oospore_modeling.R

f80872088

2025-07-30

```
require(lubridate)
```

```
## Loading required package: lubridate
```

```
##
## Attaching package: 'lubridate'
```

```
## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```
require(lme4)
```

```
## Loading required package: lme4
```

```
## Loading required package: Matrix
```

```
require(broom.mixed)
```

```
## Loading required package: broom.mixed
```

```
require(effects)
```

```
## Loading required package: effects
```

```
## Loading required package: carData
```

```
## lattice theme set by effectsTheme()
## See ?effectsTheme for details.
```

```
require(dplyr)
```

```
## Loading required package: dplyr
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
require(FactoMineR)
```

```
## Loading required package: FactoMineR
```

```
require(factoextra)
```

```
## Loading required package: factoextra

## Loading required package: ggplot2

## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
```

```
require(missMDA)
```

```
## Loading required package: missMDA
```

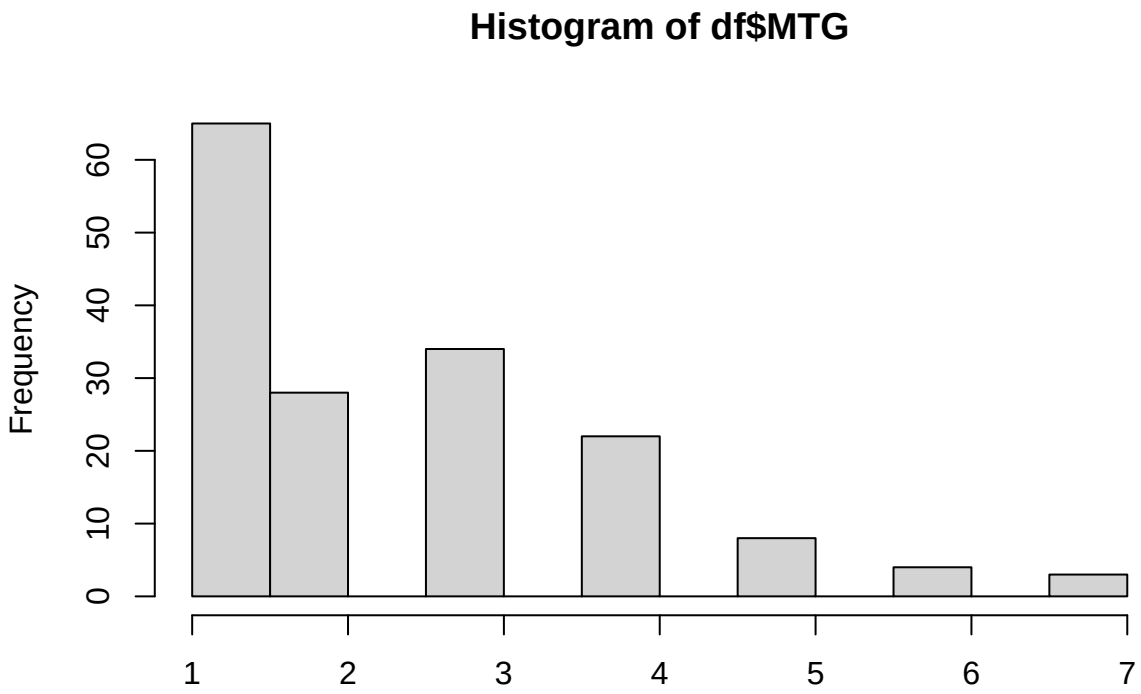
```
require(ggbiplot)
```

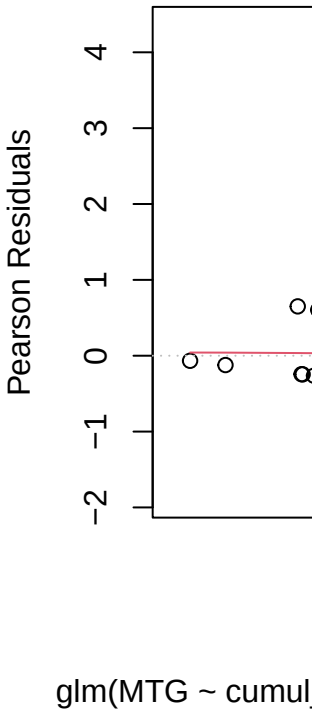
```
## Loading required package: ggbiplot
```

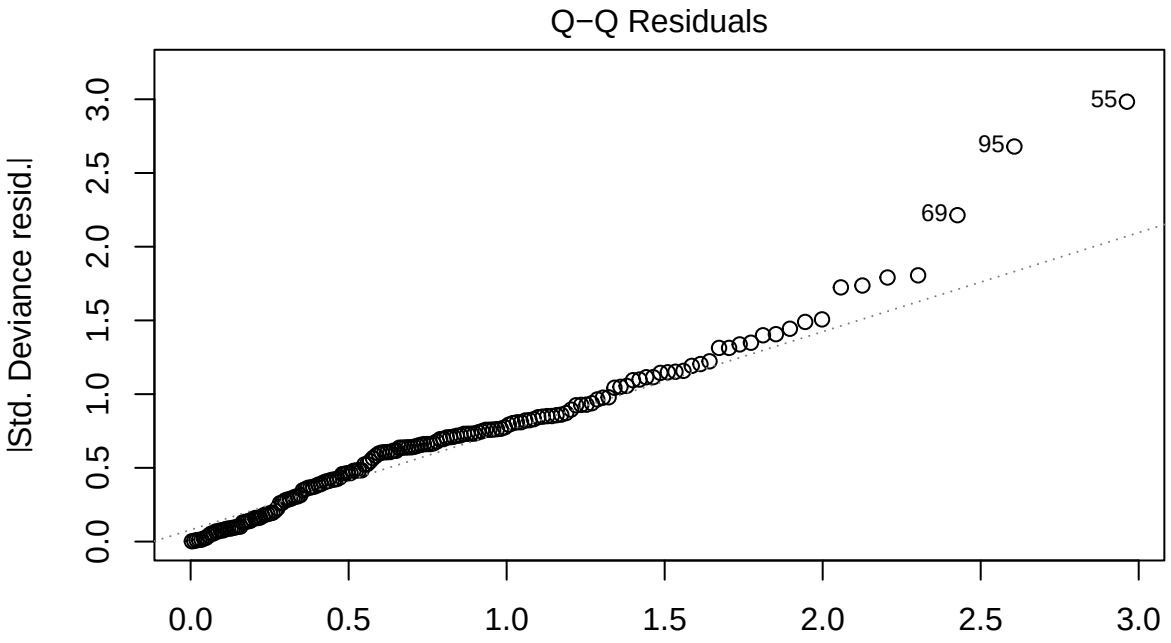
```
## DATASET PROCESSING
setwd("~/mnt/Data-Work-CH/22_Plant_Production-CH/222.6_Mycologie_protected/Projets de recherche/38_SMALA
df_all <- read.table("Oosp_not_all_2003-2024_v9.csv", sep = ";", header = T)
df_all$BBCH <- as.numeric(df_all$BBCH)
df_all$date <- as_datetime(df_all$date, format = "%d.%m.%Y")
df_all$MTG <- as.numeric(df_all$MTG)
df_all$nb_germ_oosp_1d <- as.numeric(df_all$nb_germ_oosp_1d)
df_all$cumul_precipit_1Jan <-  as.numeric(df_all$cumul_precipit_1Jan)
df_all$nb_days_rainfall_30d <-  as.numeric(df_all$nb_days_rainfall_30d)
df_all$solar_radiation_1Jan <- as.numeric(df_all$solar_radiation_1Jan)
df_all$VPD <- as.numeric(df_all$VPD)
df_all$RH <- as.numeric(df_all$RH)
df_all$temp <- as.numeric(df_all$temp)
df_all$TDD <- as.numeric(df_all$TDD)

# solda_radiation variables were ultimately not included in the model variable selection
# because they were strongly correlated with TDD, thus biasing the predictions.
# Also, they included a lot of missing values, thus making TDD a better variable choice.

### PCA FUNCTION
pca <- function(df){
  dataPCA <- cbind(df$cumul_precipit_1Jan, df$nb_days_rainfall_30d, df$VPD,
                   df$RH, df$temp, df$TDD)
  dataPCA <- matrix(as.numeric(unlist(dataPCA)),nrow = nrow(dataPCA))
  colnames(dataPCA) <- (colnames(subset(df, select = c(cumul_precipit_1Jan, nb_days_rainfall_30d, VPD,
                                                       RH, temp, TDD))))
  pca <- prcomp(dataPCA, scale. = T)
  summary(pca)
  pca$rotation
  ## PLOTS
  # specifying MTG categories for PCA groups
  MTG_cat <- df$MTG
  for (i in 1:length(MTG_cat)) {
    if (MTG_cat[i] < 3) {
      MTG_cat[i] <- "1-2"
    }
    if (MTG_cat[i] > 2) {
      MTG_cat[i] <- "3-10"
    }
  }
  p <- ggbiplot(pca, groups = MTG_cat, choices = c(1,2), ellipse = T, ellipse.prob = 0.4) + theme_bw()
  print(p)
}
```

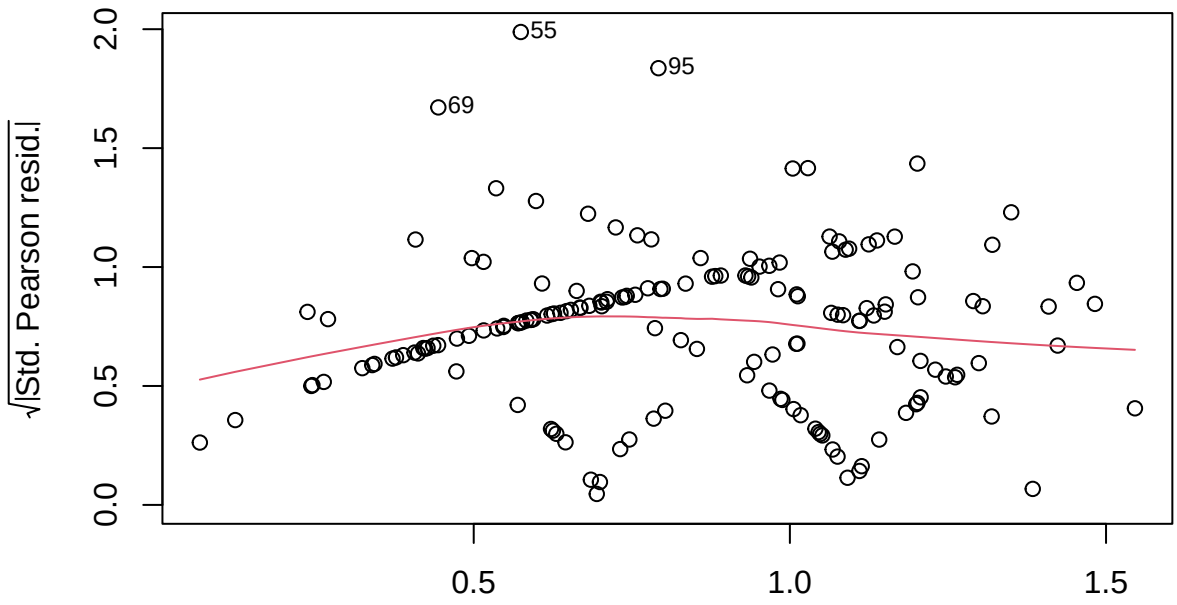
```r
### MODEL FUNCTIONS, NEEDS DATASET AS INPUT.
## the two functions creates distinct models: one with MGT as response variable,
## the other with Nspores as response variable
## they then plot the model partial plots, the QQ-residuals, the table statistics

### Average oospore maturation day
model_MGT <- function(df){
  MGT_model <- glm(data = df, formula =  MTG ~ cumul_precipit_1Jan + nb_days_rainfall_30d +
                     + VPD + RH + temp + TDD, family = "poisson")

  # SHOWING DISTRIBUTION OF MAIN RESPONSE VARIABLES OF INTEREST
  hist(df$MTG)

  # MODEL INFO AND PARTIAL EFFECTS PLOTS
  plot(MGT_model)
  plot(allEffects(MGT_model))

  # MODEL STATISTICS TABLES
  tidy(MGT_model)
  # glance(MGT_model)
}

### Number of spores 1 day after first germination
model_Nspores1d <- function(df){
  Nspores_model <- glm(data = df, formula =  nb_germ_oosp_1d ~ cumul_precipit_1Jan + nb_days_rainfall_3
                         VPD + RH + temp + TDD, family = "poisson")

  # SHOWING DISTRIBUTION OF MAIN RESPONSE VARIABLES OF INTEREST
  hist(df$nb_germ_oosp_1d)

  # MODEL INFO AND PARTIAL EFFECTS PLOTS
  plot(Nspores_model)
  plot(allEffects(Nspores_model))

  # MODEL STATISTICS TABLES
  tidy(Nspores_model)
  # glance(Nspores_model)
}

### Number of spores 10 days after first germination
model_Nspores10d <- function(df){
  Nspores_model <- glm(data = df, formula =  nb_germ_oosp_10d ~ cumul_precipit_1Jan + nb_days_rainfall_
                         VPD + RH + temp + TDD, family = "poisson")

  # SHOWING DISTRIBUTION OF MAIN RESPONSE VARIABLES OF INTEREST
  hist(df$nb_germ_oosp_10d)

  # MODEL INFO AND PARTIAL EFFECTS PLOTS
  plot(Nspores_model)
  plot(allEffects(Nspores_model))

  # MODEL STATISTICS TABLES
```

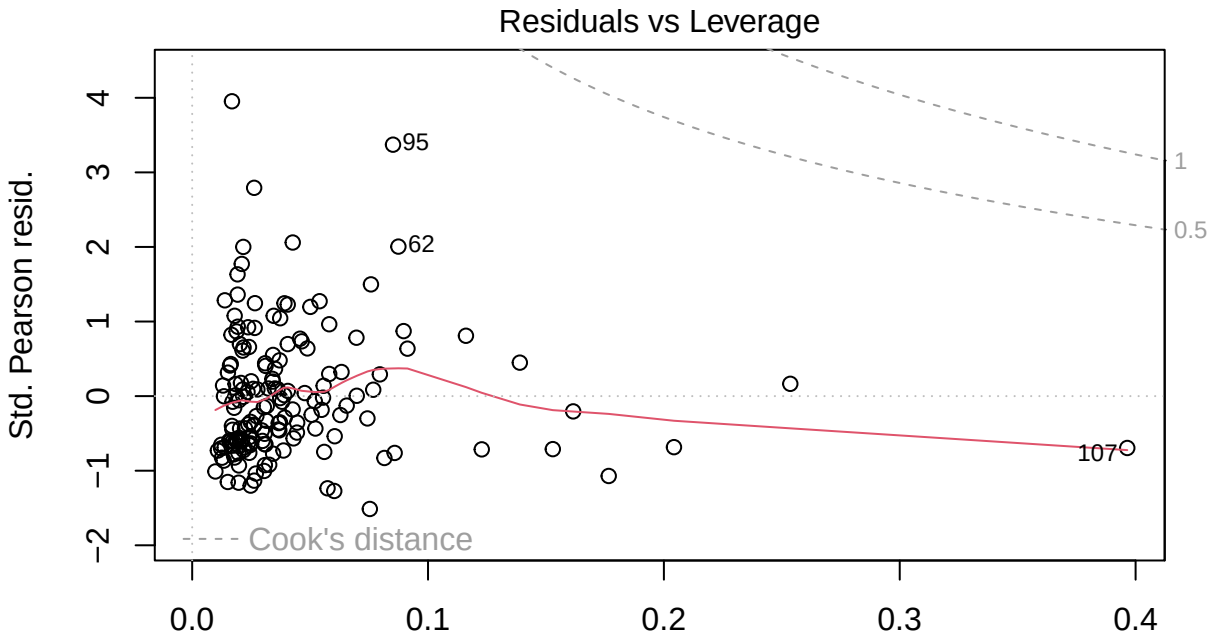
```
  tidy(Nspores_model)
  # glance(Nspores_model)
}

## ALL BBCH DATASET
model_MGT(df_all)
```





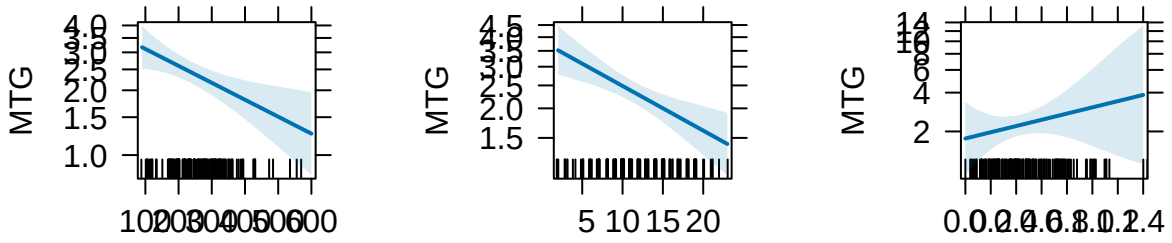

Q−Q Residuals
|Std. Deviance resid.|
55
95
69
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Theoretical Quantiles
glm(MTG ~ cumul_precipit_1Jan + nb_days_rainfall_30d + +VPD + RH + temp + T .


Scale−Location
√|Std. Pearson resid.|
55
95
69
0.0
0.5
1.0
1.5
2.0
Predicted values
glm(MTG ~ cumul_precipit_1Jan + nb_days_rainfall_30d + +VPD + RH + temp + T .

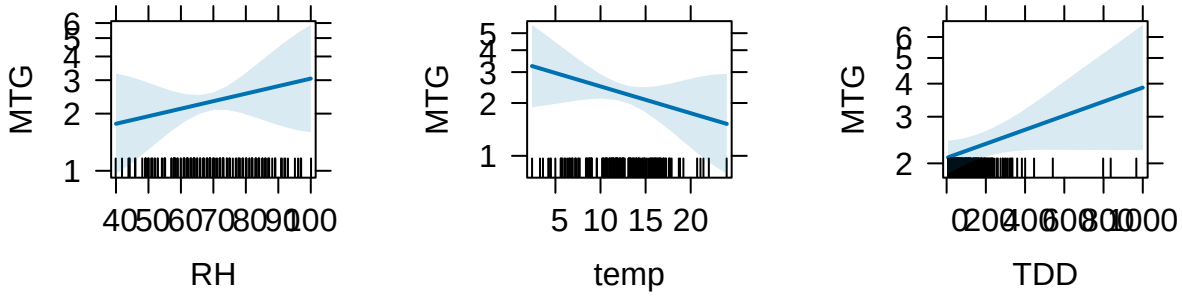


glm(MTG ~ cumul_precipit_1Jan + nb_days_rainfall_30d + +VPD + RH + temp + T .

ul_precipit_1Jan effect plot **nb_days_rainfall_30d effect plot** **VPD effect plot**



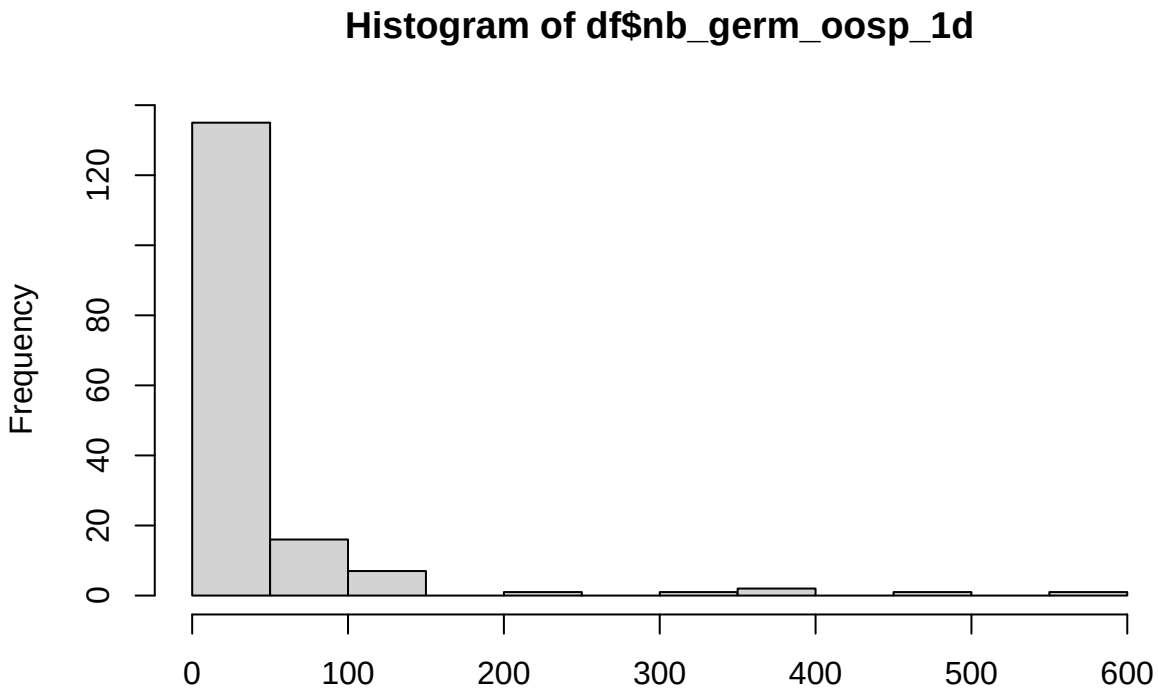
RH effect plot **temp effect plot** **TDD effect plot**

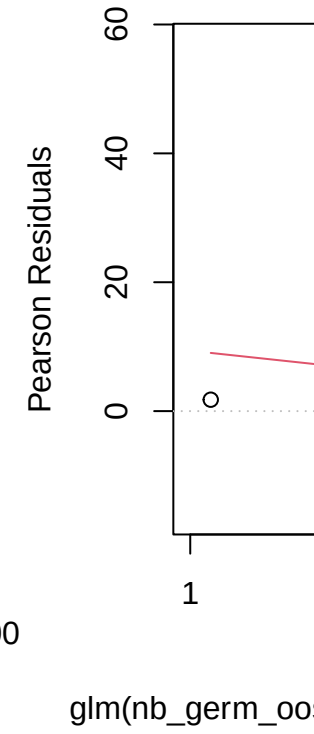


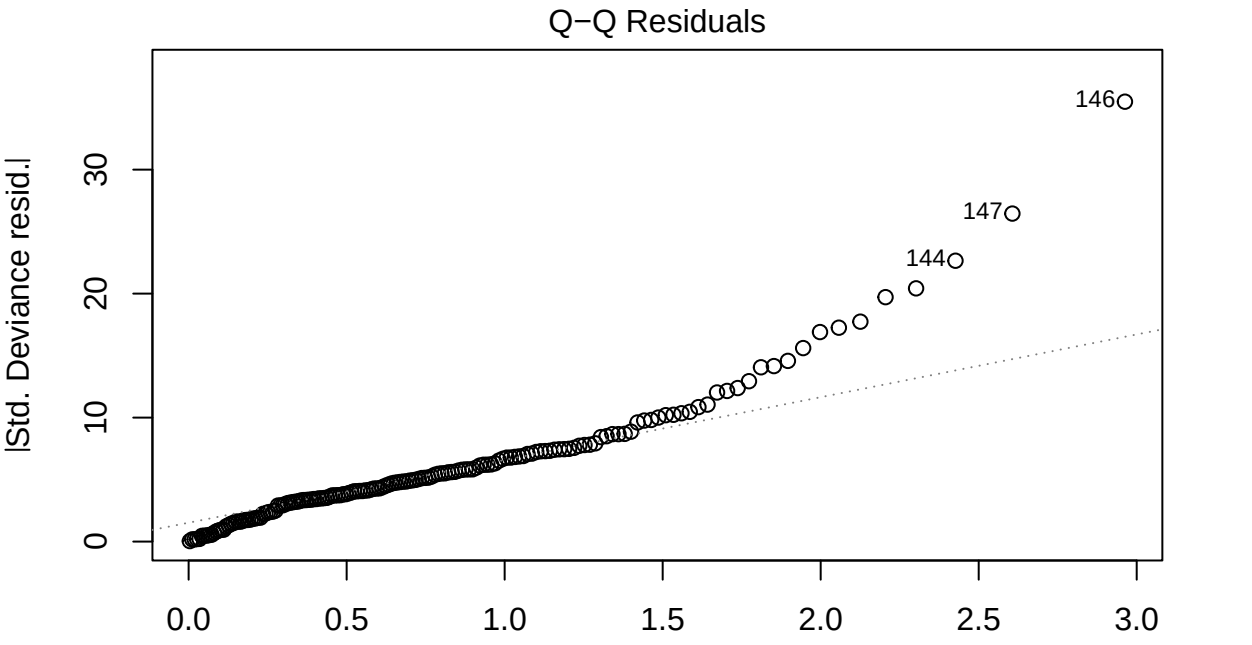

```
## # A tibble: 7 x 5
##   term                estimate std.error statistic  p.value
##   <chr>                  <dbl>     <dbl>     <dbl>    <dbl>
## 1 (Intercept)           1.27      0.813        1.57  0.117
```

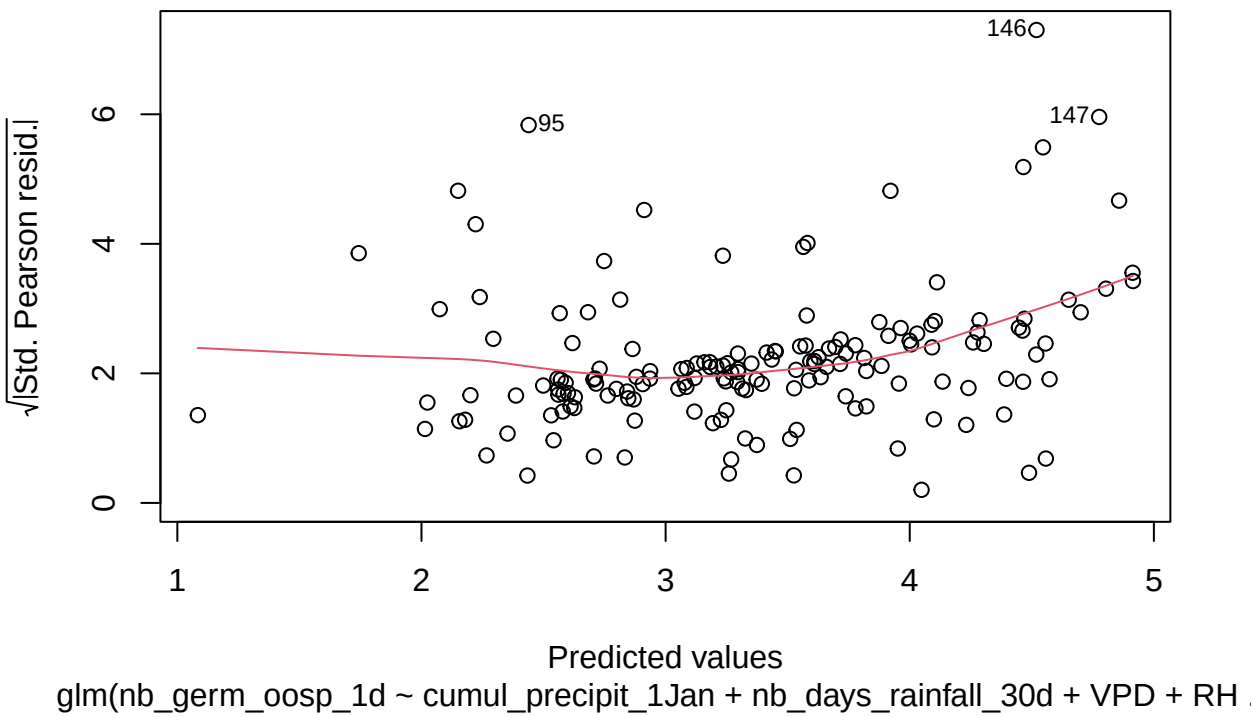
```
## 2 cumul_precipit_1Jan  -0.00181   0.000635    -2.85  0.00431
## 3 nb_days_rainfall_30d -0.0434    0.0121      -3.59  0.000335
## 4 VPD                   0.556     0.682        0.815 0.415
## 5 RH                    0.00913   0.0106       0.864 0.388
## 6 temp                 -0.0351    0.0278      -1.26  0.207
## 7 TDD                   0.000611  0.000327     1.87  0.0616
```

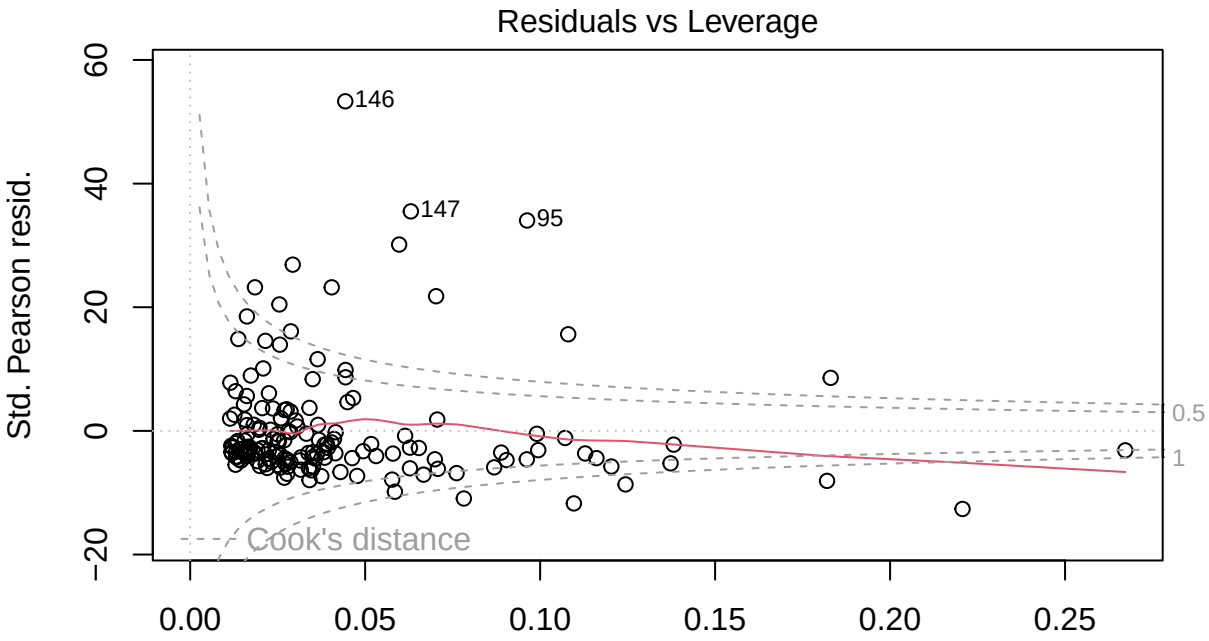
```
model_Nspores1d(df_all)
```

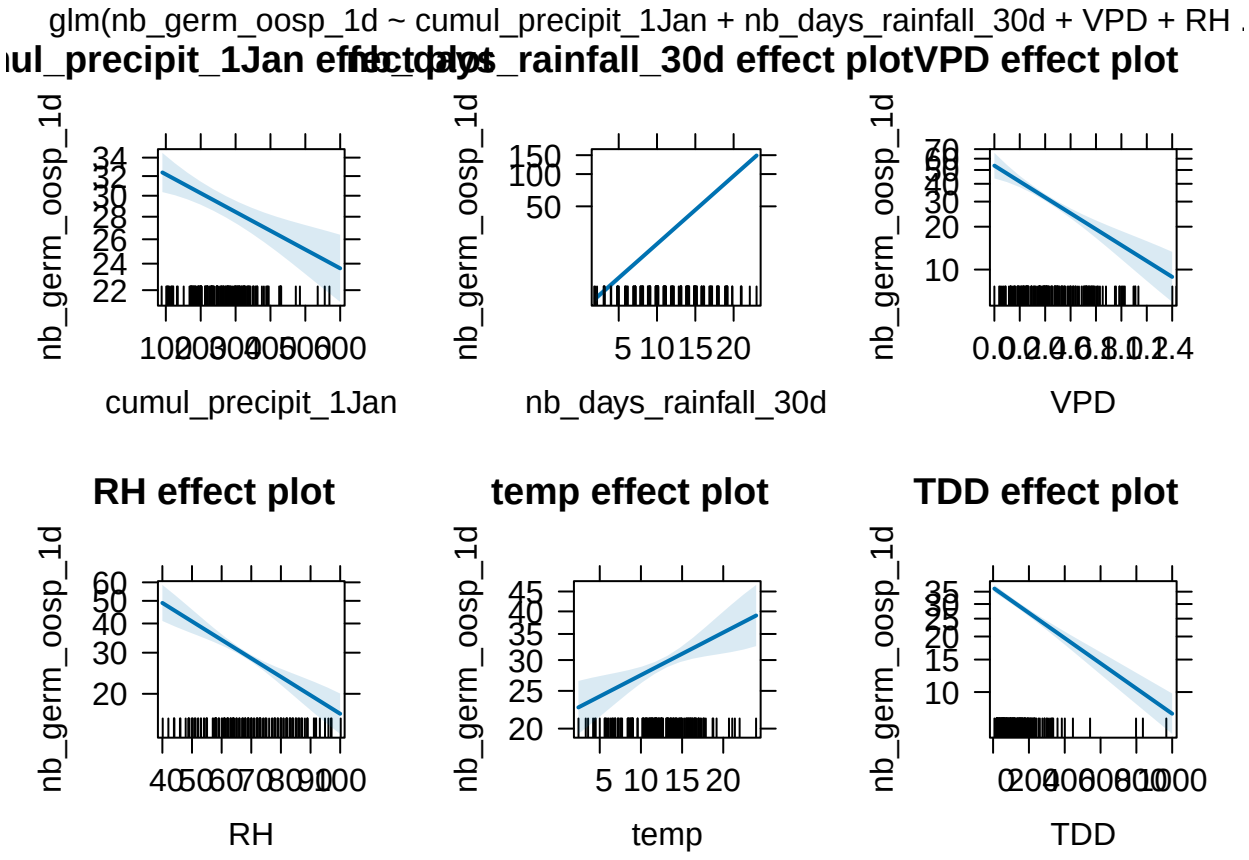





Q−Q Residuals
|Std. Deviance resid.|
146
147
144
0
10
20
30
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Theoretical Quantiles
glm(nb_germ_oosp_1d ~ cumul_precipit_1Jan + nb_days_rainfall_30d + VPD + RH


Scale−Location
√|Std. Pearson resid.|
146
147
95
0
2
4
6
1
2
3
4
5
Predicted values
glm(nb_germ_oosp_1d ~ cumul_precipit_1Jan + nb_days_rainfall_30d + VPD + RH

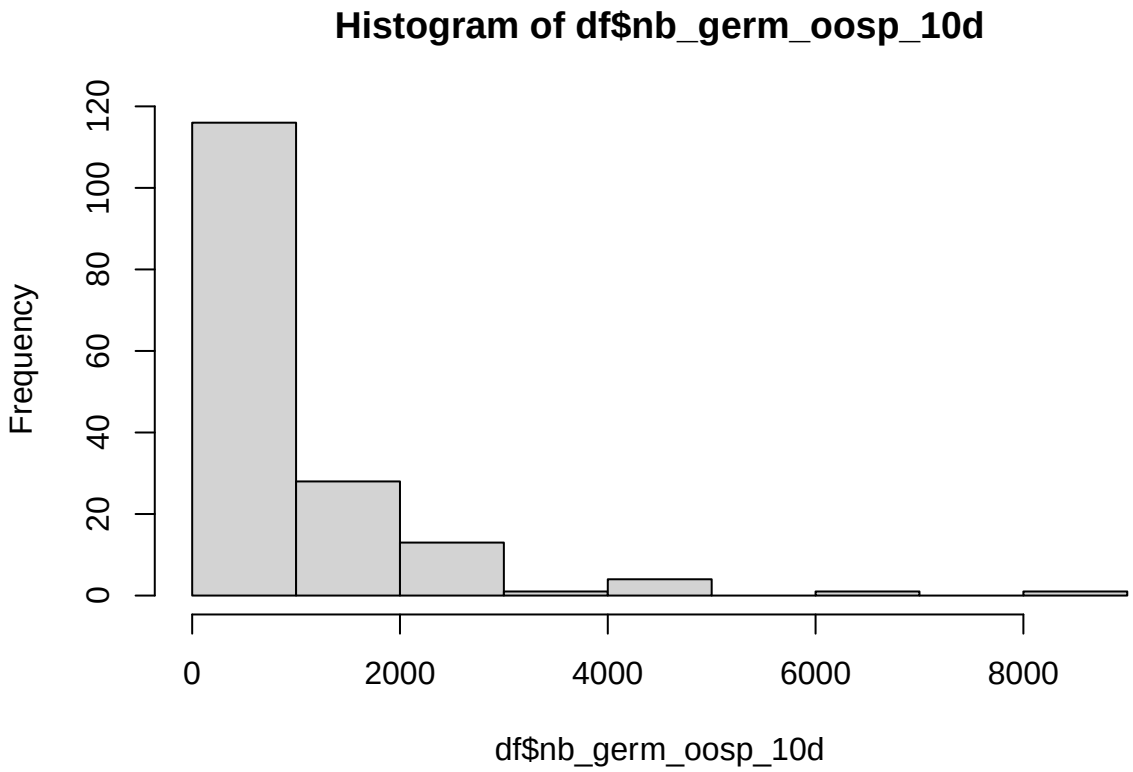


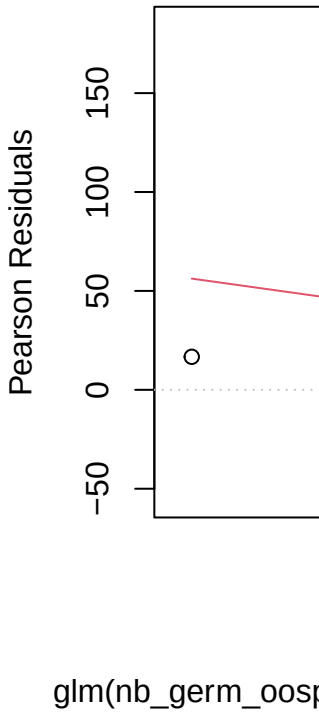


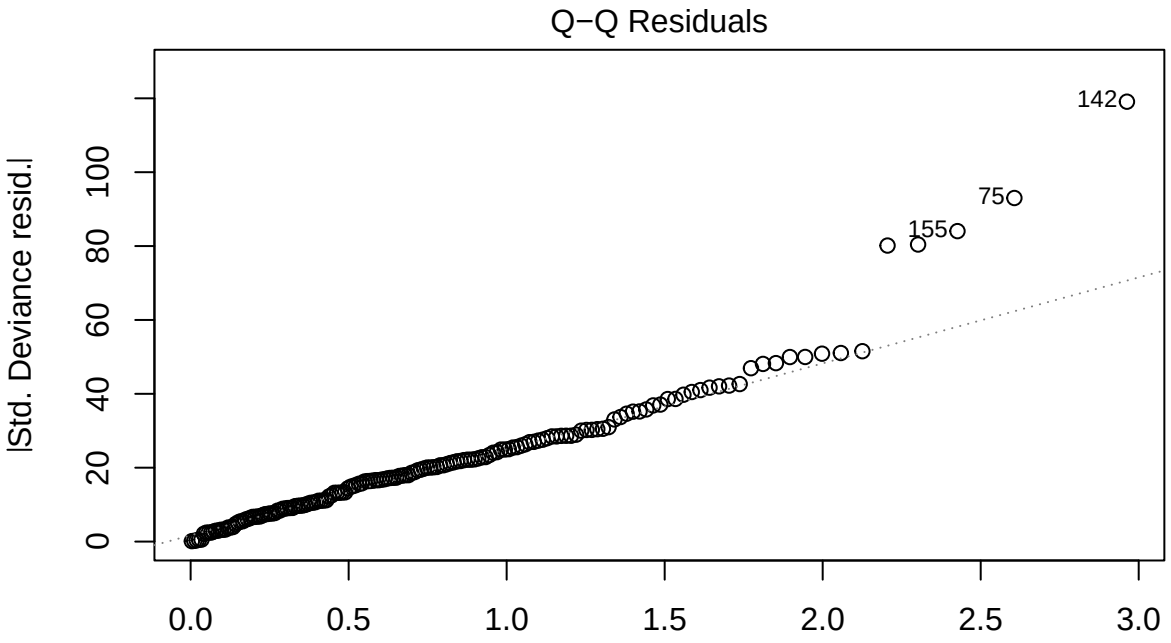

```
## # A tibble: 7 x 5
##   term                estimate std.error statistic  p.value
##   <chr>                  <dbl>     <dbl>     <dbl>    <dbl>
## 1 (Intercept)            3.60      0.255      14.1  2.87e-45
```

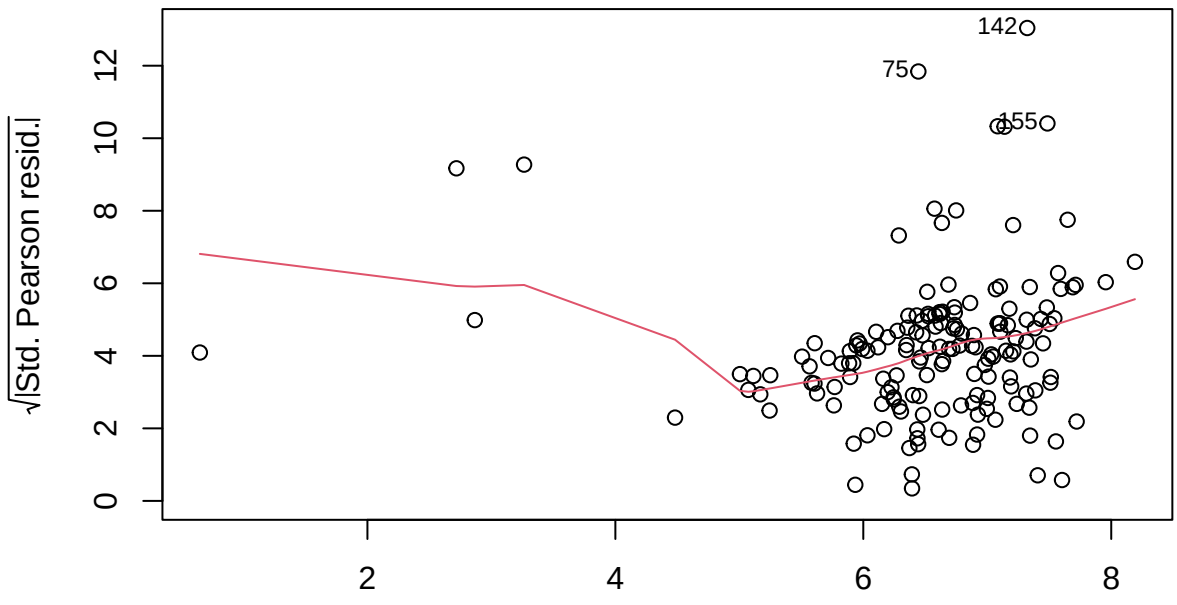
```
## 2 cumul_precipit_1Jan  -0.000617  0.000162     -3.81 1.40e- 4
## 3 nb_days_rainfall_30d  0.146     0.00301      48.5  0
## 4 VPD                  -1.28      0.221        -5.80 6.69e- 9
## 5 RH                   -0.0182    0.00316      -5.76 8.62e- 9
## 6 temp                  0.0252    0.00785       3.21 1.34e- 3
## 7 TDD                  -0.00157   0.000147    -10.7  8.59e-27
```

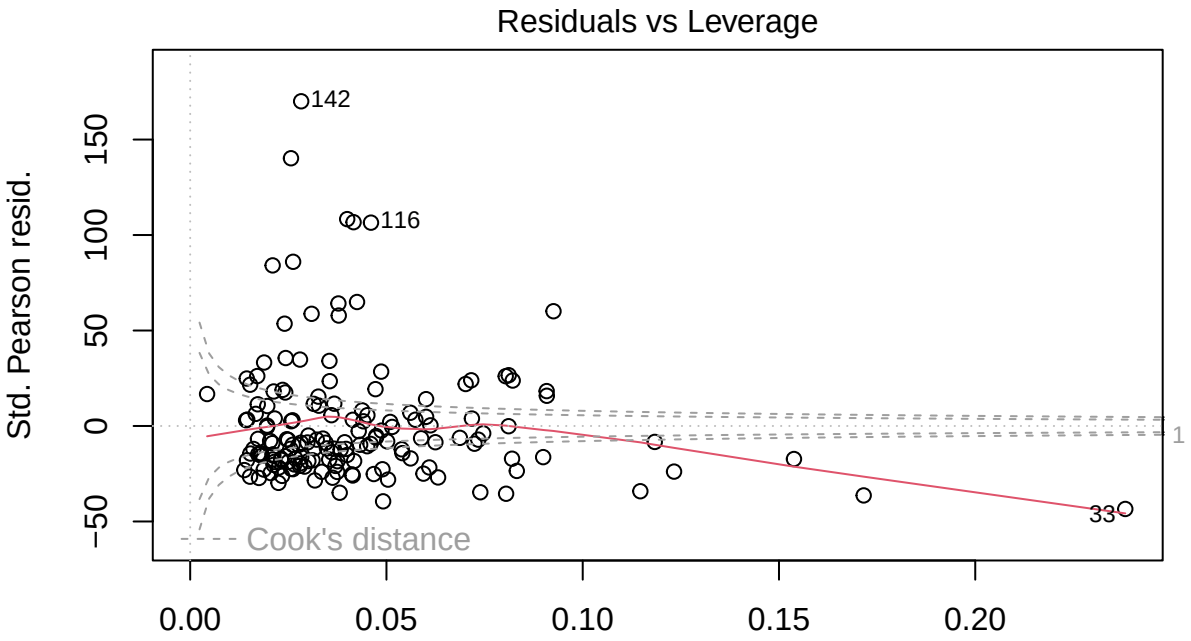
```
model_Nspores10d(df_all)
```



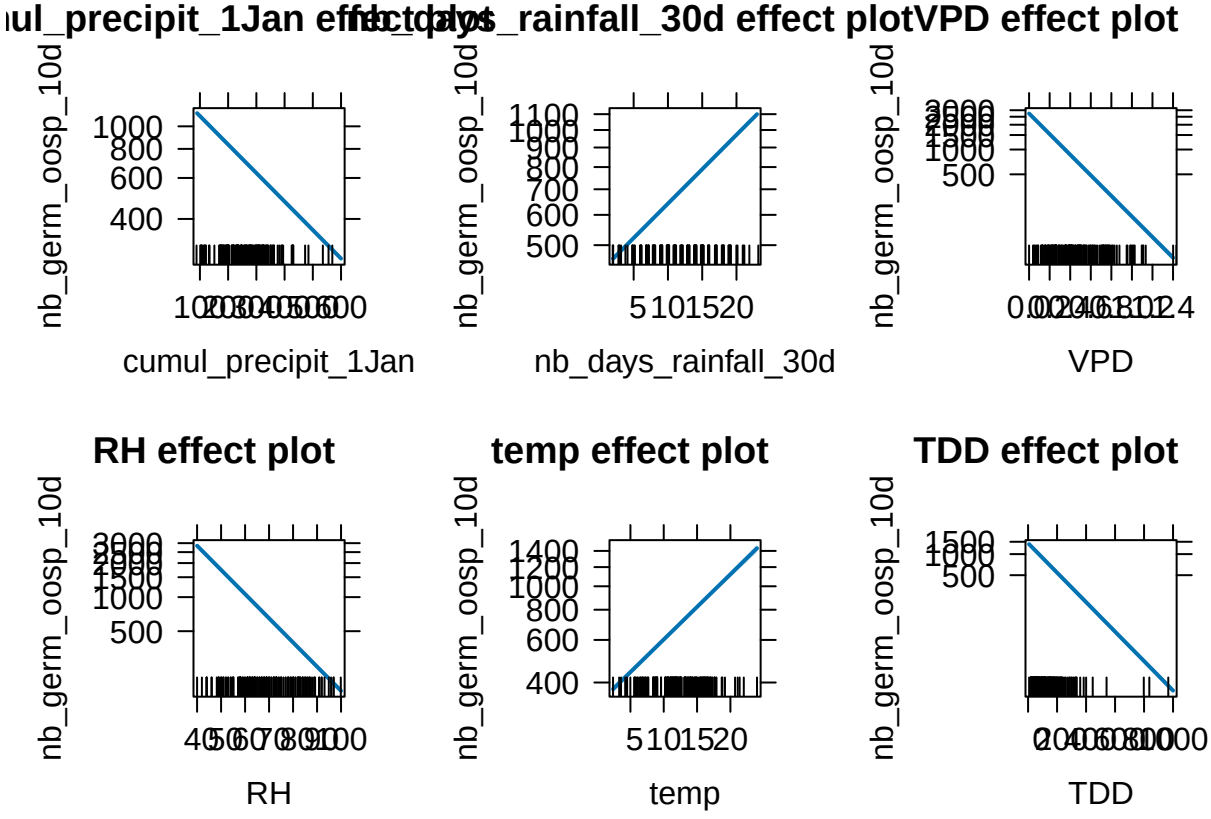



Q−Q Residuals
|Std. Deviance resid.|
0
20
40
60
80
100
142
75
155
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Theoretical Quantiles
glm(nb_germ_oosp_10d ~ cumul_precipit_1Jan + nb_days_rainfall_30d + VPD + R .


Scale−Location
√|Std. Pearson resid.|
0
2
4
6
8
10
12
142
75
155
2
4
6
8
Predicted values
glm(nb_germ_oosp_10d ~ cumul_precipit_1Jan + nb_days_rainfall_30d + VPD + R .



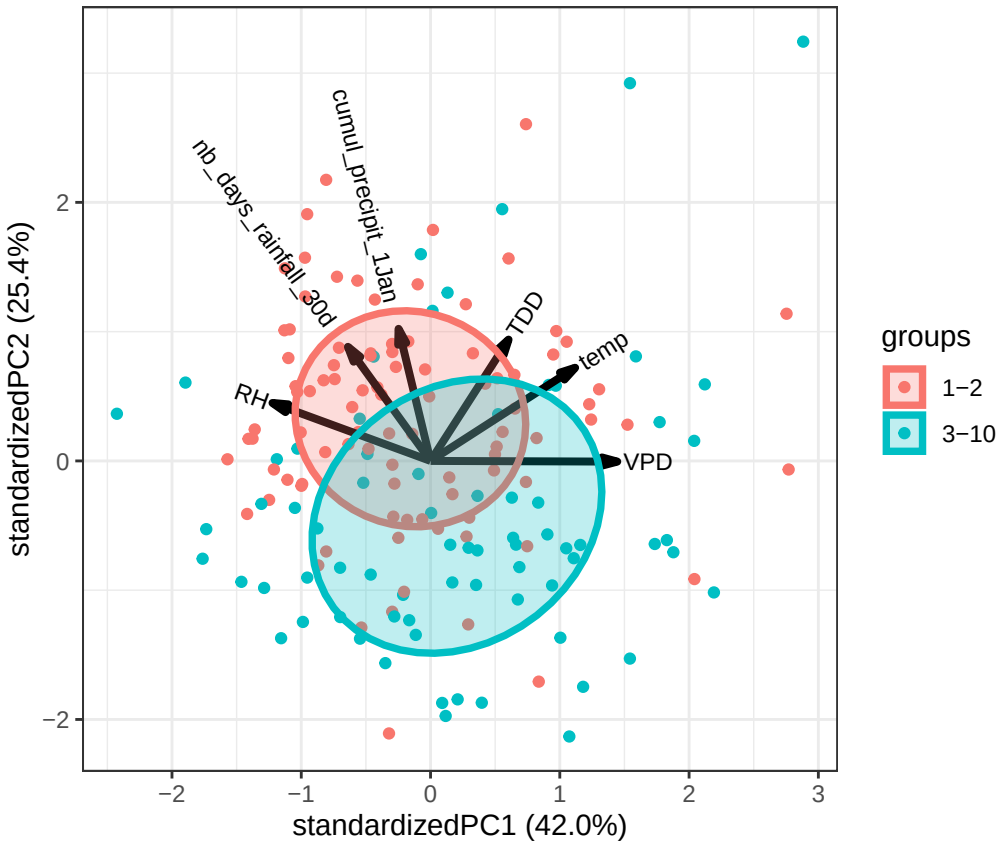
glm(nb_germ_oosp_10d ~ cumul_precipit_1Jan + nb_days_rainfall_30d + VPD + R .



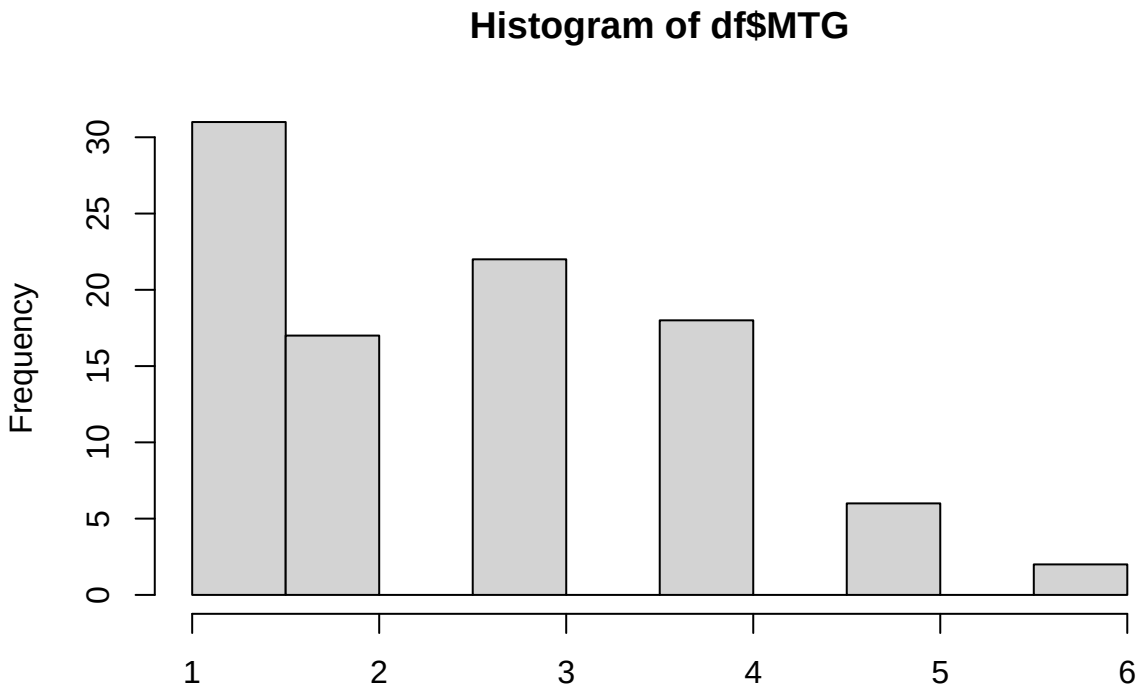

```
## # A tibble: 7 x 5
##   term                estimate std.error statistic   p.value
##   <chr>                  <dbl>     <dbl>     <dbl>     <dbl>
## 1 (Intercept)          11.5     0.0519        222.  0
```

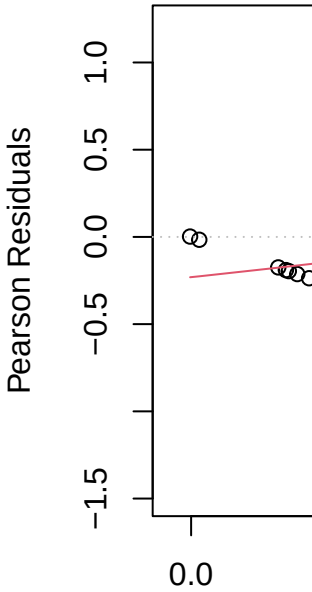
```
## 2 cumul_precipit_1Jan  -0.00282 0.0000339     -83.2 0
## 3 nb_days_rainfall_30d  0.0414  0.000595       69.6 0
## 4 VPD                  -2.88    0.0499        -57.6 0
## 5 RH                   -0.0492  0.000677      -72.8 0
## 6 temp                  0.0621  0.00183        33.9 1.70e-251
## 7 TDD                  -0.00485 0.0000444    -109.  0
```

```
pca(df_all)
```

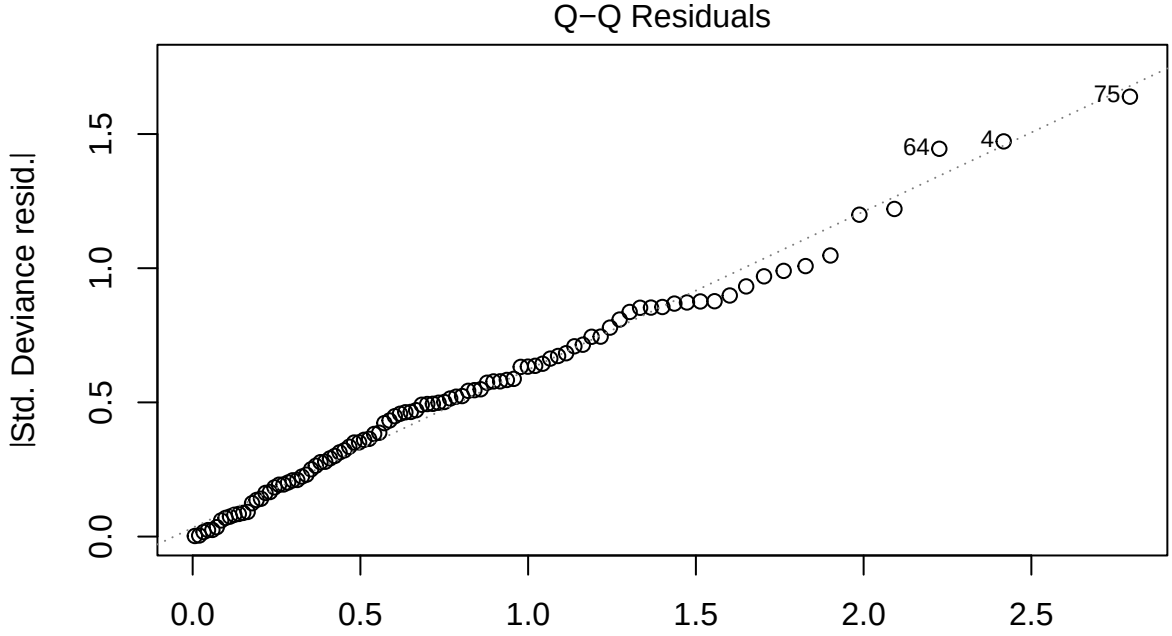



```
## DATASET BBCH 0:12
df <- df_all %>% filter(df_all$BBCH < 13)
model_MGT(df)
```

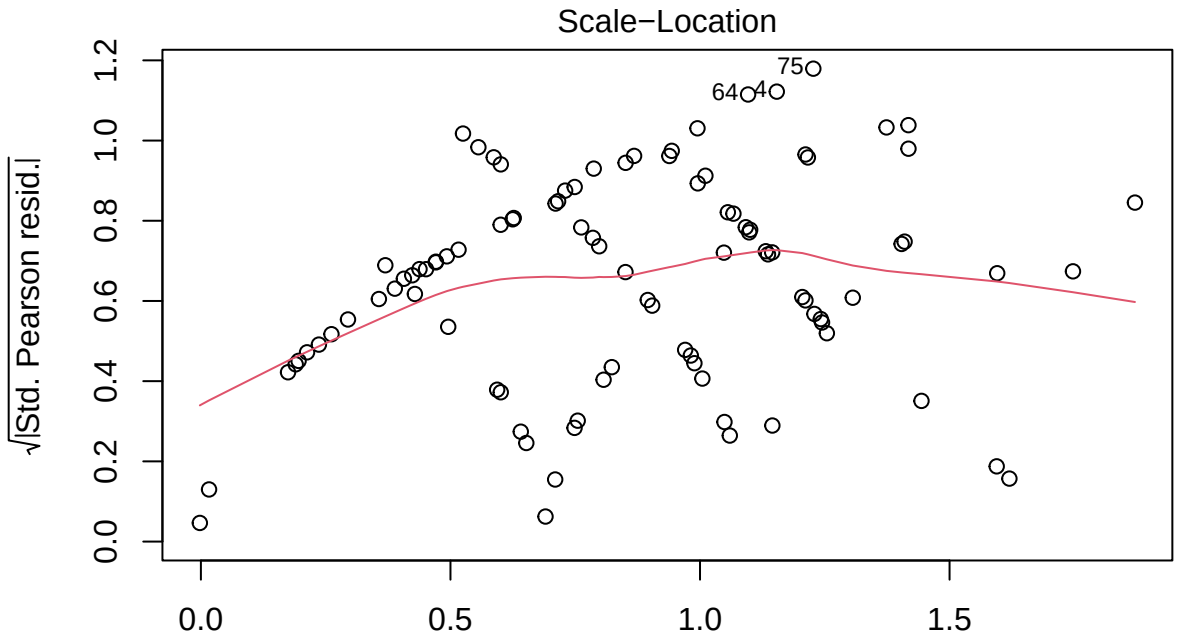


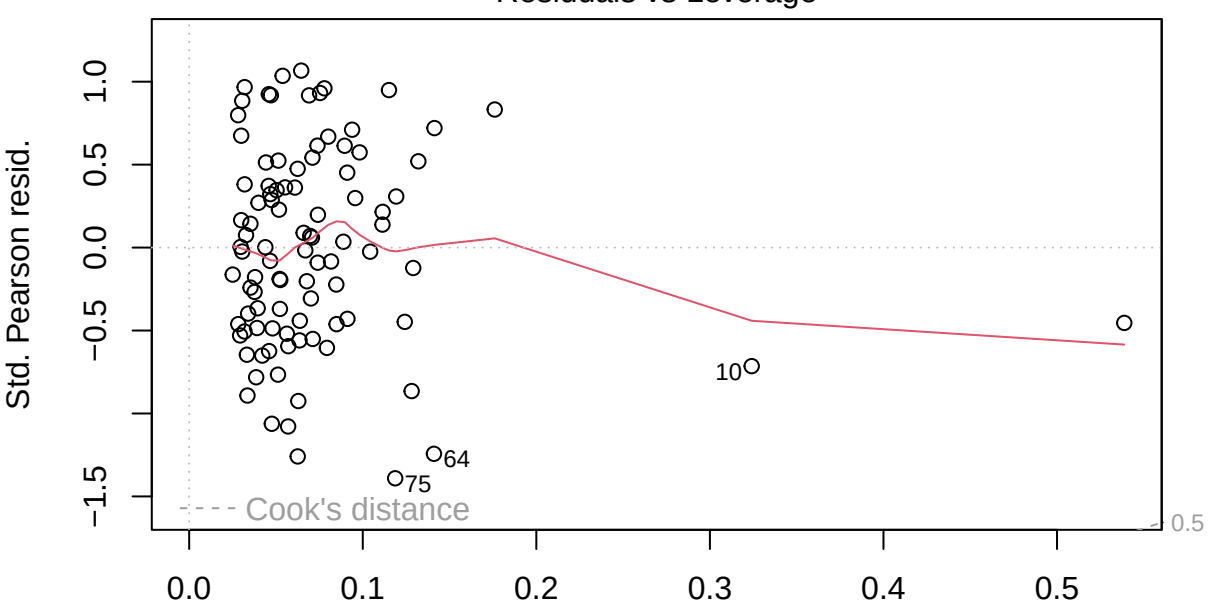


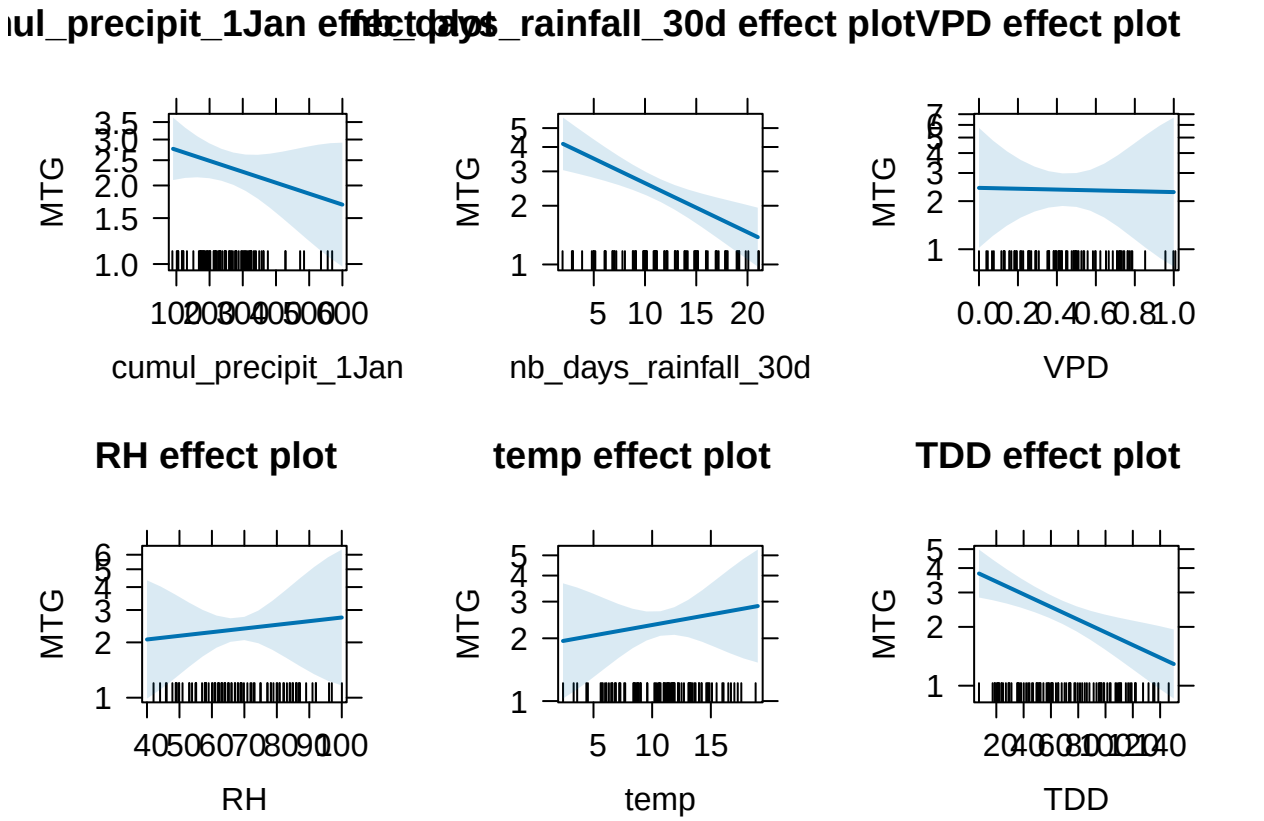
glm(MTG ~ cumul



glm(MTG ~ cumul_precipit_1Jan + nb_days_rainfall_30d + +VPD + RH + temp + T .

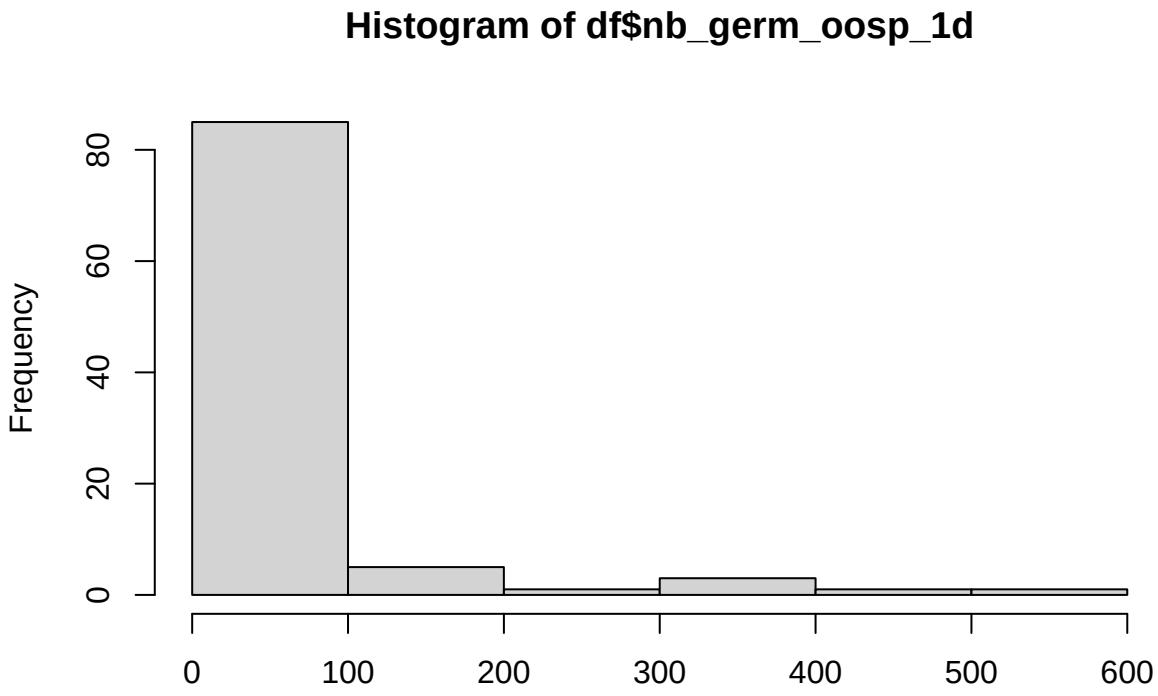

Scale−Location
√|Std. Pearson resid.|
0.0
0.2
0.4
0.6
0.8
1.0
1.2
75
64
4
0.0
0.5
1.0
1.5
Predicted values
glm(MTG ~ cumul_precipit_1Jan + nb_days_rainfall_30d + +VPD + RH + temp + T .

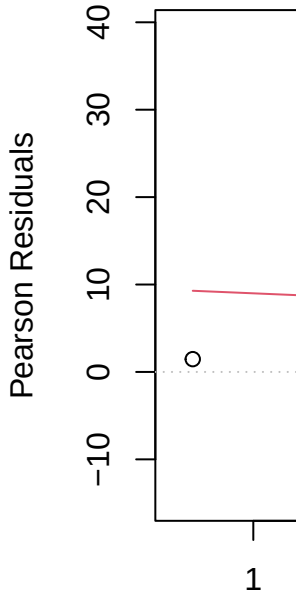

Residuals vs Leverage
Std. Pearson resid.
1.0
0.5
0.0
−0.5
−1.5
10
64
75
Cook's distance
0.5
0.0
0.1
0.2
0.3
0.4
0.5
Leverage
glm(MTG ~ cumul_precipit_1Jan + nb_days_rainfall_30d + +VPD + RH + temp + T .

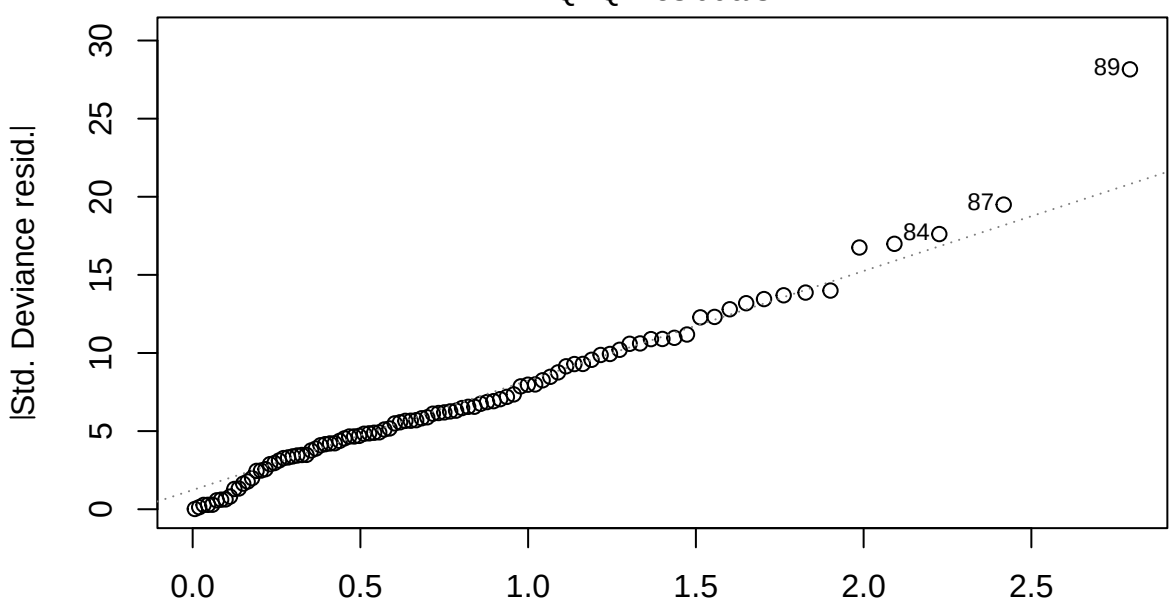


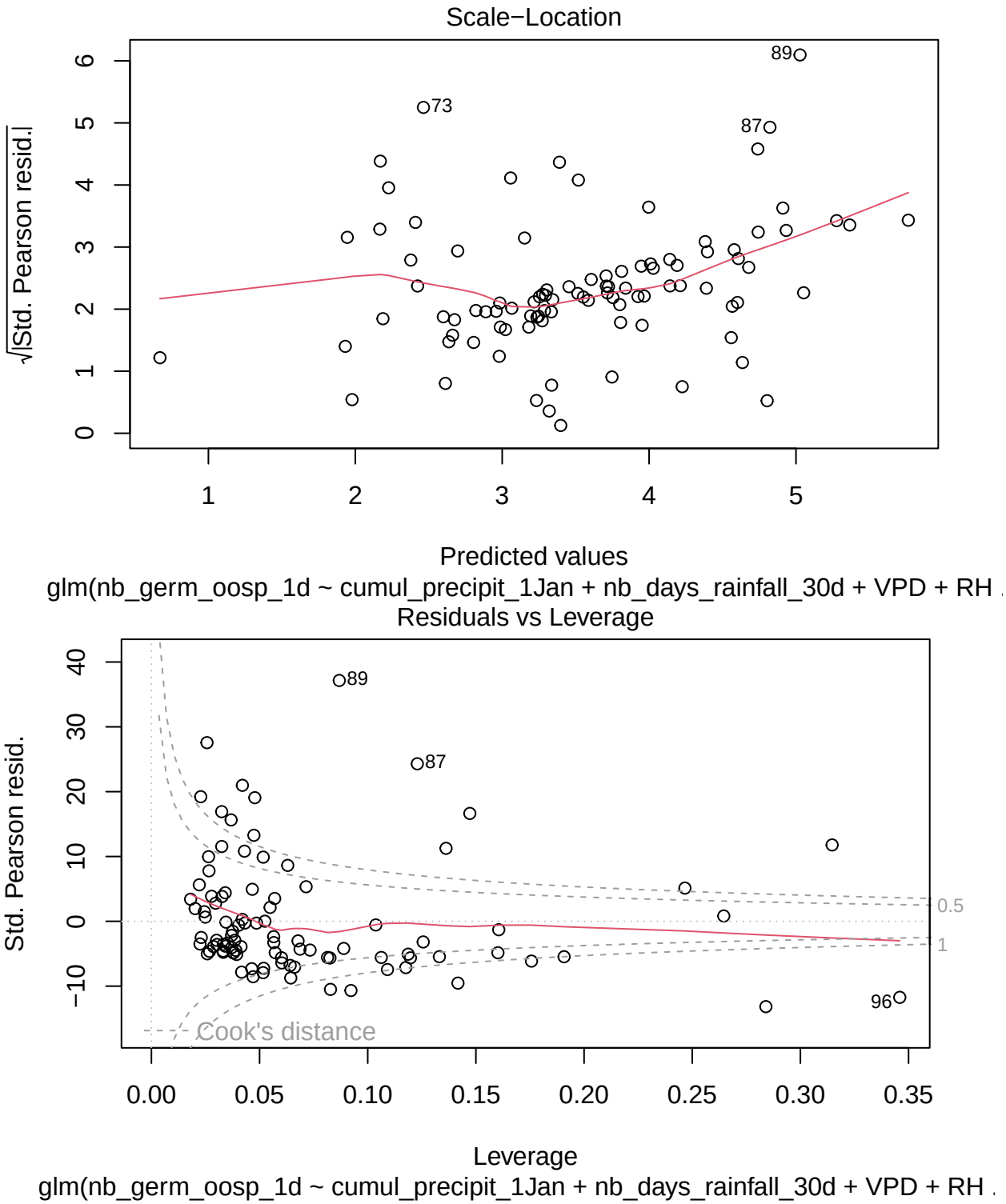

```
## # A tibble: 7 x 5
##   term                  estimate std.error statistic  p.value
##   <chr>                    <dbl>     <dbl>     <dbl>    <dbl>
## 1 (Intercept)            1.77      1.10        1.62   0.106
## 2 cumul_precipit_1Jan   -0.000966  0.000772   -1.25   0.211
## 3 nb_days_rainfall_30d  -0.0579    0.0162     -3.58   0.000348
## 4 VPD                   -0.0607    0.980      -0.0620 0.951
## 5 RH                     0.00459   0.0134      0.344  0.731
## 6 temp                   0.0232    0.0379      0.613  0.540
## 7 TDD                   -0.00747   0.00227    -3.29   0.00101
```

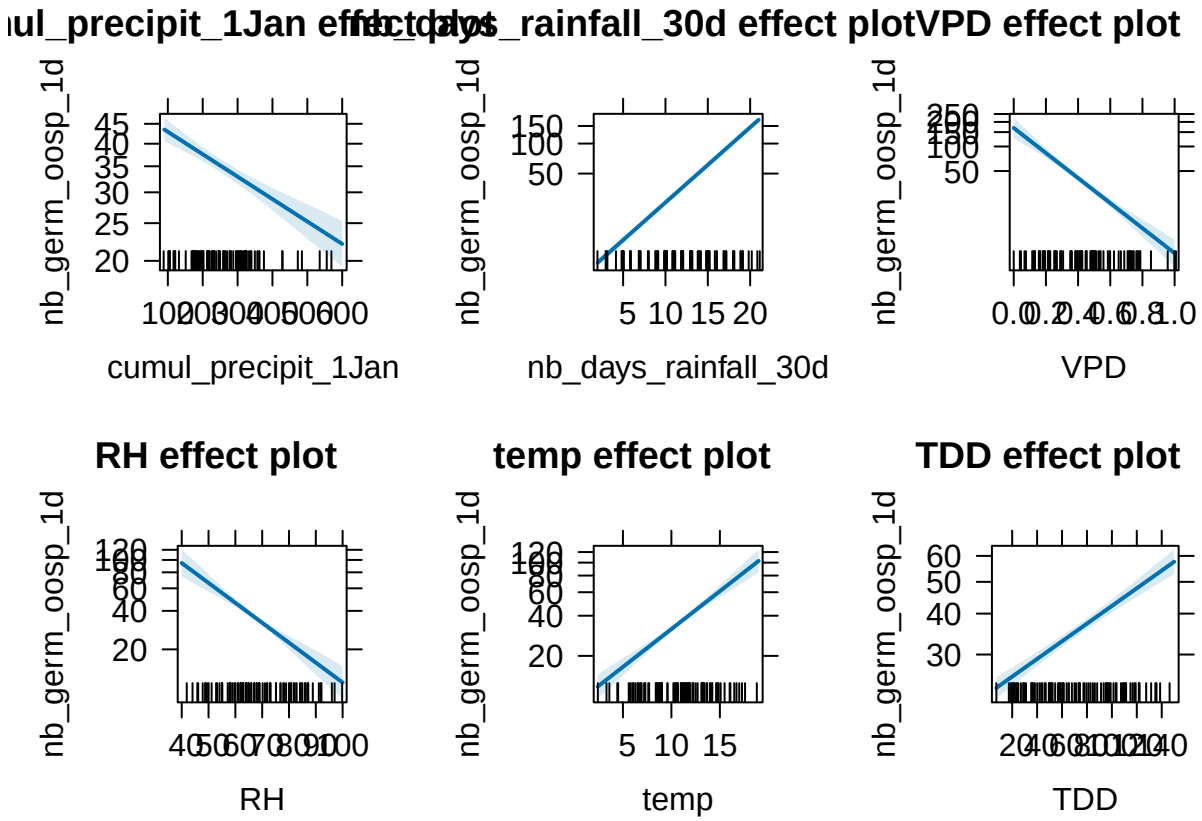
```
model_Nspores1d(df)
```


Histogram of df$nb_germ_oosp_1d
Frequency
0
20
40
60
80
0
100
200
300
400
500
600
df$nb_germ_oosp_1d


Pearson Residuals
40
30
20
10
0
−10
1
glm(nb_germ_oos

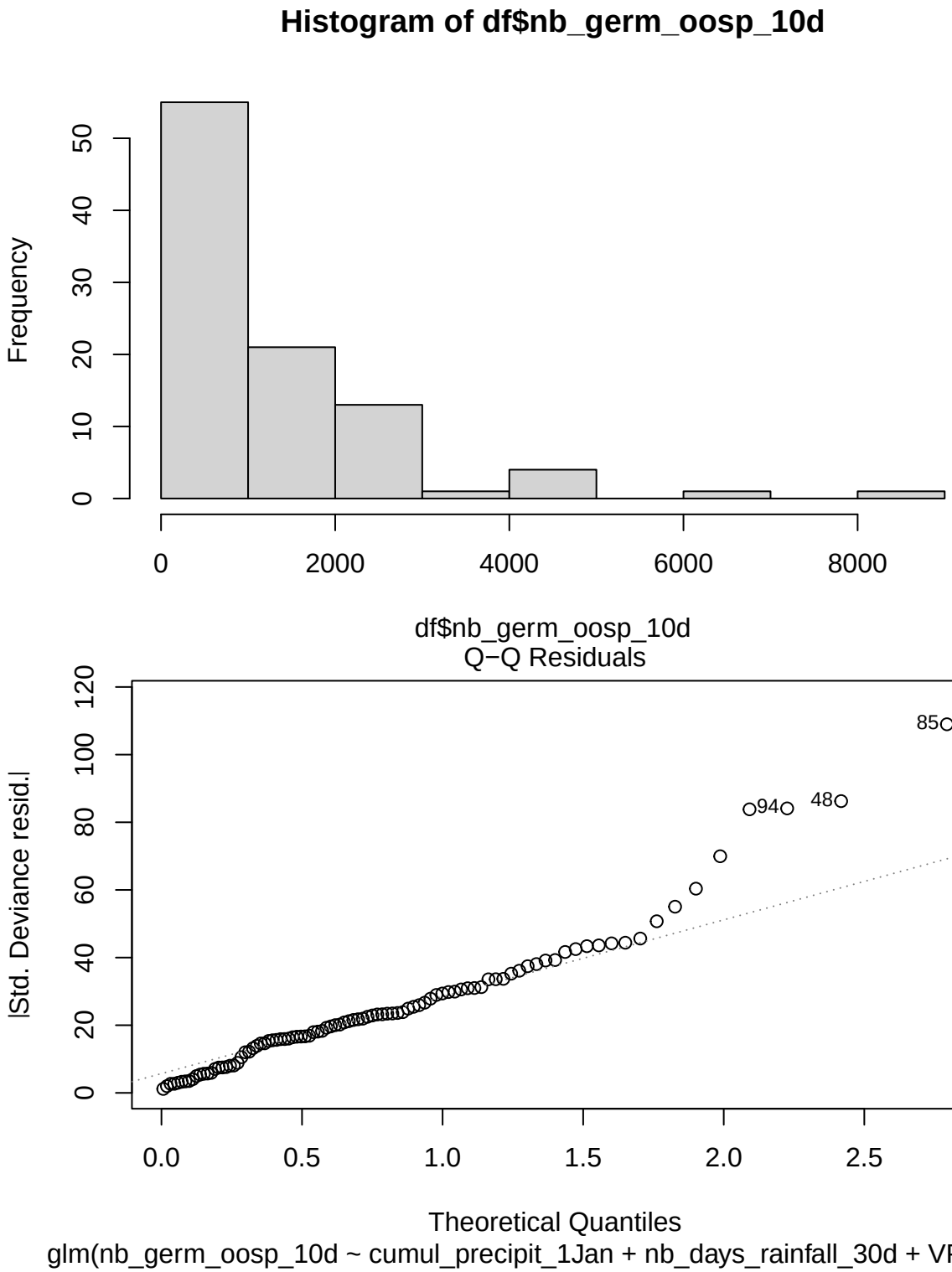

Q−Q Residuals
|Std. Deviance resid.|
0
5
10
15
20
25
30
89
87
84
0.0
0.5
1.0
1.5
2.0
2.5
Theoretical Quantiles
glm(nb_germ_oosp_1d ~ cumul_precipit_1Jan + nb_days_rainfall_30d + VPD + RH

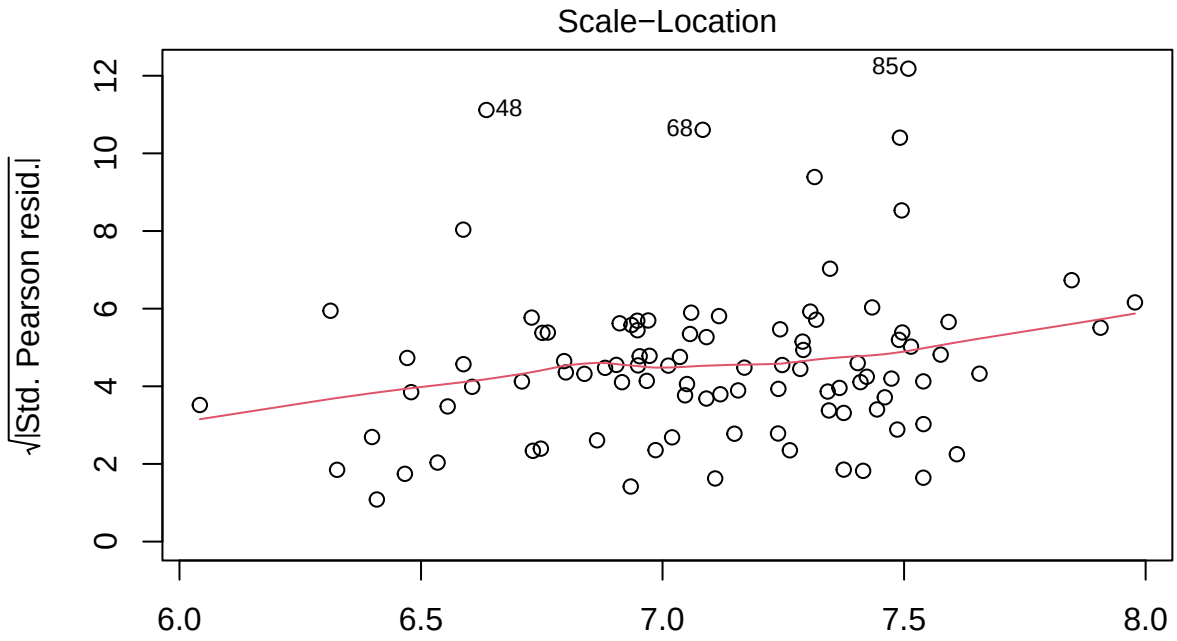

Scale−Location
√|Std. Pearson resid.|
89
73
87
0
1
2
3
4
5
6
1
2
3
4
5
Predicted values
glm(nb_germ_oosp_1d ~ cumul_precipit_1Jan + nb_days_rainfall_30d + VPD + RH
Residuals vs Leverage
Std. Pearson resid.
89
87
96
−10
0
10
20
30
40
0.5
1
Cook's distance
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
Leverage
glm(nb_germ_oosp_1d ~ cumul_precipit_1Jan + nb_days_rainfall_30d + VPD + RH

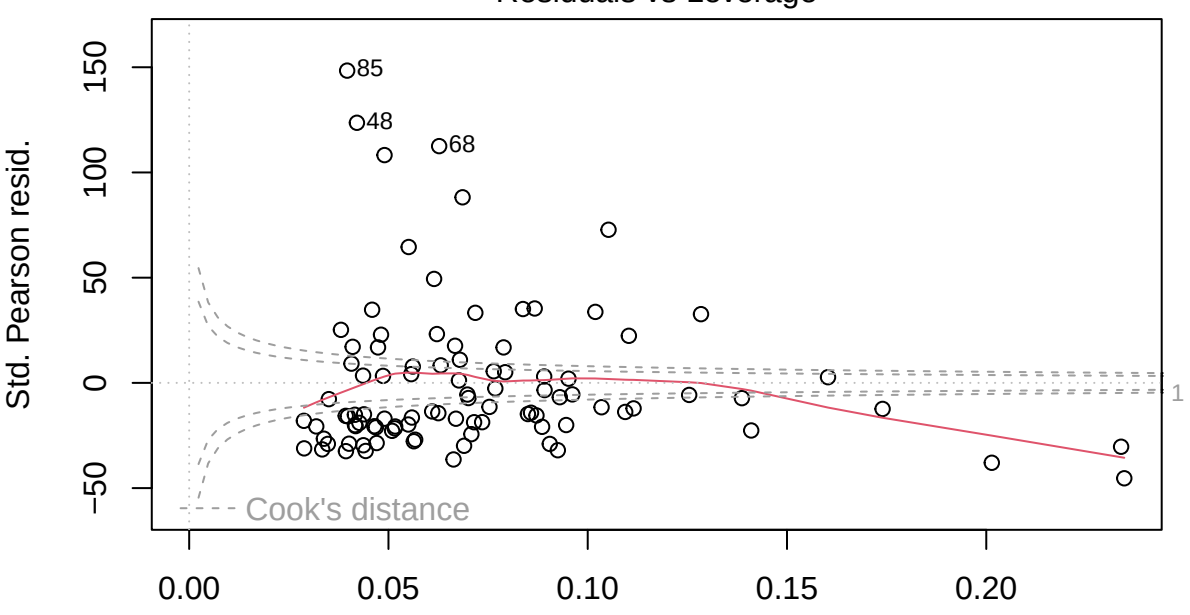


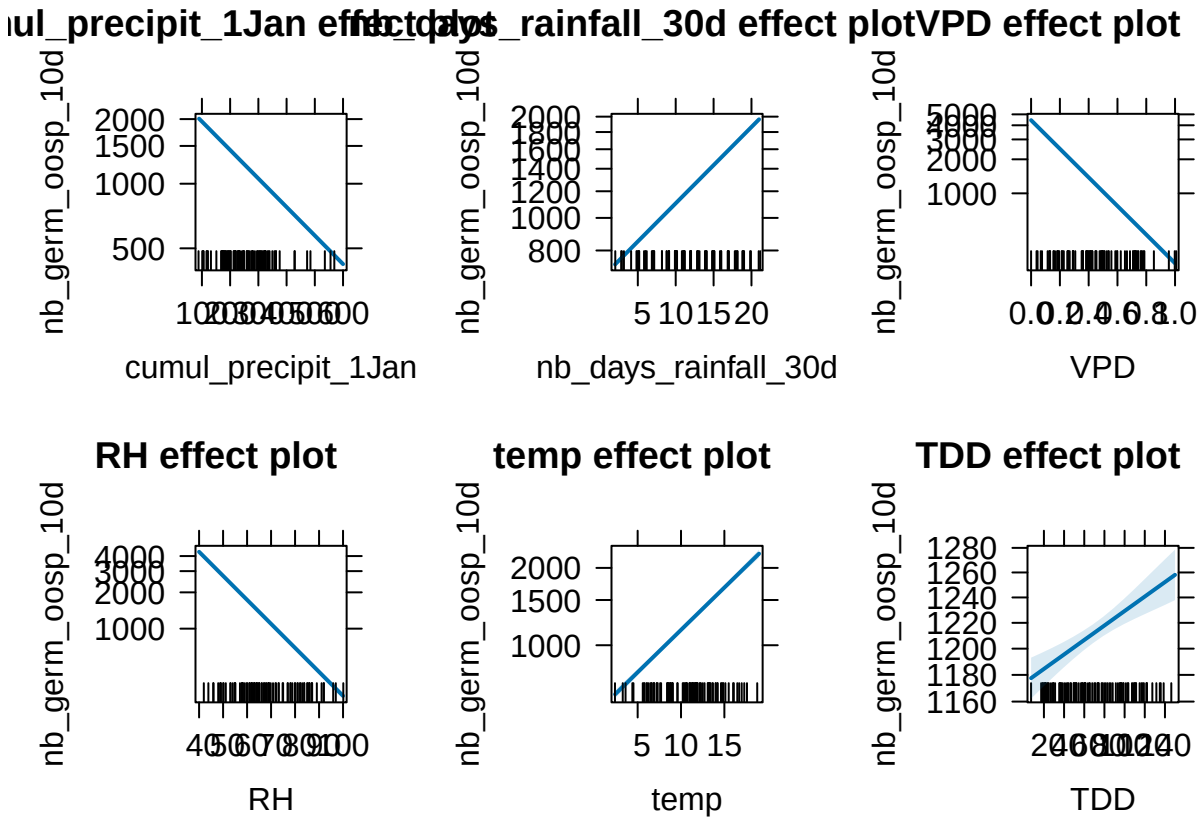

```
## # A tibble: 7 x 5
##   term                 estimate std.error statistic  p.value
##   <chr>                   <dbl>     <dbl>     <dbl>    <dbl>
## 1 (Intercept)           4.02     0.358        11.2  2.46e-29
## 2 cumul_precipit_1Jan  -0.00133  0.000189     -7.01 2.37e-12
## 3 nb_days_rainfall_30d  0.173    0.00350      49.5  0
## 4 VPD                  -3.53     0.342       -10.3  4.48e-25
## 5 RH                   -0.0358   0.00448      -7.99 1.31e-15
## 6 temp                  0.131    0.0122       10.7  1.01e-26
## 7 TDD                   0.00621  0.000510     12.2  3.90e-34
```

```
model_Nspores10d(df)
```


Histogram of df$nb_germ_oosp_10d
Frequency
0
10
20
30
40
50
0
2000
4000
6000
8000
df$nb_germ_oosp_10d
Pearson Residuals
−50
0
50
100
150
6.0
glm(nb_germ_oosp
Q−Q Residuals
|Std. Deviance resid.|
0
20
40
60
80
100
120
85
94
48
0.0
0.5
1.0
1.5
2.0
2.5
Theoretical Quantiles
glm(nb_germ_oosp_10d ~ cumul_precipit_1Jan + nb_days_rainfall_30d + VPD + R .

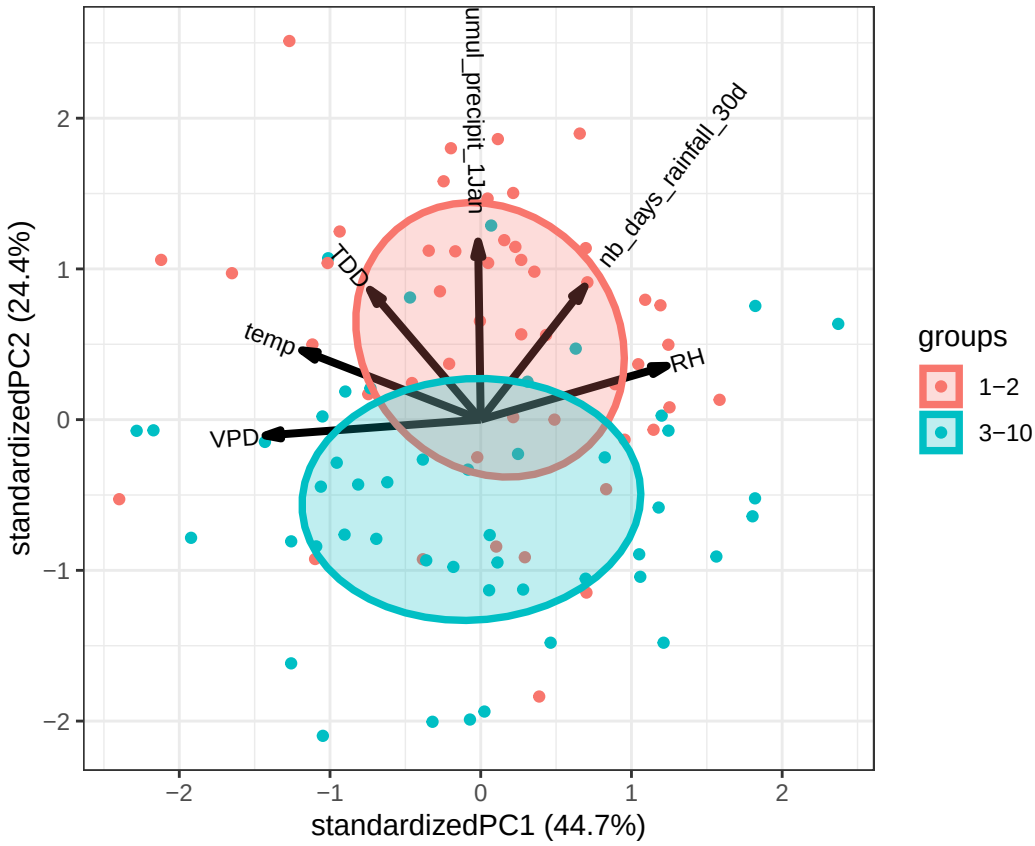

Scale−Location
√|Std. Pearson resid.|
85
48
68
0
2
4
6
8
10
12
6.0
6.5
7.0
7.5
8.0
Predicted values
glm(nb_germ_oosp_10d ~ cumul_precipit_1Jan + nb_days_rainfall_30d + VPD + R .


Residuals vs Leverage
Std. Pearson resid.
85
48
68
150
100
50
0
−50
1
Cook's distance
0.00
0.05
0.10
0.15
0.20
Leverage
glm(nb_germ_oosp_10d ~ cumul_precipit_1Jan + nb_days_rainfall_30d + VPD + R .

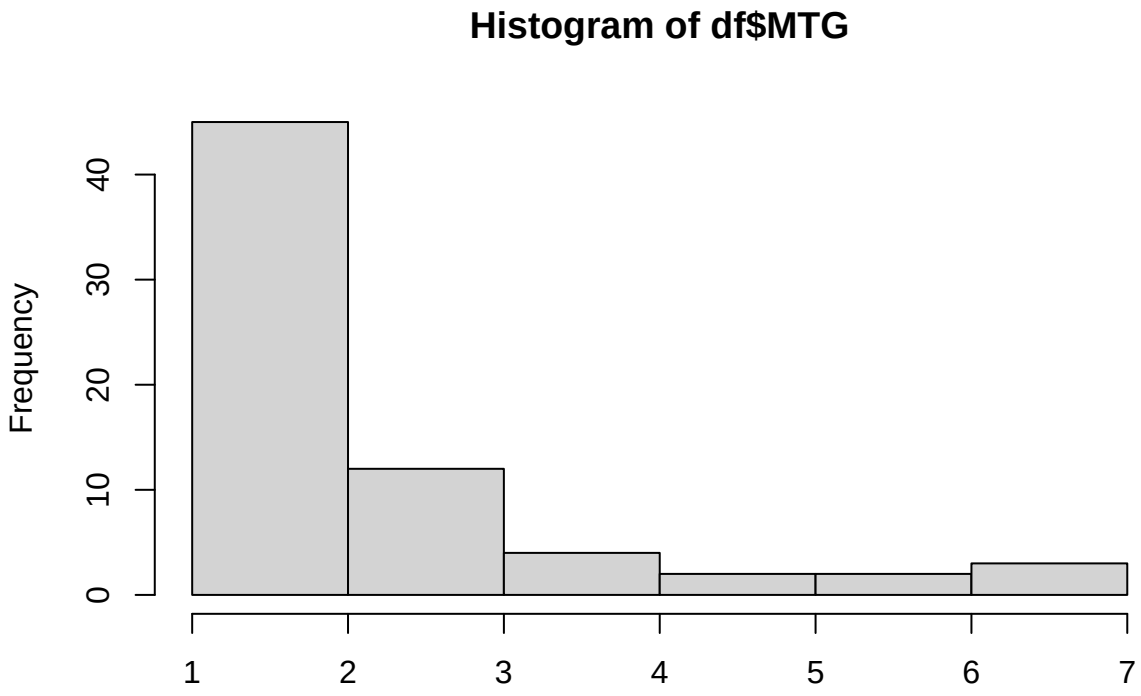


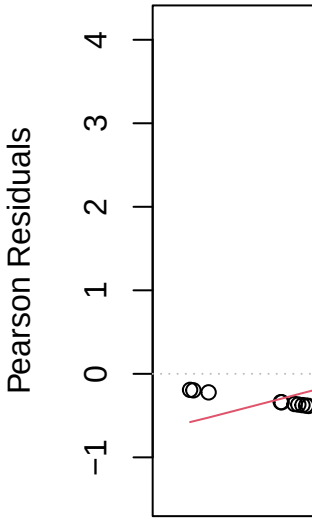

```
## # A tibble: 7 x 5
##   term                  estimate std.error statistic   p.value
##   <chr>                    <dbl>     <dbl>     <dbl>     <dbl>
## 1 (Intercept)          10.8      0.0623       174.   0
## 2 cumul_precipit_1Jan  -0.00306  0.0000374    -81.8  0
## 3 nb_days_rainfall_30d  0.0523   0.000674      77.5  0
## 4 VPD                  -2.91     0.0627       -46.5  0
## 5 RH                   -0.0459   0.000808     -56.7  0
## 6 temp                  0.0760   0.00230       33.1  1.06e-239
## 7 TDD                   0.000464 0.0000957      4.85 1.26e-  6
```

```
pca(df)
```

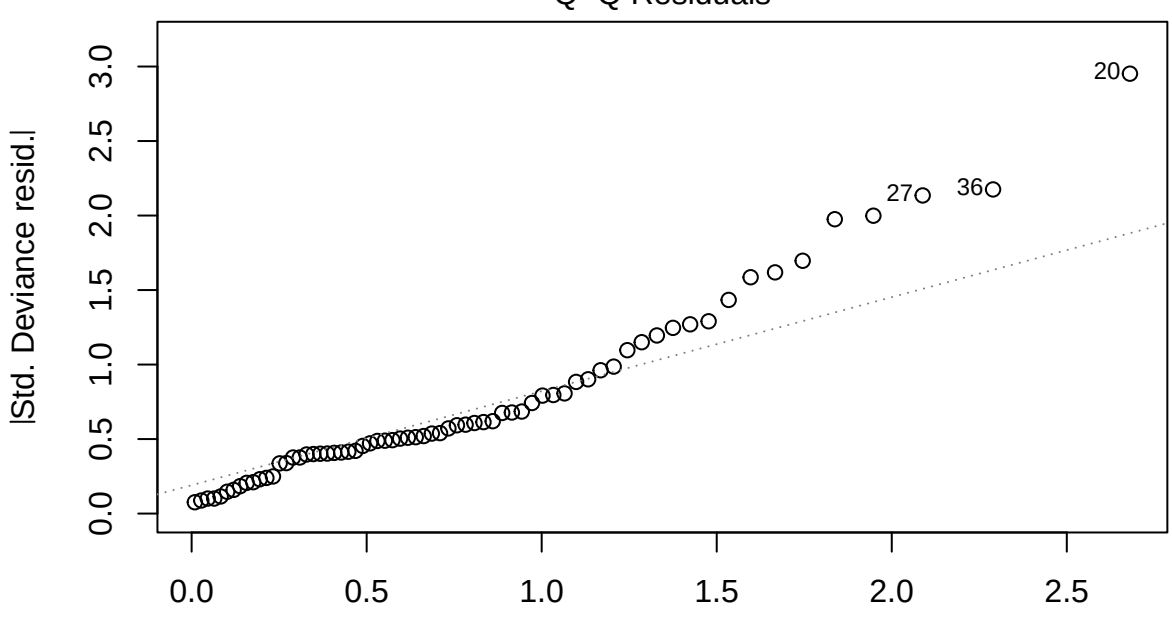



```
## DATASET BBCH 13:+
df <- df_all %>% filter(df_all$BBCH >= 13)
model_MGT(df)
```

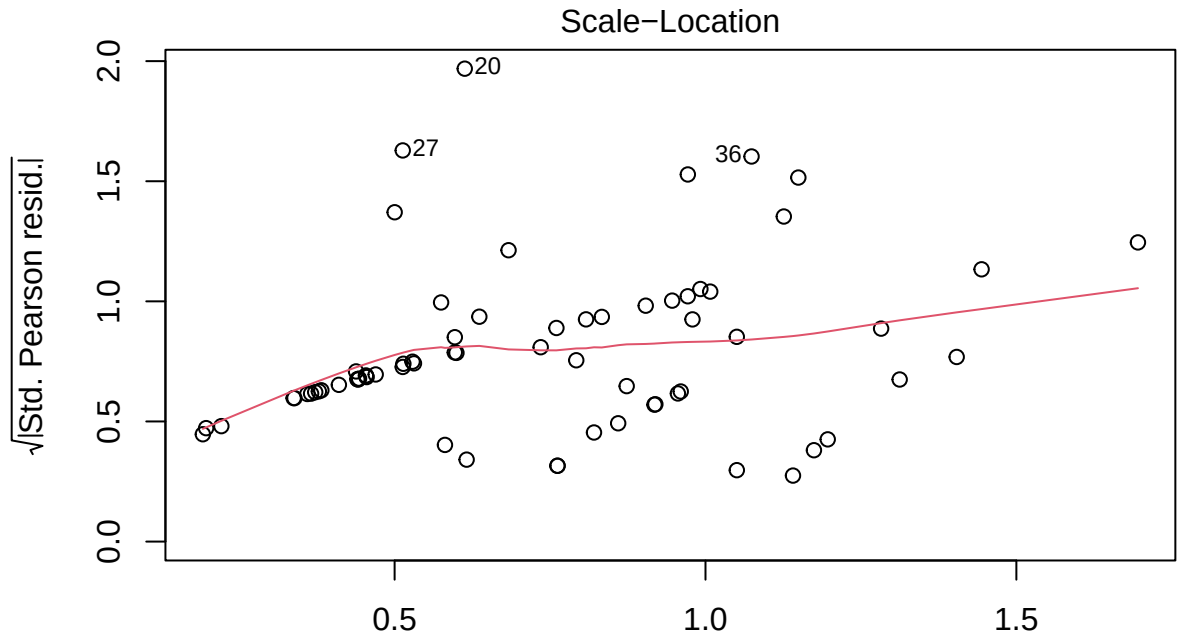




glm(MTG ~ cumul

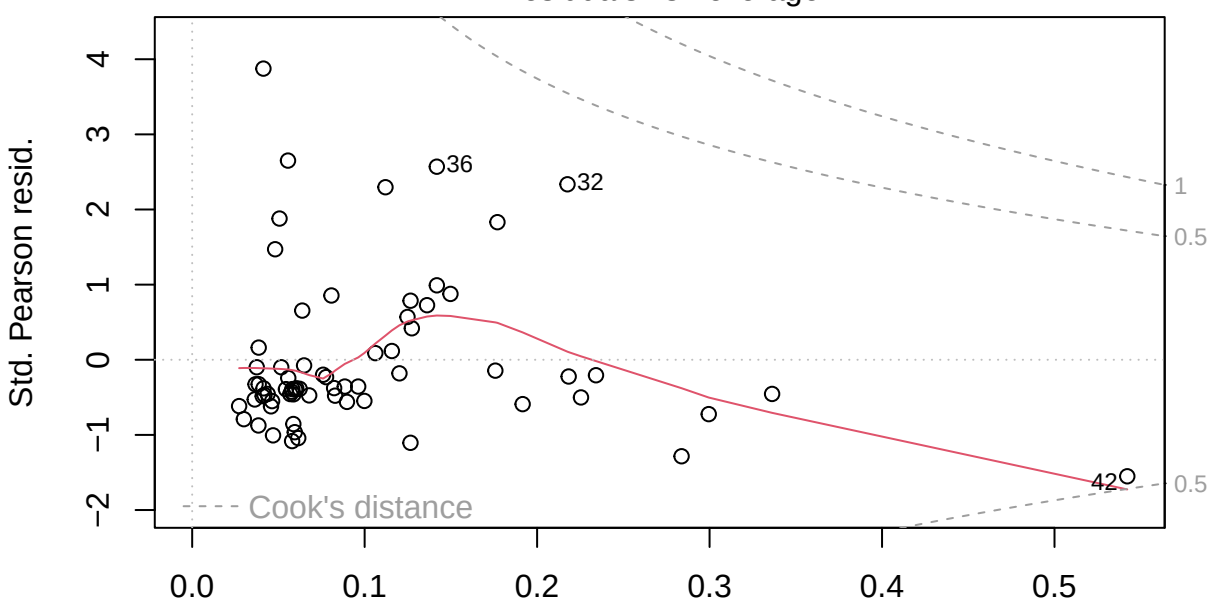


glm(MTG ~ cumul_precipit_1Jan + nb_days_rainfall_30d + +VPD + RH + temp + T .



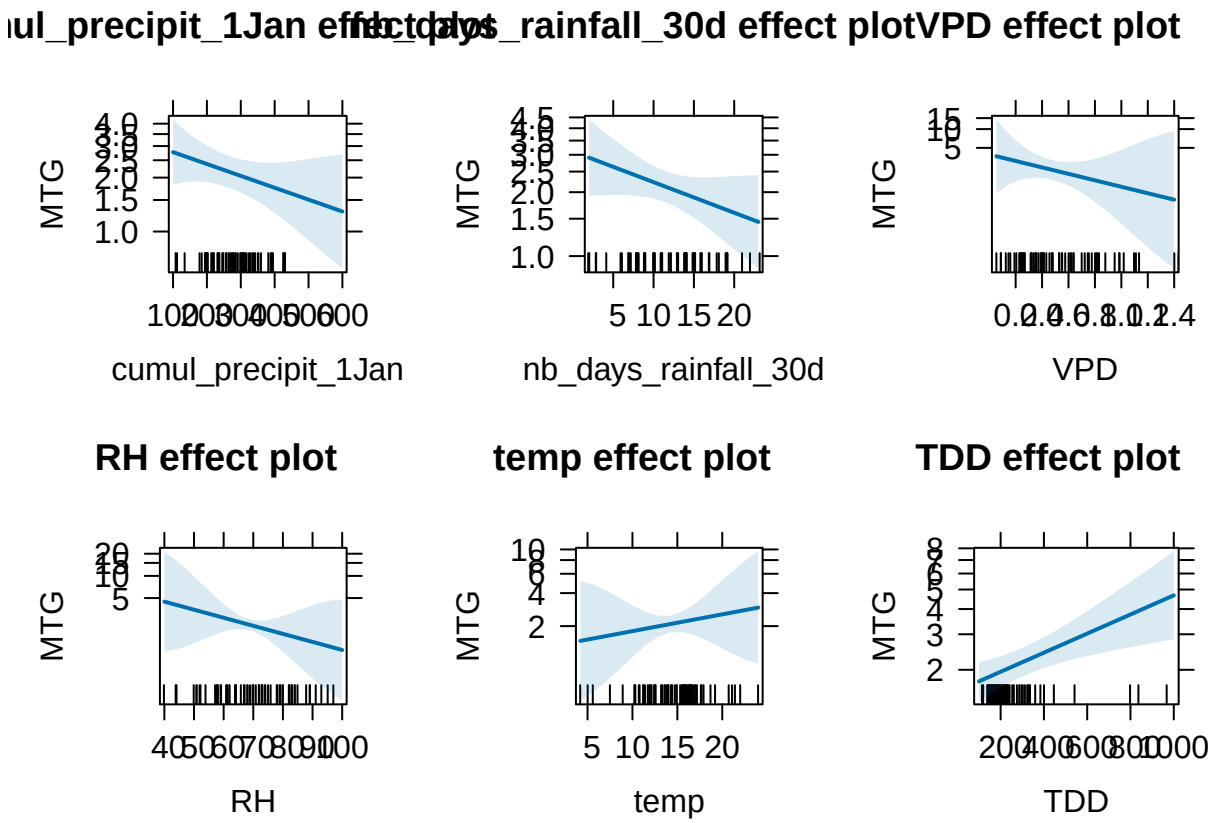
Predicted values

glm(MTG ~ cumul_precipit_1Jan + nb_days_rainfall_30d + +VPD + RH + temp + T .



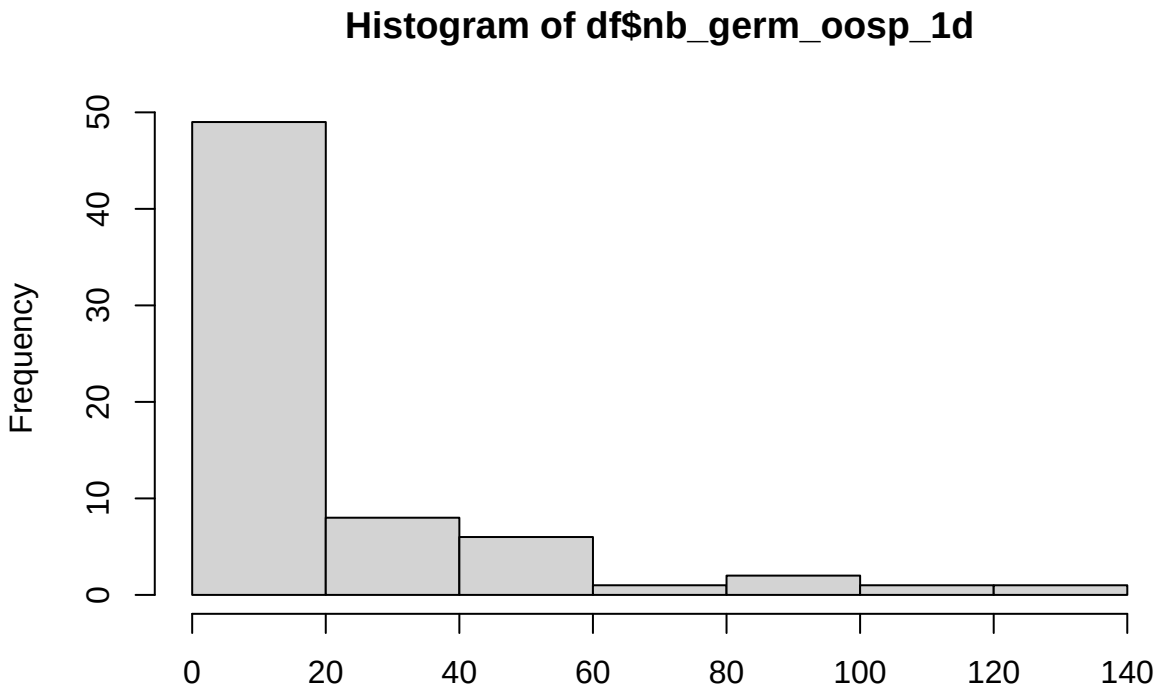
Leverage

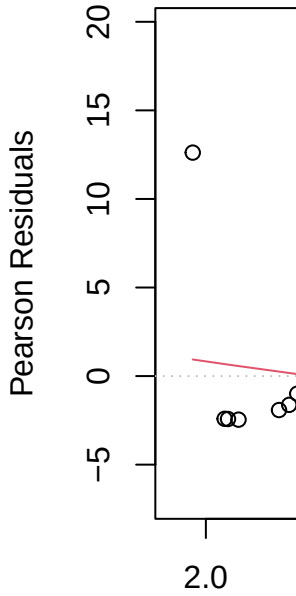
glm(MTG ~ cumul_precipit_1Jan + nb_days_rainfall_30d + +VPD + RH + temp + T .

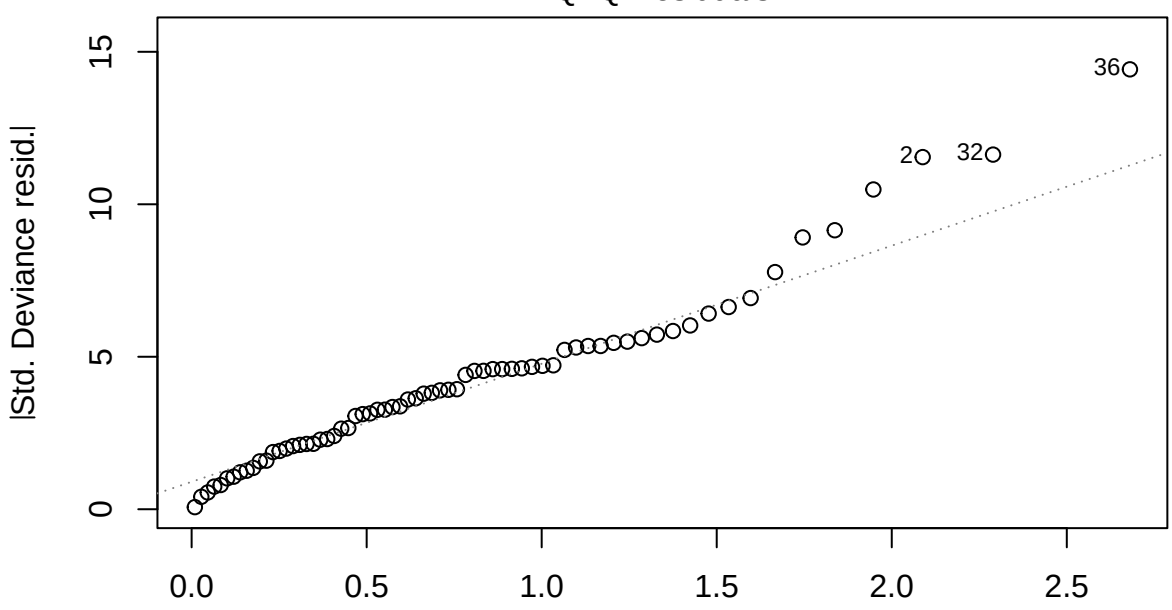



```
## # A tibble: 7 x 5
##   term                 estimate std.error statistic p.value
##   <chr>                   <dbl>     <dbl>     <dbl>   <dbl>
## 1 (Intercept)           3.16     1.89         1.67  0.0941
## 2 cumul_precipit_1Jan  -0.00153  0.00112     -1.36  0.174
## 3 nb_days_rainfall_30d -0.0332   0.0208      -1.59  0.111
## 4 VPD                  -1.19     1.47        -0.810 0.418
## 5 RH                   -0.0252   0.0266      -0.946 0.344
## 6 temp                  0.0352   0.0621       0.567 0.571
## 7 TDD                   0.00109  0.000350     3.11  0.00186
```

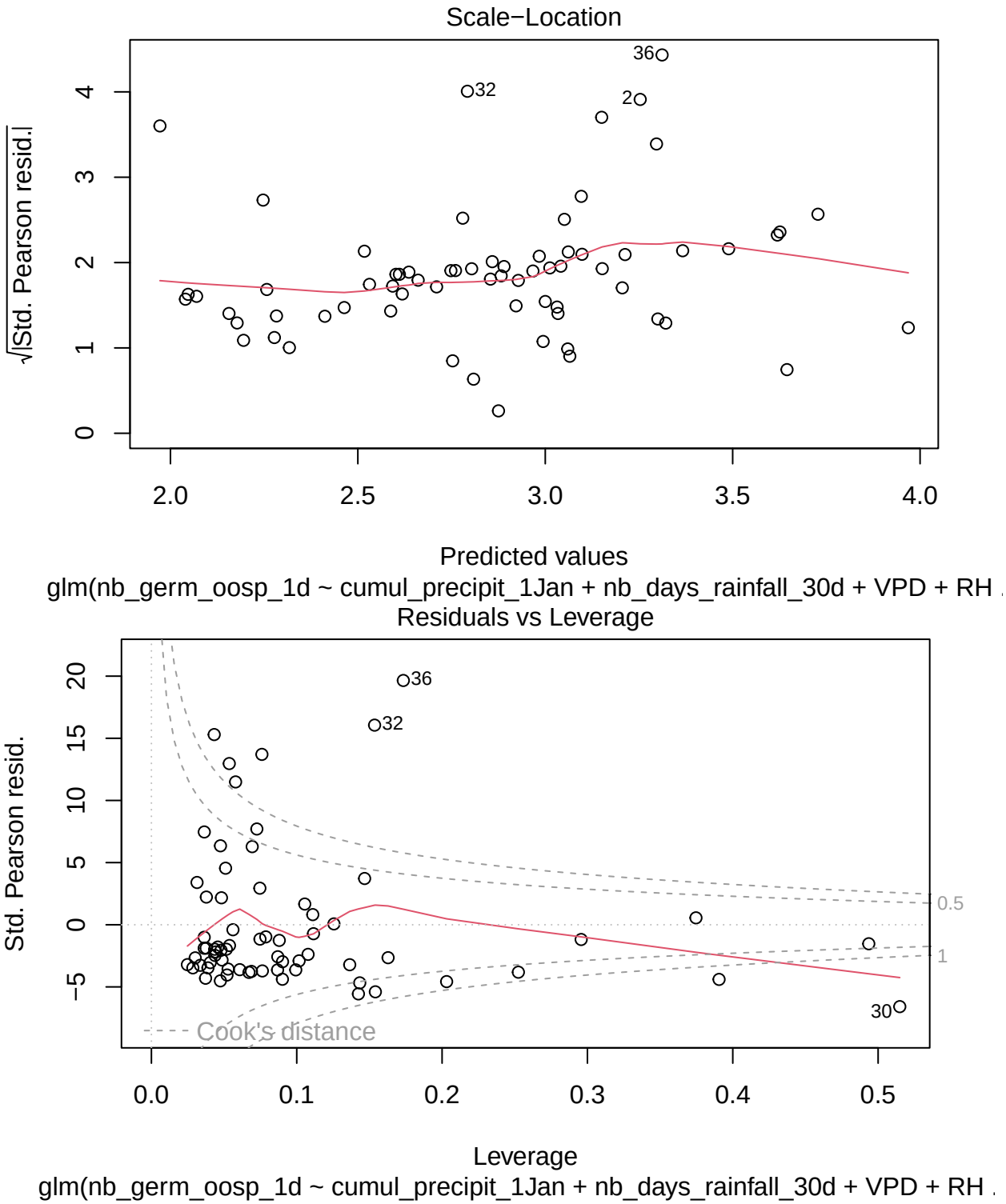
```
model_Nspores1d(df)
```

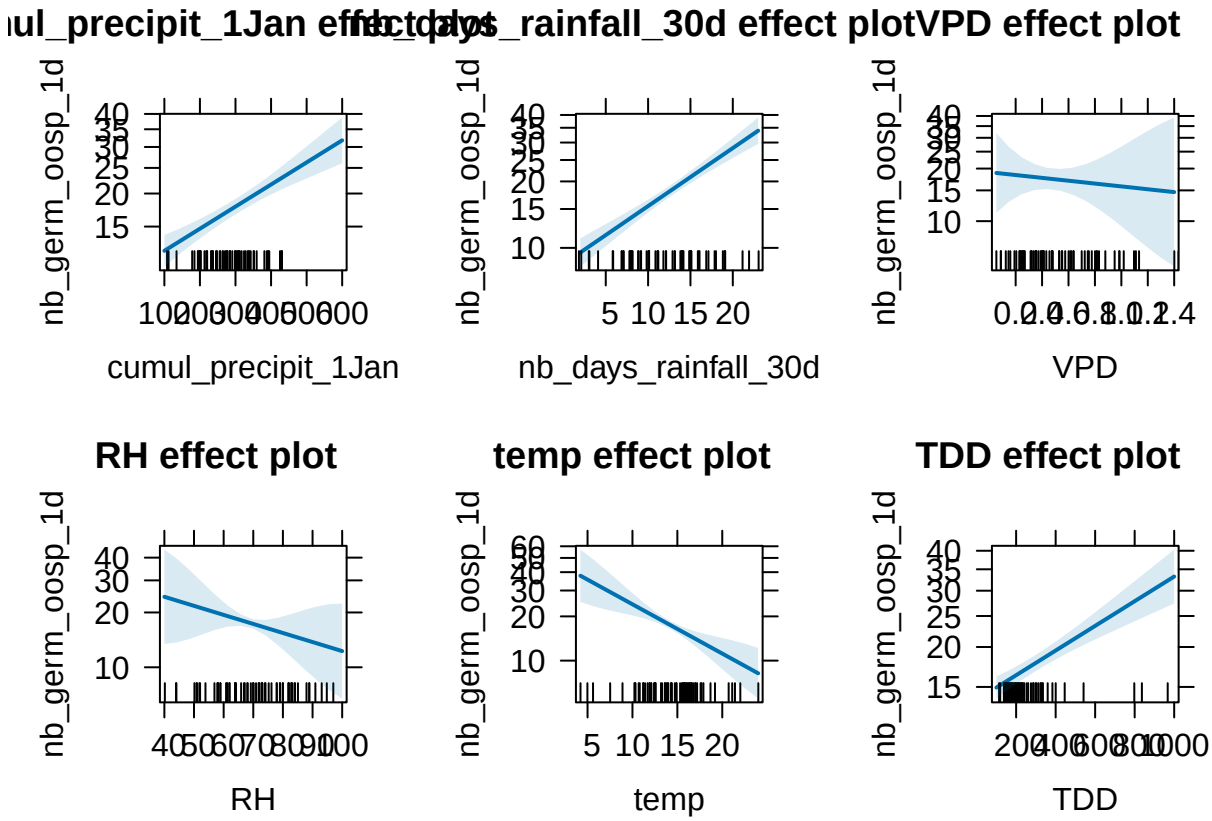






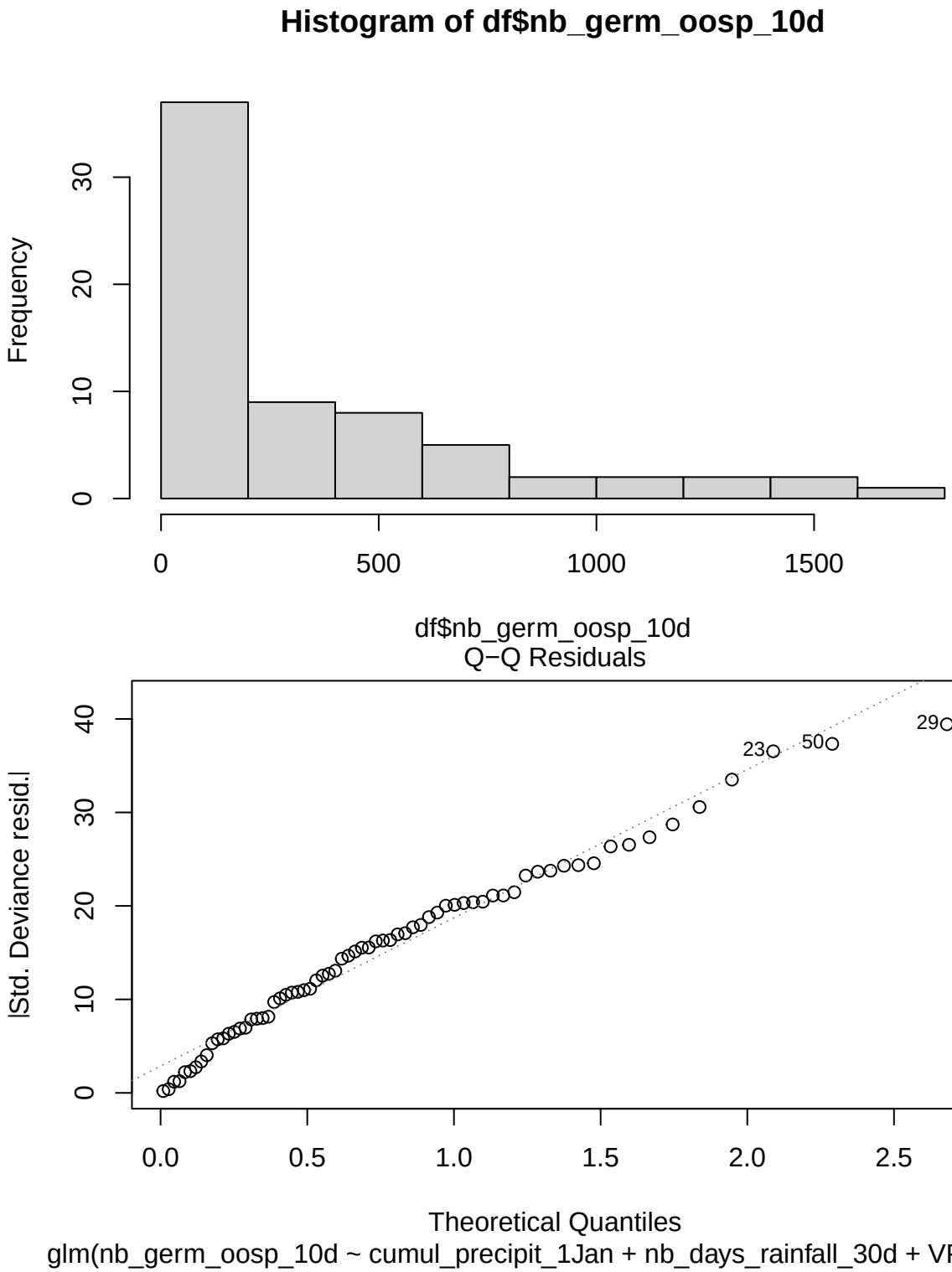
glm(nb_germ_oosp_1d ~ cumul_precipit_1Jan + nb_days_rainfall_30d + VPD + RH .

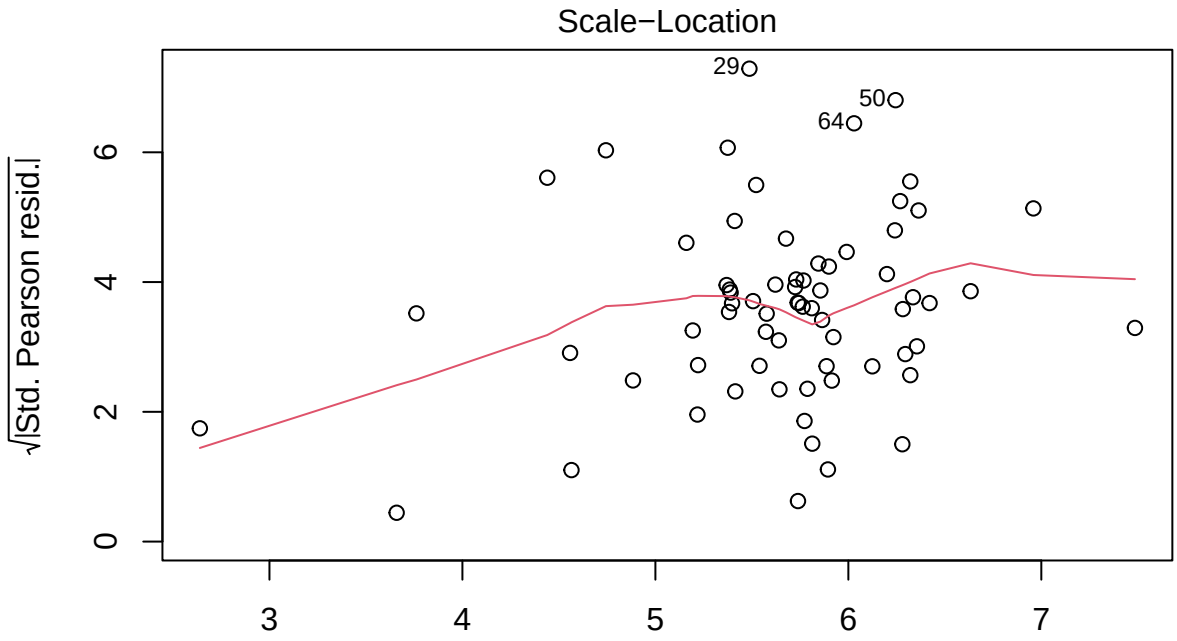

Scale−Location
√|Std. Pearson resid.|
36
32
2
0
1
2
3
4
2.0
2.5
3.0
3.5
4.0
Predicted values
glm(nb_germ_oosp_1d ~ cumul_precipit_1Jan + nb_days_rainfall_30d + VPD + RH
Residuals vs Leverage
Std. Pearson resid.
36
32
30
−5
0
5
10
15
20
0.5
1
Cook's distance
0.0
0.1
0.2
0.3
0.4
0.5
Leverage
glm(nb_germ_oosp_1d ~ cumul_precipit_1Jan + nb_days_rainfall_30d + VPD + RH

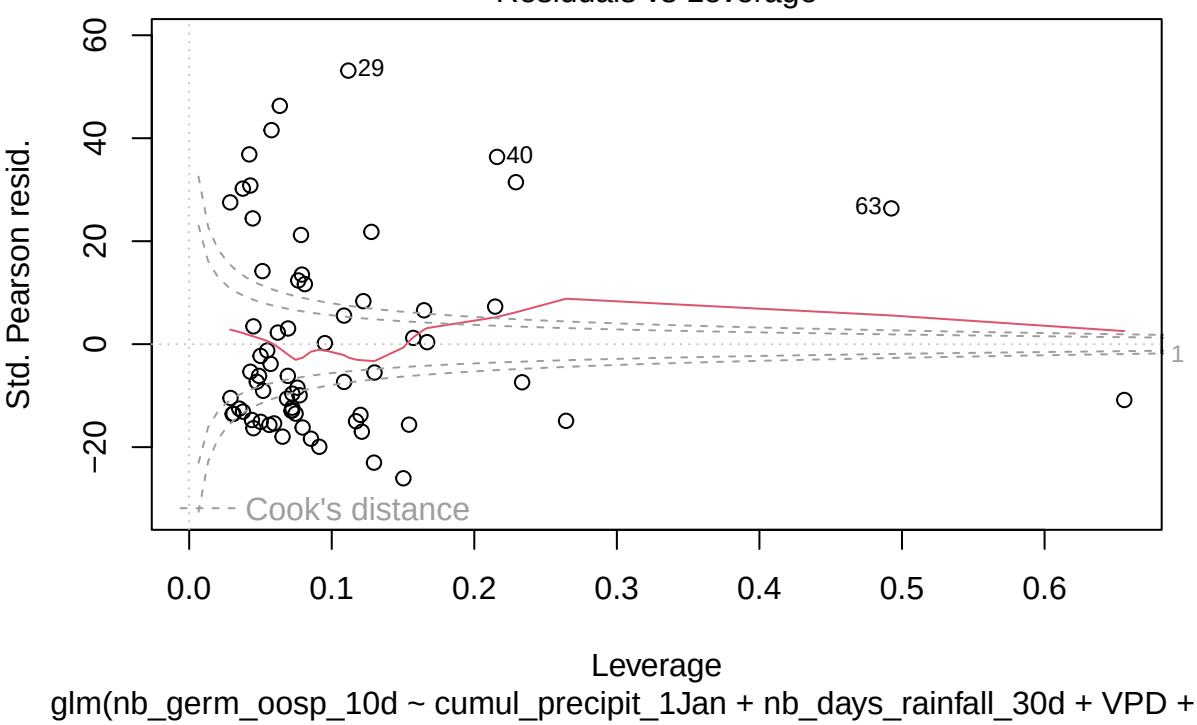


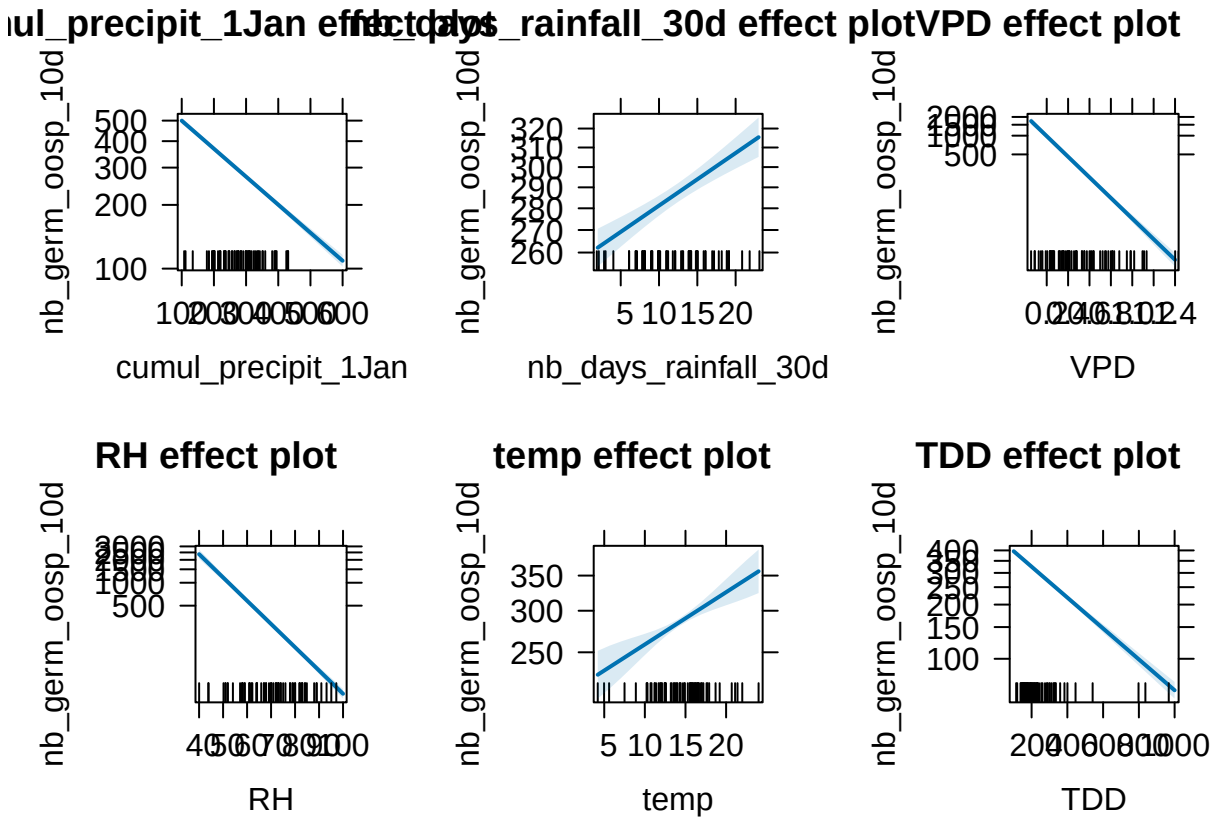

```
## # A tibble: 7 x 5
##   term                 estimate std.error statistic  p.value
##   <chr>                   <dbl>     <dbl>     <dbl>    <dbl>
## 1 (Intercept)           3.36      0.763        4.41  1.05e- 5
## 2 cumul_precipit_1Jan   0.00192   0.000321     6.00  1.97e- 9
## 3 nb_days_rainfall_30d  0.0605    0.00632      9.58  1.01e-21
## 4 VPD                  -0.187     0.569       -0.329 7.42e- 1
## 5 RH                   -0.0114    0.0101      -1.13  2.59e- 1
## 6 temp                 -0.0773    0.0205      -3.76  1.70e- 4
## 7 TDD                   0.000888  0.000136     6.55  5.82e-11
```

```
model_Nspores10d(df)
```


Histogram of df$nb_germ_oosp_10d
Frequency
0
10
20
30
0
500
1000
1500
df$nb_germ_oosp_10d
Pearson Residuals
40
20
0
−20
3
glm(nb_germ_oosp
Q−Q Residuals
|Std. Deviance resid.|
0
10
20
30
40
29
23
50
0.0
0.5
1.0
1.5
2.0
2.5
Theoretical Quantiles
glm(nb_germ_oosp_10d ~ cumul_precipit_1Jan + nb_days_rainfall_30d + VPD + R .

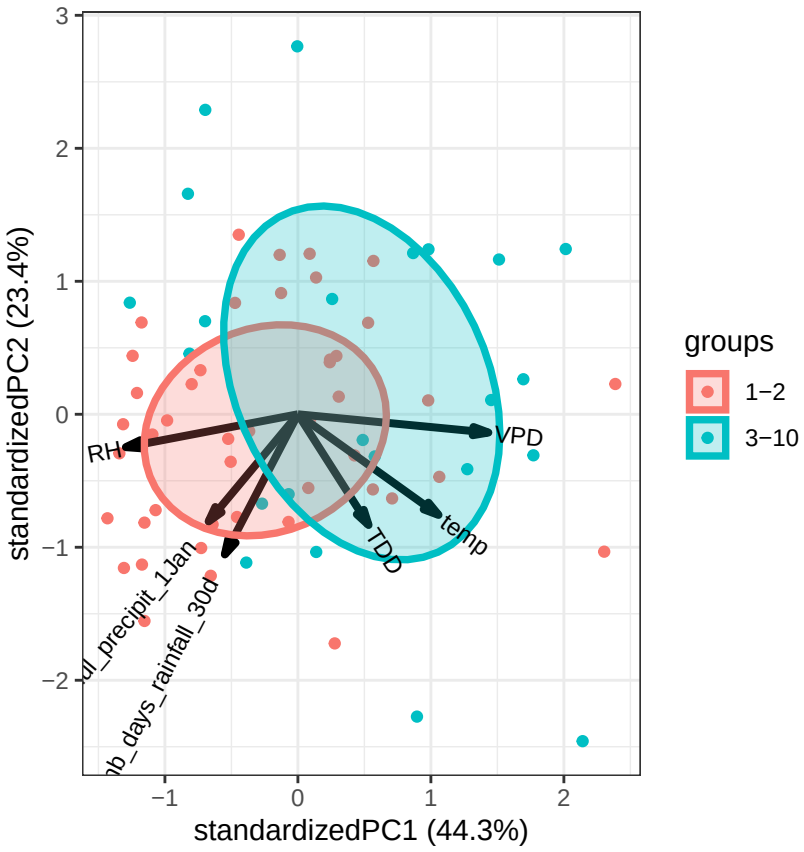

Scale−Location
29
50
64
√|Std. Pearson resid.|
0
2
4
6
3
4
5
6
7
Predicted values
glm(nb_germ_oosp_10d ~ cumul_precipit_1Jan + nb_days_rainfall_30d + VPD + R .


Residuals vs Leverage
29
40
63
Std. Pearson resid.
60
40
20
0
−20
1
Cook's distance
0.0
0.1
0.2
0.3
0.4
0.5
0.6
Leverage
glm(nb_germ_oosp_10d ~ cumul_precipit_1Jan + nb_days_rainfall_30d + VPD + R .




```
## # A tibble: 7 x 5
##   term                 estimate std.error statistic   p.value
##   <chr>                   <dbl>     <dbl>     <dbl>     <dbl>
## 1 (Intercept)          13.5     0.178         76.2  0
## 2 cumul_precipit_1Jan  -0.00305 0.0000947    -32.2  8.25e-227
## 3 nb_days_rainfall_30d  0.00881 0.00138        6.40 1.54e- 10
## 4 VPD                  -3.82    0.145        -26.3  8.26e-153
## 5 RH                   -0.0703  0.00243      -28.9  4.02e-184
## 6 temp                  0.0229  0.00510        4.49 7.00e-  6
## 7 TDD                  -0.00198 0.0000693    -28.5  1.34e-178
```

```
pca(df)
```


standardizedPC2 (23.4%)
standardizedPC1 (44.3%)
3
2
1
0
−1
−2
−1
0
1
2
RH
VPD
TDD
temp
groups
1−2
3−10