SpaceAndTime

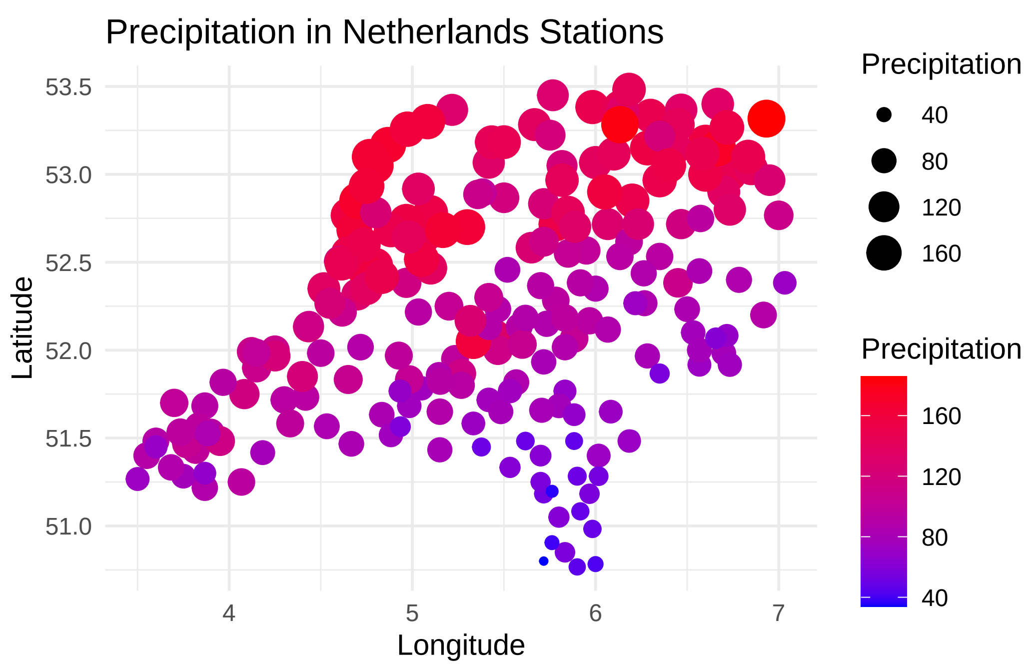
Table of contents

TODO check if I checked all the residuals for all models

Question 1 Spatial modelling Kingdom of the Netherlands

1 a)

```
ggplot(data = netherlandsDF) + geom_point(aes(x = longitude, y = latitude,
    size = precip, color = precip)) + scale_color_continuous(low = "blue",
    high = "red") + labs(title = "Precipitation in Netherlands Stations",
    x = "Longitude", y = "Latitude", size = "Precipitation", color = "Precipitation") +
    theme_minimal()
```



From what we can see from the data it does seem to be spatially correlated as we can that the Dutch provinces of north Holland, Friesland and Groningen has higher precipitation and as we go south the precipitation does decrease as we can see from the Dutch provinces of Zeeland, north Brabant and Limburg where precipitation is significantly lower than their northern counterparts.

From this data, latitude seems to be the biggest factor in the variation of the precipitation as the longitude only suggests some slight variations in the data.

```
# Create geodata object
precipitationNetherland_geoR = as.geodata(netherlandsDF, coords.col = c("longitude",
    "latitude"), data.col = "precip")

summary(precipitationNetherland_geoR)
```

```
Number of data points: 220

Coordinates summary
    longitude latitude
min     3.500   50.767
max     7.033   53.483

Distance summary
     min      max
0.001000 3.998498

Data summary
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
 33.9000  80.8500 100.1500 106.5705 137.3500 185.6000
```

As we can see from the numerical summary of the data the median is different from the mean, which indicates it is not a symmetric distribution of data points and is instead positively skewed since the mean is bigger than the median. As such there are more values on the left side of the distribution.

1 b)

```
# set seed for reproducibility
set.seed(26041999)

# Select 3 random rows from the data frame
randomRowsPrecipitation = netherlandsDF %>%
    sample_n(3)

# Add a new column with labels
randomRowsPrecipitation$label = c("A", "B", "C")
```

```r
# Print the randomly selected rows
randomRowsPrecipitation
```

```
# A tibble: 3 x 5
  station_name   longitude latitude precip label
  <chr>              <dbl>    <dbl>  <dbl> <chr>
1 NIJKERK             5.47     52.2   89.1 A
2 WOLPHAARTSDIJK      3.73     51.5   95.9 B
3 EEXT                6.73     53    147.  C
```

```r
# Remove the selected rows from the original dataset
netherlandsDF_filtered = netherlandsDF %>%
    anti_join(randomRowsPrecipitation)
```

```
Joining, by = c("station_name", "longitude", "latitude", "precip")
```

```r
# Print the resulting dataframe
netherlandsDF_filtered
```

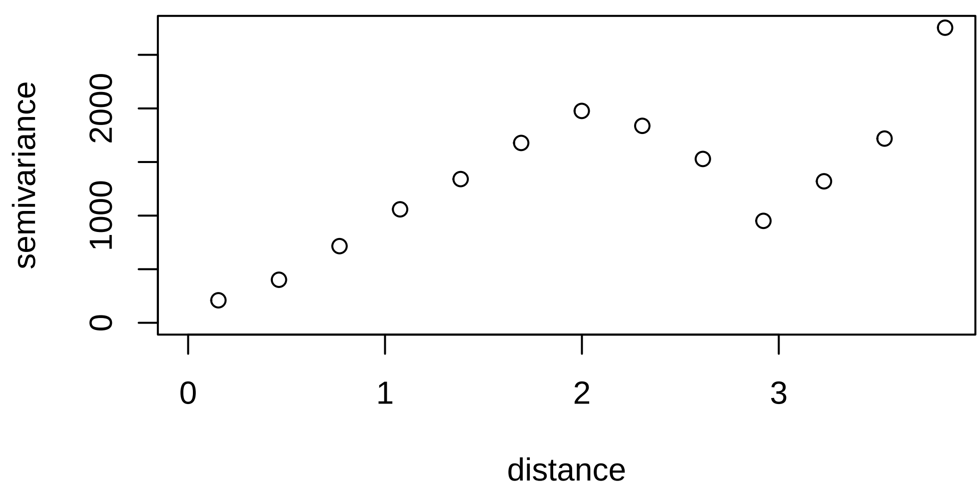
```
# A tibble: 217 x 4
   station_name      longitude latitude precip
   <chr>                 <dbl>    <dbl>  <dbl>
 1 WEST TERSCHELLING      5.22     53.4  130.
 2 GRONINGEN-1            6.6      53.2  157.
 3 HOORN                  5.07     52.6  146.
 4 HOOFDDORP              4.7      52.3  130.
 5 WINTERSWIJK            6.7      52.0   77.7
 6 KERKWERVE              3.87     51.7   91.8
 7 WESTDORPE-1            3.87     51.2   87.7
 8 OUDENBOSCH             4.53     51.6   84.2
 9 ROERMOND               5.97     51.2   56.4
10 PETTEN                 4.65     52.8  158.
# ... with 207 more rows
```

1 c)

```r
# # Calculate empirical variogram
variogramPrecipitationNetherlands = variog(precipitationNetherland_geoR)
```

```
variog: computing omnidirectional variogram
```

```r
# Plot empirical variogram
plot(variogramPrecipitationNetherlands)
```




```r
variogramPrecipitationNetherlands$n
```

```
 [1] 1108 2543 3474 3915 3726 3293 2615 1597  811  497  346  147   17
```

From the plotted variogram we can see there a very clear need for a nugget as there is a non-zero value around zero distance, this values seems to be around 75 to 100 at the zero distance from how much is it decreasing.

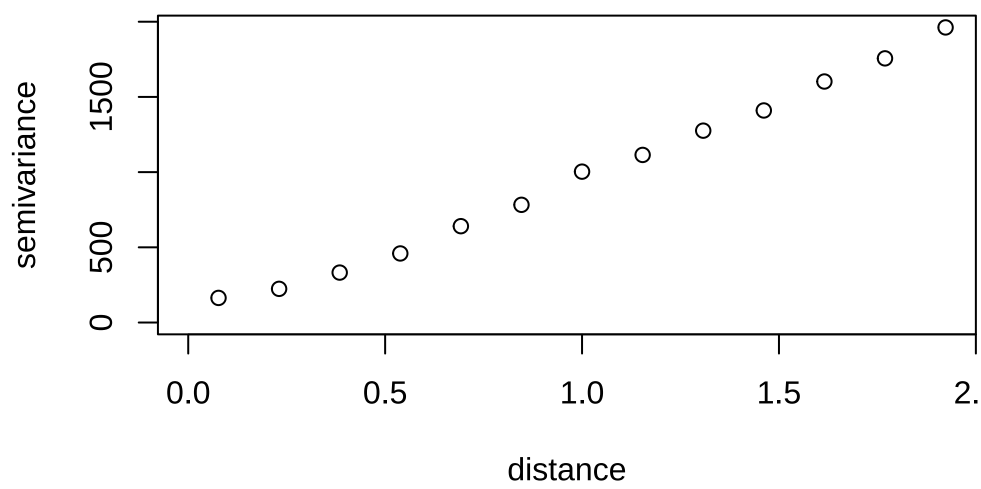
The semi variance continuous to increase with distance till around the distance of 2 degrees distance wise, after this there is a decrease in variance that is not representative of the data as we are more and more uncertain the further we are from our known points, as such we will choose the distance of two as the cut off for the maximum distance.

we know change the maximum distance change and recut our previous variogram.

```
variogramPrecipitationNetherlands = variog(precipitationNetherland_geoR,
    option = "bin", max.dist = 2)
```

```
variog: computing omnidirectional variogram
```

```
plot(variogramPrecipitationNetherlands)
```



As we can see from the newly updated variogram the increase is almost linear with a curve near 0 where we can see the need for the nugget.

1 d)

Now that we have the variogram we will start by fitting a model to estimate the covariance via weighted least squares. Fitting this variogram we get the estimated values of σ^2, ϕ and τ^2 also known as the nugget

We will first start with the default Matrén = 0,5 which is equivalent to an exponential as the form observed in the previous variogram seems to not fully linear and therefore require the curvature from a function like the exponential function to account for the behaviour at the near 0 distance.

From this we will try different models to search for the model with the best fit.

```
# ?variofit thau = nugget variability sigmasq = if the model can
# capture more or less of the total variability phi = if the
# correlation extends over a bigger or smaller distance loss value =
# goodness of fit (smaller means better fit)

krigingVariogramFittedDefault = variofit(variogramPrecipitationNetherlands,
    nugget = 85)
```

```
variofit: covariance model used is matern
variofit: weights used: npairs
variofit: minimisation function used: optim

Warning in variofit(variogramPrecipitationNetherlands, nugget = 85): initial
values not provided - running the default search

variofit: searching for best initial value ... selected values:
              sigmasq   phi    tausq kappa
initial.value "1962.06" "1.54" "85"  "0.5"
status        "est"     "est"  "est" "fix"
loss value: 818053329.562714
```

```
krigingVariogramFittedDefault
```

```
variofit: model parameters estimated by WLS (weighted least squares):
covariance model is: matern with fixed kappa = 0.5 (exponential)
parameter estimates:
      tausq     sigmasq         phi
       0.00 57246963.03    58260.33
Practical Range with cor=0.05 for asymptotic range: 174532.4

variofit: minimised weighted sum of squares = 29450180
```

Now we will increase the kappa of the Matrén to see if the increased flexibility and smoothness leads to a better fit

```r
krigingVariogramFittedMatrén1.5 = variofit(variogramPrecipitationNetherlands,
    kappa = 1.5, nugget = 85)
```

```
variofit: covariance model used is matern
variofit: weights used: npairs
variofit: minimisation function used: optim
```

```
Warning in variofit(variogramPrecipitationNetherlands, kappa = 1.5, nugget =
85): initial values not provided - running the default search
```

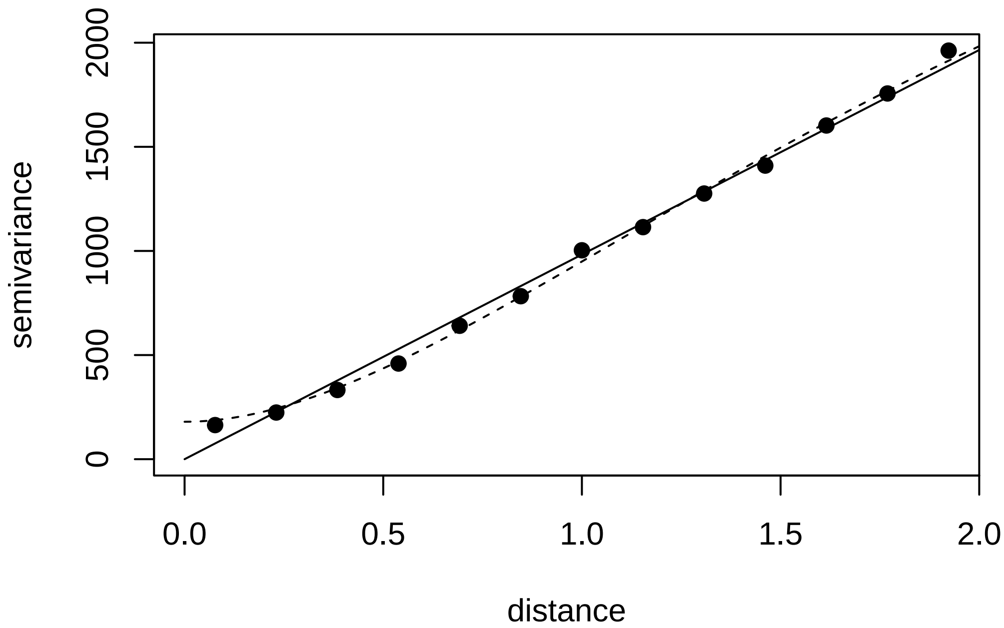
```
variofit: searching for best initial value ... selected values:
              sigmasq   phi    tausq kappa
initial.value "1962.06" "0.62" "85"  "1.5"
status        "est"     "est"  "est" "fix"
loss value: 181359537.270584
```

```r
krigingVariogramFittedMatrén1.5
```

```
variofit: model parameters estimated by WLS (weighted least squares):
covariance model is: matern with fixed kappa = 1.5
parameter estimates:
    tausq   sigmasq       phi
 179.6825 3303.4310    1.0931
Practical Range with cor=0.05 for asymptotic range: 5.185748

variofit: minimised weighted sum of squares = 15252946
```

We will first visually compare these 2 models to see which one has a better

```r
par(mar = c(4, 4, 2, 2))
plot(variogramPrecipitationNetherlands, pch = 19)
lines(krigingVariogramFittedDefault)
lines(krigingVariogramFittedMatrén1.5, lty = 2)
```




```r
# lines(krigingVariogramFittedMatrén2.5, lty = 3)
```

Immediately we can see that the extra flexibility of the Matrén 1,5 not only better follows the actual data, it actually accounts correctly for the initial variance from the nugget which the Matrén 0,5 does not as it simply decreases to 0.

We now will test if any additional flexibility changes can improve the model fit

We will first try to again increase the kappa to see if the model again benefits from the extra flexibility

```r
krigingVariogramFittedMatrén2.0 = variofit(variogramPrecipitationNetherlands,
    kappa = 2, nugget = 85)
```

```
variofit: covariance model used is matern
variofit: weights used: npairs
variofit: minimisation function used: optim
```

```
Warning in variofit(variogramPrecipitationNetherlands, kappa = 2, nugget = 85):
initial values not provided - running the default search
```

```
variofit: searching for best initial value ... selected values:
              sigmasq   phi    tausq    kappa
initial.value "1962.06" "0.62" "196.21" "2"
status        "est"     "est"  "est"    "fix"
loss value: 237942406.426042
```

```
kriging VariogramFittedMatrén2.0
```

```
variofit: model parameters estimated by WLS (weighted least squares):
covariance model is: matern with fixed kappa = 2
parameter estimates:
    tausq   sigmasq       phi
 195.4398 2916.3473    0.8144
Practical Range with cor=0.05 for asymptotic range: 4.371936

variofit: minimised weighted sum of squares = 18113597
```

Here we will instead see if the model will benefit instead form a cut of flexibility to make it less smooth

```
krigingVariogramFittedMatrén1.0 = variofit(variogramPrecipitationNetherlands,
    kappa = 1, nugget = 85)
```

```
variofit: covariance model used is matern
variofit: weights used: npairs
variofit: minimisation function used: optim
```

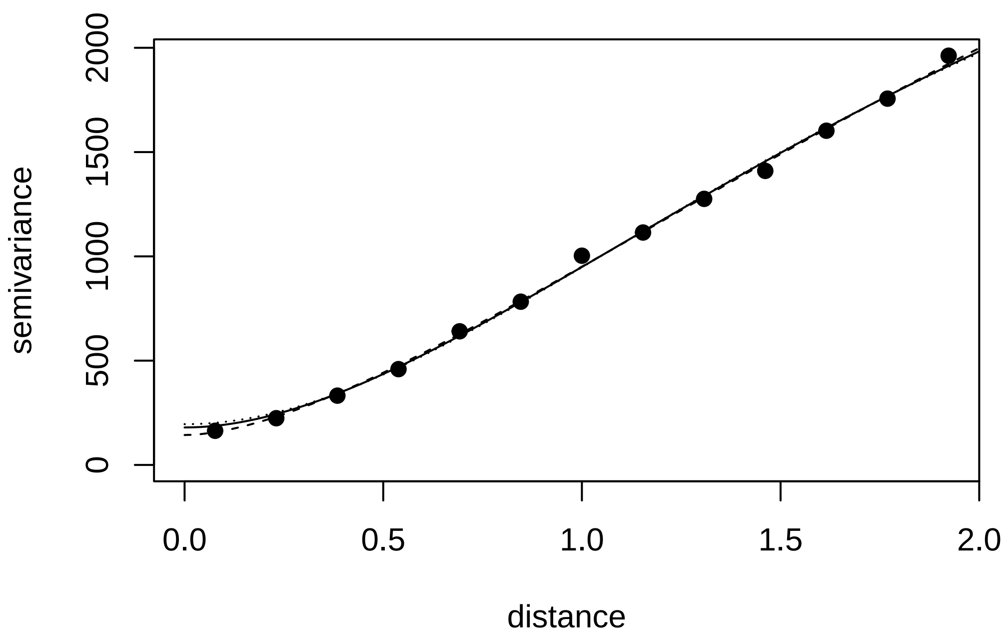
```
Warning in variofit(variogramPrecipitationNetherlands, kappa = 1, nugget = 85):
initial values not provided - running the default search
```

```
variofit: searching for best initial value ... selected values:
              sigmasq   phi    tausq    kappa
initial.value "1962.06" "0.92" "196.21" "1"
status        "est"     "est"  "est"    "fix"
loss value: 340447379.700866
```

```
krigingVariogramFittedMatrén1.0
```

```
variofit: model parameters estimated by WLS (weighted least squares):
covariance model is: matern with fixed kappa = 1
parameter estimates:
    tausq   sigmasq       phi
 143.3735 4519.3807    1.9424
Practical Range with cor=0.05 for asymptotic range: 7.766862

variofit: minimised weighted sum of squares = 11764275
```

```
par(mar = c(4, 4, 2, 2))
plot(variogramPrecipitationNetherlands, pch = 19)
lines(krigingVariogramFittedMatrén1.5)
lines(krigingVariogramFittedMatrén1.0, lty = 2)
lines(krigingVariogramFittedMatrén2.0, lty = 3)
```



As we can see from the new graph it does seem that actually a lower flexibility Matrén has a better fit since the extra flexibility near the start and end of the data points made the models deviate too much from the points.

Lastly we will make very small changes in kappa and use the goodness of fit to evaluate each of the models to determine which one is the better fit.

TODO

1 e)

To fit a model using the maximum likelihood we will have to try multiple initial values to make sure this is indeed the maximum likelihood and not just a local maximum.

```r
# ?likest

maximumLikelihoodNetherlandsInitial10.1 = likfit(precipitationNetherland_geoR,
    ini.cov.pars = c(10, 1))
```

```
---------------------------------------------------------------
likfit: likelihood maximisation using the function optim.
likfit: Use control() to pass additional
         arguments for the maximisation function.
        For further details see documentation for optim.
likfit: It is highly advisable to run this function several
        times with different initial values for the parameters.
likfit: WARNING: This step can be time demanding!
---------------------------------------------------------------
likfit: end of numerical maximisation.
```

```r
maximumLikelihoodNetherlandsInitial1.10 = likfit(precipitationNetherland_geoR,
    ini.cov.pars = c(1, 10))
```

```
---------------------------------------------------------------
likfit: likelihood maximisation using the function optim.
likfit: Use control() to pass additional
         arguments for the maximisation function.
        For further details see documentation for optim.
likfit: It is highly advisable to run this function several
        times with different initial values for the parameters.
likfit: WARNING: This step can be time demanding!
---------------------------------------------------------------
likfit: end of numerical maximisation.
```

```r
maximumLikelihoodNetherlandsInitial100.10 = likfit(precipitationNetherland_geoR,
    ini.cov.pars = c(100, 10))
```

```
---------------------------------------------------------------
likfit: likelihood maximisation using the function optim.
likfit: Use control() to pass additional
         arguments for the maximisation function.
        For further details see documentation for optim.
likfit: It is highly advisable to run this function several
        times with different initial values for the parameters.
likfit: WARNING: This step can be time demanding!
---------------------------------------------------------------
likfit: end of numerical maximisation.
```

```
maximumLikelihoodNetherlandsInitial10.100 = likfit(precipitationNetherland_geoR,
    ini.cov.pars = c(10, 100))
```

```
---------------------------------------------------------------
likfit: likelihood maximisation using the function optim.
likfit: Use control() to pass additional
         arguments for the maximisation function.
        For further details see documentation for optim.
likfit: It is highly advisable to run this function several
        times with different initial values for the parameters.
likfit: WARNING: This step can be time demanding!
---------------------------------------------------------------
likfit: end of numerical maximisation.

WARNING: estimated range is more than 10 times bigger than the biggest distance between two p
 1) excluding spatial dependence if estimated sill is too low and/or
 2) taking trends (covariates) into account
```

```
maximumLikelihoodNetherlandsInitial1.1 = likfit(precipitationNetherland_geoR,
    ini.cov.pars = c(1, 1))
```

```
---------------------------------------------------------------
likfit: likelihood maximisation using the function optim.
likfit: Use control() to pass additional
         arguments for the maximisation function.
        For further details see documentation for optim.
likfit: It is highly advisable to run this function several
        times with different initial values for the parameters.
likfit: WARNING: This step can be time demanding!
```

```
----------------------------------------------------------------
likfit: end of numerical maximisation.
```

```r
maximumLikelihoodNetherlandsInitial1000.1000 = likfit(precipitationNetherland_geoR,
    ini.cov.pars = c(1000, 1000))
```

```
----------------------------------------------------------------
likfit: likelihood maximisation using the function optim.
likfit: Use control() to pass additional
         arguments for the maximisation function.
        For further details see documentation for optim.
likfit: It is highly advisable to run this function several
        times with different initial values for the parameters.
likfit: WARNING: This step can be time demanding!
----------------------------------------------------------------
likfit: end of numerical maximisation.

WARNING: estimated range is more than 10 times bigger than the biggest distance between two p
 1) excluding spatial dependence if estimated sill is too low and/or
 2) taking trends (covariates) into account
```

```r
maximumLikelihoodNetherlandsInitial500.500 = likfit(precipitationNetherland_geoR,
    ini.cov.pars = c(500, 500))
```

```
----------------------------------------------------------------
likfit: likelihood maximisation using the function optim.
likfit: Use control() to pass additional
         arguments for the maximisation function.
        For further details see documentation for optim.
likfit: It is highly advisable to run this function several
        times with different initial values for the parameters.
likfit: WARNING: This step can be time demanding!
----------------------------------------------------------------
likfit: end of numerical maximisation.

WARNING: estimated range is more than 10 times bigger than the biggest distance between two p
 1) excluding spatial dependence if estimated sill is too low and/or
 2) taking trends (covariates) into account
```

```r
maximumLikelihoodNetherlandsInitial10.1
```

```
likfit: estimated model parameters:
      beta      tausq    sigmasq        phi 
" 102.303" " 110.901" "2998.290" "   6.828" 
Practical Range with cor=0.05 for asymptotic range: 20.45549

likfit: maximised log-likelihood = -908.1
```

```r
maximumLikelihoodNetherlandsInitial1.10
```

```
likfit: estimated model parameters:
      beta      tausq    sigmasq        phi 
" 102.383" " 111.658" "4307.936" "   9.991" 
Practical Range with cor=0.05 for asymptotic range: 29.92928

likfit: maximised log-likelihood = -908.2
```

```r
maximumLikelihoodNetherlandsInitial100.10
```

```
likfit: estimated model parameters:
      beta      tausq    sigmasq        phi 
" 102.383" " 111.658" "4307.936" "   9.991" 
Practical Range with cor=0.05 for asymptotic range: 29.92928

likfit: maximised log-likelihood = -908.2
```

```r
maximumLikelihoodNetherlandsInitial10.100
```

```
likfit: estimated model parameters:
     beta     tausq   sigmasq       phi 
"  102.6" "  114.7" "40700.5" "  100.0" 
Practical Range with cor=0.05 for asymptotic range: 299.5729

likfit: maximised log-likelihood = -909
```

```r
maximumLikelihoodNetherlandsInitial1.1
```

```
likfit: estimated model parameters:
      beta      tausq    sigmasq        phi 
" 102.303" " 110.901" "2998.290" "   6.828" 
Practical Range with cor=0.05 for asymptotic range: 20.45549

likfit: maximised log-likelihood = -908.1
```

```
maximumLikelihoodNetherlandsInitial1000.1000
```

```
likfit: estimated model parameters:
      beta      tausq    sigmasq        phi 
"   103.7" "   169.9" "173718.2" "  1000.0" 
Practical Range with cor=0.05 for asymptotic range: 2995.732

likfit: maximised log-likelihood = -916.2
```

```
maximumLikelihoodNetherlandsInitial500.500
```

```
likfit: estimated model parameters:
      beta      tausq    sigmasq        phi 
"   103.2" "   144.7" "128228.2" "   500.0" 
Practical Range with cor=0.05 for asymptotic range: 1497.866

likfit: maximised log-likelihood = -911.7
```

As we can see from the maximised log-likelihoods it does seem that have indeed reached the maximum log-likelihood as none of the values are too fastly different and the likelihood worsedned as we started to increase much more our starting values.

Next we will try the REML that takes into account the fact that some of the parameters of the model are related to the variance of the residuals and not the mean.

```
REMLmaximumLikelihoodNetherlandsInitial10.1 = likfit(precipitationNetherland_geoR,
    ini.cov.pars = c(10, 1), lik.method = "REML")
```

```
---------------------------------------------------------------
likfit: likelihood maximisation using the function optim.
likfit: Use control() to pass additional
         arguments for the maximisation function.
```

```
        For further details see documentation for optim.
likfit: It is highly advisable to run this function several
        times with different initial values for the parameters.
likfit: WARNING: This step can be time demanding!
---------------------------------------------------------------
likfit: end of numerical maximisation.

WARNING: estimated range is more than 10 times bigger than the biggest distance between two p
 1) excluding spatial dependence if estimated sill is too low and/or
 2) taking trends (covariates) into account
```

```r
REMLmaximumLikelihoodNetherlandsInitial1.10 = likfit(precipitationNetherland_geoR,
    ini.cov.pars = c(1, 10), lik.method = "REML")
```

```
---------------------------------------------------------------
likfit: likelihood maximisation using the function optim.
likfit: Use control() to pass additional
         arguments for the maximisation function.
        For further details see documentation for optim.
likfit: It is highly advisable to run this function several
        times with different initial values for the parameters.
likfit: WARNING: This step can be time demanding!
---------------------------------------------------------------
likfit: end of numerical maximisation.

WARNING: estimated range is more than 10 times bigger than the biggest distance between two p
 1) excluding spatial dependence if estimated sill is too low and/or
 2) taking trends (covariates) into account
```

```r
REMLmaximumLikelihoodNetherlandsInitial100.1 = likfit(precipitationNetherland_geoR,
    ini.cov.pars = c(100, 1), lik.method = "REML")
```

```
---------------------------------------------------------------
likfit: likelihood maximisation using the function optim.
likfit: Use control() to pass additional
         arguments for the maximisation function.
        For further details see documentation for optim.
likfit: It is highly advisable to run this function several
        times with different initial values for the parameters.
likfit: WARNING: This step can be time demanding!
```

```
---------------------------------------------------------------
likfit: end of numerical maximisation.

WARNING: estimated range is more than 10 times bigger than the biggest distance between two p
 1) excluding spatial dependence if estimated sill is too low and/or
 2) taking trends (covariates) into account
```

```r
REMLmaximumLikelihoodNetherlandsInitial1.100 = likfit(precipitationNetherland_geoR,
    ini.cov.pars = c(1, 100), lik.method = "REML")
```

```
---------------------------------------------------------------
likfit: likelihood maximisation using the function optim.
likfit: Use control() to pass additional
         arguments for the maximisation function.
        For further details see documentation for optim.
likfit: It is highly advisable to run this function several
        times with different initial values for the parameters.
likfit: WARNING: This step can be time demanding!
---------------------------------------------------------------
likfit: end of numerical maximisation.

WARNING: estimated range is more than 10 times bigger than the biggest distance between two p
 1) excluding spatial dependence if estimated sill is too low and/or
 2) taking trends (covariates) into account
```

```r
REMLmaximumLikelihoodNetherlandsInitial10.1
```

```
likfit: estimated model parameters:
      beta      tausq    sigmasq        phi
"  102.55" "  111.99" "18004.23" "   41.92"
Practical Range with cor=0.05 for asymptotic range: 125.5882

likfit: maximised log-likelihood = -900.2
```

```r
REMLmaximumLikelihoodNetherlandsInitial1.10
```

```
likfit: estimated model parameters:
      beta      tausq    sigmasq        phi
```

```
"  102.55" "  111.93" "18056.69" "   42.01"
Practical Range with cor=0.05 for asymptotic range: 125.8445

likfit: maximised log-likelihood = -900.2
```

```r
REMLmaximumLikelihoodNetherlandsInitial100.1
```

```
likfit: estimated model parameters:
      beta      tausq    sigmasq        phi
"  102.55" "  111.99" "18004.23" "   41.92"
Practical Range with cor=0.05 for asymptotic range: 125.5882

likfit: maximised log-likelihood = -900.2
```

```r
REMLmaximumLikelihoodNetherlandsInitial1.100
```

```
likfit: estimated model parameters:
     beta     tausq   sigmasq       phi
"  102.6" "  112.4" "42582.9" "  100.0"
Practical Range with cor=0.05 for asymptotic range: 299.5729

likfit: maximised log-likelihood = -900.1
```

As we can see from these new models, the new likelihood method actually did improve our model as we have a lower log-likelihood.

Since the data did not seem to be perfectly stationary as seen in the previous questions, we will now check if adding a linear trend improves our model

```r
linearREMLmaximumLikelihoodNetherlandsInitial10.1 = likfit(precipitationNetherland_geoR,
    trend = "1st", ini.cov.pars = c(10, 1), lik.method = "REML")
```

```
---------------------------------------------------------------
likfit: likelihood maximisation using the function optim.
likfit: Use control() to pass additional
         arguments for the maximisation function.
        For further details see documentation for optim.
likfit: It is highly advisable to run this function several
        times with different initial values for the parameters.
```

```
likfit: WARNING: This step can be time demanding!
---------------------------------------------------------------
likfit: end of numerical maximisation.
```

```r
linearREMLmaximumLikelihoodNetherlandsInitial1.10 = likfit(precipitationNetherland_geoR,
    trend = "1st", ini.cov.pars = c(1, 10), lik.method = "REML")
```

```
---------------------------------------------------------------
likfit: likelihood maximisation using the function optim.
likfit: Use control() to pass additional
         arguments for the maximisation function.
        For further details see documentation for optim.
likfit: It is highly advisable to run this function several
        times with different initial values for the parameters.
likfit: WARNING: This step can be time demanding!
---------------------------------------------------------------
likfit: end of numerical maximisation.
```

```r
linearREMLmaximumLikelihoodNetherlandsInitial100.10 = likfit(precipitationNetherland_geoR,
    trend = "1st", ini.cov.pars = c(100, 10), lik.method = "REML")
```

```
---------------------------------------------------------------
likfit: likelihood maximisation using the function optim.
likfit: Use control() to pass additional
         arguments for the maximisation function.
        For further details see documentation for optim.
likfit: It is highly advisable to run this function several
        times with different initial values for the parameters.
likfit: WARNING: This step can be time demanding!
---------------------------------------------------------------
likfit: end of numerical maximisation.
```

```r
linearREMLmaximumLikelihoodNetherlandsInitial10.100 = likfit(precipitationNetherland_geoR,
    trend = "1st", ini.cov.pars = c(10, 100), lik.method = "REML")
```

```
---------------------------------------------------------------
likfit: likelihood maximisation using the function optim.
likfit: Use control() to pass additional
         arguments for the maximisation function.
```

```
        For further details see documentation for optim.
likfit: It is highly advisable to run this function several
        times with different initial values for the parameters.
likfit: WARNING: This step can be time demanding!
---------------------------------------------------------------
likfit: end of numerical maximisation.

WARNING: estimated range is more than 10 times bigger than the biggest distance between two p
 1) excluding spatial dependence if estimated sill is too low and/or
 2) taking trends (covariates) into account
```

```
linearREMLmaximumLikelihoodNetherlandsInitial10.1
```

```
likfit: estimated model parameters:
       beta0         beta1         beta2         tausq       sigmasq           phi
"-2196.7161" "  -11.5303" "   45.2810" "  108.3604" "  194.1407" "    0.3866"
Practical Range with cor=0.05 for asymptotic range: 1.158135

likfit: maximised log-likelihood = -883.1
```

```
linearREMLmaximumLikelihoodNetherlandsInitial1.10
```

```
likfit: estimated model parameters:
      beta0       beta1       beta2       tausq     sigmasq         phi
"-2081.678" "   -6.931" "   42.621" "  121.167" " 3409.622" "    9.925"
Practical Range with cor=0.05 for asymptotic range: 29.7319

likfit: maximised log-likelihood = -884.3
```

```
linearREMLmaximumLikelihoodNetherlandsInitial100.10
```

```
likfit: estimated model parameters:
      beta0       beta1       beta2       tausq     sigmasq         phi
"-2081.678" "   -6.931" "   42.621" "  121.167" " 3409.622" "    9.925"
Practical Range with cor=0.05 for asymptotic range: 29.7319

likfit: maximised log-likelihood = -884.3
```

```r
linearREMLmaximumLikelihoodNetherlandsInitial10.100
```

```
likfit: estimated model parameters:
      beta0       beta1       beta2       tausq     sigmasq         phi 
"-2082.220" "   -6.886" "   42.632" "  122.827" "33350.771" "   99.999" 
Practical Range with cor=0.05 for asymptotic range: 299.5701

likfit: maximised log-likelihood = -884.3
```

From these models we can see that the linear trend does indeed improve our model, now we will check if there are any other covariance functions that can improve the model further.

```r
Matren0.5linearREMLmaximumLikelihoodNetherlandsInitial10.1 = likfit(precipitationNetherlan
    trend = "1st", ini.cov.pars = c(10, 1), lik.method = "REML", cov.model = "matern",
    kappa = 0.5)
```

```
---------------------------------------------------------------
likfit: likelihood maximisation using the function optim.
likfit: Use control() to pass additional
         arguments for the maximisation function.
        For further details see documentation for optim.
likfit: It is highly advisable to run this function several
        times with different initial values for the parameters.
likfit: WARNING: This step can be time demanding!
---------------------------------------------------------------
likfit: end of numerical maximisation.
```

```r
Matren1.0linearREMLmaximumLikelihoodNetherlandsInitial10.1 = likfit(precipitationNetherlan
    trend = "1st", ini.cov.pars = c(10, 1), lik.method = "REML", cov.model = "matern",
    kappa = 1)
```

```
---------------------------------------------------------------
likfit: likelihood maximisation using the function optim.
likfit: Use control() to pass additional
         arguments for the maximisation function.
        For further details see documentation for optim.
likfit: It is highly advisable to run this function several
        times with different initial values for the parameters.
likfit: WARNING: This step can be time demanding!
---------------------------------------------------------------
likfit: end of numerical maximisation.
```

```r
Matren1.5linearREMLmaximumLikelihoodNetherlandsInitial10.1 = likfit(precipitationNetherlan
    trend = "1st", ini.cov.pars = c(10, 1), lik.method = "REML", cov.model = "matern",
    kappa = 1.5)
```

```
---------------------------------------------------------------
likfit: likelihood maximisation using the function optim.
likfit: Use control() to pass additional
         arguments for the maximisation function.
        For further details see documentation for optim.
likfit: It is highly advisable to run this function several
        times with different initial values for the parameters.
likfit: WARNING: This step can be time demanding!
---------------------------------------------------------------
likfit: end of numerical maximisation.
```

```r
Matren2.0linearREMLmaximumLikelihoodNetherlandsInitial10.1 = likfit(precipitationNetherlan
    trend = "1st", ini.cov.pars = c(10, 1), lik.method = "REML", cov.model = "matern",
    kappa = 2)
```

```
---------------------------------------------------------------
likfit: likelihood maximisation using the function optim.
likfit: Use control() to pass additional
         arguments for the maximisation function.
        For further details see documentation for optim.
likfit: It is highly advisable to run this function several
        times with different initial values for the parameters.
likfit: WARNING: This step can be time demanding!
---------------------------------------------------------------
likfit: end of numerical maximisation.
```

```r
Matren2.5linearREMLmaximumLikelihoodNetherlandsInitial10.1 = likfit(precipitationNetherlan
    trend = "1st", ini.cov.pars = c(10, 1), lik.method = "REML", cov.model = "matern",
    kappa = 2.5)
```

```
---------------------------------------------------------------
likfit: likelihood maximisation using the function optim.
likfit: Use control() to pass additional
         arguments for the maximisation function.
        For further details see documentation for optim.
```

```
likfit: It is highly advisable to run this function several
        times with different initial values for the parameters.
likfit: WARNING: This step can be time demanding!
---------------------------------------------------------------
likfit: end of numerical maximisation.
```

```
Matren0.5linearREMLmaximumLikelihoodNetherlandsInitial10.1
```

```
likfit: estimated model parameters:
       beta0         beta1         beta2         tausq       sigmasq           phi
"-2196.7161" "  -11.5303" "   45.2810" "  108.3604" "  194.1407" "    0.3866"
Practical Range with cor=0.05 for asymptotic range: 1.158135

likfit: maximised log-likelihood = -883.1
```

```
Matren1.0linearREMLmaximumLikelihoodNetherlandsInitial10.1
```

```
likfit: estimated model parameters:
       beta0         beta1         beta2         tausq       sigmasq           phi
"-2223.8920" "  -12.2266" "   45.8698" "  120.8871" "  163.6950" "    0.2008"
Practical Range with cor=0.05 for asymptotic range: 0.8029735

likfit: maximised log-likelihood = -882.4
```

```
Matren1.5linearREMLmaximumLikelihoodNetherlandsInitial10.1
```

```
likfit: estimated model parameters:
       beta0         beta1         beta2         tausq       sigmasq           phi
"-2235.6567" "  -12.4723" "   46.1193" "  125.2375" "  153.6762" "    0.1478"
Practical Range with cor=0.05 for asymptotic range: 0.7009258

likfit: maximised log-likelihood = -882.2
```

```
Matren2.0linearREMLmaximumLikelihoodNetherlandsInitial10.1
```

```
likfit: estimated model parameters:
      beta0        beta1        beta2        tausq      sigmasq          phi
"-2242.200" "  -12.600" "   46.257" "  127.251" "  148.795" "    0.121"
```

```
Practical Range with cor=0.05 for asymptotic range: 0.6496077

likfit: maximised log-likelihood = -882.1
```

```
Matren2.5linearREMLmaximumLikelihoodNetherlandsInitial10.1
```

```
likfit: estimated model parameters:
       beta0        beta1        beta2        tausq      sigmasq          phi
"-2246.3667" "  -12.6795" "   46.3448" "  128.3372" "  145.9609" "    0.1044"
Practical Range with cor=0.05 for asymptotic range: 0.6180795

likfit: maximised log-likelihood = -882.1
```

It does seem that the matrén covariance function did indeed slightly improved the model so we will compare it to a model using a spherical covariance function. The spherical covariance function is appropriate for this scenario has the spatial correlation between data points decreases rapidly as the distance between the points increases and we are limited with the range of correlation has after 2 degrees of distance we loose sensible correlation, hence the cut in the variogram.

```
SphericallinearREMLmaximumLikelihoodNetherlandsInitial10.1 = likfit(precipitationNetherlan
    trend = "1st", ini.cov.pars = c(10, 1), lik.method = "REML", cov.model = "spherical")
```

```
kappa not used for the spherical correlation function
---------------------------------------------------------------
likfit: likelihood maximisation using the function optim.
likfit: Use control() to pass additional
         arguments for the maximisation function.
        For further details see documentation for optim.
likfit: It is highly advisable to run this function several
        times with different initial values for the parameters.
likfit: WARNING: This step can be time demanding!
---------------------------------------------------------------
likfit: end of numerical maximisation.
```

```
SphericallinearREMLmaximumLikelihoodNetherlandsInitial1.10 = likfit(precipitationNetherlan
    trend = "1st", ini.cov.pars = c(1, 10), lik.method = "REML", cov.model = "spherical")
```

```
kappa not used for the spherical correlation function
---------------------------------------------------------------
likfit: likelihood maximisation using the function optim.
likfit: Use control() to pass additional
         arguments for the maximisation function.
        For further details see documentation for optim.
likfit: It is highly advisable to run this function several
        times with different initial values for the parameters.
likfit: WARNING: This step can be time demanding!
---------------------------------------------------------------
likfit: end of numerical maximisation.
```

```
SphericallinearREMLmaximumLikelihoodNetherlandsInitial100.10 = likfit(precipitationNetherl
    trend = "1st", ini.cov.pars = c(100, 10), lik.method = "REML", cov.model = "spherical"
```

```
kappa not used for the spherical correlation function
---------------------------------------------------------------
likfit: likelihood maximisation using the function optim.
likfit: Use control() to pass additional
         arguments for the maximisation function.
        For further details see documentation for optim.
likfit: It is highly advisable to run this function several
        times with different initial values for the parameters.
likfit: WARNING: This step can be time demanding!
---------------------------------------------------------------
likfit: end of numerical maximisation.
```

```
SphericallinearREMLmaximumLikelihoodNetherlandsInitial10.100 = likfit(precipitationNetherl
    trend = "1st", ini.cov.pars = c(10, 100), lik.method = "REML", cov.model = "spherical"
```

```
kappa not used for the spherical correlation function
---------------------------------------------------------------
likfit: likelihood maximisation using the function optim.
likfit: Use control() to pass additional
         arguments for the maximisation function.
        For further details see documentation for optim.
likfit: It is highly advisable to run this function several
        times with different initial values for the parameters.
likfit: WARNING: This step can be time demanding!
---------------------------------------------------------------
```

```
likfit: end of numerical maximisation.

WARNING: estimated range is more than 10 times bigger than the biggest distance between two p
 1) excluding spatial dependence if estimated sill is too low and/or
 2) taking trends (covariates) into account
```

```
SphericallinearREMLmaximumLikelihoodNetherlandsInitial10.1
```

```
likfit: estimated model parameters:
      beta0       beta1       beta2       tausq     sigmasq         phi
"-2176.793" "  -11.433" "   44.882" "  122.897" "  214.731" "    1.012"
Practical Range with cor=0.05 for asymptotic range: 1.012405

likfit: maximised log-likelihood = -883.5
```

```
SphericallinearREMLmaximumLikelihoodNetherlandsInitial1.10
```

```
likfit: estimated model parameters:
      beta0       beta1       beta2       tausq     sigmasq         phi
"-2080.429" "   -6.642" "   42.570" "  121.277" " 2261.238" "    9.905"
Practical Range with cor=0.05 for asymptotic range: 9.904762

likfit: maximised log-likelihood = -884.3
```

```
SphericallinearREMLmaximumLikelihoodNetherlandsInitial100.10
```

```
likfit: estimated model parameters:
      beta0       beta1       beta2       tausq     sigmasq         phi
"-2080.429" "   -6.642" "   42.570" "  121.277" " 2261.238" "    9.905"
Practical Range with cor=0.05 for asymptotic range: 9.904762

likfit: maximised log-likelihood = -884.3
```

```
SphericallinearREMLmaximumLikelihoodNetherlandsInitial10.100
```

```
likfit: estimated model parameters:
      beta0        beta1        beta2        tausq      sigmasq          phi
"-2080.968" "   -6.844" "   42.603" "  121.192" "22882.919" "   99.999"
Practical Range with cor=0.05 for asymptotic range: 99.99893

likfit: maximised log-likelihood = -884.3
```

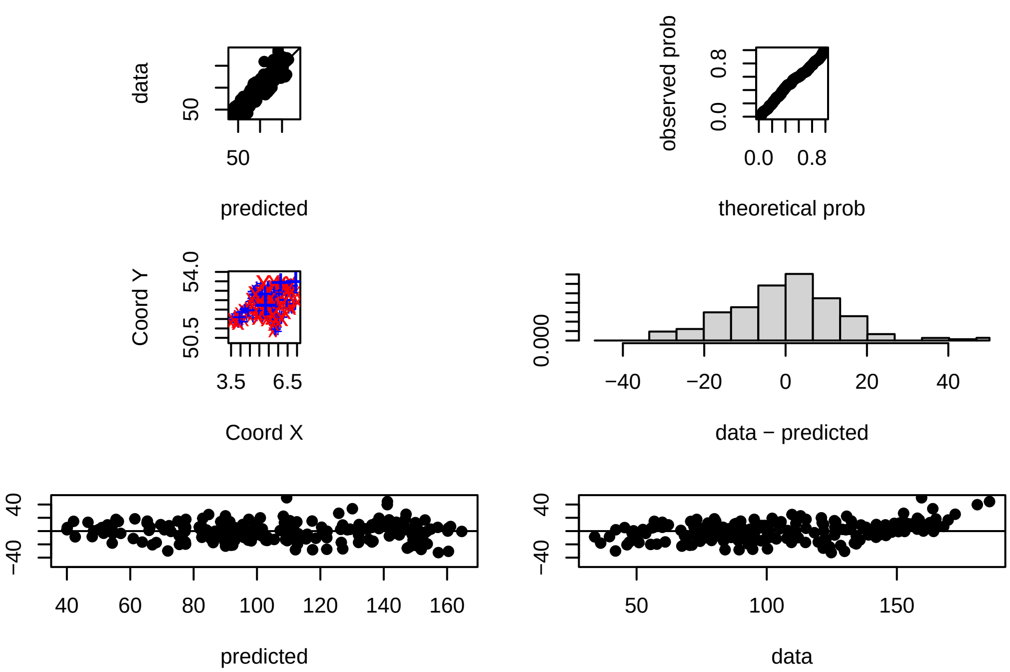
As we can see the spherical covariance function does not provide as good of a fit as the Matrén.

We will now validate the model by doing cross-validation on the model.

```
xv.ml <- xvalid(precipitationNetherland_geoR, model = Matren2.5linearREMLmaximumLikelihood
```

```
xvalid: number of data locations       = 220
xvalid: number of validation locations = 220
xvalid: performing cross-validation at location ... 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 1
xvalid: end of cross-validation
```

```
par(mfrow = c(3, 2), mar = c(4, 2, 2, 2))
plot(xv.ml, error = TRUE, std.error = FALSE, pch = 19)
```



From these plots we can see that the residuals seem mostly normal without any quickly identifiable patterns or bias.

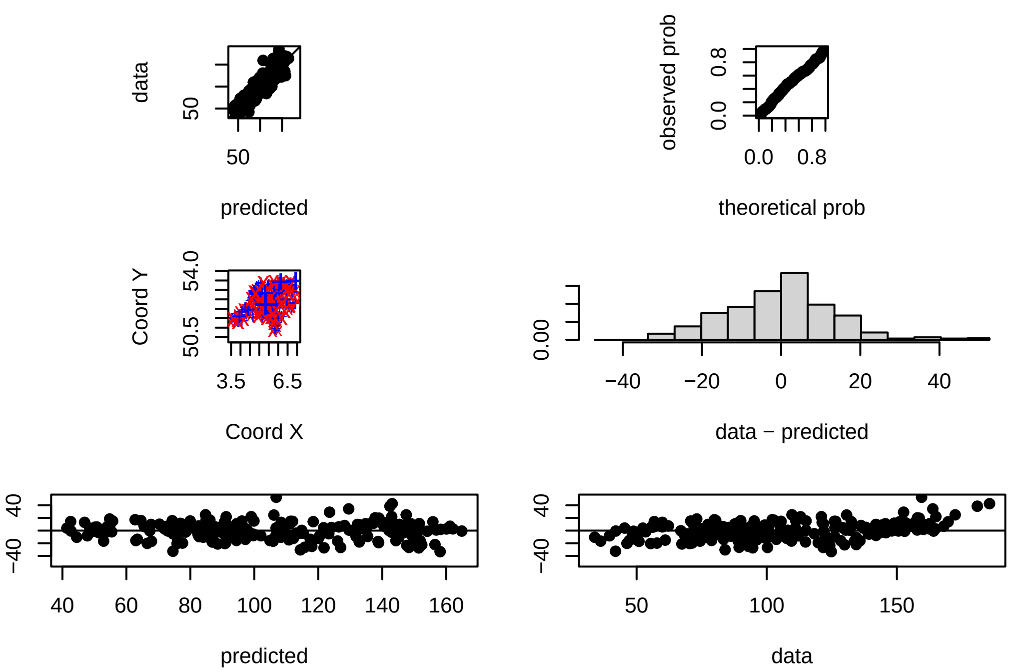
From the first top left graph we can see however that we seem to sightly underestimate more data points.

To account for bias we will also perform cross-validation to the next best performing model using the spherical function instead

```
xv.ml <- xvalid(precipitationNetherland_geoR, model = SphericallinearREMLmaximumLikelihood
```

```
xvalid: number of data locations       = 220
xvalid: number of validation locations = 220
xvalid: performing cross-validation at location ... 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13
xvalid: end of cross-validation
```

```
par(mfrow = c(3, 2), mar = c(4, 2, 2, 2))
plot(xv.ml, error = TRUE, std.error = FALSE, pch = 19)
```



As we can see the data at the start seems to be systematically underestimated and at the end it seems to overestimated. Furthermore the theoretical data plot seems to be less linear.

This confirms that the spherical covariance function is indeed a worse model than the Matrén model.

1 f)

1 g)

1 h)

he predicts using the grid but we have to change it to our own coordinate and the prior use previous results

Question 2

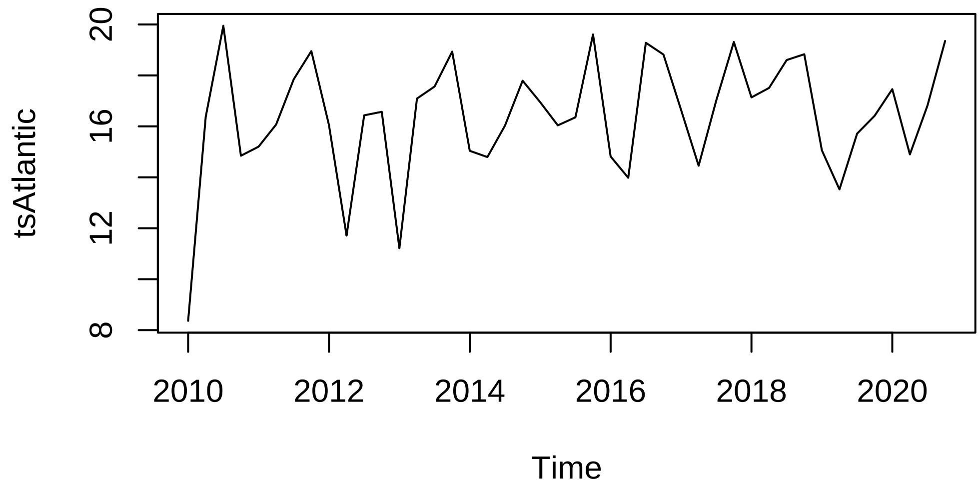
2 a)

We fist start by making the appropriate changes in the data to average the data to quarterly means

```
AMOCDF$Date = as.Date(AMOCDF$Date, format = "%d/%m/%Y")

## I will now make a column with the quarter and year that I will use
## to create the averages per quarter
AMOCDF$YearQuarter = paste(AMOCDF$Year, AMOCDF$Quarter, sep = "-")

YearQuarterAverage = AMOCDF %>%
    group_by(YearQuarter) %>%
    summarise(AverageStrength = mean(Strength))
```

Now we will convert the average data to a time series object to be able to plot it

```
tsAtlantic = ts(YearQuarterAverage, start = c(2010, 1), frequency = 4)

tsAtlantic = tsAtlantic[, "AverageStrength"]

plot.ts(tsAtlantic)
```



Trend analysis

From this graph we can see a yearly oscillation of Sverdrups. We can also identify that the peaks in Sverdrups are usually in the last quarter before the start of a new year and the valleys are on the second quarter of the year.

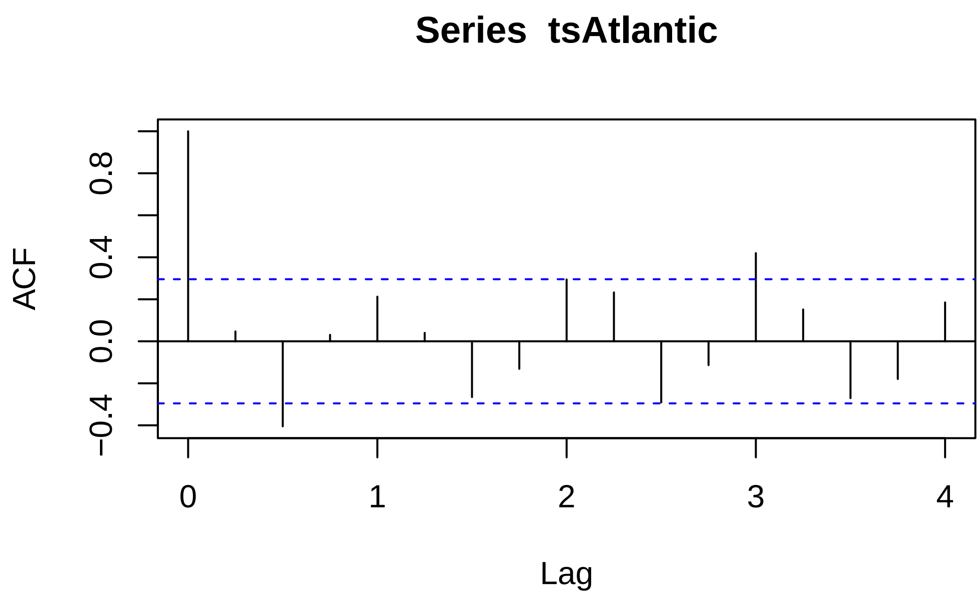
The data does seem stationary enough that if we were to differentiate we would start losing some of the structure.

2 b)

ACF

First we will start by checking the ACF(Autocorrelation Function) and PACF(Partial Autocorrelation Function) to check for if we have stationary data or not to help us decide between an ARMA or an ARIMA model.

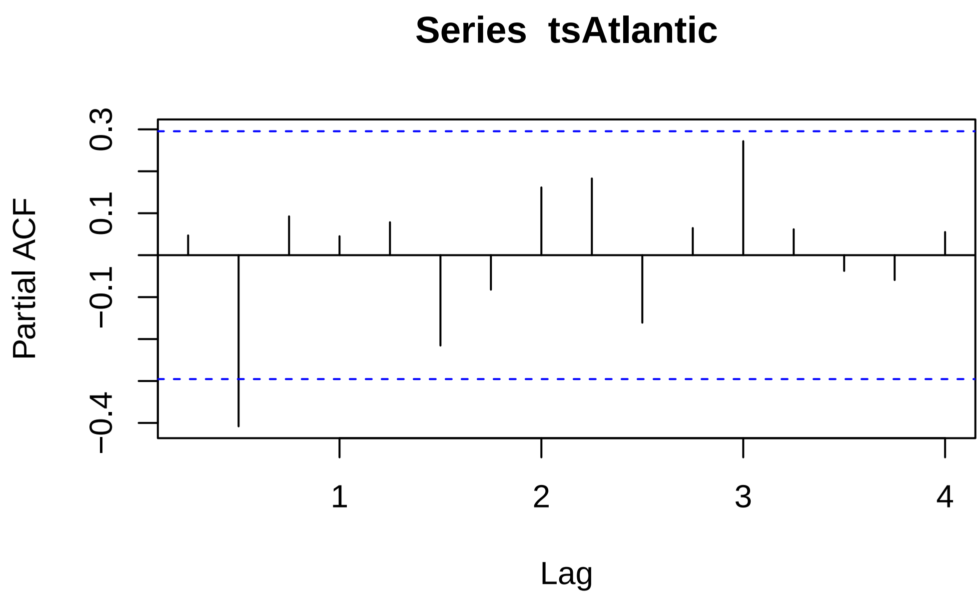
```
acf(tsAtlantic)
```



We can see that for ACF OF Average strength slowly decreases as lag increases to infinity with lag = 3 still being a significant values, meaning it is not a simple MA model as AR is clearly not quickly cut-off.

PACF

```
pacf(tsAtlantic)
```



The PACF seems to be cut-off at lag 0,5 indicating an AR model might be a best fit for our data to be a but with some almost significant values after the cut it might be also appropriate to some non-zero q values to confirm our initial assumption

As such we will now proceed to fit multiple model firstly with the initial assumption that, then I will both use models with non-zero q and the model given by the auto.arima function to double check that the assumptions made by the previous analyses is correct.

```
# it is always a good practice to try multiple values of p,d and q to
# see if we can do better we then obviously compare via the AIC of the
# models and their log likelihoods it is never enough to check those we
# also need to check the residuals

## order is p, d ,q

## initial models under our assumptions

model100 = Arima(tsAtlantic, order = c(1, 0, 0))
model200 = Arima(tsAtlantic, order = c(2, 0, 0))
model300 = Arima(tsAtlantic, order = c(3, 0, 0))

## now I will add postive q values
```

```r
model101 = Arima(tsAtlantic, order = c(1, 0, 1))
model102 = Arima(tsAtlantic, order = c(1, 0, 2))
model103 = Arima(tsAtlantic, order = c(1, 0, 3))

model201 = Arima(tsAtlantic, order = c(2, 0, 1))
model202 = Arima(tsAtlantic, order = c(2, 0, 2))
model203 = Arima(tsAtlantic, order = c(2, 0, 3))

model301 = Arima(tsAtlantic, order = c(3, 0, 1))
model302 = Arima(tsAtlantic, order = c(3, 0, 2))
model303 = Arima(tsAtlantic, order = c(3, 0, 3))



## lastly we will use auto.arima without seasonality to confirm our
## inital assumptions
modelAuto = auto.arima(tsAtlantic, max.d = 0, max.p = 5, max.q = 5, seasonal = FALSE)
```

best model selection

```r
model100
```

```
Series: tsAtlantic
ARIMA(1,0,0) with non-zero mean

Coefficients:
         ar1     mean
      0.0665  16.3878
s.e.  0.1788   0.3726

sigma^2 = 5.572:  log likelihood = -99.2
AIC=204.41   AICc=205.01   BIC=209.76
```

```r
model200
```

```
Series: tsAtlantic
ARIMA(2,0,0) with non-zero mean

Coefficients:
```

```
         ar1      ar2     mean
      0.0990  -0.5565  16.4298
s.e.  0.1576   0.1488   0.2113

sigma^2 = 4.321:  log likelihood = -93.45
AIC=194.9   AICc=195.92   BIC=202.04
```

```
model300
```

```
Series: tsAtlantic
ARIMA(3,0,0) with non-zero mean

Coefficients:
         ar1      ar2     ar3     mean
      0.1626  -0.5690  0.1464  16.4227
s.e.  0.1729   0.1479  0.1708   0.2409

sigma^2 = 4.35:  log likelihood = -93.09
AIC=196.17   AICc=197.75   BIC=205.1
```

As we can see from these inital models ARIMA(2,0,0) is the model that has the best fit has we can see from its lower AIC score of 194,9.

Now we will check against the other models to check the validity of our assumptions.

```
model101
```

```
Series: tsAtlantic
ARIMA(1,0,1) with non-zero mean

Coefficients:
          ar1     ma1     mean
      -0.4204  0.7718  16.3721
s.e.   0.2390  0.1466   0.4067

sigma^2 = 5.045:  log likelihood = -96.64
AIC=201.29   AICc=202.31   BIC=208.43
```

```
model102
```

```
Series: tsAtlantic
ARIMA(1,0,2) with non-zero mean

Coefficients:
         ar1     ma1      ma2     mean
      0.0230  0.1275  -0.4485  16.4289
s.e.  0.3051  0.2420   0.1348   0.2224

sigma^2 = 4.651:  log likelihood = -94.41
AIC=198.81   AICc=200.39   BIC=207.73
```

```
model103
```

```
Series: tsAtlantic
ARIMA(1,0,3) with non-zero mean

Coefficients:
          ar1     ma1      ma2      ma3     mean
      -0.5284  0.7214  -0.3646  -0.3072  16.4299
s.e.   0.9228  0.8649   0.2077   0.3545   0.2195

sigma^2 = 4.733:  log likelihood = -94.25
AIC=200.5   AICc=202.77   BIC=211.21
```

```
model201
```

```
Series: tsAtlantic
ARIMA(2,0,1) with non-zero mean

Coefficients:
          ar1      ar2     ma1     mean
      -0.0669  -0.5475  0.2187  16.4255
s.e.   0.2740   0.1555  0.2883   0.2300

sigma^2 = 4.366:  log likelihood = -93.15
AIC=196.31   AICc=197.88   BIC=205.23
```

```
model202
```

```
Series: tsAtlantic
ARIMA(2,0,2) with non-zero mean

Coefficients:
         ar1      ar2      ma1     ma2     mean
      0.0787  -0.9982  -0.0255  0.9999  16.4015
s.e.  0.0285   0.0066   0.0899  0.1158   0.2684

sigma^2 = 3.378:  log likelihood = -89.46
AIC=190.91   AICc=193.18   BIC=201.62
```

```
model203
```

```
Series: tsAtlantic
ARIMA(2,0,3) with non-zero mean

Coefficients:
         ar1      ar2     ma1     ma2     ma3     mean
      0.0325  -0.9621  0.0499  0.8487  0.4147  16.4028
s.e.  0.0645   0.0442  0.1987  0.2041  0.2511   0.3044

sigma^2 = 3.315:  log likelihood = -89.07
AIC=192.13   AICc=195.25   BIC=204.62
```

```
model301
```

```
Series: tsAtlantic
ARIMA(3,0,1) with non-zero mean

Coefficients:
         ar1      ar2     ar3      ma1     mean
      0.4092  -0.5931  0.2864  -0.2467  16.4191
s.e.  0.6330   0.1651  0.3580   0.6291   0.2537

sigma^2 = 4.449:  log likelihood = -93.03
AIC=198.06   AICc=200.34   BIC=208.77
```

```
model302
```

```
Series: tsAtlantic
ARIMA(3,0,2) with non-zero mean

Coefficients:
         ar1      ar2     ar3      ma1     ma2     mean
      0.2684  -0.9851  0.2222  -0.3030  1.0000  16.4144
s.e.  0.1999   0.0305  0.1995   0.1453  0.1921   0.2922

sigma^2 = 3.392:  log likelihood = -89.53
AIC=193.06   AICc=196.17   BIC=205.54
```

```
model303
```

```
Series: tsAtlantic
ARIMA(3,0,3) with non-zero mean

Coefficients:
          ar1      ar2      ar3     ma1     ma2     ma3     mean
      -0.3983  -0.9518  -0.4263  0.4381  0.7816  0.7352  16.4197
s.e.   0.3385   0.0412   0.3442  0.2946  0.1690  0.2422   0.2691

sigma^2 = 3.352:  log likelihood = -88.54
AIC=193.08   AICc=197.19   BIC=207.35
```

In this initial analysis we have found models that do have a lower AIC lower log likelihood than our previous best model, however these model ma's standard error are to close the the ma values indicating that while we are getting a better fit we might be overfitting to our data.

As such this does confirm our initial assumption for the choice of a zero q value.

Now lastly we will check if the auto.arima function does comfirm our initial assumptions.

```
modelAuto
```

```
Series: tsAtlantic
ARIMA(2,0,0) with non-zero mean

Coefficients:
         ar1      ar2     mean
      0.0990  -0.5565  16.4298
s.e.  0.1576   0.1488   0.2113
```

```
sigma^2 = 4.321:  log likelihood = -93.45
AIC=194.9   AICc=195.92   BIC=202.04
```

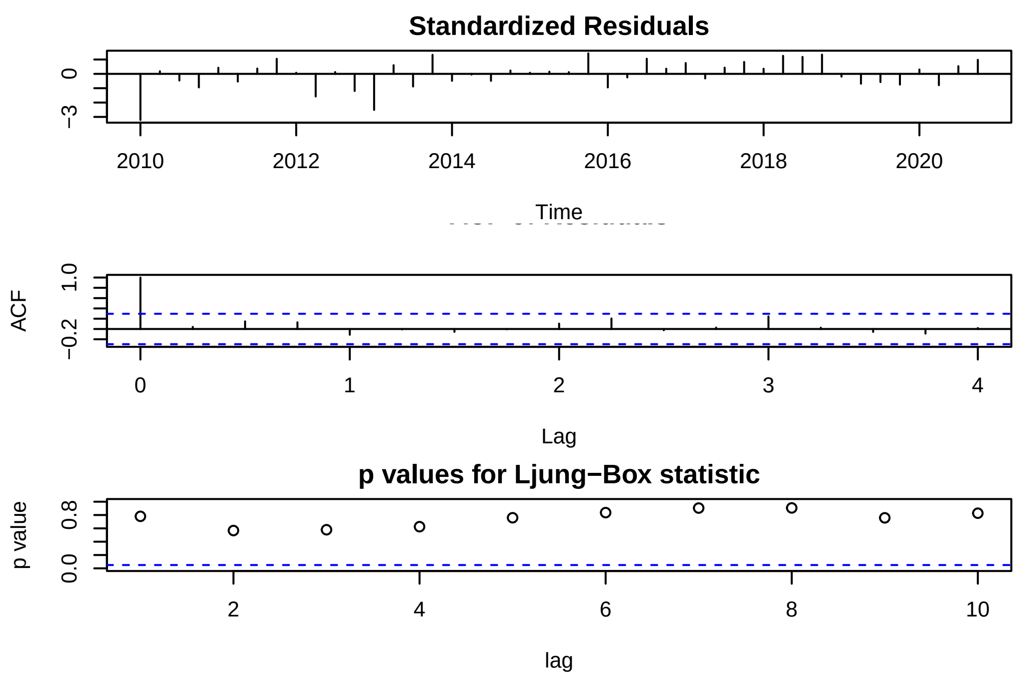
The function does confirm our assumption that ARIMA(2,0,0) is indeed the best model.

We will now check the residuals to verify if any of ou previously selected model validates well or if it is simply the best of bad models.

talk about the model being more easily explainability becaues MA = 0

Best model residual validation

```r
# Set smaller margins
par(mar = c(4, 4, 2, 2))

tsdiag(model200)
```




```r
# Reset margins
par(mar = c(5, 4, 4, 2) + 0.1)
```

Initially from the standardised residuals plot we can identify some sort of sinusoidal pattern, this implies that there is a seasonal trend that is not being accounted for in our model and as such this trends needs to be accounted in future models to better explain and increase the prediction power of a new model.

Forecasting

Now using the forecast function we will forecast the next 4 quarters of 2021

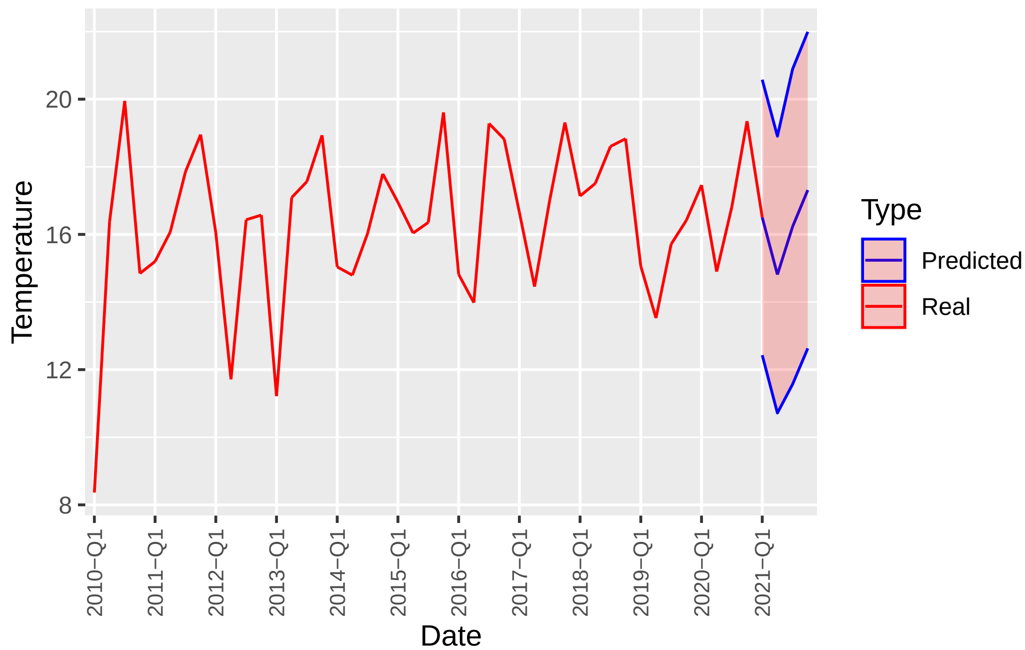
```
forecast(model200, 4)
```

```
        Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
2021 Q1       16.50240 13.83841 19.16639 12.42818 20.57662
2021 Q2       14.81104 12.13403 17.48804 10.71691 18.90517
2021 Q3       16.22919 13.18168 19.27669 11.56843 20.88994
2021 Q4       17.31076 14.24941 20.37212 12.62882 21.99271
```

But this data is better visualized in a graph to better understand if the predictions are sensible compared to our real data.

```
predictedArimaDF = data.frame(forecast(model200, 4))

predictedArimaDF$YearQuarter = c("2021-Q1", "2021-Q2", "2021-Q3", "2021-Q4")

# Combine real_data and pred_data into a single data frame
combinedDataframeAMOC = rbind(data.frame(Date = YearQuarterAverage$YearQuarter,
    Temperature = YearQuarterAverage$AverageStrength, Type = "Real"), data.frame(Date = pr
    Temperature = predictedArimaDF$Point.Forecast, Type = "Predicted"))

predictedArimaDF$Temperature = predictedArimaDF$Point.Forecast

predictedArimaDF$Type = "Predicted"


# Create the ggplot
plotARIMA = ggplot(combinedDataframeAMOC, aes(x = Date, y = Temperature,
    color = Type, group = 1)) + geom_line() + scale_color_manual(values = c("blue",
    "red"))

# Add the 95% confidence interval
plotARIMA = plotARIMA + geom_ribbon(data = predictedArimaDF, aes(x = YearQuarter,
    ymin = Lo.95, ymax = Hi.95), fill = "red", alpha = 0.2)
```

```r
# Adjust the x-axis labels
plotARIMA = plotARIMA + scale_x_discrete(breaks = combinedDataframeAMOC$Date[c(TRUE,
    rep(FALSE, 3))], labels = combinedDataframeAMOC$Date[c(TRUE, rep(FALSE,
    3))])

plotARIMA = plotARIMA + theme(axis.text.x = element_text(angle = 90, vjust = 0.5,
    size = 8))

# Display the plot
print(plotARIMA)
```



As we can see from the graph the ARIMA (2,0,0) seems to give us a sensible forecast for the 2021 quarter values, however as we can see the interval of the prediction accuracy our model is not too certain on the values most likely due to our model not accounting for the seasonal cycle of our data.

2 c)

Initial assumptions

From the previous exploratory analysis of the data we have established that the data did not need to be differentiated since it was constant, this translates to polynomial DLM component of order 2 that will use linear model to account for this type of changes in the data.

Furthermore, from the residual analysis we have inferred that there is an underlying seasonal trend present on the data, this seasonal trend will be represented by a seasonal component of frequency 4 to represent the 4 quarters per year.

model fitting

```
## linear model, order = 2, quadratic order = 3 , etc

## what we want is a linear model with a seasonal component so we add
## the 2 components together in a model

## things to try, another term like quadratic, or a arma component
## stacked on top of this

## Initial model with a linear polynomial and a seasonal component

buildFun = function(x) {
    dlmModPoly(order = 2, dV = exp(x[1]), dW = c(0, exp(x[2]))) + dlmModSeas(frequency = 4
        dV = 0, dW = c(exp(x[3]), rep(0, 2)))
}

linearDLM = dlmMLE(tsAtlantic, parm = c(0, 0, 0), build = buildFun)

linearDLM$par
```

```
[1]   1.151339 -18.078101  -2.189479
```

```
fittedLinearDLM = buildFun(linearDLM$par)

V(fittedLinearDLM)
```

```
         [,1]
[1,] 3.162425
```

```
W(fittedLinearDLM)
```

```
     [,1]         [,2]     [,3] [,4] [,5]
[1,]    0 0.000000e+00 0.000000    0    0
[2,]    0 1.408576e-08 0.000000    0    0
[3,]    0 0.000000e+00 0.111975    0    0
[4,]    0 0.000000e+00 0.000000    0    0
[5,]    0 0.000000e+00 0.000000    0    0
```

```
## second model with a quadratic polynomial and a seasonal component

buildFunQuad = function(x) {
    dlmModPoly(order = 3, dV = exp(x[1]), dW = c(0, exp(x[2]), exp(x[3]))) +
        dlmModSeas(frequency = 4, dV = 0, dW = c(exp(x[4]), rep(0, 2)))
}

quadraticDLM = dlmMLE(tsAtlantic, parm = c(0, 0, 0, 0), build = buildFunQuad)

quadraticDLM$par
```

```
[1]   1.161355 -17.807081 -28.603103  -2.352292
```

```
fittedQuadraticDLM = buildFunQuad(quadraticDLM$par)

V(fittedQuadraticDLM)
```

```
         [,1]
[1,] 3.194257
```

```
W(fittedQuadraticDLM)
```

```
     [,1]         [,2]         [,3]       [,4] [,5] [,6]
[1,]    0 0.000000e+00 0.000000e+00 0.00000000    0    0
[2,]    0 1.847069e-08 0.000000e+00 0.00000000    0    0
[3,]    0 0.000000e+00 3.782948e-13 0.00000000    0    0
[4,]    0 0.000000e+00 0.000000e+00 0.09515082    0    0
[5,]    0 0.000000e+00 0.000000e+00 0.00000000    0    0
[6,]    0 0.000000e+00 0.000000e+00 0.00000000    0    0
```

TODO include dlm with arima?

Now we will compare both models through their log likelihood using the dlmLL function and see if the extra flexibility from the extra polynomial function is providing a better fit

```
dlmLL(tsAtlantic, fittedLinearDLM)
```

```
[1] 94.98804
```

```
dlmLL(tsAtlantic, fittedQuadraticDLM)
```

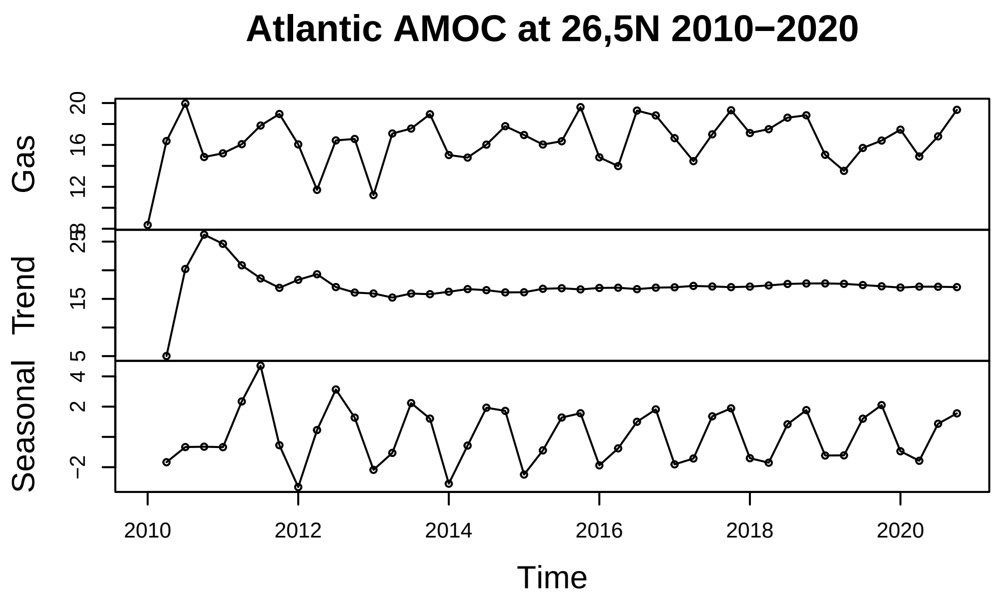
```
[1] 108.043
```

As we can see the dlm model using only a linear polynomial has a lower log likelihood than the model with an extra quadratic term, meaning this extra flexibility does not contribute to a better model fit and as such we will use the linear fitted model to do our forecasting.

```
amocPredict = dlmFilter(tsAtlantic, mod = fittedLinearDLM)
summary(amocPredict)
```

```
    Length Class  Mode
y    44    ts     numeric
mod  10    dlm    list
m   225    mts    numeric
U.C  45    -none- list
D.C 225    -none- numeric
a   220    mts    numeric
U.R  44    -none- list
D.R 220    -none- numeric
f    44    ts     numeric
```

```
x = cbind(tsAtlantic, dropFirst(amocPredict$a[, c(1, 3)]))
x = window(x, start = c(2010, 1))
colnames(x) = c("Gas", "Trend", "Seasonal")
plot(x, type = "o", main = "Atlantic AMOC at 26,5N 2010-2020")
```




```r
amocForecast = dlmForecast(amocPredict, nAhead = 4)
summary(amocForecast)
```

```
  Length Class  Mode
a 20     mts    numeric
R  4     -none- list
f  4     ts     numeric
Q  4     -none- list
```

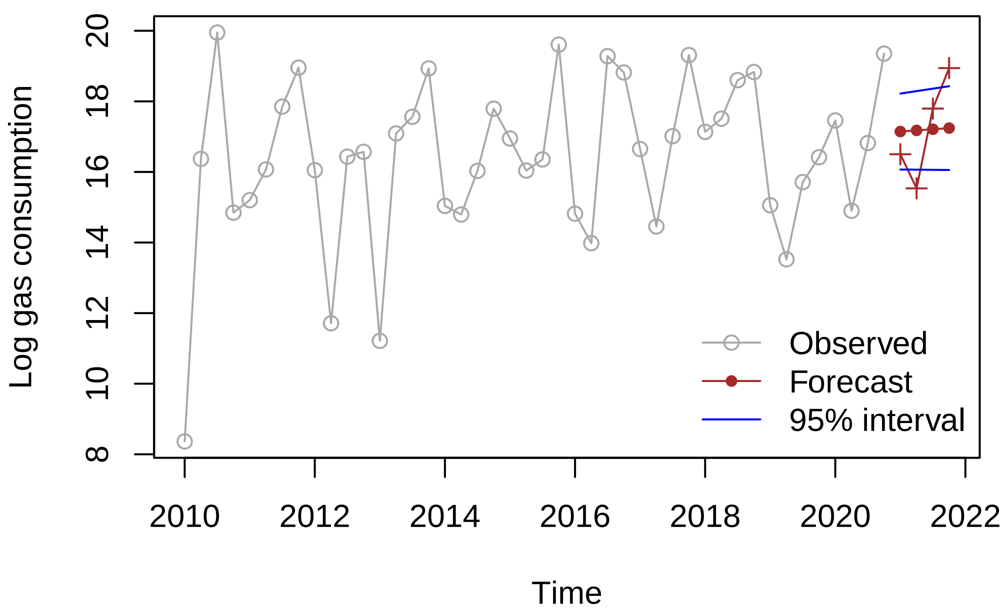
```r
dim(amocForecast$a)
```

```
[1] 4 5
```

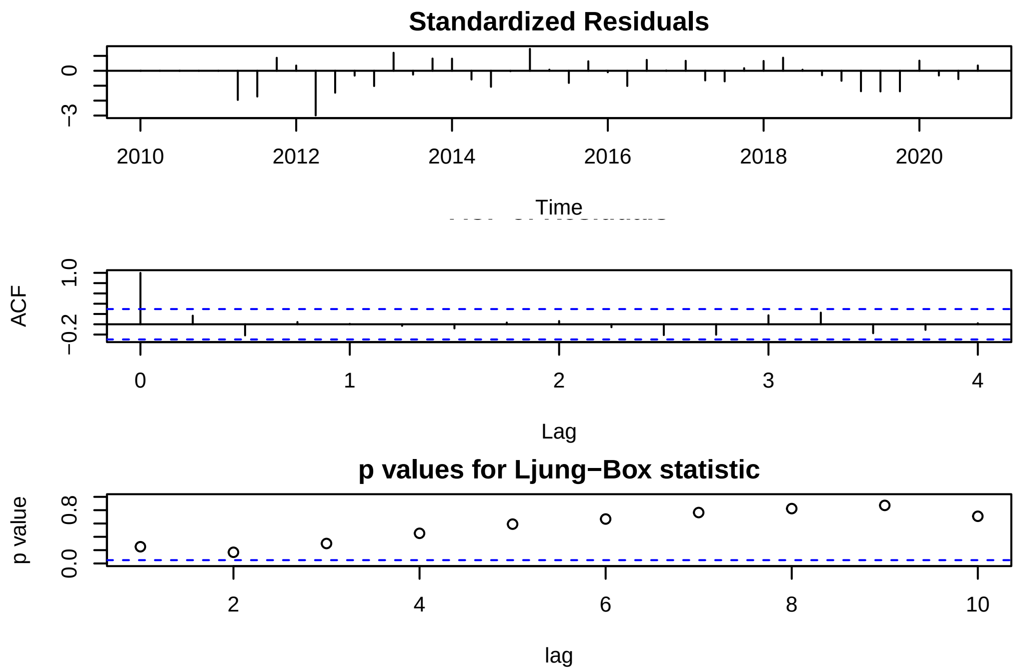
```r
dim(amocForecast$f)
```

```
[1] 4 1
```

```
sqrtR = sapply(amocForecast$R, function(x) sqrt(x[1, 1]))
pl = amocForecast$a[, 1] + qnorm(0.025, sd = sqrtR)
pu = amocForecast$a[, 1] + qnorm(0.975, sd = sqrtR)
x = ts.union(window(tsAtlantic, start = c(2010, 1)), amocForecast$a[, 1],
    amocForecast$f, pl, pu)
par(mar = c(4, 4, 2, 2))
plot(x, plot.type = "single", type = "o", pch = c(1, 20, 3, NA, NA), col = c("darkgrey",
    "brown", "brown", "blue", "blue"), ylab = "Log gas consumption")

legend("bottomright", legend = c("Observed", "Forecast", "95% interval"),
    bty = "n", pch = c(1, 20, NA), lty = 1, col = c("darkgrey", "brown",
        "blue"))
```




```
# Set smaller margins
par(mar = c(4, 4, 2, 2))

tsdiag(amocPredict)
```

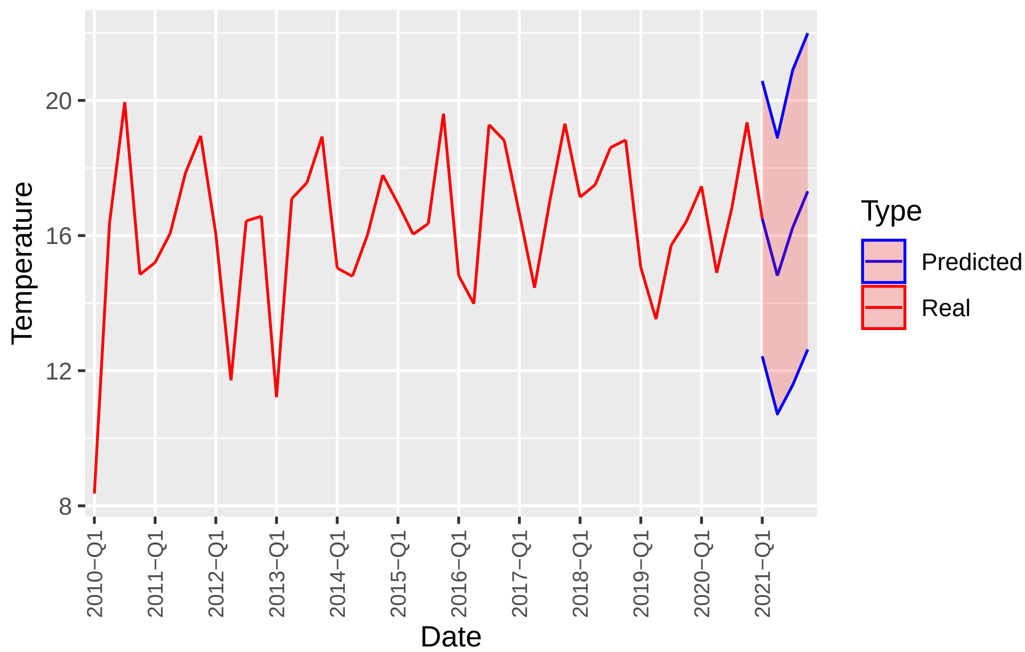



```
# Reset margins
par(mar = c(5, 4, 4, 2) + 0.1)
```

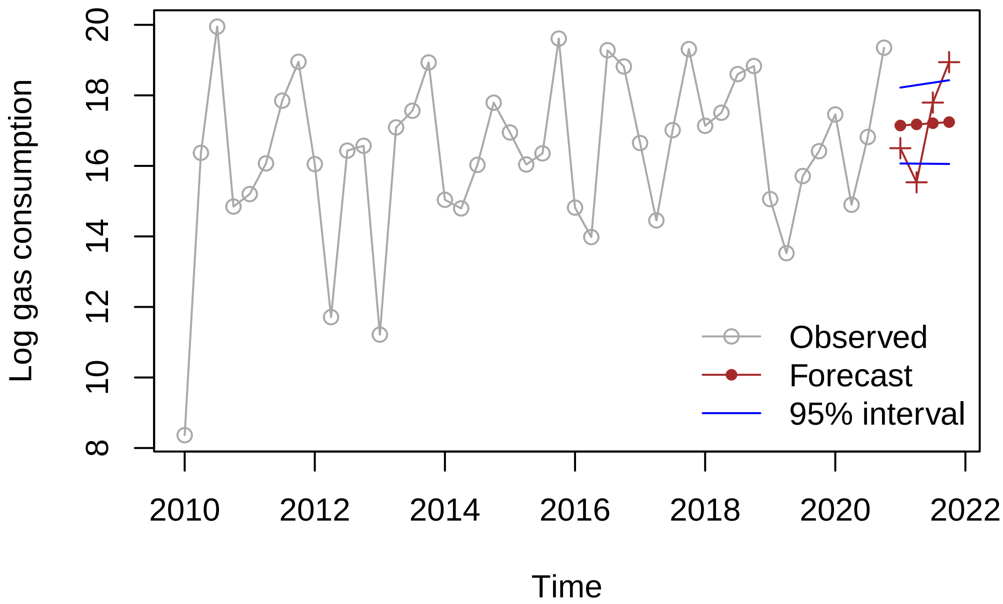
2 d)

Again comparing the forecast values and their respective prediction intervals as we can see from the graphs bellow the dlm model has smaller prediction intervals, most likely due to being able to explain the underlying seasonal trend reducing therefore the uncertainty in comparison the ARIMA model.

```
print(plotARIMA)
```




```
sqaretRoot = sapply(amocForecast$R, function(x) sqrt(x[1, 1]))
predictionLow = amocForecast$a[, 1] + qnorm(0.025, sd = sqaretRoot)  ## Low
predictionUpper = amocForecast$a[, 1] + qnorm(0.975, sd = sqaretRoot)  ## Upper
x = ts.union(window(tsAtlantic, start = c(2010, 1)), amocForecast$a[, 1],
    amocForecast$f, predictionLow, predictionUpper)
par(mar = c(4, 4, 2, 2))
plot(x, plot.type = "single", type = "o", pch = c(1, 20, 3, NA, NA), col = c("darkgrey",
    "brown", "brown", "blue", "blue"), ylab = "Log gas consumption")

legend("bottomright", legend = c("Observed", "Forecast", "95% interval"),
    bty = "n", pch = c(1, 20, NA), lty = 1, col = c("darkgrey", "brown",
        "blue"))
```



2 e)

```
# AMOCDFMonthly =AMOCDF %>% mutate(YearMonth = paste0(year(Date), '-',
# month(Date, label = TRUE, abbr = FALSE)))
```

Question 3

Question 3 a)

I will start with the time series analysis of the temperature in California

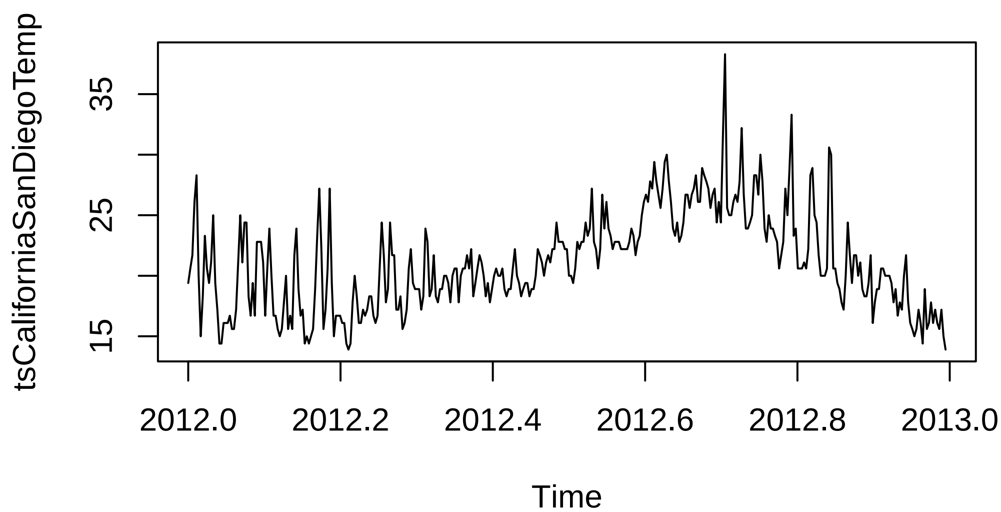
other approach see max temp in the entire state with 8 cities

TODO WARNING

For a dataset of daily data with only 1 year of a cycle data available a daily frequency won't be a very good fit because we only have one observation per cycle, we need a hidden entry to capture the 12 months instead

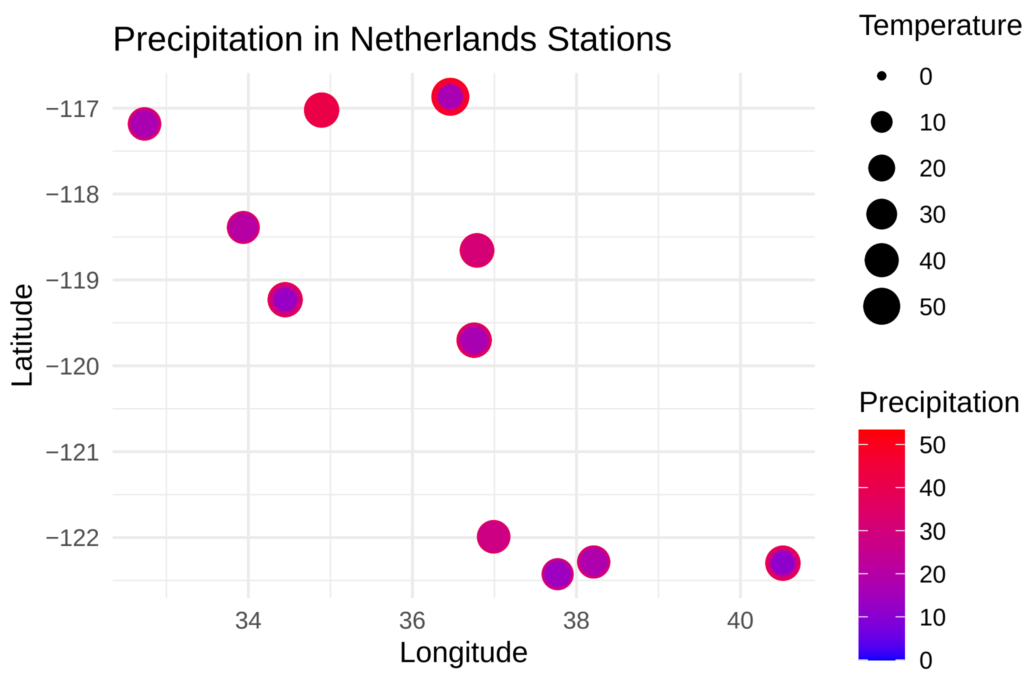
```r
californiaTempDF$Date = as.Date(as.character(californiaTempDF$Date), format = "%Y%m%d",
    origin = "1970-01-01")

## its better to just get a time series object for each city and plot
## each of those in the same plot

tsCaliforniaSanDiegoTemp = ts(californiaTempDF$`San Diego`, start = c(2012,
    1), frequency = 366)
```

```r
plot.ts(tsCaliforniaSanDiegoTemp)
```




```r
californiaLongTempDF = pivot_longer(californiaTempDF, cols = -Date, names_to = "Location",
    values_to = "Temperature")

spatialTemperatureCaliforniaDF = merge(californiaLongTempDF, californiaSpatialDataDF)


ggplot(data = spatialTemperatureCaliforniaDF) + geom_point(aes(x = Lat, y = Long,
    color = Temperature, size = Temperature)) + scale_color_continuous(low = "blue",
    high = "red") + labs(title = "Precipitation in Netherlands Stations",
```

```
    x = "Longitude", y = "Latitude", color = "Precipitation") + theme_minimal()
```



3 b)

```
geoDataCalifornia = as.geodata(spatialTemperatureCaliforniaDF, coords.col = 4:5,
    data.col = "Temperature", covar.col = "Elev")
```

```
as.geodata: 4004 replicated data locations found.
 Consider using jitterDupCoords() for jittering replicated locations.
WARNING: there are data at coincident or very closed locations, some of the geoR's functions
 Use function dup.coords() to locate duplicated coordinates.
 Consider using jitterDupCoords() for jittering replicated locations
```

```
variogramCalifornia = variog(geoDataCalifornia)
```

```
variog: computing omnidirectional variogram
variog: co-locatted data found, adding one bin at the origin
```

```r
plot(variogramCalifornia)
```