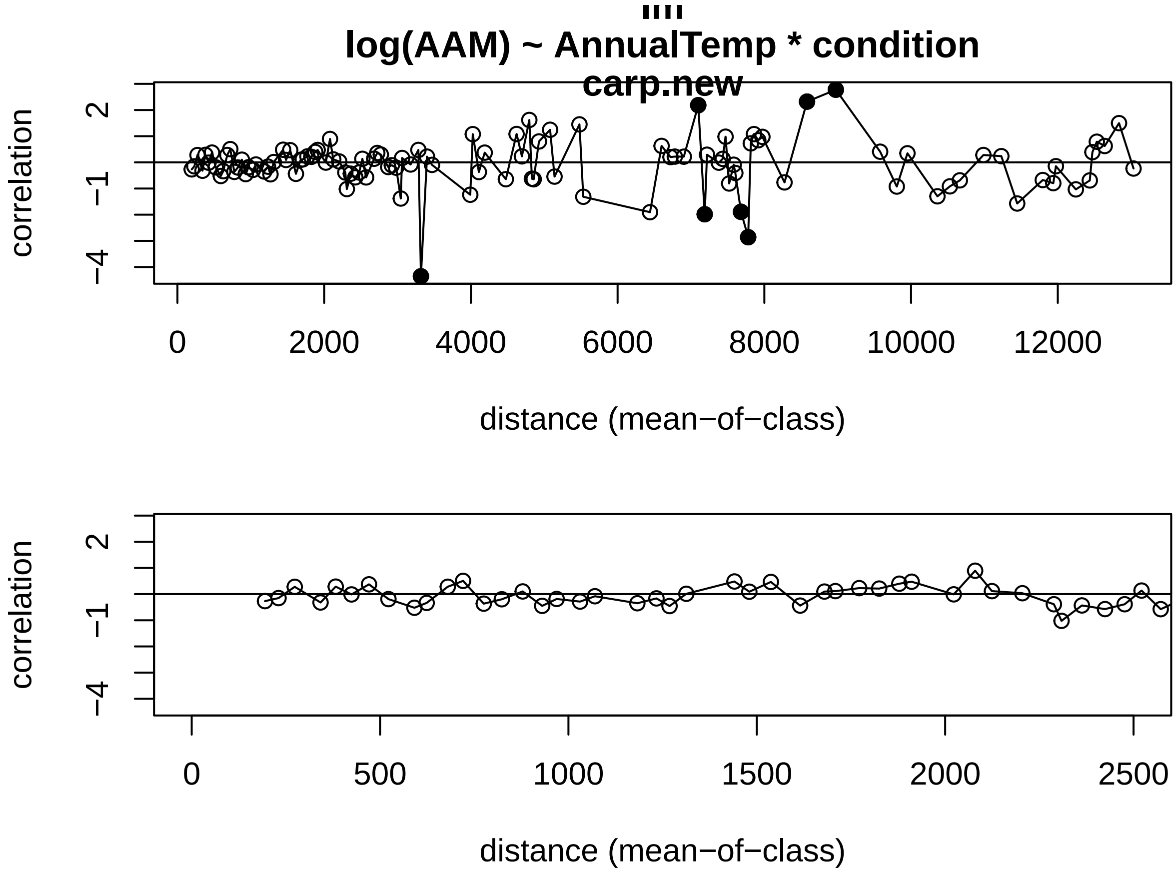

```
plot_correlog(annual.add)
```

```
## 50   of   500 100   of   500 150   of   500 200   of   500 250   of   500 300   of   500 350   of   500 400   of   50
```



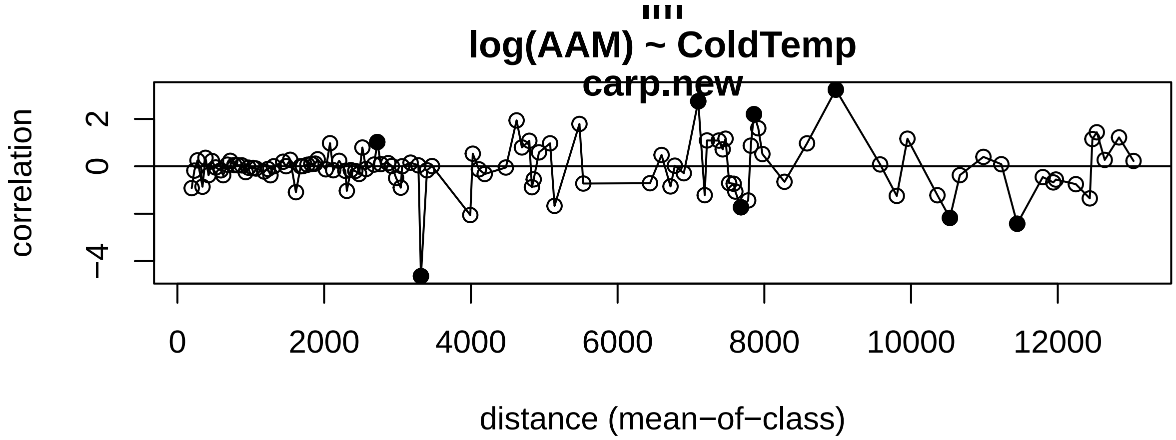

```
plot_correlog(annual.int)
```

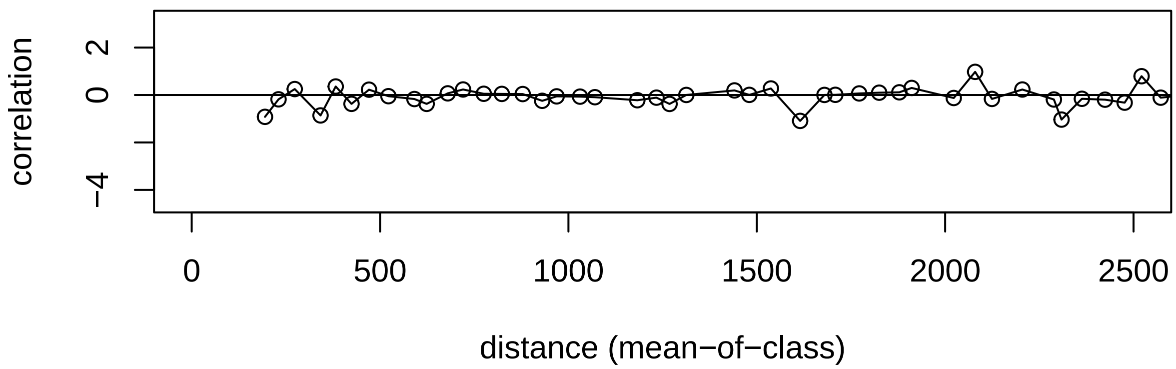
```
## 50  of  500 100  of  500 150  of  500 200  of  500 250  of  500 300  of  500 350  of  500 400  of  5(
```




```
## Cold temperature
plot_correlog(cold.lin)
```

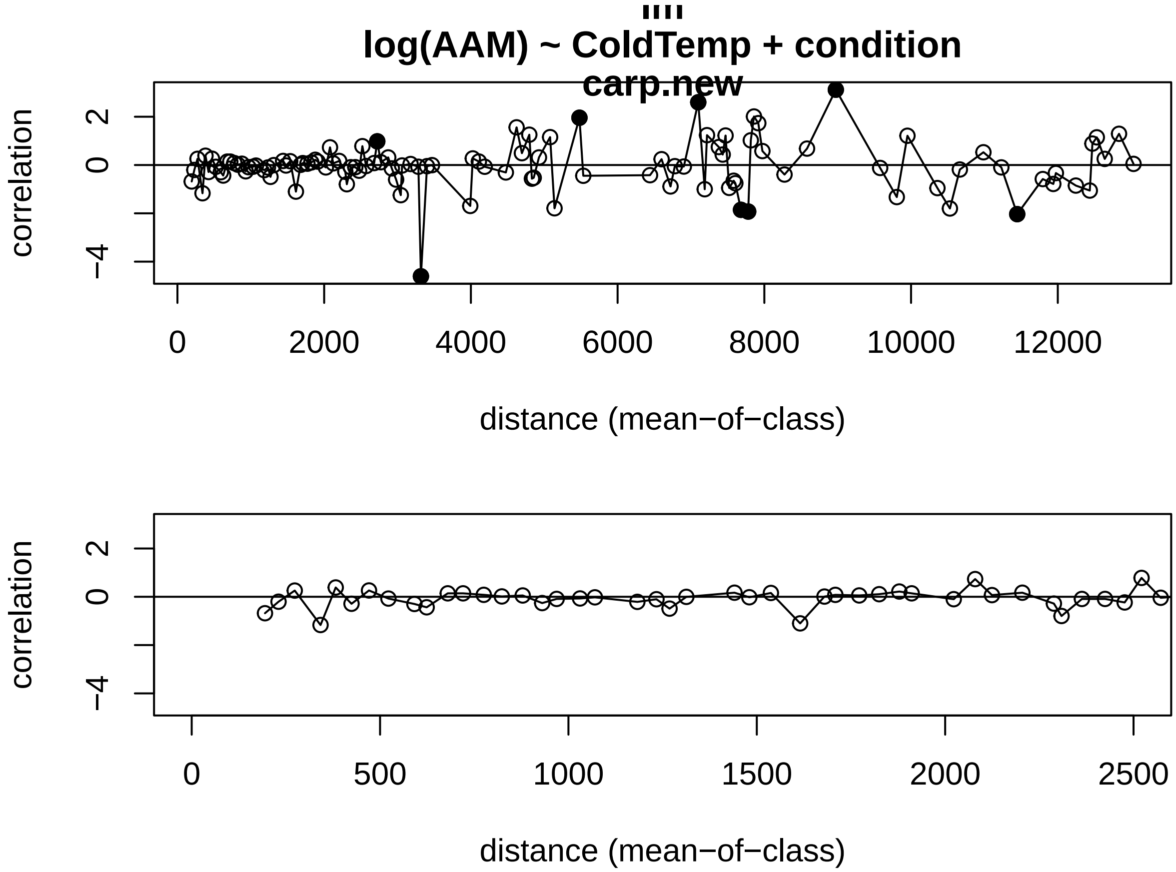
```
## 50   of   500 100  of   500 150  of   500 200  of   500 250  of   500 300  of   500 350  of   500 400  of   5
```





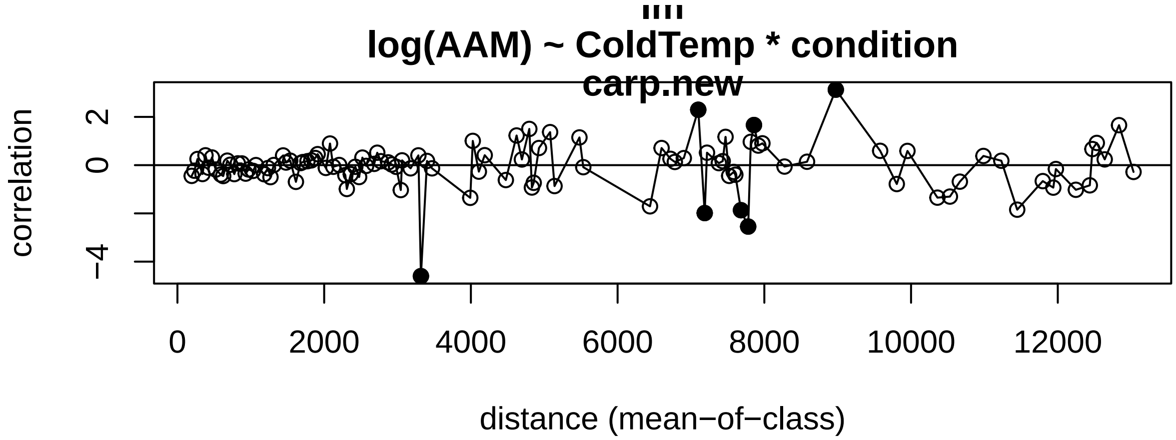

```
plot_correlog(cold.add)
```

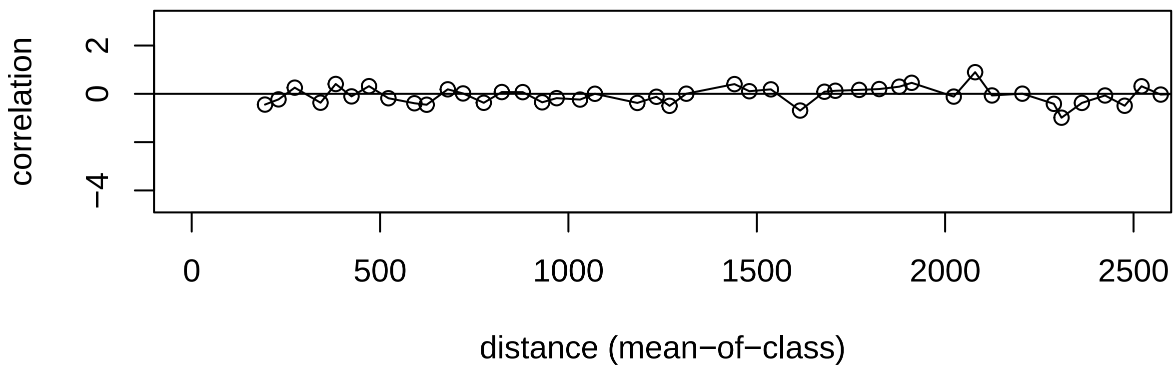
```
## 50   of   500 100   of   500 150   of   500 200   of   500 250   of   500 300   of   500 350   of   500 400   of   5(
```




```r
plot_correlog(cold.int)
```

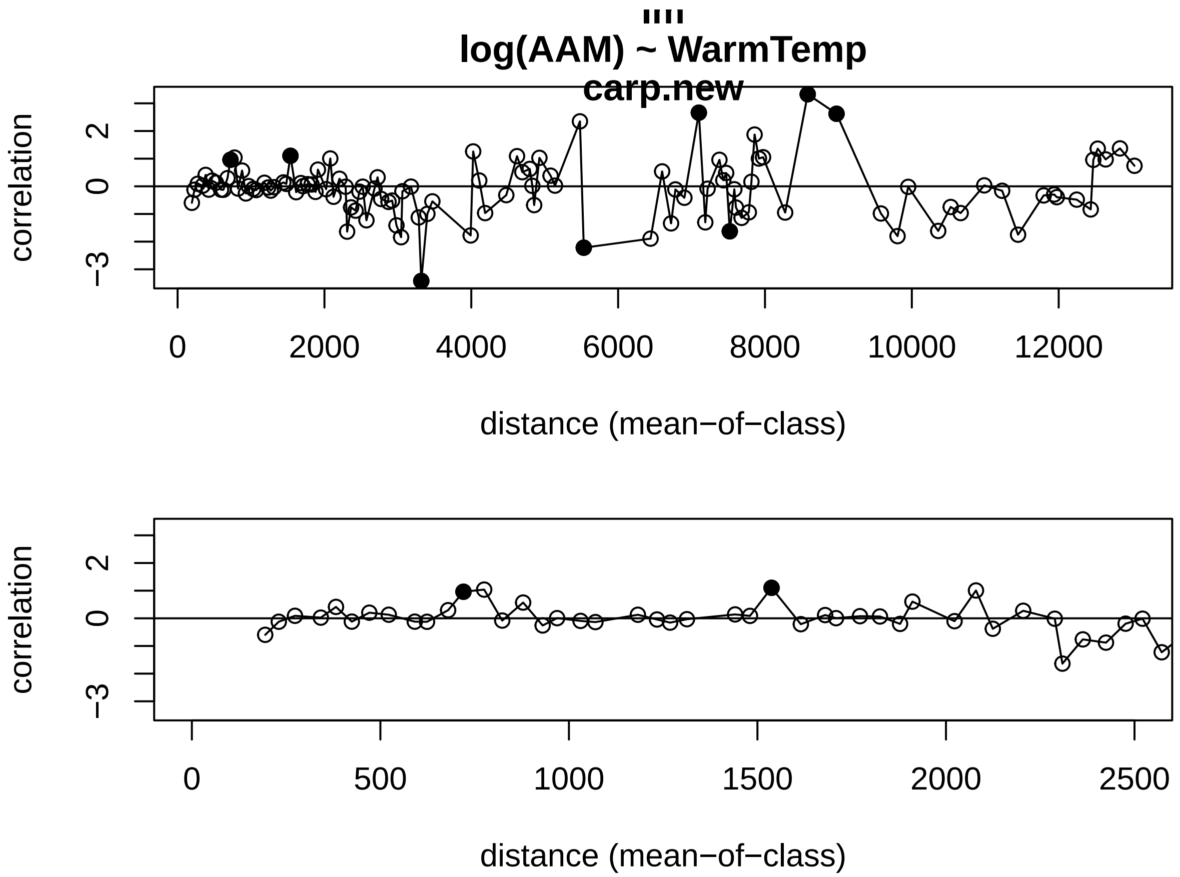
```
## 50   of   500 100   of   500 150   of   500 200   of   500 250   of   500 300   of   500 350   of   500 400   of   5
```





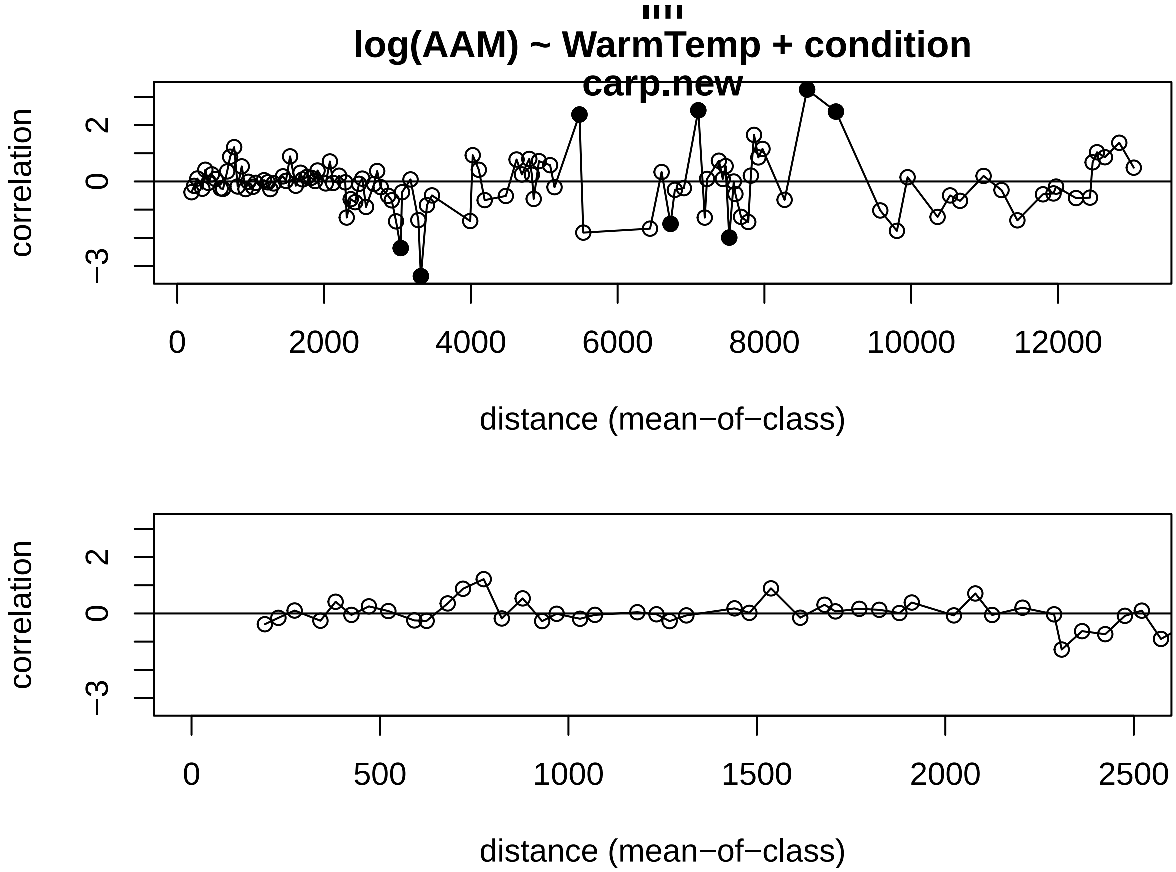

```r
## Warm temperature
plot_correlog(warm.lin)
```

```
## 50   of   500 100   of   500 150   of   500 200   of   500 250   of   500 300   of   500 350   of   500 400   of   50
```



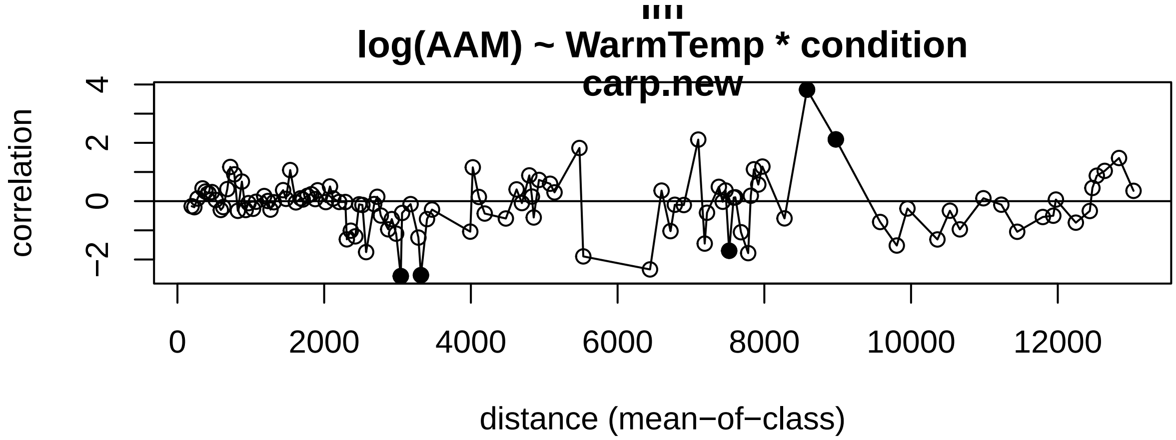

```r
plot_correlog(warm.add)
```

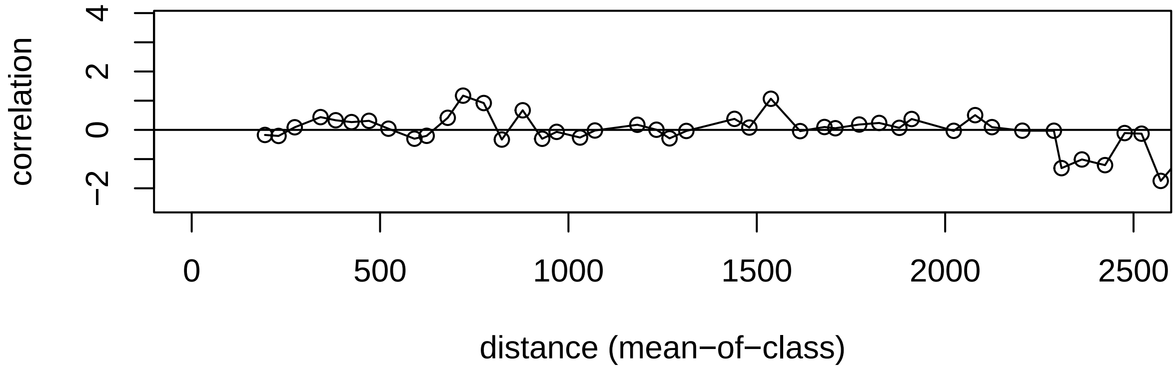
```
## 50  of  500 100  of  500 150  of  500 200  of  500 250  of  500 300  of  500 350  of  500 400  of  50
```




```r
plot_correlog(warm.int)
```

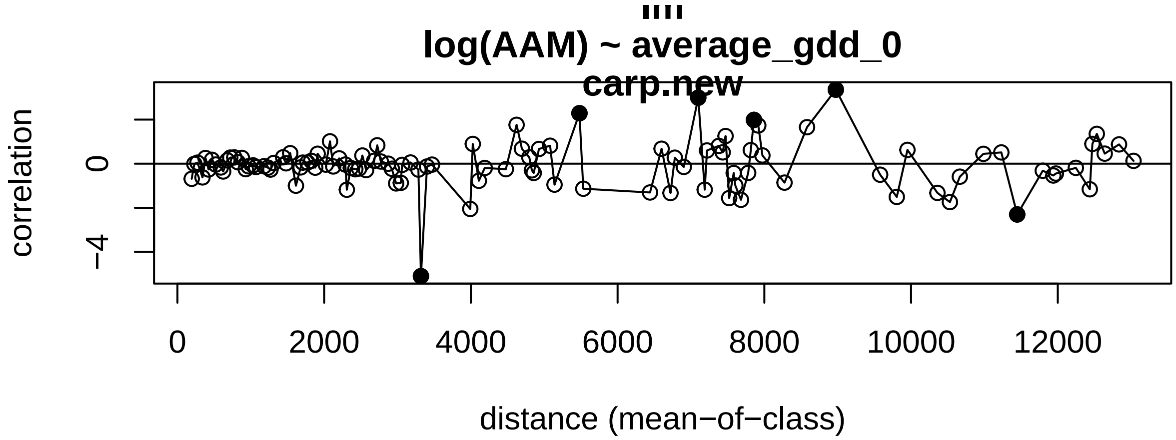
```
## 50   of   500 100   of   500 150   of   500 200   of   500 250   of   500 300   of   500 350   of   500 400   of   5
```

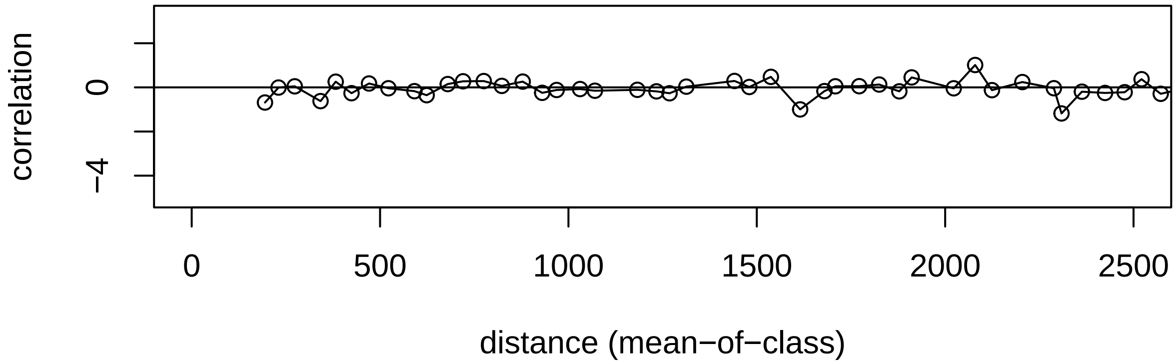





```
## GDD0
plot_correlog(gdd.lin)
```

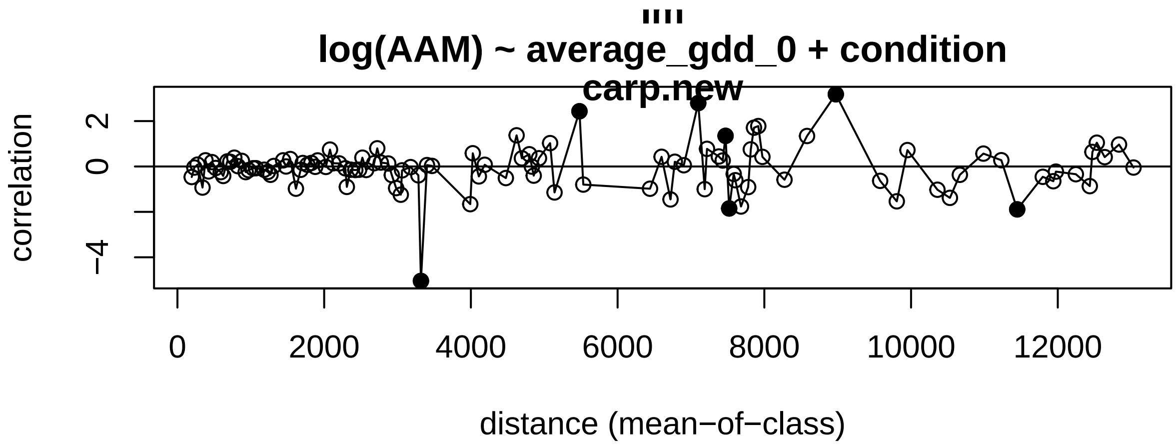
```
## 50  of  500 100  of  500 150  of  500 200  of  500 250  of  500 300  of  500 350  of  500 400  of  50
```

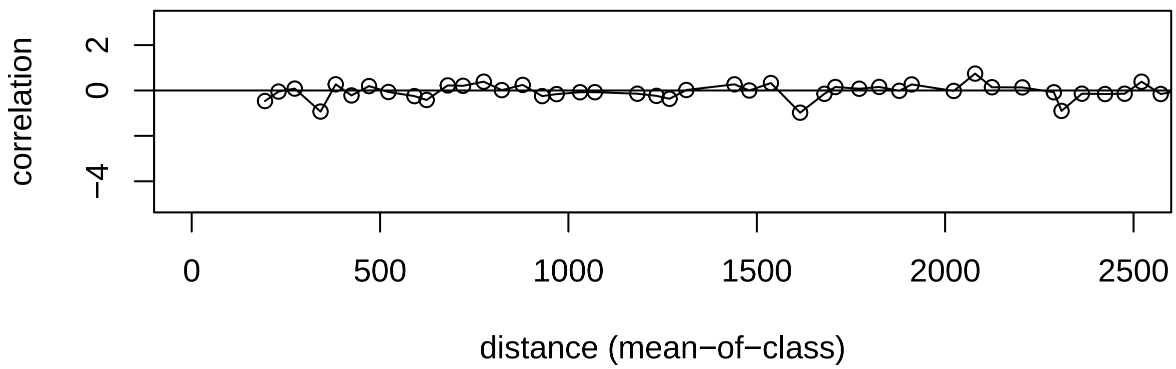





```
plot_correlog(gdd.add)
```

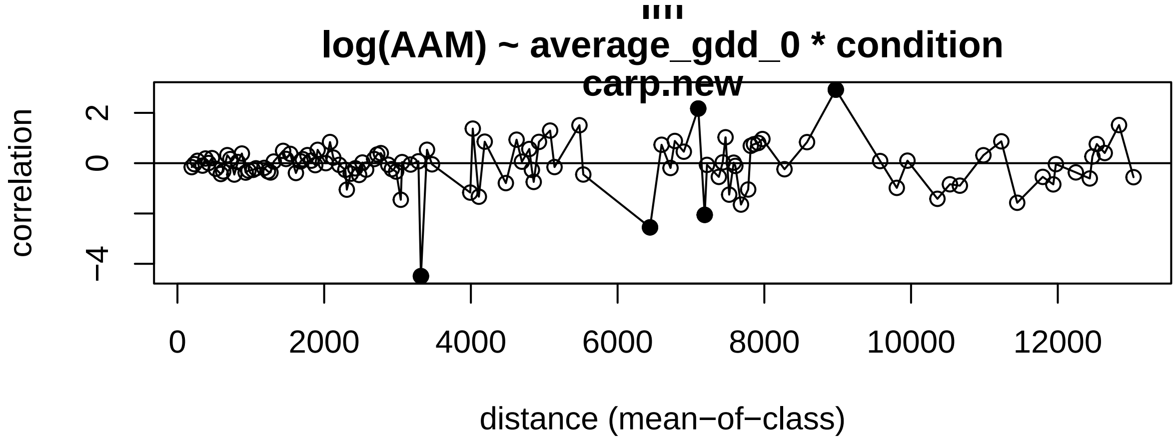
```
## 50  of  500 100  of  500 150  of  500 200  of  500 250  of  500 300  of  500 350  of  500 400  of  50
```

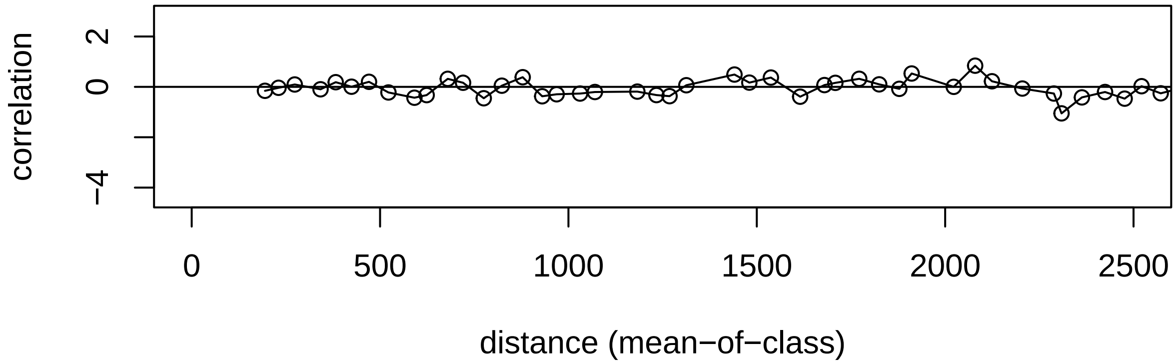





```
plot_correlog(gdd.int)
```

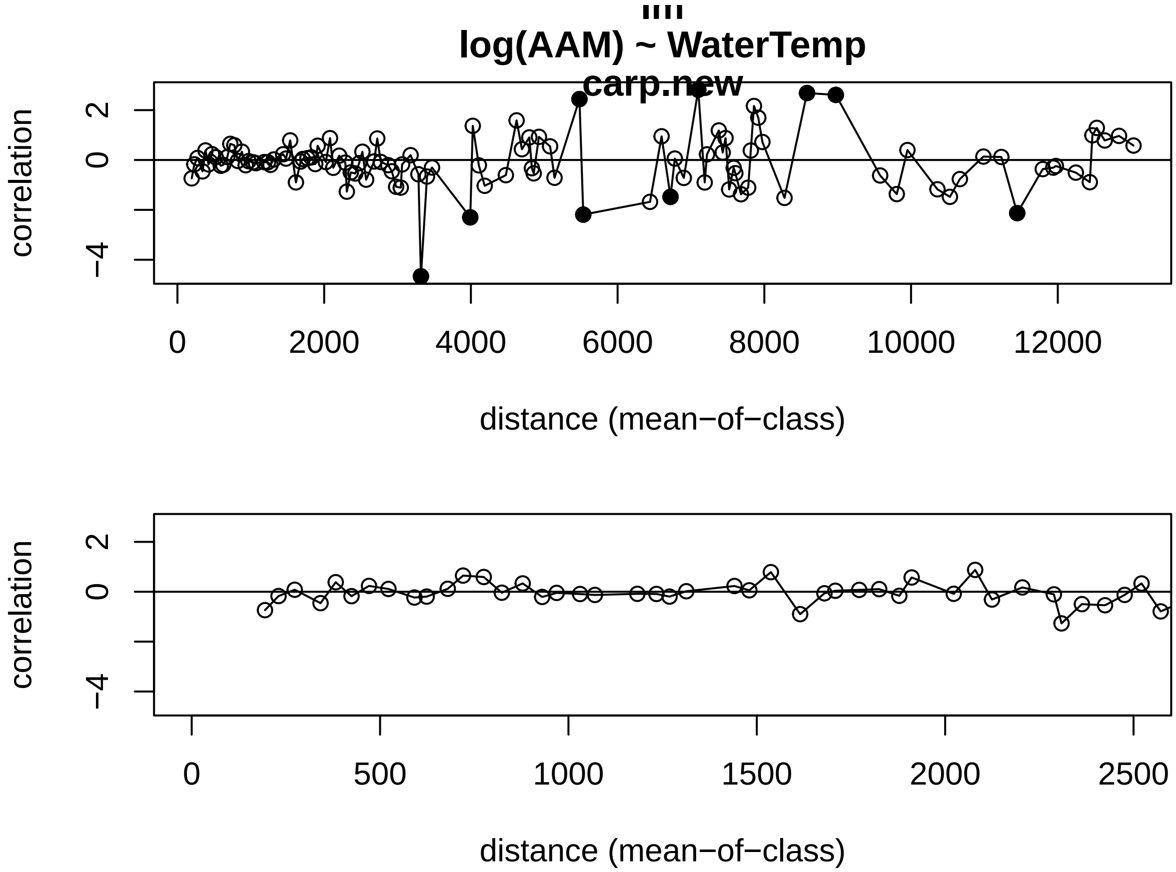
```
## 50   of   500 100   of   500 150   of   500 200   of   500 250   of   500 300   of   500 350   of   500 400   of   50
```





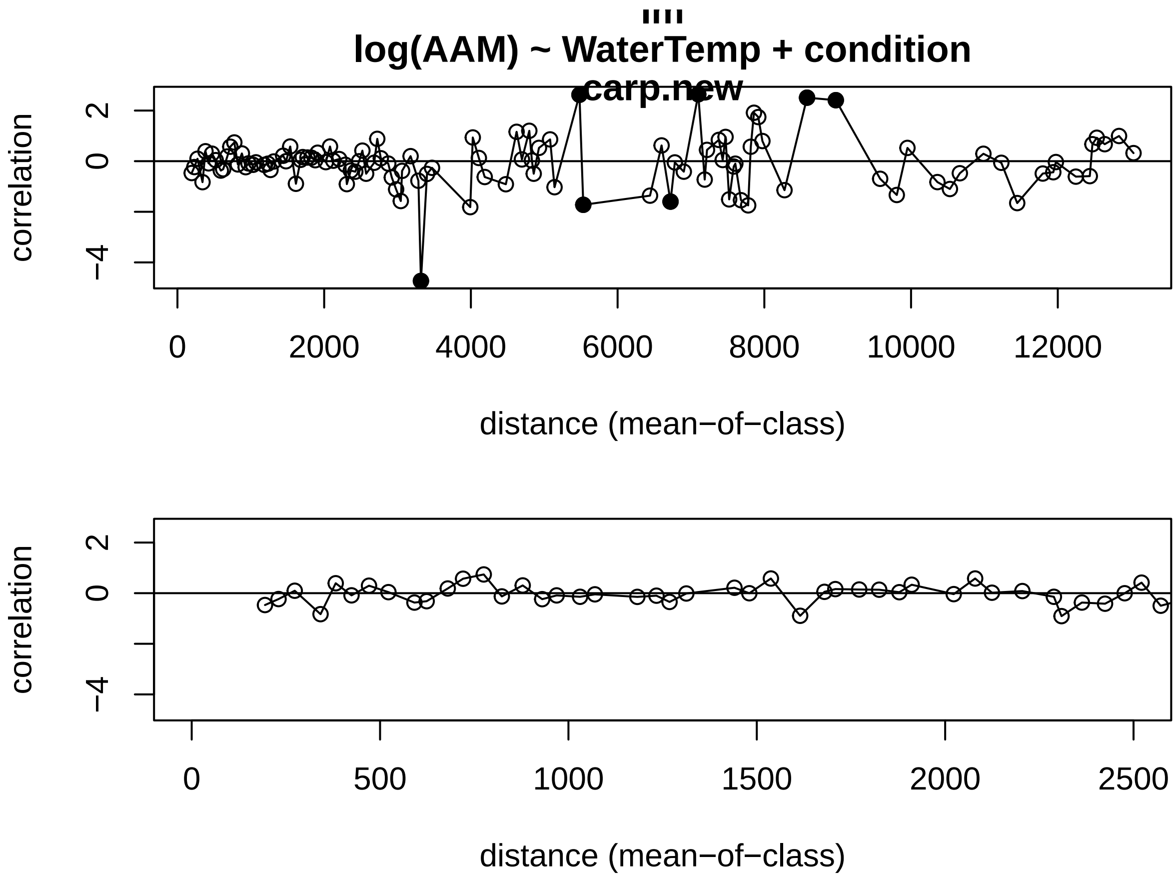

```
## Annual water
plot_correlog(water.lin)
```

```
## 50  of  500 100  of  500 150  of  500 200  of  500 250  of  500 300  of  500 350  of  500 400  of  5(
```



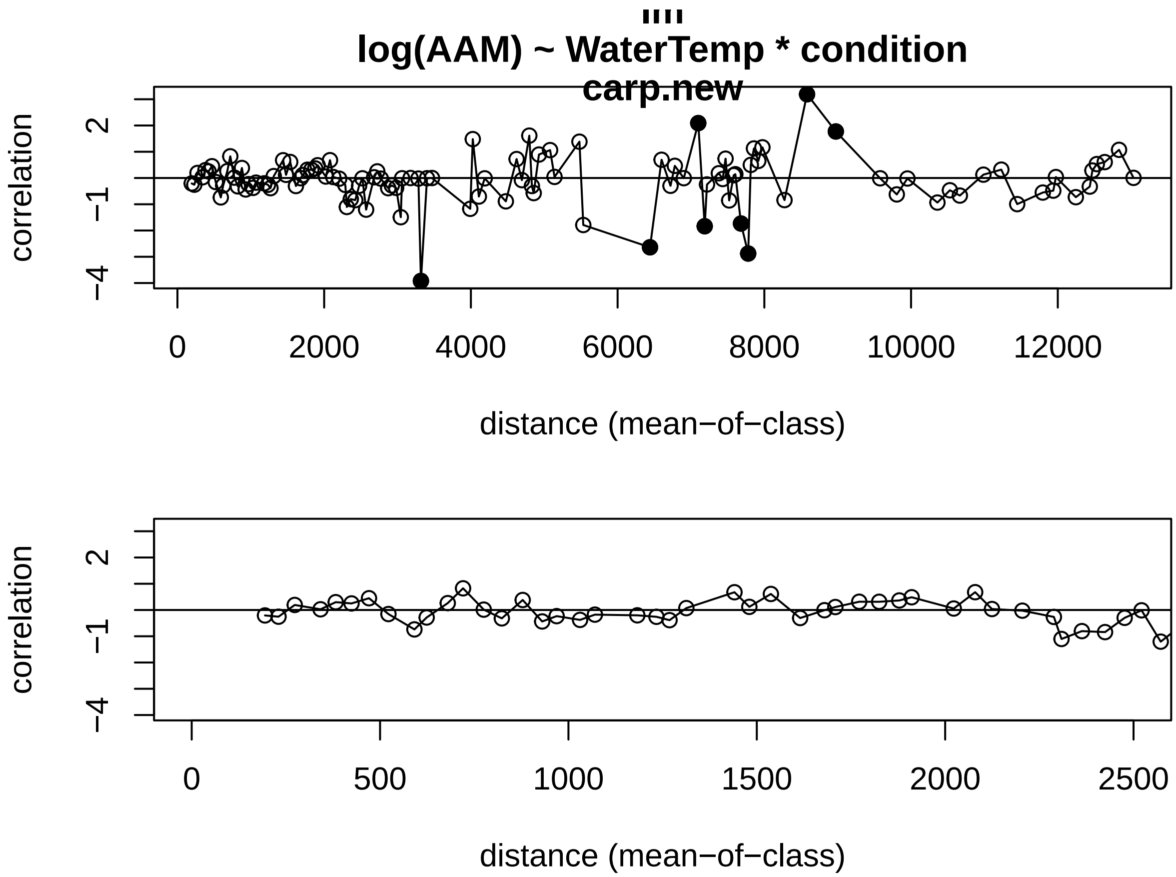

```r
plot_correlog(water.add)
```

```
## 50  of  500 100  of  500 150  of  500 200  of  500 250  of  500 300  of  500 350  of  500 400  of  5(
```



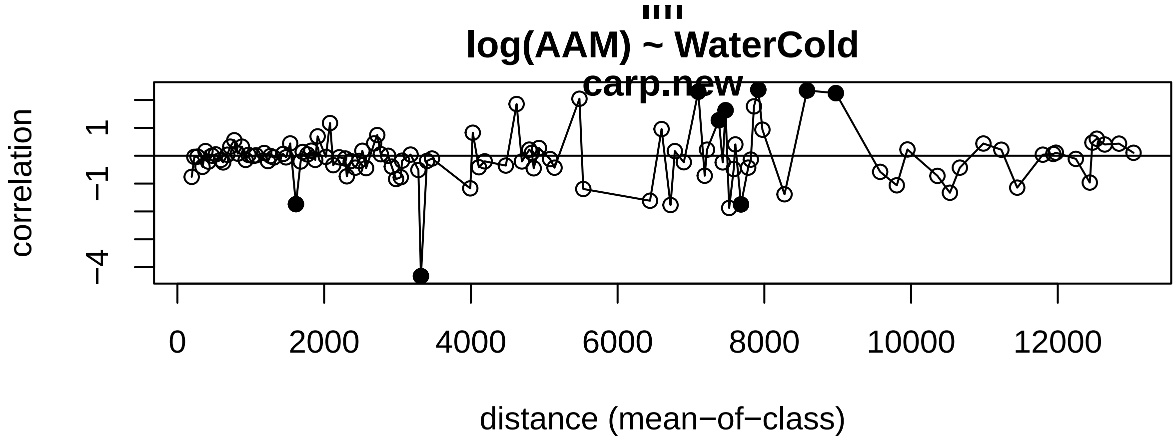

```
plot_correlog(water.int)
```

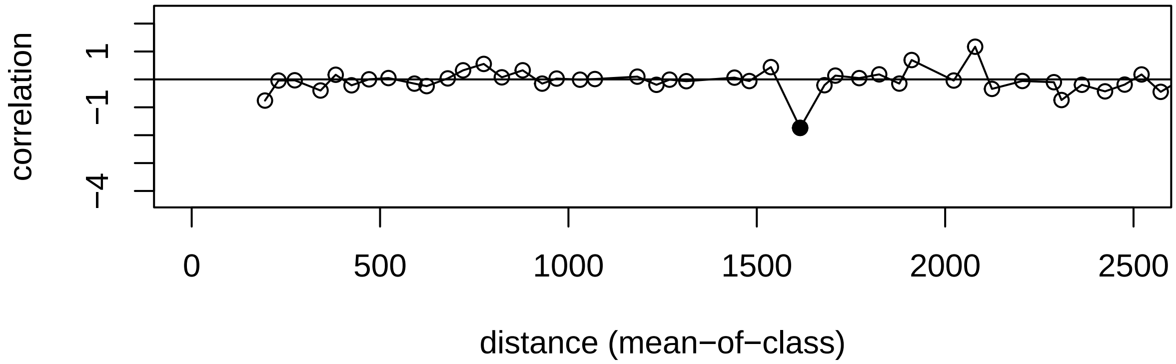
```
## 50  of  500 100  of  500 150  of  500 200  of  500 250  of  500 300  of  500 350  of  500 400  of  50
```




```
## Cold water
plot_correlog(waterC.lin)
```

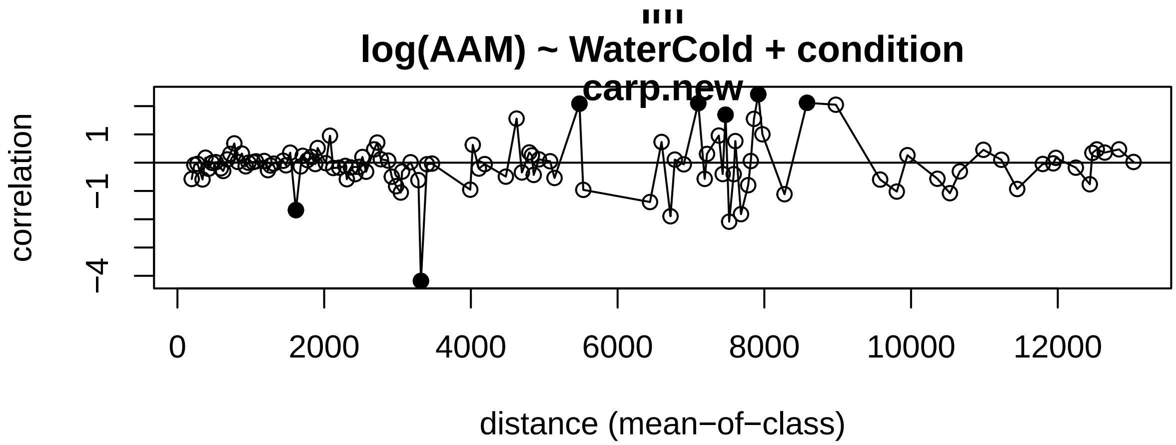
```
## 50  of  500 100  of  500 150  of  500 200  of  500 250  of  500 300  of  500 350  of  500 400  of  5(
```

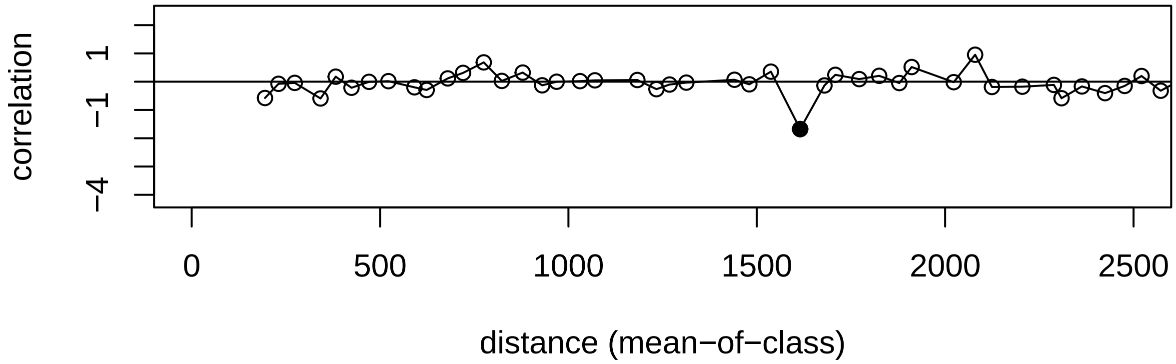





```
plot_correlog(waterC.add)
```

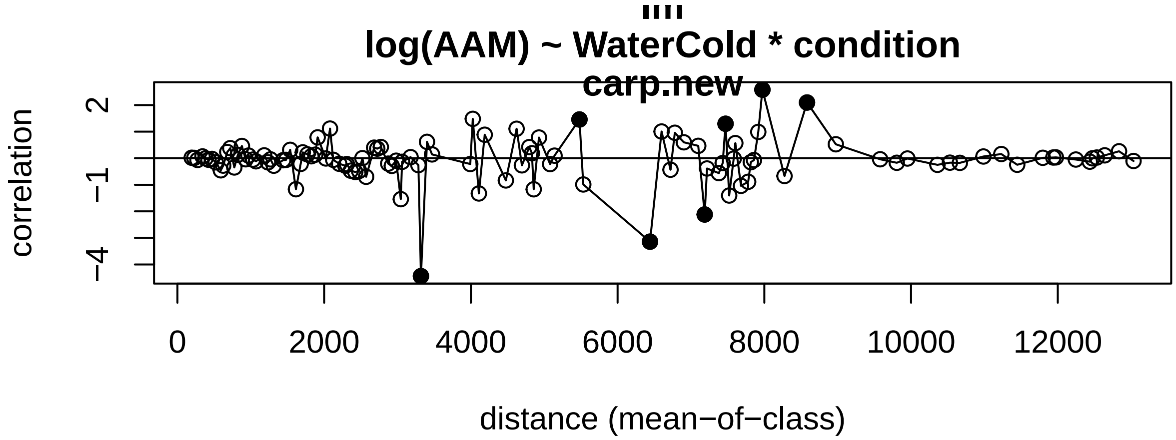
```
## 50  of  500 100  of  500 150  of  500 200  of  500 250  of  500 300  of  500 350  of  500 400  of  50
```

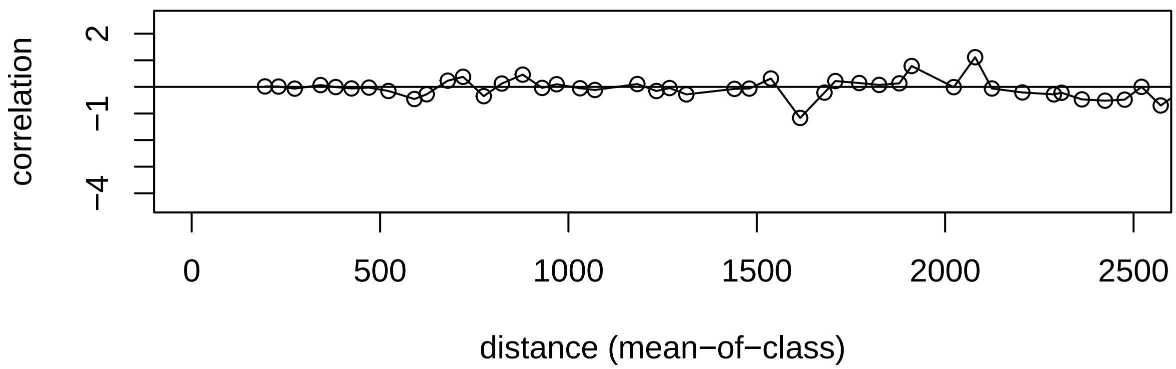





```
plot_correlog(waterC.int)
```

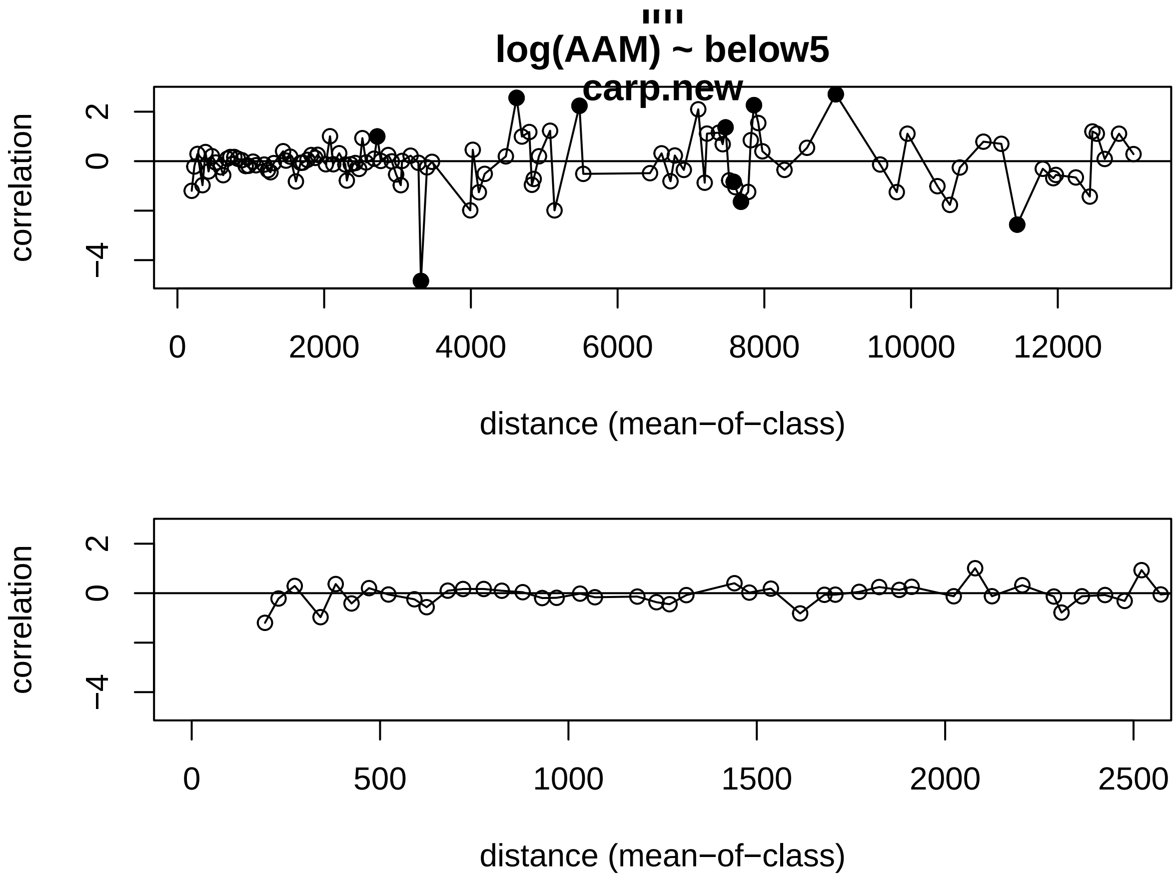
```
## 50  of  500 100  of  500 150  of  500 200  of  500 250  of  500 300  of  500 350  of  500 400  of  50
```





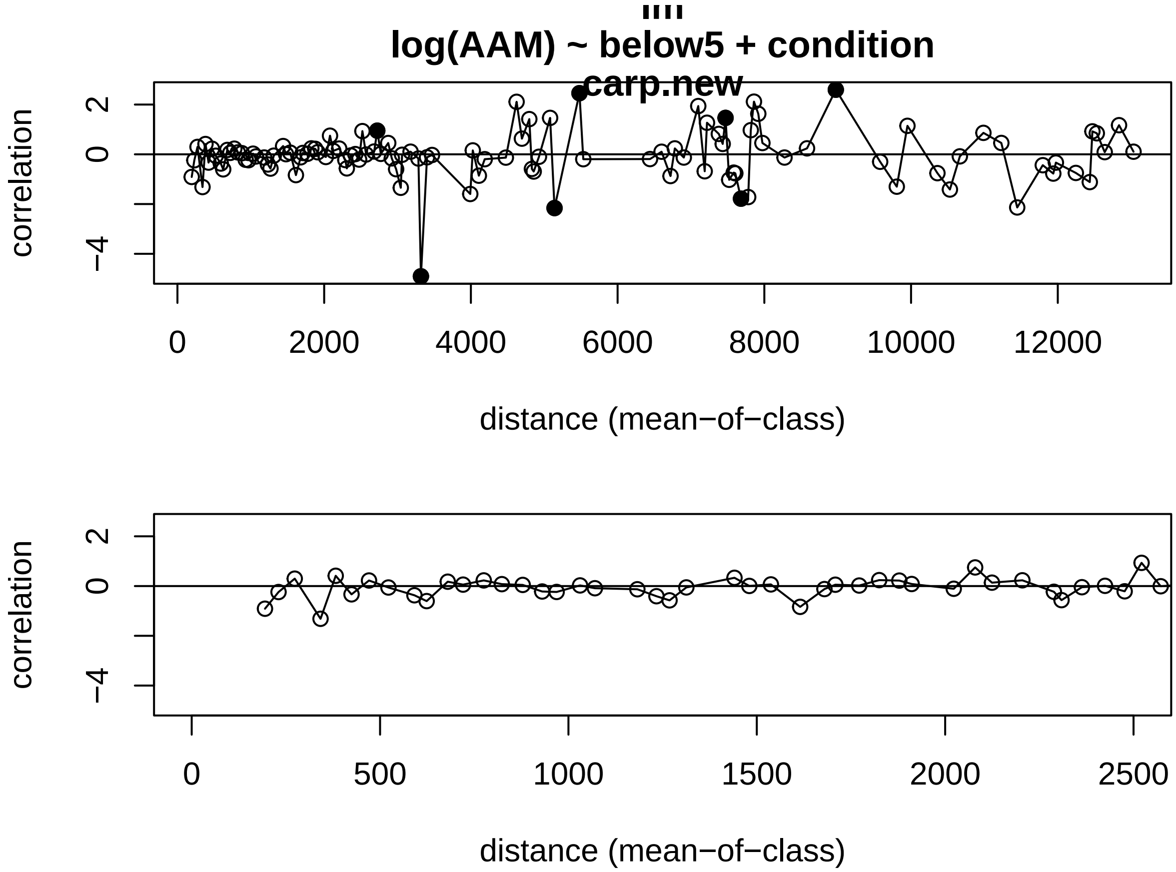

```
## below5 for 5
plot_correlog(below5.lin)
```

```
## 50  of  500 100  of  500 150  of  500 200  of  500 250  of  500 300  of  500 350  of  500 400  of  5
```



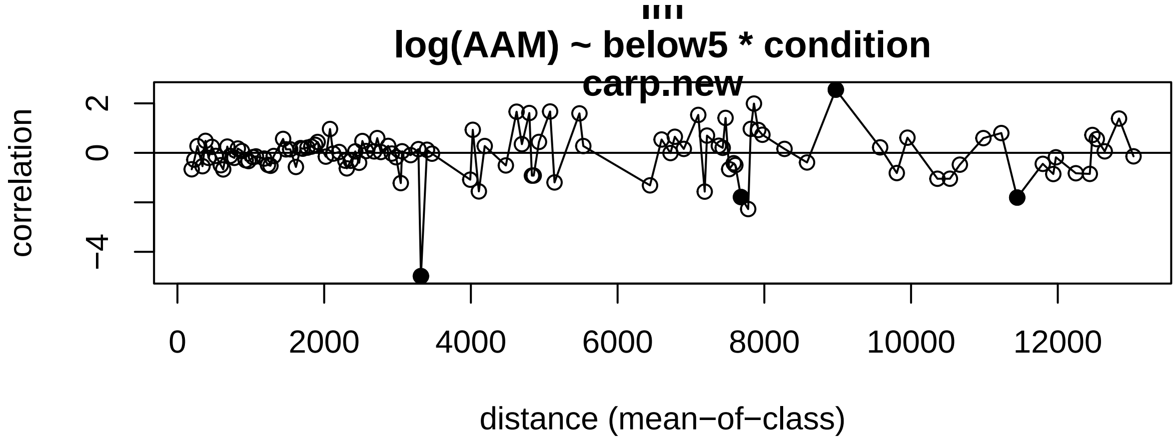

```
plot_correlog(below5.add)
```

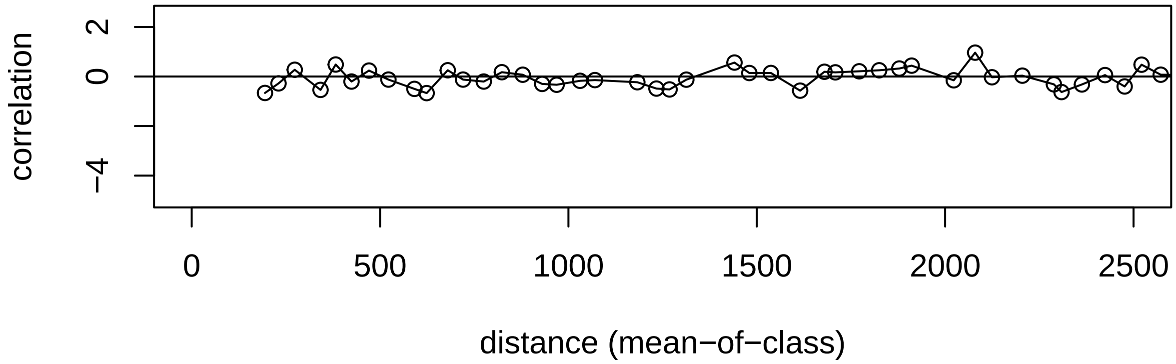
```
## 50  of  500 100  of  500 150  of  500 200  of  500 250  of  500 300  of  500 350  of  500 400  of  50
```




```r
plot_correlog(below5.int)
```

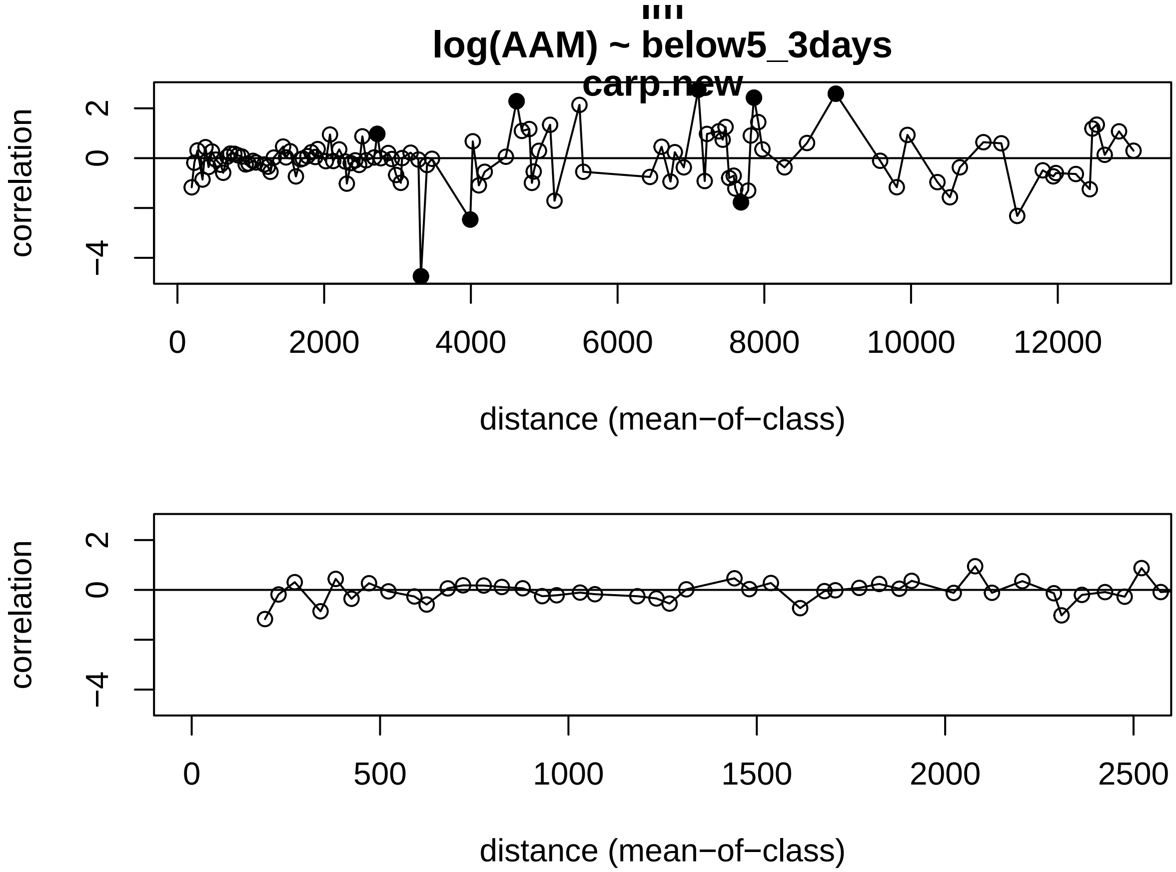
```
## 50   of   500 100   of   500 150   of   500 200   of   500 250   of   500 300   of   500 350   of   500 400   of   5
```





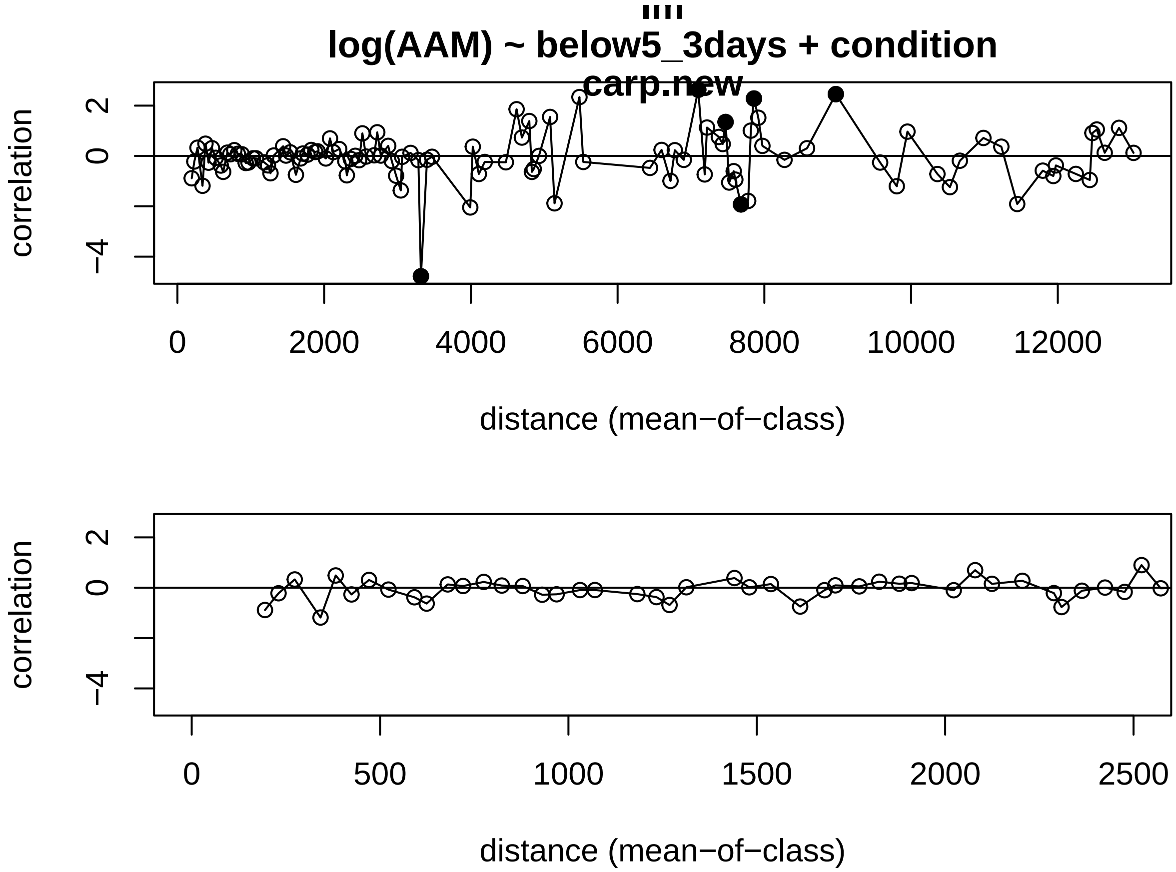

```
## below5 for 3
plot_correlog(below5.3d.lin)
```

```
## 50 of 500 100 of 500 150 of 500 200 of 500 250 of 500 300 of 500 350 of 500 400 of 50
```



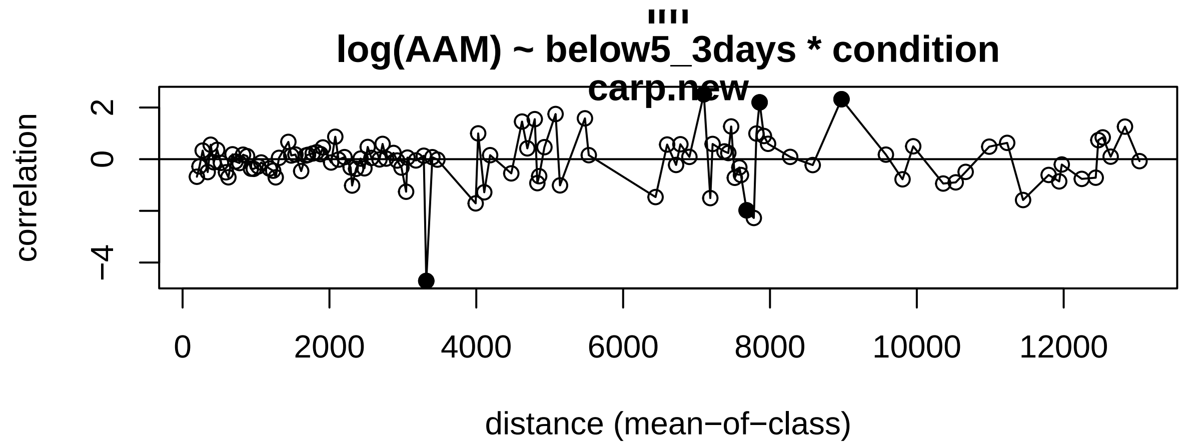

```
plot_correlog(below5.3d.add)
```

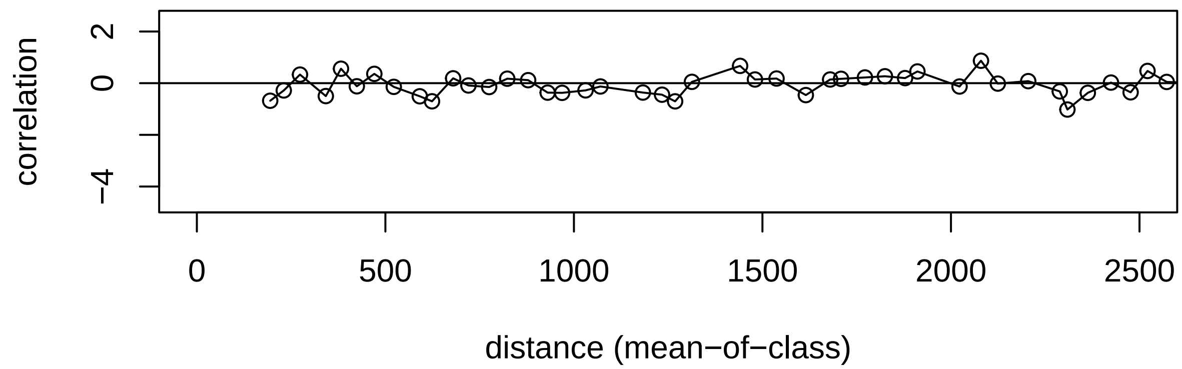
```
## 50   of   500 100   of   500 150   of   500 200   of   500 250   of   500 300   of   500 350   of   500 400   of   50
```




```
plot_correlog(below5.3d.int)
```

```
## 50   of   500 100   of   500 150   of   500 200   of   500 250   of   500 300   of   500 350   of   500 400   of   50
```

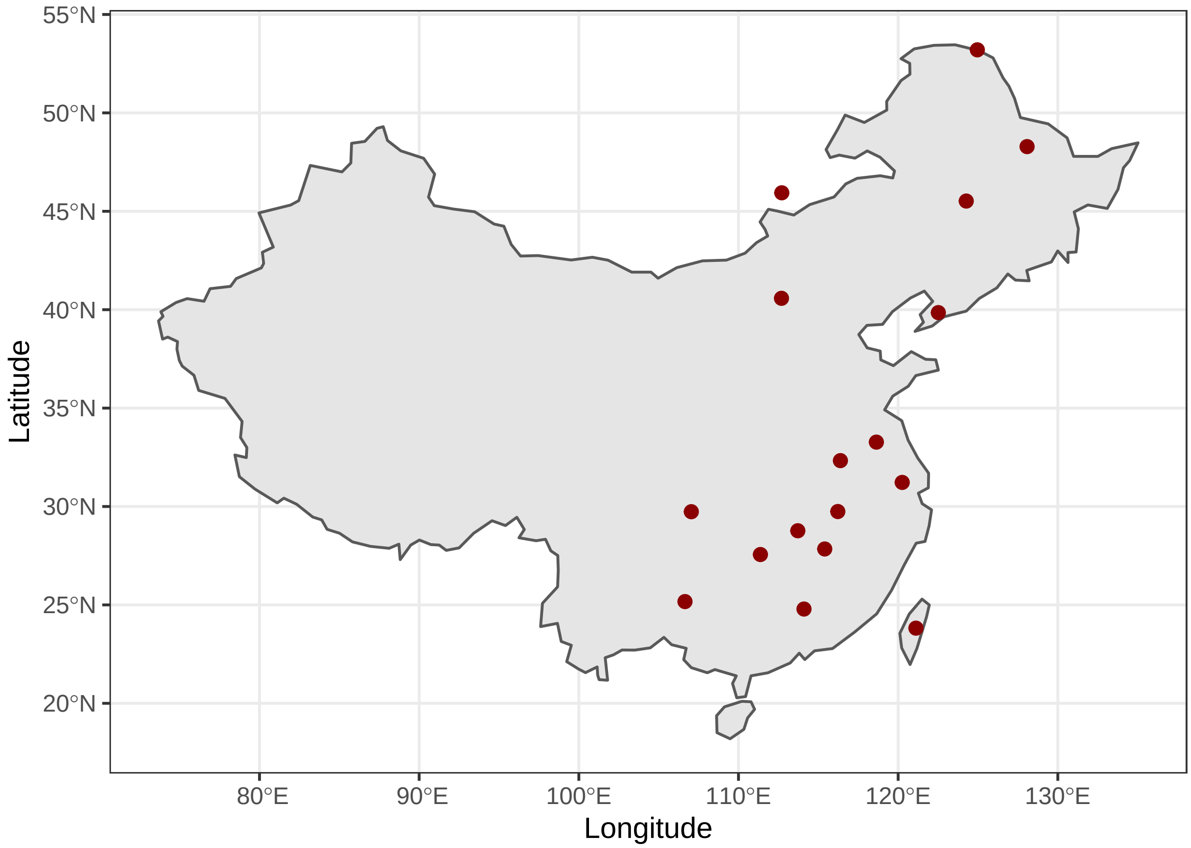




Section 4: China Dataset analyses

Now let's look at the datapoints in China. There are 17 datapoints.

```
## Subset the Chinese dataset
black.china <- carp.r[carp.r$china == "y",]

## Get the world data
world <- ne_countries(scale = "small", returnclass = "sf")

## Plots (china plot)
world %>%
  filter(admin == "China" | admin =="Taiwan") %>%
  # select only China
  ggplot()+
  geom_sf()+
  geom_point(aes(x=longitude, y=latitude), data = black.china,
             size = 2, color = "darkred")+
  theme_bw()+
  xlab("Longitude") + ylab("Latitude")
```



Note that the datapoint outside of the Chinese region is “Northern China”. It is calculated through geographical center.

Global moran’s I

```
# Run the models
lm.annual.china <- lm(log(AAM)~AnnualTemp, data = black.china)
summary(lm.annual.china)
```

```
##
## Call:
## lm(formula = log(AAM) ~ AnnualTemp, data = black.china)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.166813 -0.109074 -0.008175  0.090822  0.184010
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.997652   0.054821  36.440 4.67e-16 ***
## AnnualTemp  -0.018779   0.003967  -4.734 0.000266 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1174 on 15 degrees of freedom
## Multiple R-squared:  0.5991, Adjusted R-squared:  0.5724
```

```
## F-statistic: 22.42 on 1 and 15 DF,  p-value: 0.000266
```

```r
lm.cold.china <- lm(log(AAM)~ColdTemp, data = black.china)
summary(lm.cold.china)
```

```
##
## Call:
## lm(formula = log(AAM) ~ ColdTemp, data = black.china)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.16135 -0.04743 -0.01076  0.07864  0.17702
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.759579   0.026158  67.268  < 2e-16 ***
## ColdTemp    -0.012227   0.002237  -5.467  6.5e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1072 on 15 degrees of freedom
## Multiple R-squared:  0.6658, Adjusted R-squared:  0.6435
## F-statistic: 29.88 on 1 and 15 DF,  p-value: 6.5e-05
```

```r
## Make spatial dataframe
coords <- data.frame("long"=black.china[,14],"lat"=black.china[,13])
df <- data.frame(a = 1:nrow(black.china[14]))
spatial.data <- SpatialPointsDataFrame(coords,df,proj4string = tmin.1979@crs)

# Get a distance matrix from all points
dists <- spDists(spatial.data, longlat = TRUE)

## Run the Moran.I test on the residuals
Moran.annual <- Moran.I(lm.annual.china$residuals, dists)
Moran.cold <- Moran.I(lm.cold.china$residuals, dists)

global.moran <- data.frame(
  Model = c("Moran.annual", "Moran.cold"),
  Observed = c(Moran.annual$observed, Moran.cold$observed),
  Expected = c(Moran.annual$expected, Moran.cold$expected),
  sd = c(Moran.annual$sd, Moran.cold$sd),
  p.value = c(Moran.annual$p.value, Moran.cold$p.value)
)

kable(global.moran)
```

Model	Observed	Expected	sd	p.value
Moran.annual	-0.0594853	-0.0625	0.0497950	0.9517230
Moran.cold	-0.0450195	-0.0625	0.0495955	0.7244924

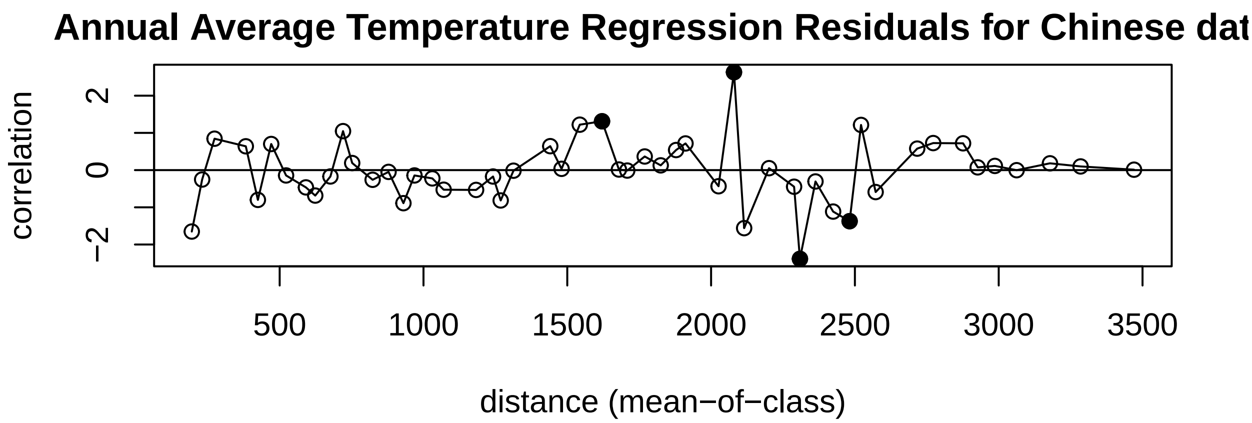
No global spatial autocorrelation in the Chinese dataset.

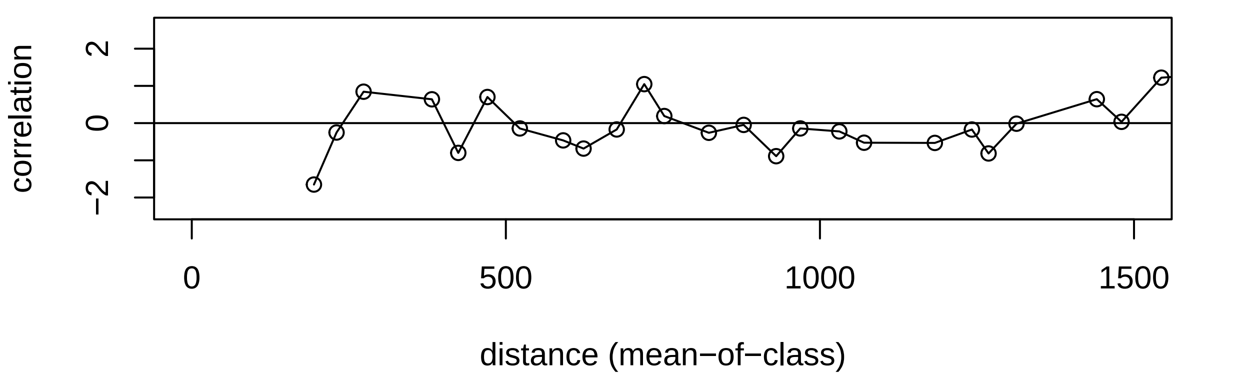
Local moran's I - correlog

```
## Annual temperature
par(mfrow=c(2,1),mar=c(4,4,2,2))

# Use the coorelog function to develop the relationship
test <- correlog(black.china$longitude,
                 black.china$latitude, lm.annual.china$residuals,
                 increment=50, resamp=500, latlon=T)
```

```
## 50  of  500 100  of  500 150  of  500 200  of  500 250  of  500 300  of  500 350  of  500 400  of  50
```

```
# Plot with the entire distance range
plot(test,
     main="Annual Average Temperature Regression Residuals for Chinese data")
abline(h=0)
# Reduce the distance range to 1500 km
plot(test, main="", xlim=c(0,1500))
abline(h=0)
```

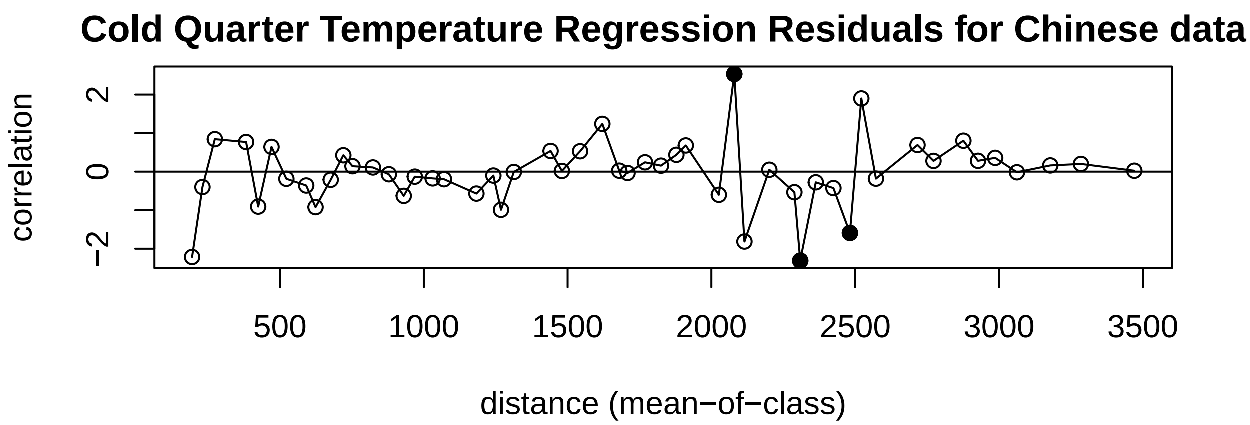


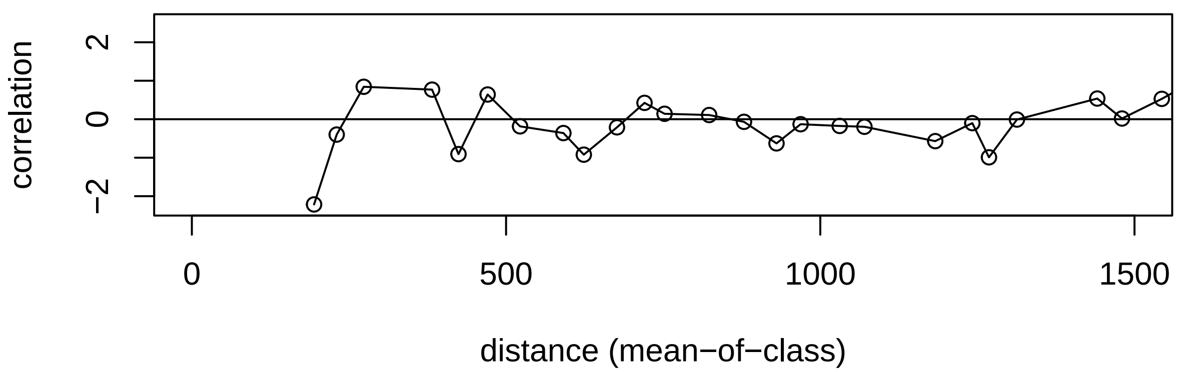



```
## Cold temperature
par(mfrow=c(2,1),mar=c(4,4,2,2))
# Use the coorelog function to develop the relationship
test <- correlog(black.china$longitude,
                 black.china$latitude, lm.cold.china$residuals,
                 increment=50, resamp=500, latlon=T)
```

```
## 50  of  500 100  of  500 150  of  500 200  of  500 250  of  500 300  of  500 350  of  500 400  of  50
```

```r
# Plot with the entire distance range
plot(test,
     main="Cold Quarter Temperature Regression Residuals for Chinese data")
abline(h=0)

# Reduce the distance range to 1500 km
plot(test, main="", xlim=c(0,1500))
abline(h=0)
```

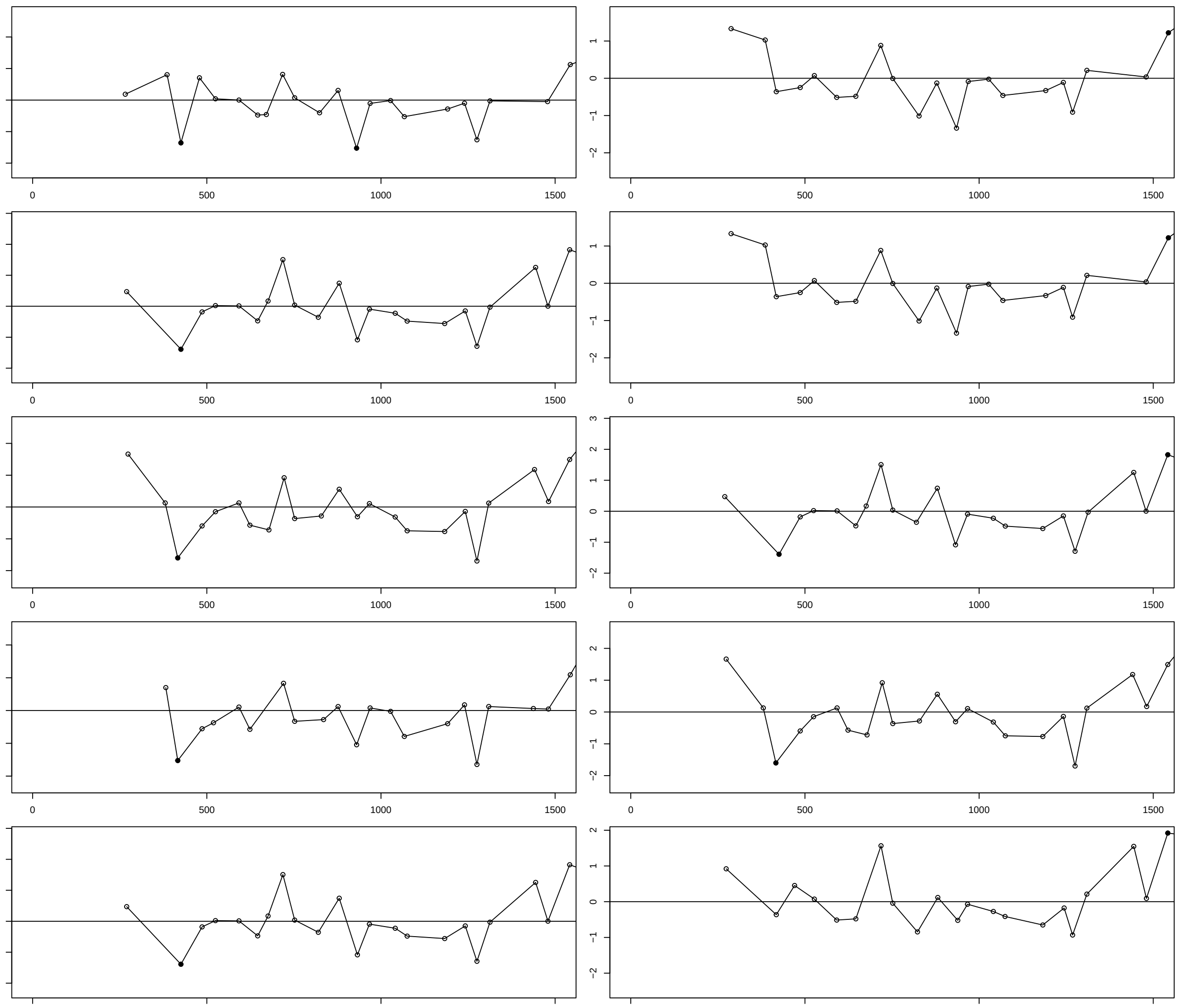




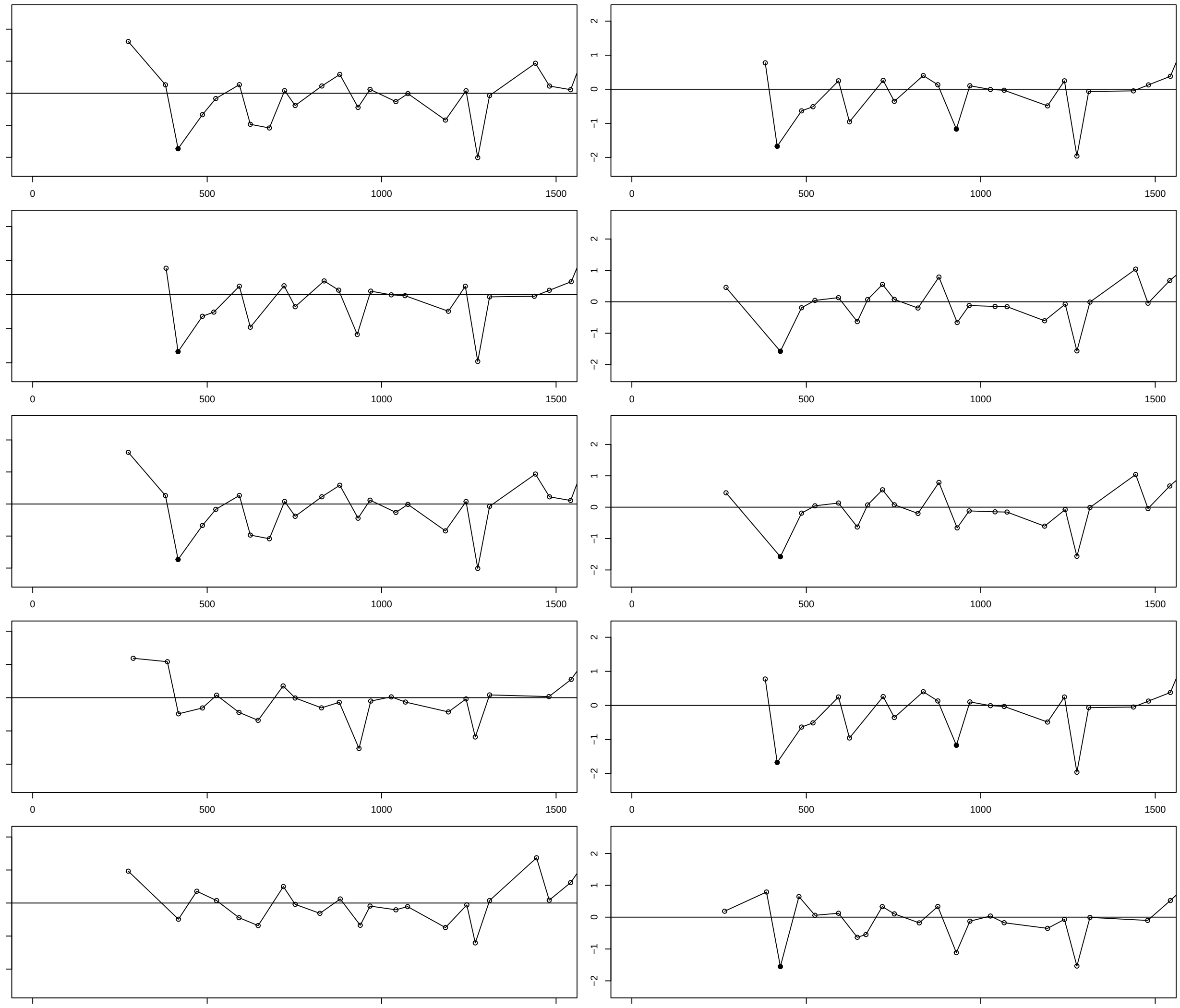
Subsample at 250km for Chinese dataset

Now we subsample at 250km to reduce spatial autocorrelation.

```r
table(black.china$spatial.code.250)

## Annual temperature
par(mfrow=c(5,2), mar=c(1,1,2,2))
for (i in 1:10){
  sub <- black.china %>% group_by(spatial.code.250) %>% sample_n(size=1)
  reg.sub.annual <- lm(log(sub$AAM)~sub$AnnualTemp)

  test <- correlog(sub$longitude, sub$latitude, reg.sub.annual$residuals,
                   increment=50, resamp=500, latlon=T)
  plot(test, main="", xlim=c(0,1500))
  abline(h=0)
}
```




```r
## Cold temperature
par(mfrow=c(5,2), mar=c(1,1,2,2))
for (i in 1:10){
  sub <- black.china %>% group_by(spatial.code.250) %>% sample_n(size=1)
  reg.sub.cold <- lm(log(sub$AAM)~sub$ColdTemp)

  test <- correlog(sub$longitude, sub$latitude, reg.sub.cold$residuals,
                   increment=50, resamp=500, latlon=T)
  plot(test, main="", xlim=c(0,1500))
  abline(h=0)
}
```



Conclusions

- There was no global spatial autocorrelation in the entire dataset, but when looking at local clustering, we saw a strong negative autocorrelation at around 550 km.
- We tried to subsample at 250 km, and 550 km. 220 km did not reduce the spatial autocorrelation we saw. 550 km did reduce that, but left us with too few data to fit a model.
- We then tried to subsample by latitudinal straitification. We found that the negative autocorrelation was reduced in some iterations (when South Ukarine was not selected). We hypothesized that particular datapoint is the cause.
- We removed the SU data point and ran the local moran's I again. This time, we have no local clustering at 550 km!