

# 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_SMAL")
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)

# solar_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)
  print(summary(pca))
  print(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)
}

```

```

### 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_30d +
    + 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_30d +
    + 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)
}

### Random Forest and Partition Trees (decision trees)
random_forest <- function(df){

  require(randomForest)
  require(caret)

  ### RANDOM FOREST COMPUTAION ON ALL EXPLANATORY VARIABLES

  # dataRF <- subset(df, select = -c(date, solar_radiation_30d, solar_radiation_1Jan, MTG, maturity, nb
  dataRF <- as.data.frame(cbind(df$cumul_precipit_1Jan, df$nb_days_rainfall_30d, df$VPD,
                                df$RH, df$temp, df$TDD))
  # 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"
    }
  }
}

colnames(dataRF) <- colnames(subset(df, select = c(cumul_precipit_1Jan, nb_days_rainfall_30d, VPD,
                                                    RH, temp, TDD)))

dataRF$MTG_cat <- MTG_cat

dataRF$MTG_cat <- as.factor(dataRF$MTG_cat)
set.seed(111)
ind <- sample(2, nrow(dataRF), replace = TRUE, prob = c(0.7, 0.3))
train <- dataRF[ind == 1,]
train$MTG_cat <- factor(train$MTG_cat)
test <- dataRF[ind == 2,]
test$MTG <- factor(test$MTG_cat)

rf <- randomForest(MTG_cat~., data = train, proximity = TRUE, mtry = 3)
print(rf)
plot(rf)

p1 <- predict(rf, train)
confusionMatrix(p1, train$MTG_cat)

p2 <- predict(rf, test)
confusionMatrix(p2, test$MTG_cat)

t <- tuneRF(subset(train, select = -c(MTG_cat)), train[, "MTG_cat"],

```

```

        stepFactor = 0.5,
        plot = TRUE,
        ntreeTry = 150,
        trace = TRUE,
        improve = 0.05)

hist(treesize(rf),
     main = "No. of Nodes for the Trees",
     col = "green")
# Variable Importance
varImpPlot(rf,
           sort = T,
           n.var = 6,
           main = "Ranked Variable Importance")
print(importance(rf))
#MeanDecreaseGini

partialPlot(rf, train, TDD, as.factor("1-2"))
partialPlot(rf, train, TDD, "3-10")
partialPlot(rf, train, nb_days_rainfall_30d, "1-2")
partialPlot(rf, train, nb_days_rainfall_30d, "3-10")
partialPlot(rf, train, TDD, "1-2")
partialPlot(rf, train, TDD, "3-10")

# devtools::install_github("araastat/reprtree")

require(devtools)
install_github("araastat/reprtree")
library(reprtree)

## Repartition tree / Decision tree
reprtree::plot.getTree(rf)

## MDS plot
MDSplot(rf, train$MTG_cat)

require("rpart")
require("rpart.plot")

### PARTITION TREES COMPUTAION ON SELECTED VARIABLES ONLY

dataRpart <- subset(df, select = c(cumul_precipit_1Jan, nb_days_rainfall_30d, VPD,
                                RH, temp, TDD))
dataRpart <- matrix(as.numeric(unlist(dataRpart)),nrow = nrow(dataRpart))
colnames(dataRpart) <- colnames(subset(df, select = c(cumul_precipit_1Jan, nb_days_rainfall_30d, VPD,
                                                    RH, temp, TDD)))

dataRpart <- as.data.frame(dataRpart)

## Decision tree for MTG
rf1 <- rpart(df$MTG ~., data = dataRpart, method = "poisson")
rpart.plot(rf1)

## Decision tree for Nspores1d

```

```

rf2 <- rpart(df$nb_germ_oosp_1d ~., data = dataRpart, method = "poisson")
rpart.plot(rf2)

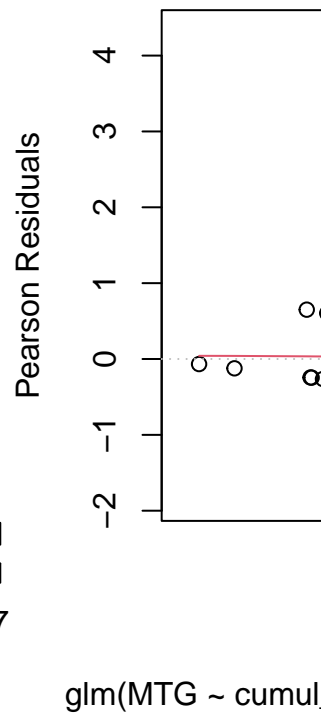
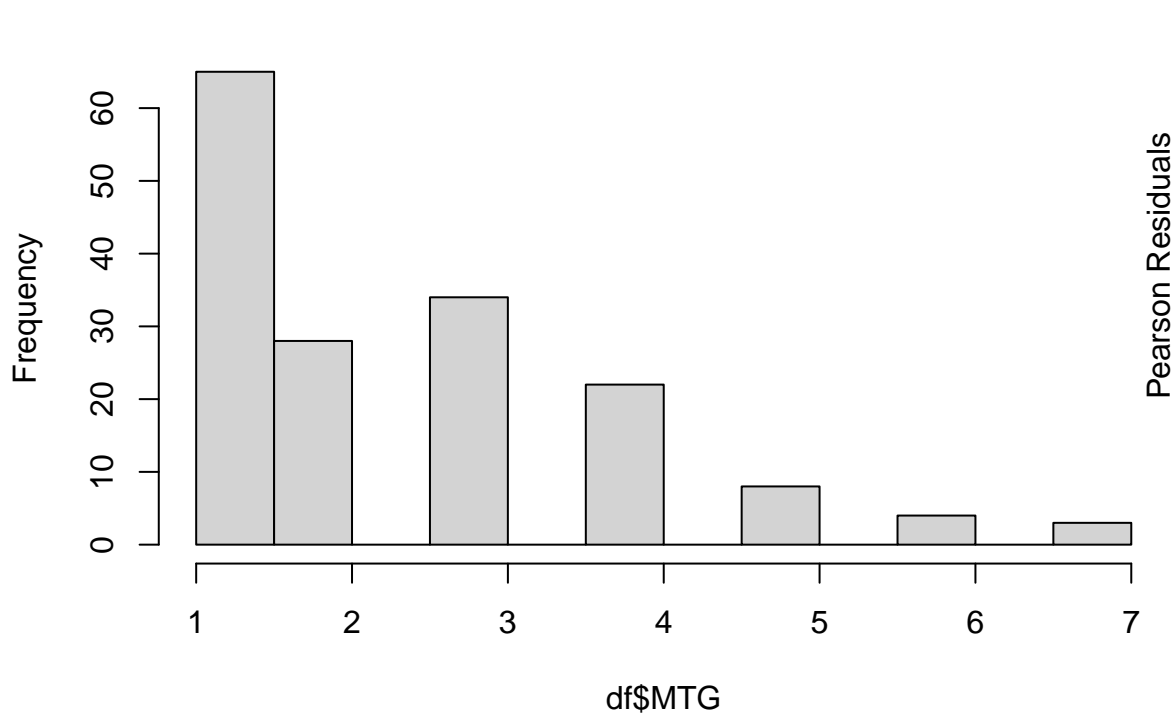
## Decision tree for Nspores10d
rf3 <- rpart(df$nb_germ_oosp_10d ~., data = dataRpart, method = "poisson")
rpart.plot(rf3)

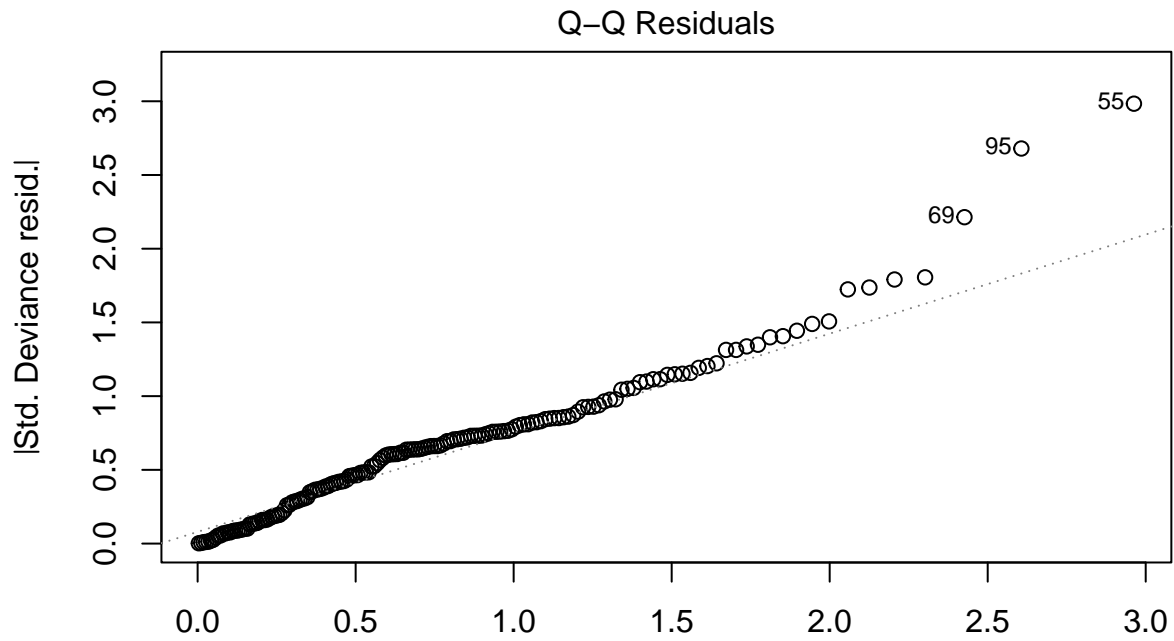
}

## ALL BBCH DATASET
model_MGT(df_all)

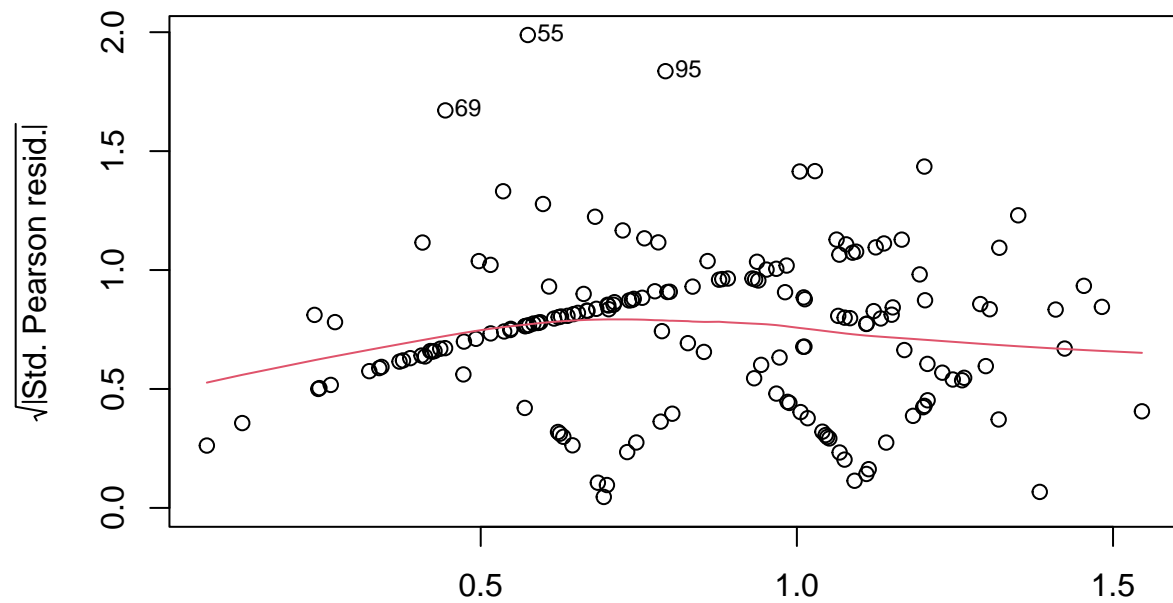
```

**Histogram of df\$MTG**

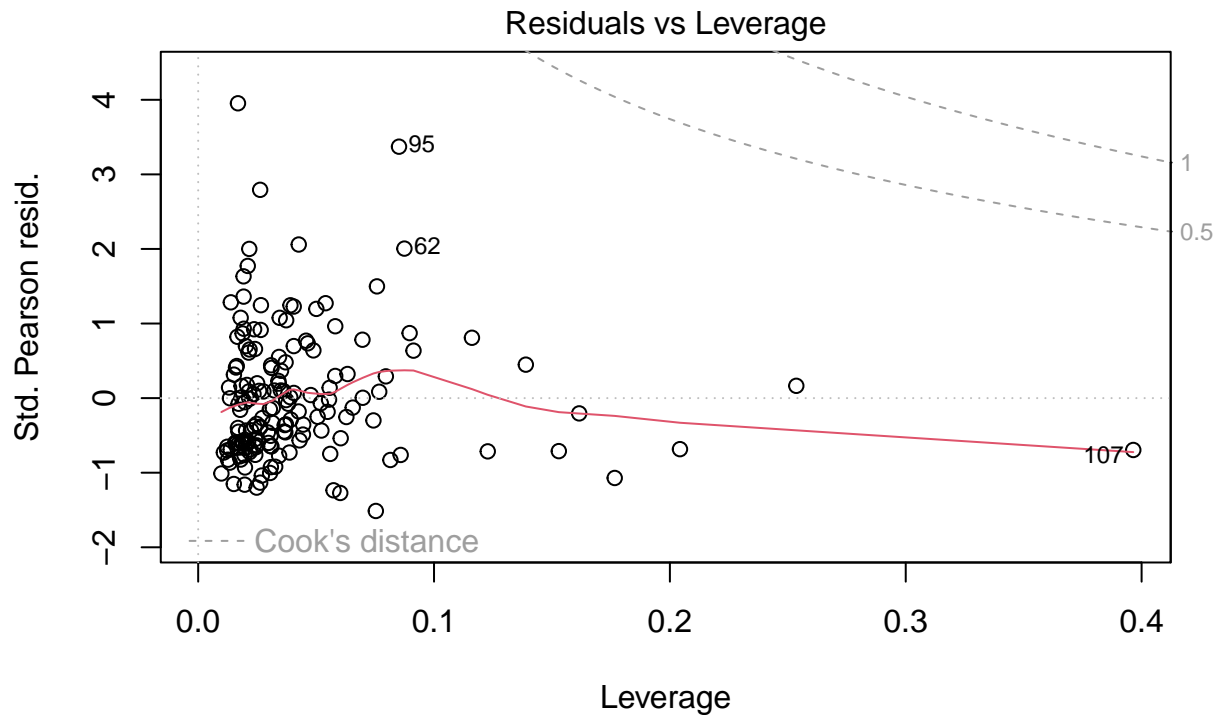




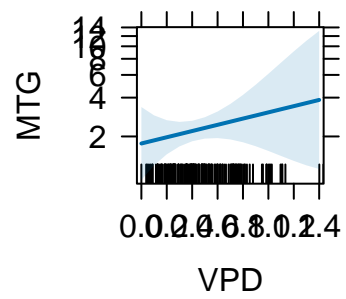
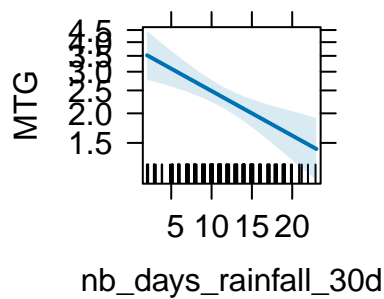
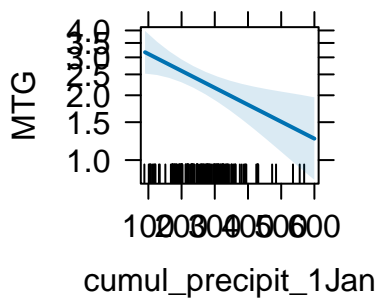
Theoretical Quantiles  
 $\text{glm}(\text{MTG} \sim \text{cumul\_precipit\_1Jan} + \text{nb\_days\_rainfall\_30d} + \text{VPD} + \text{RH} + \text{temp} + \text{T} .$   
 Scale-Location



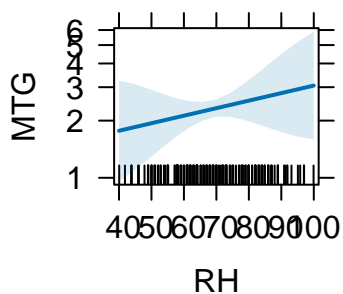
Predicted values  
 $\text{glm}(\text{MTG} \sim \text{cumul\_precipit\_1Jan} + \text{nb\_days\_rainfall\_30d} + \text{VPD} + \text{RH} + \text{temp} + \text{T} .$



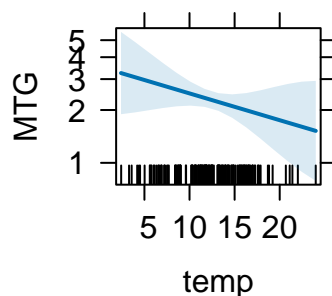
glm(MTG ~ cumul\_precipit\_1Jan + nb\_days\_rainfall\_30d + +VPD + RH + temp + TDD .  
 cumul\_precipit\_1Jan effect plot nb\_days\_rainfall\_30d effect plot VPD effect plot



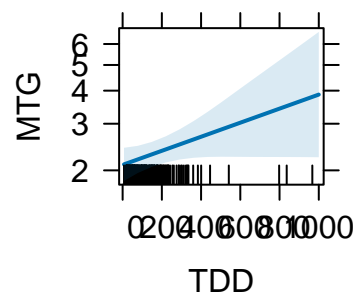
**RH effect plot**



**temp effect plot**



**TDD effect plot**



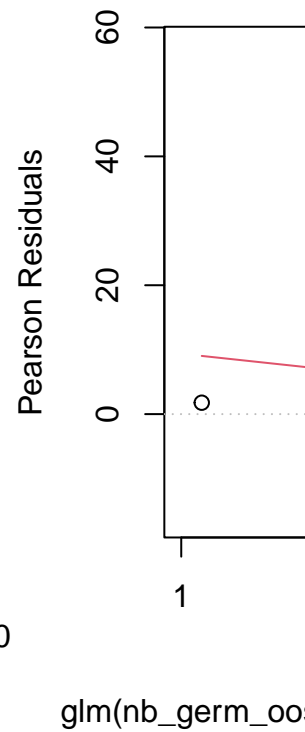
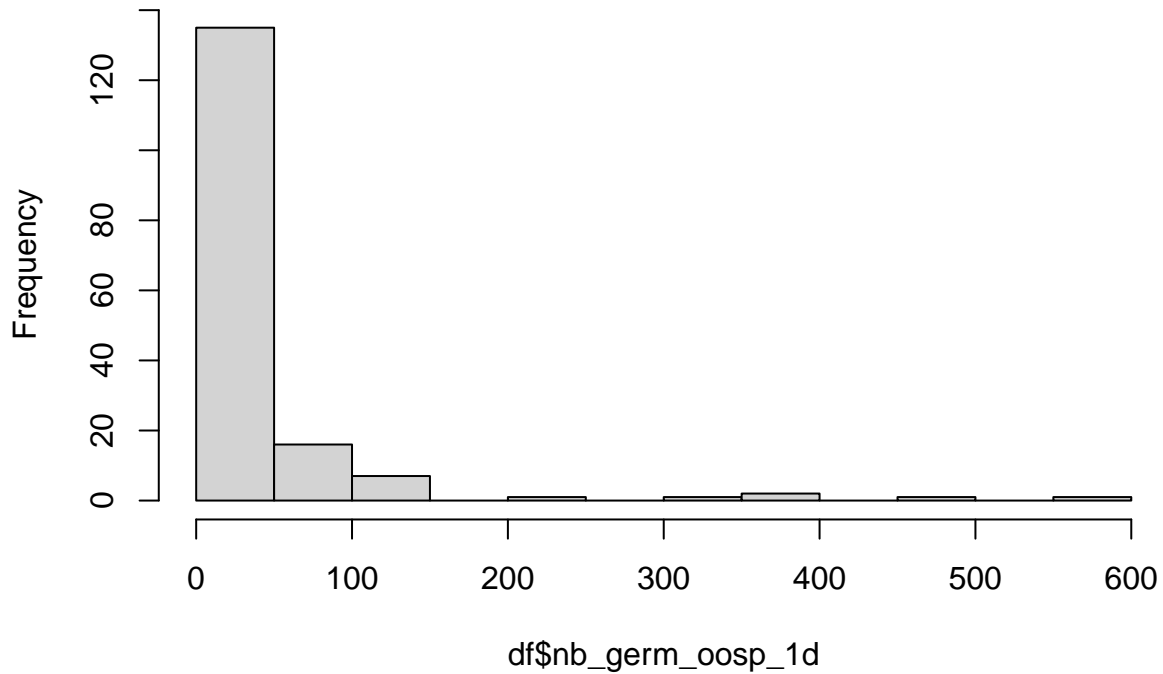
```
## # A tibble: 7 x 5
##   term                estimate std.error statistic  p.value
##   <chr>              <dbl>    <dbl>    <dbl>    <dbl>
## 1 (Intercept)        1.27      0.813      1.57  0.117
```

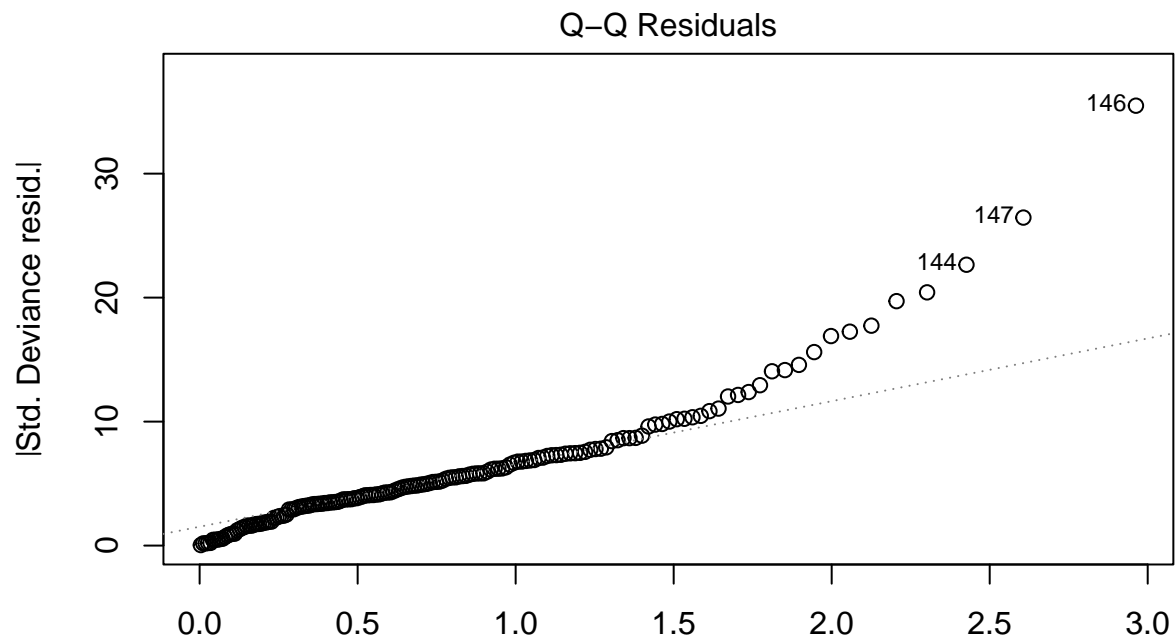


```
## 2 cumul_precipit_1Jan -0.00181 0.000635 -2.85 0.00431
## 3 nb_days_rainfall_30d -0.0434 0.0121 -3.59 0.000335
## 4 VPD 0.556 0.682 0.815 0.415
## 5 RH 0.00913 0.0106 0.864 0.388
## 6 temp -0.0351 0.0278 -1.26 0.207
## 7 TDD 0.000611 0.000327 1.87 0.0616
```

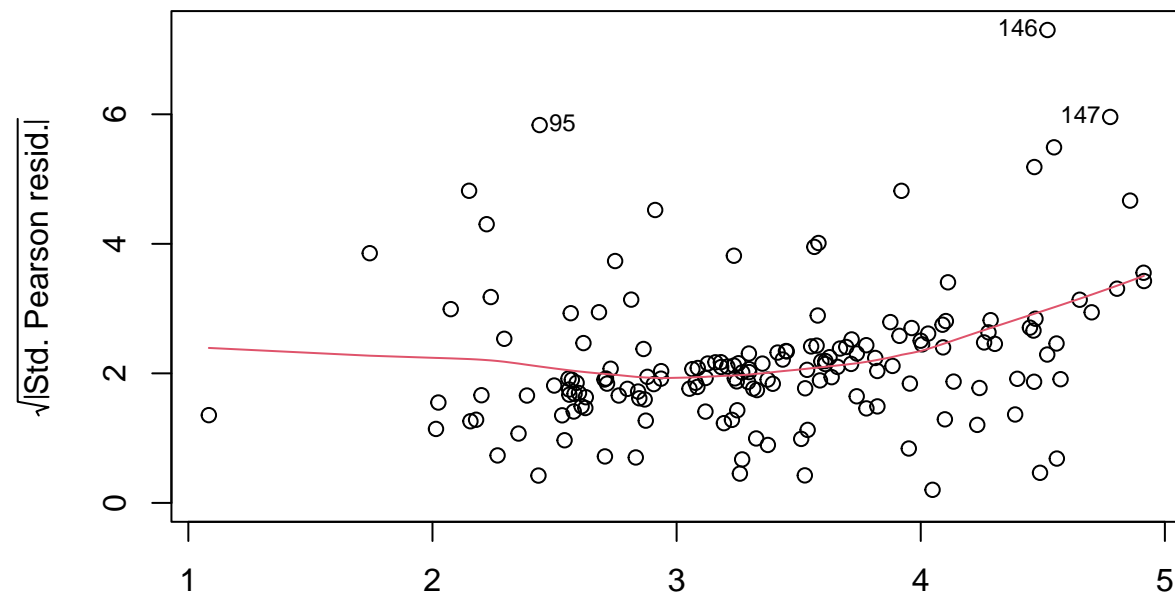
```
model_Nspores1d(df_all)
```

**Histogram of df\$nb\_germ\_oosp\_1d**

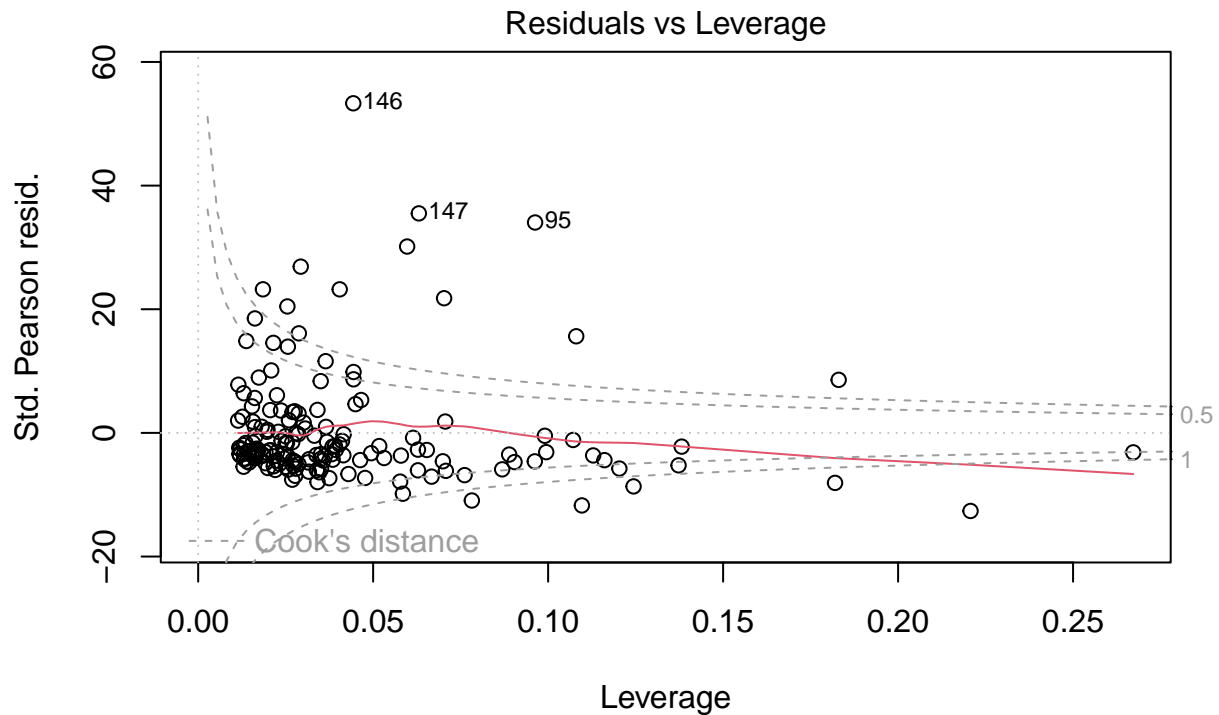




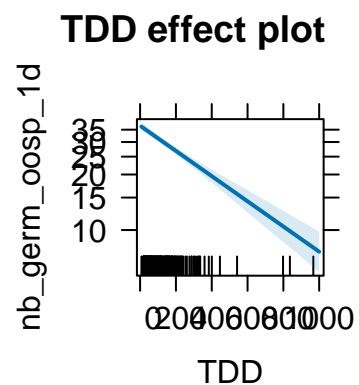
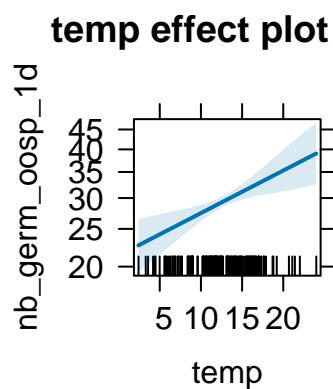
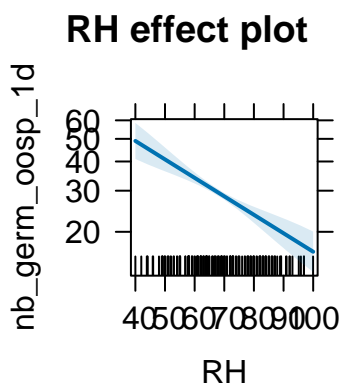
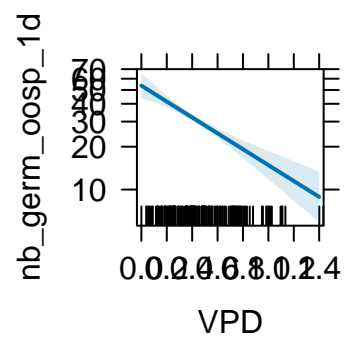
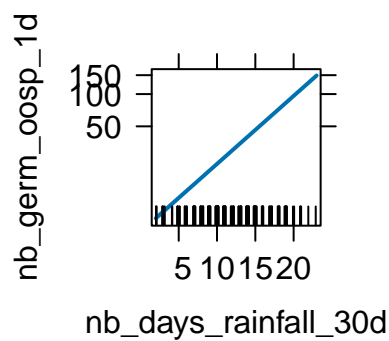
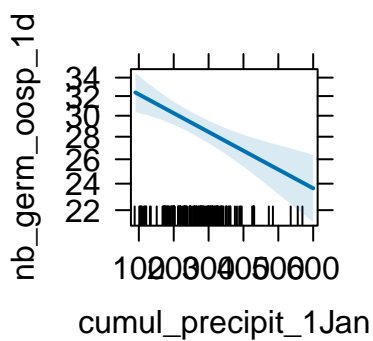
glm(nb\_germ\_oosp\_1d ~ cumul\_precipit\_1Jan + nb\_days\_rainfall\_30d + VPD + RH)  
Scale-Location



glm(nb\_germ\_oosp\_1d ~ cumul\_precipit\_1Jan + nb\_days\_rainfall\_30d + VPD + RH)



glm(nb\_germ\_oosp\_1d ~ cumul\_precipit\_1Jan + nb\_days\_rainfall\_30d + VPD + RH +  
 cumul\_precipit\_1Jan effect plot nb\_days\_rainfall\_30d effect plot VPD effect plot

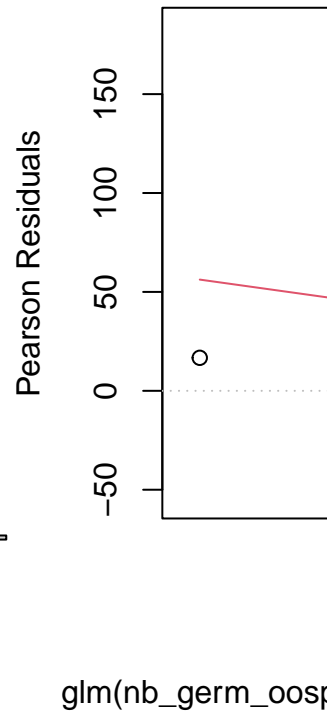
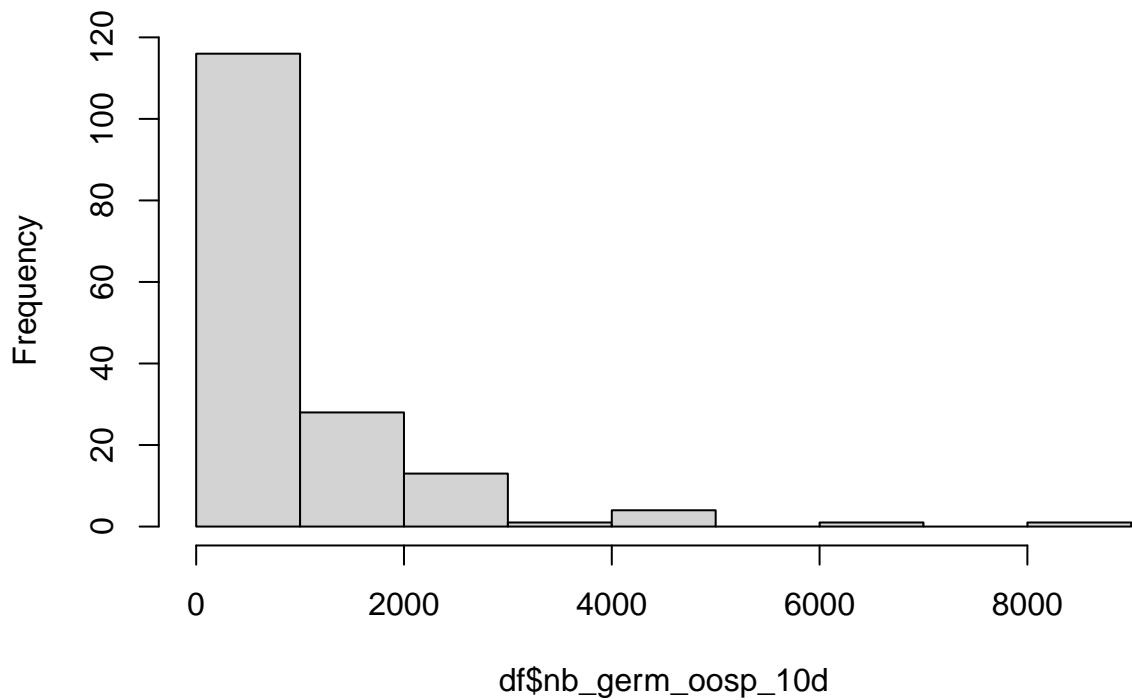


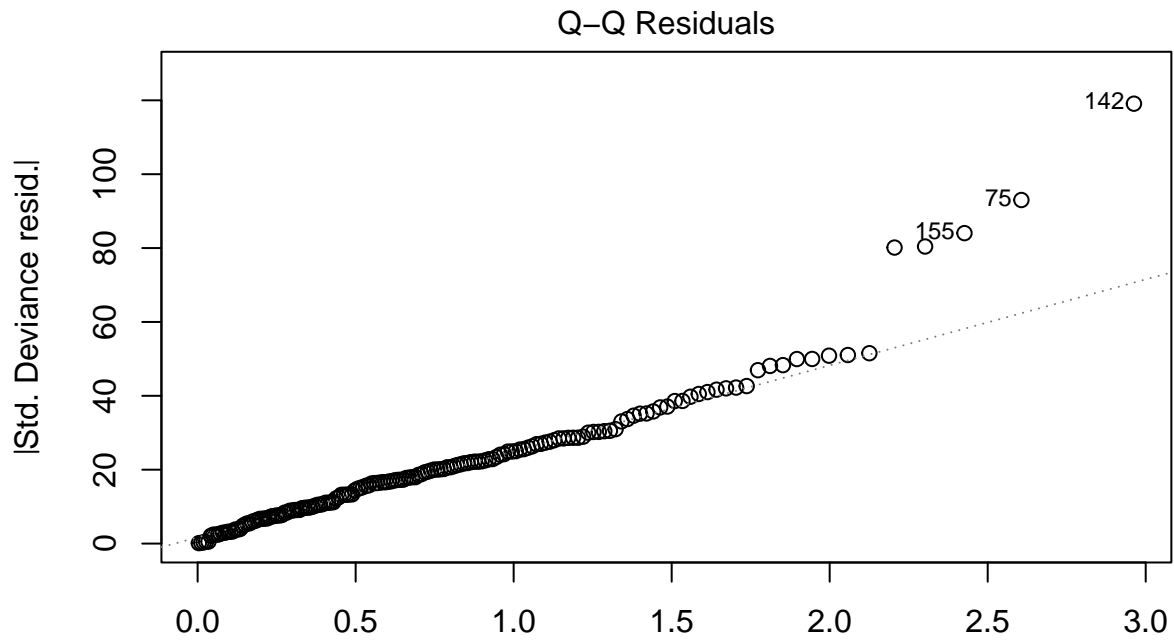
```
## # A tibble: 7 x 5
##   term                estimate std.error statistic  p.value
##   <chr>              <dbl>    <dbl>     <dbl>    <dbl>
## 1 (Intercept)        3.60      0.255      14.1 2.87e-45
```

```
## 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
```

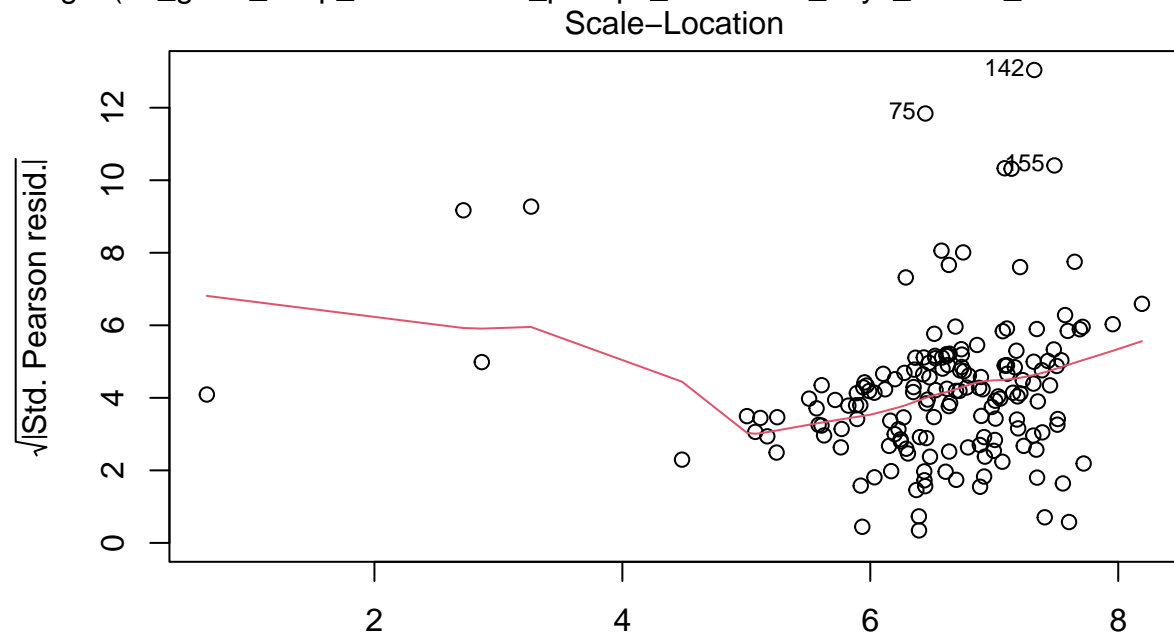
```
model_Nspores10d(df_all)
```

**Histogram of df\$nb\_germ\_oosp\_10d**

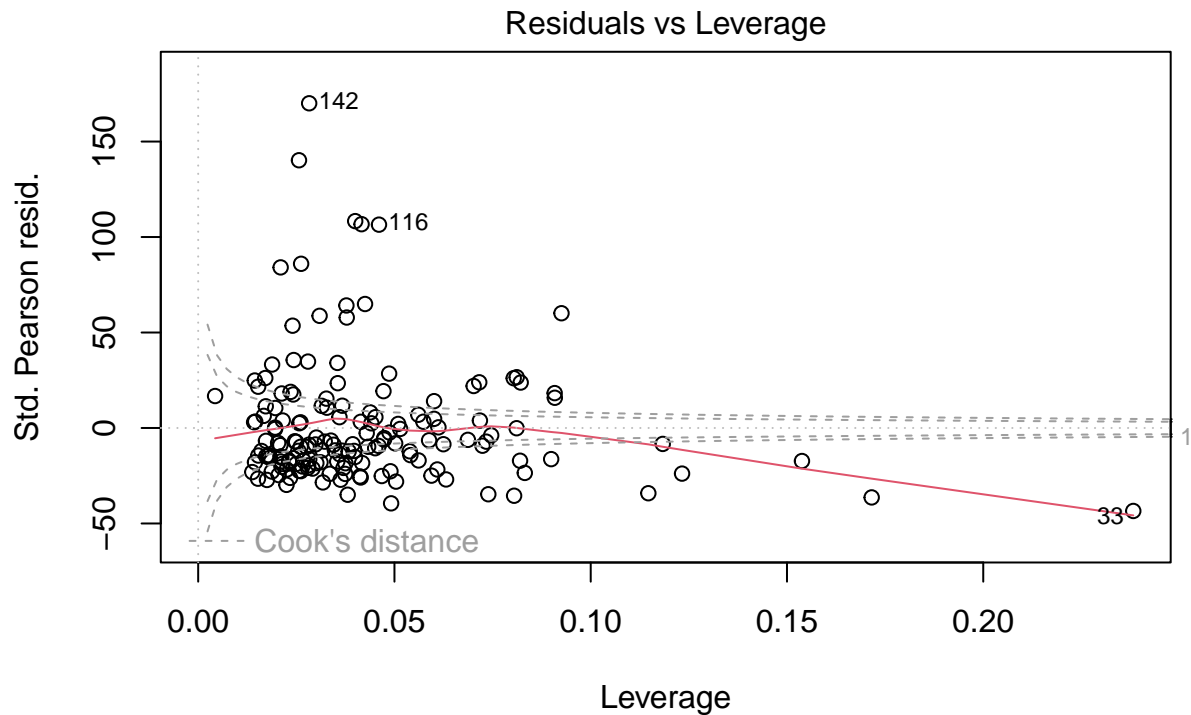




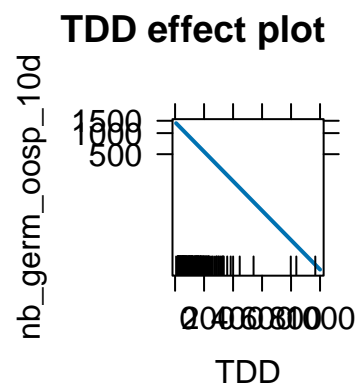
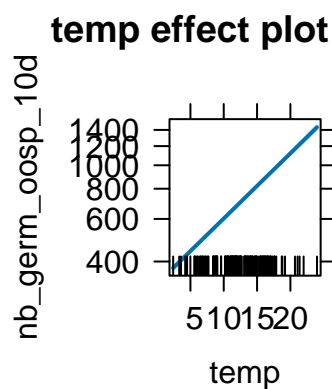
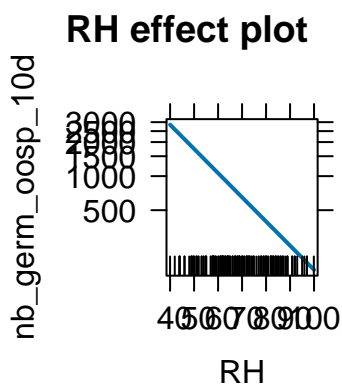
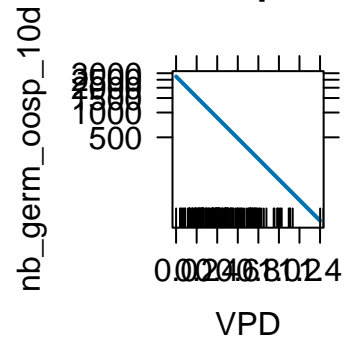
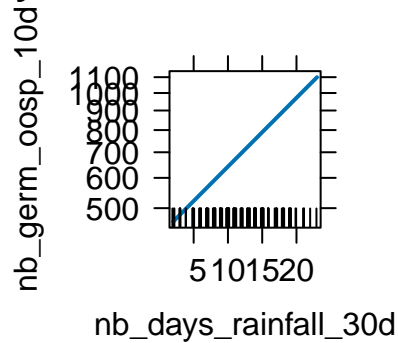
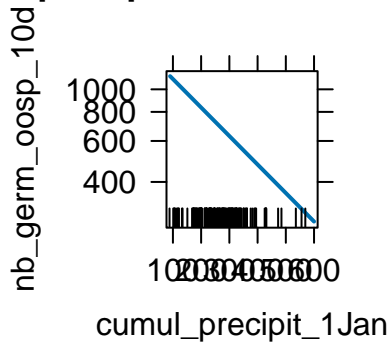
glm(nb\_germ\_oosp\_10d ~ cumul\_precipit\_1Jan + nb\_days\_rainfall\_30d + VPD + R .



glm(nb\_germ\_oosp\_10d ~ cumul\_precipit\_1Jan + nb\_days\_rainfall\_30d + VPD + R .



glm(nb\_germ\_oosp\_10d ~ cumul\_precipit\_1Jan + nb\_days\_rainfall\_30d + VPD + R .  
 cumul\_precipit\_1Jan effect plot nb\_days\_rainfall\_30d effect plot VPD effect plot



```
## # 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
```

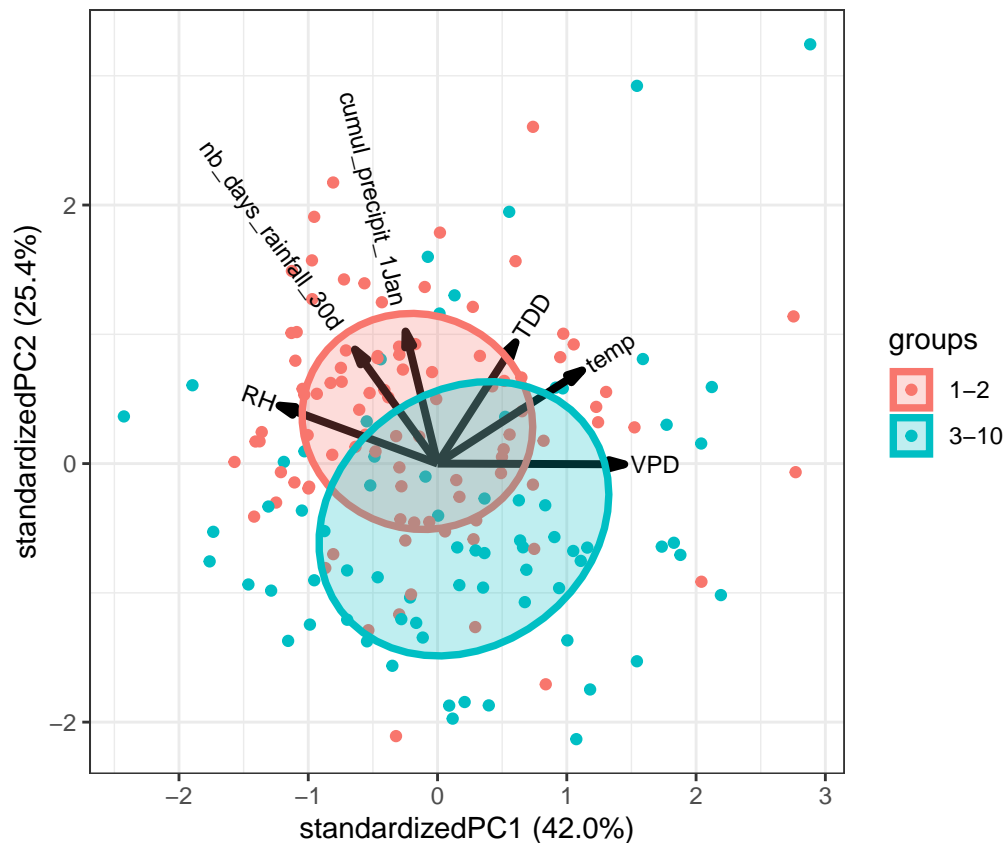
```
pca(df_all)
```

```
## Importance of components:
```

```
##          PC1    PC2    PC3    PC4    PC5    PC6
## Standard deviation 1.5867 1.2334 0.9297 0.8062 0.64221 0.18526
## Proportion of Variance 0.4196 0.2536 0.1441 0.1083 0.06874 0.00572
## Cumulative Proportion 0.4196 0.6732 0.8172 0.9255 0.99428 1.00000
```

```
##          PC1          PC2          PC3          PC4          PC5
## cumul_precipit_1Jan -0.1038664 0.552216020 0.5562542 0.60562463 0.08633395
## nb_days_rainfall_30d -0.2684060 0.476897626 0.3031477 -0.77366900 0.10031175
## VPD 0.6068608 -0.002912689 0.2227853 -0.12603455 -0.01876719
## RH -0.5136219 0.242323314 -0.4105287 0.10962499 -0.44822056
## temp 0.4698884 0.389439568 -0.1131377 -0.05726112 -0.68843392
## TDD 0.2535496 0.507182287 -0.6063955 0.05893983 0.55433641
```

```
##          PC6
## cumul_precipit_1Jan -0.0248137065
## nb_days_rainfall_30d -0.0014767400
## VPD -0.7522242632
## RH -0.5440760981
## temp 0.3708388455
## TDD -0.0007121924
```



```

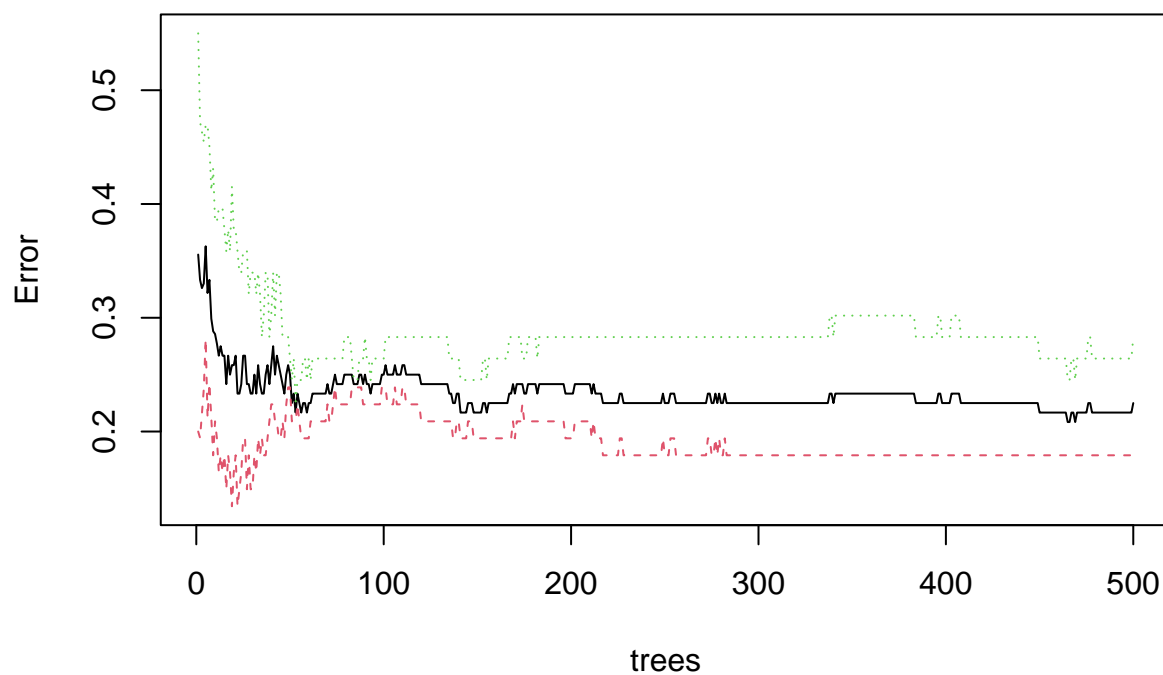
random_forest(df_all)

## Loading required package: randomForest
## randomForest 4.7-1.2
## Type rfNews() to see new features/changes/bug fixes.
##
## Attaching package: 'randomForest'
## The following object is masked from 'package:ggplot2':
##
##     margin
## The following object is masked from 'package:dplyr':
##
##     combine
## Loading required package: caret
## Loading required package: lattice
##
## Call:
## randomForest(formula = MTG_cat ~ ., data = train, proximity = TRUE,      mtry = 3)
##           Type of random forest: classification
##           Number of trees: 500
## No. of variables tried at each split: 3
##
##           OOB estimate of  error rate: 22.5%
## Confusion matrix:
##           1-2 3-10 class.error
## 1-2      55   12   0.1791045
## 3-10     15   38   0.2830189

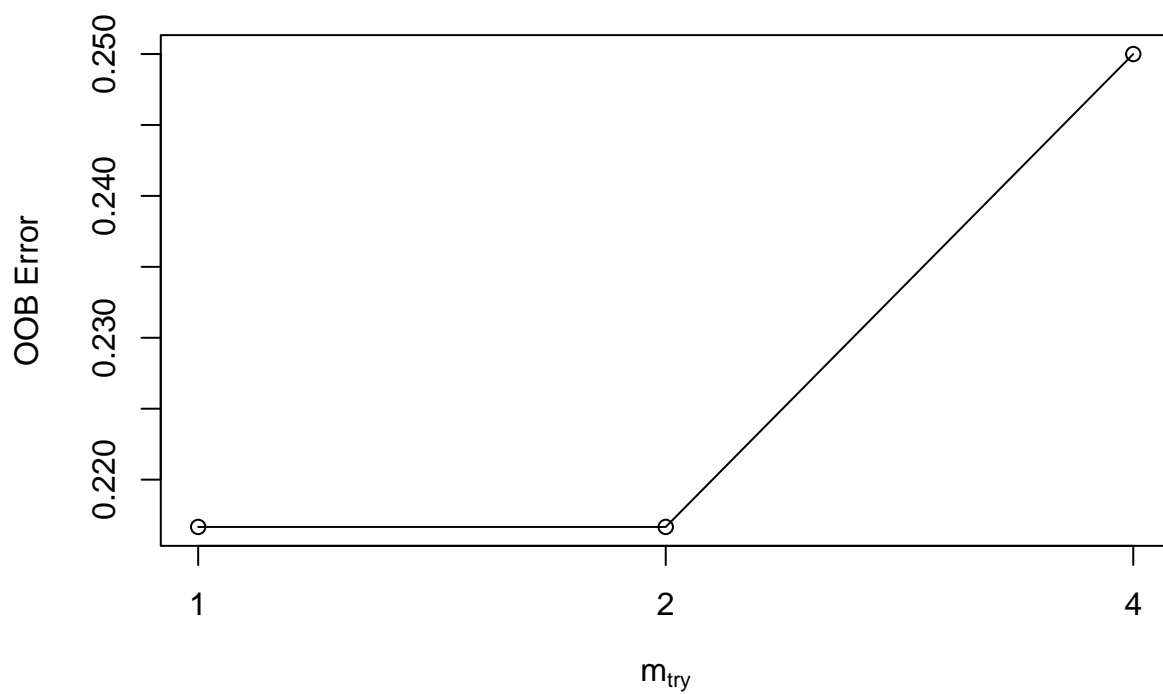
```



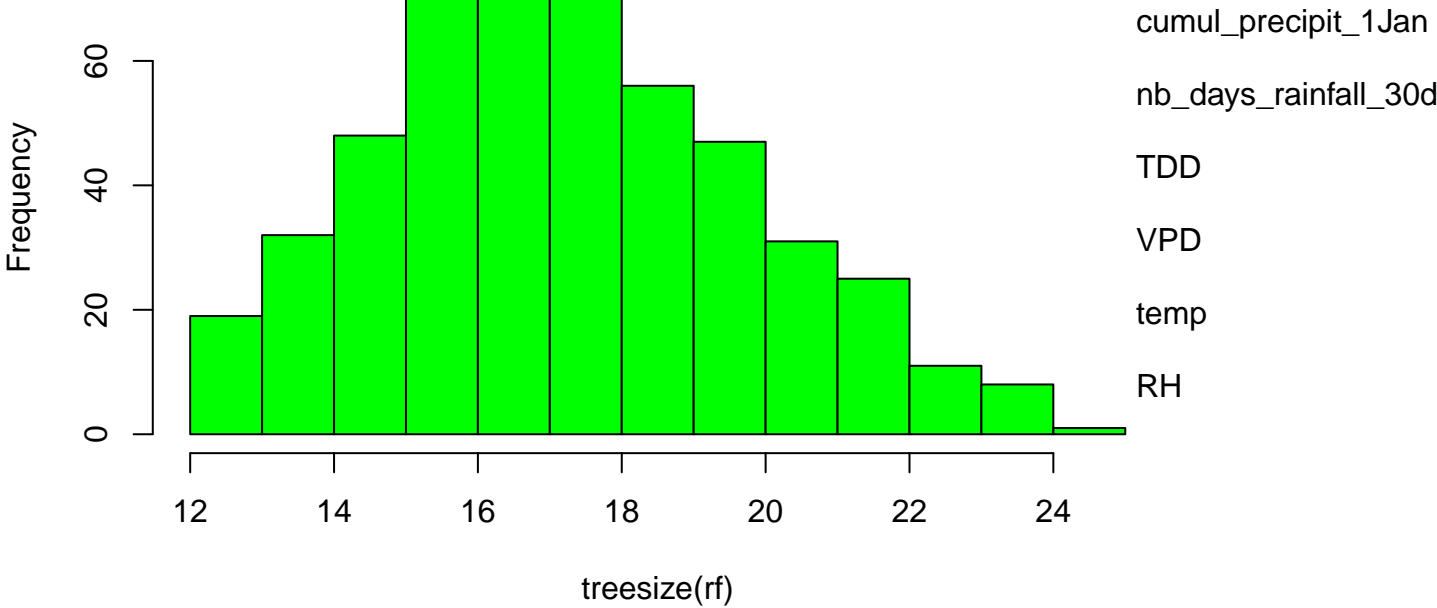
rf



```
## mtry = 2  OOB error = 21.67%
## Searching left ...
## mtry = 4    OOB error = 25%
## -0.1538462 0.05
## Searching right ...
## mtry = 1    OOB error = 21.67%
## 0 0.05
```

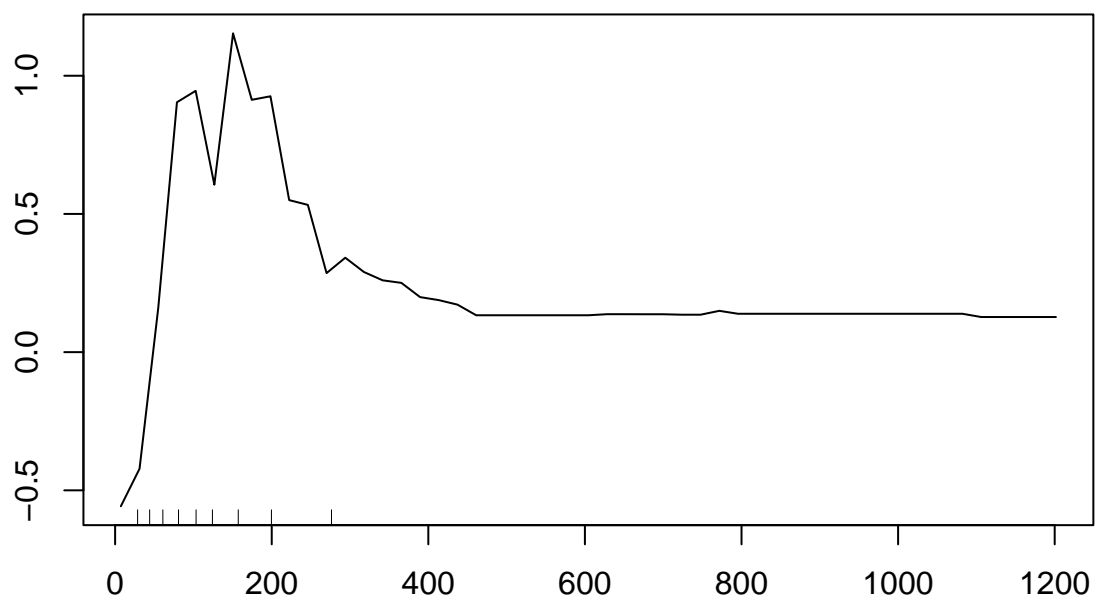


No. of Nodes for the Trees

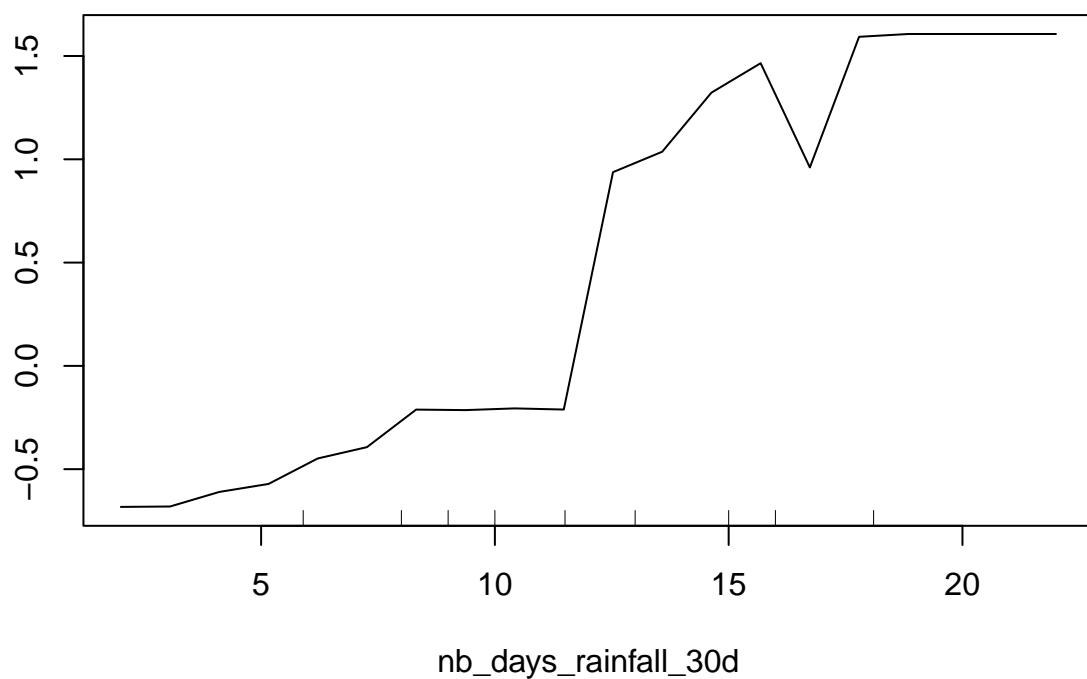


##	MeanDecreaseGini
## cumul_precipit_1Jan	14.864343
## nb_days_rainfall_30d	14.701057
## VPD	6.959436
## RH	5.627265
## temp	6.163640
## TDD	10.338158

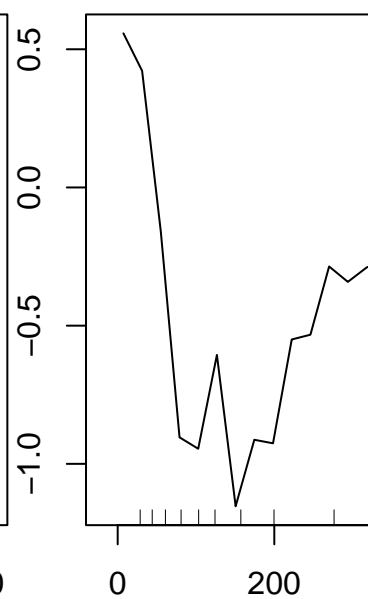
**Partial Dependence on TDD**



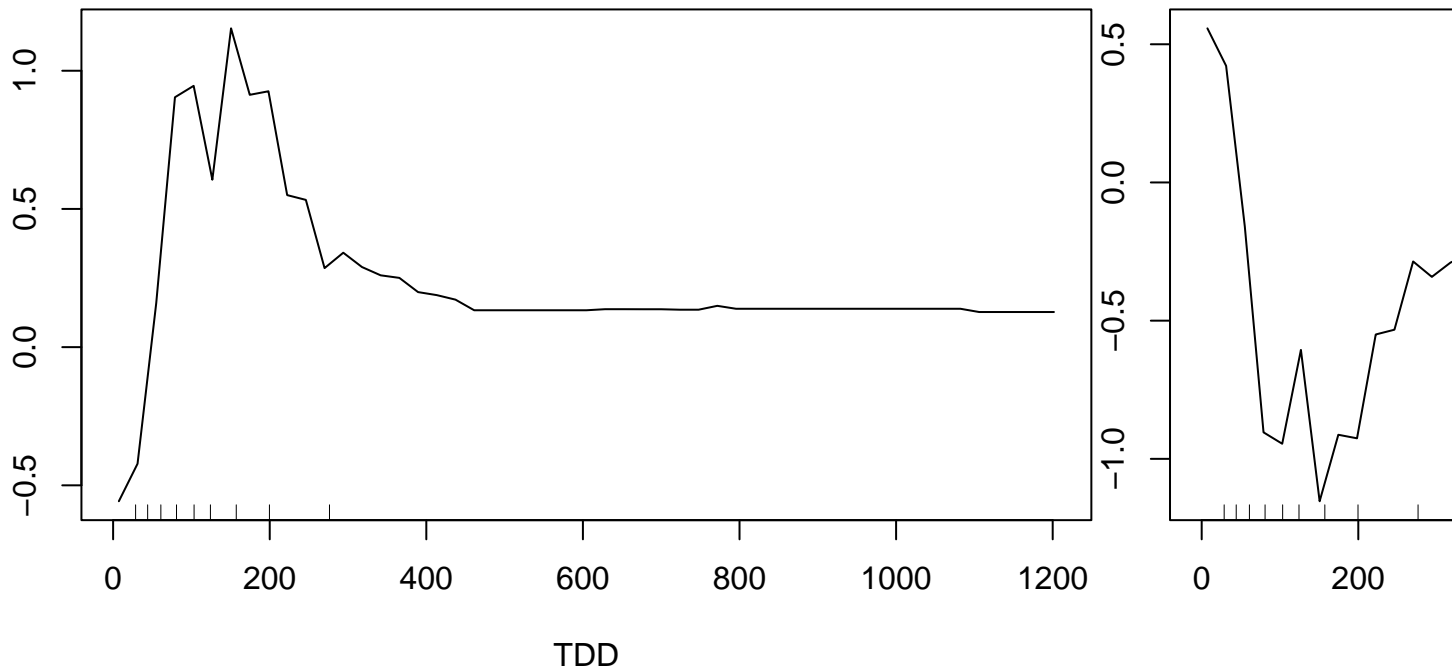
**Partial Dependence on nb\_days\_rainfall\_30d**



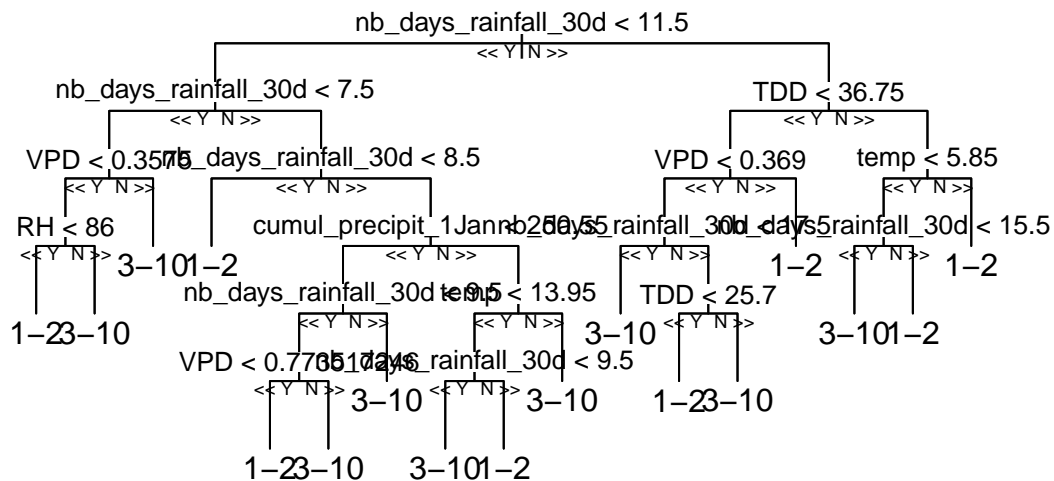
**Partial De**



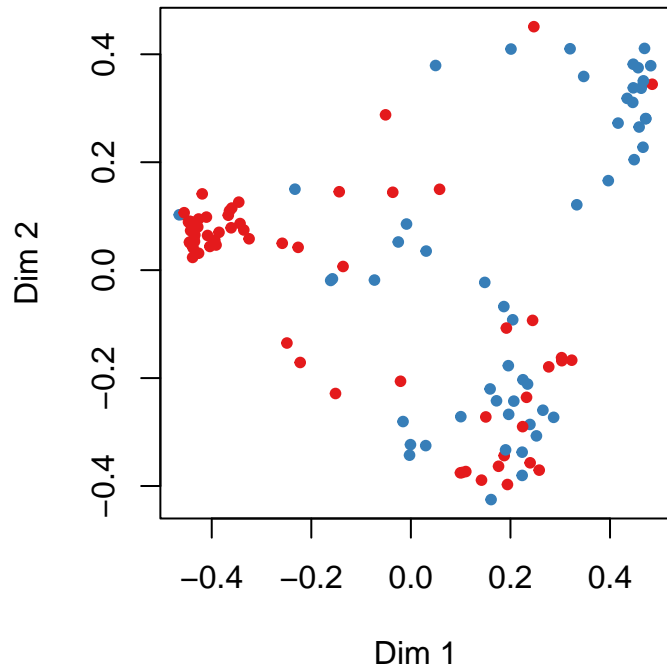
## Partial Dependence on TDD



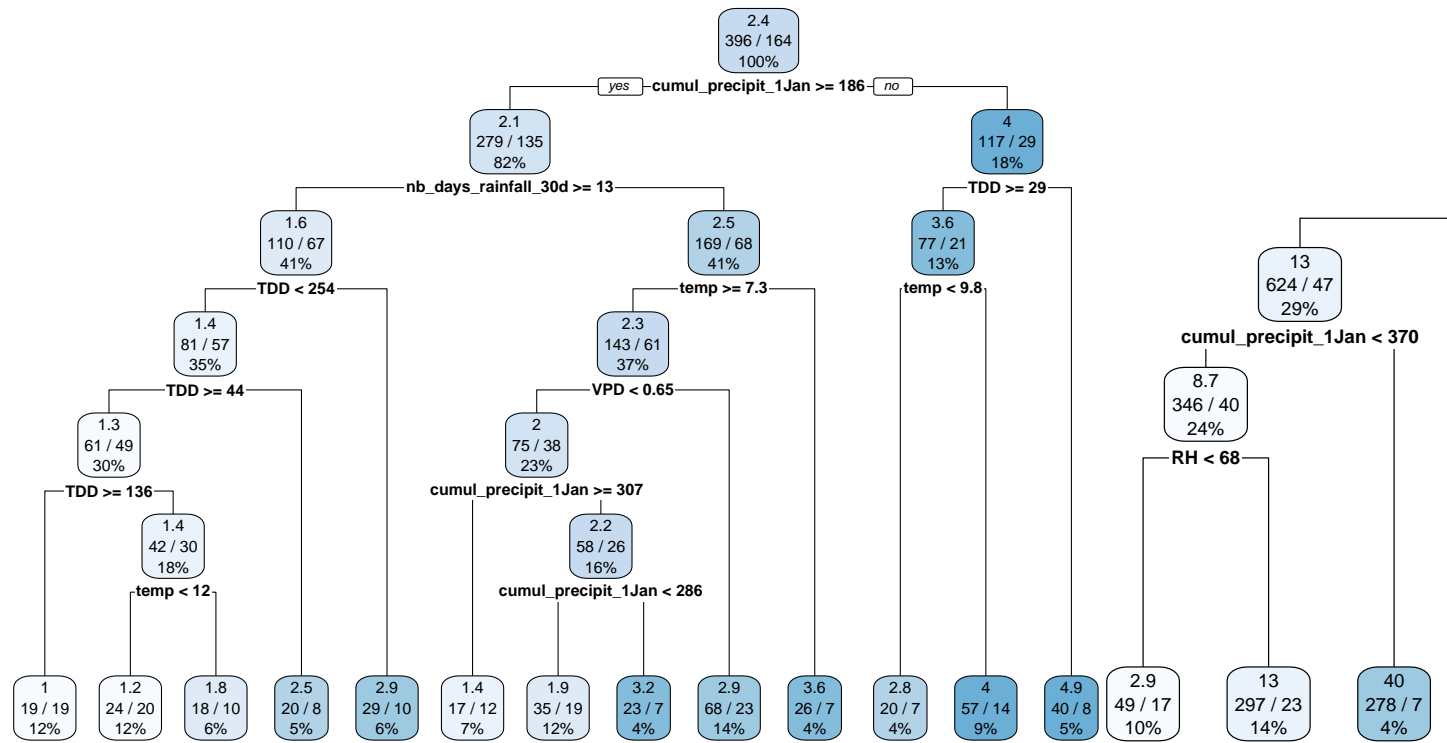
```
## Loading required package: devtools
## Loading required package: usethis
## Skipping install of 'reprtree' from a github remote, the SHA1 (7ebb9ff7) has not changed since last :
##   Use `force = TRUE` to force installation
## Loading required package: tree
## Loading required package: plotrix
## Registered S3 method overwritten by 'reprtree':
##   method      from
##   text.tree    tree
```

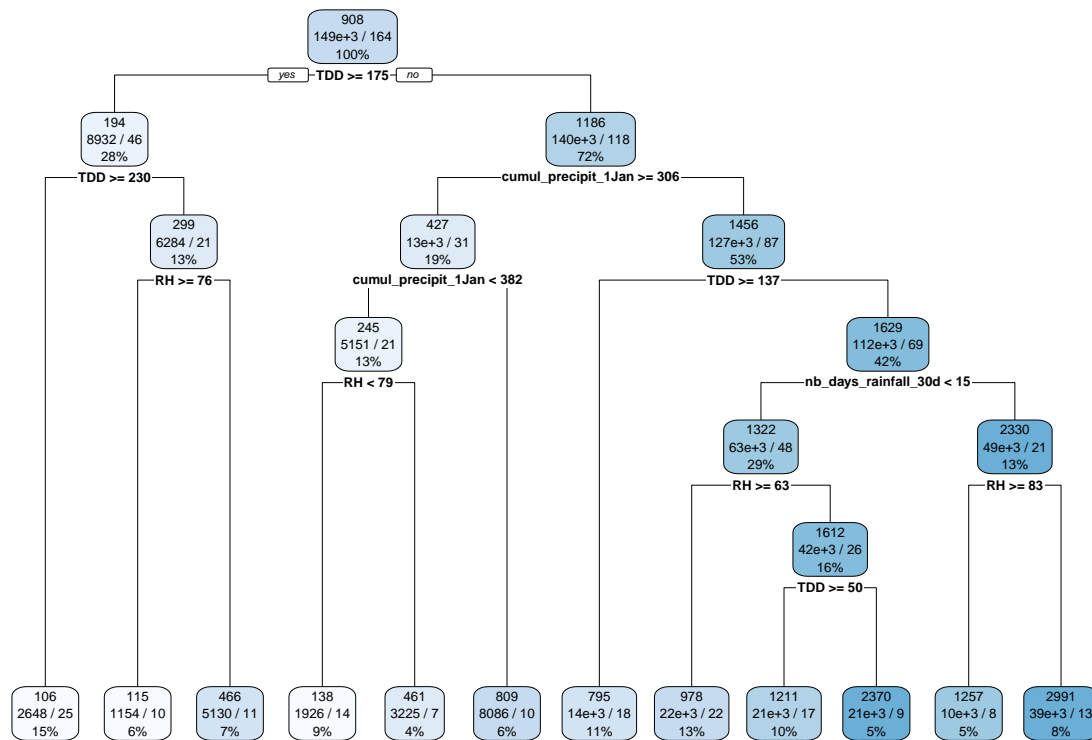


```
## Warning in RColorBrewer::brewer.pal(nlevs, "Set1"): minimal value for n is 3, returning requested palette
```



```
## Loading required package: rpart
## Loading required package: rpart.plot
```



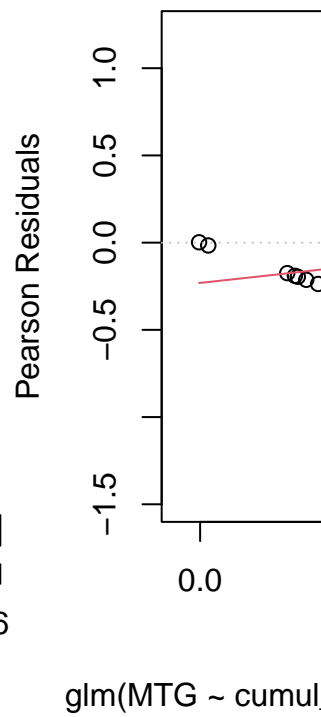
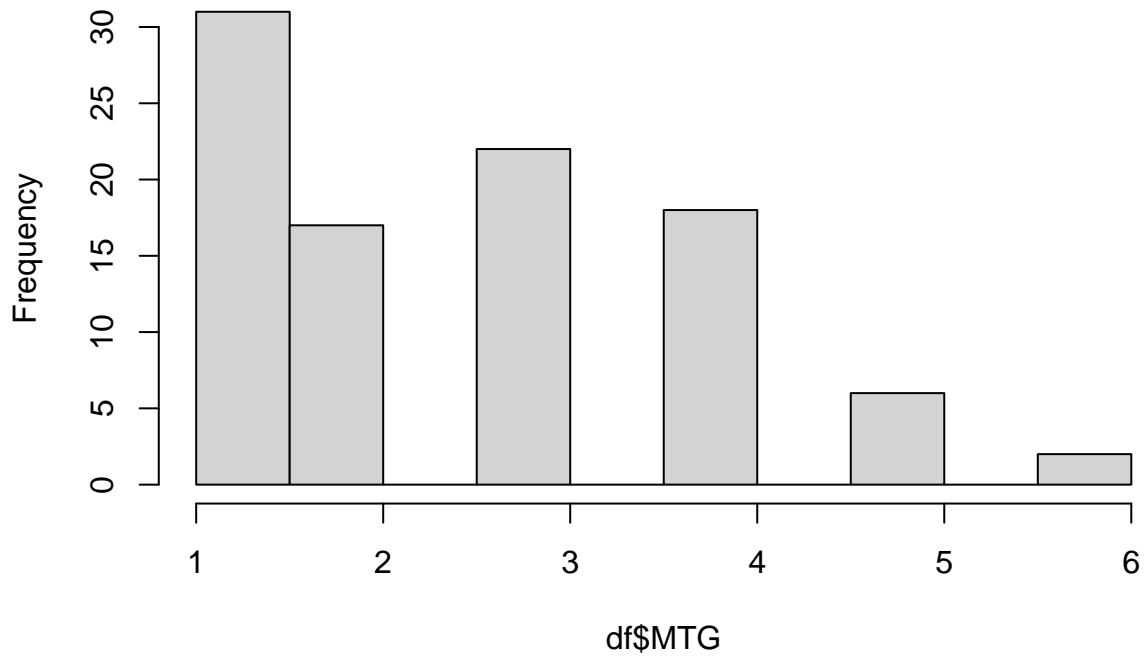


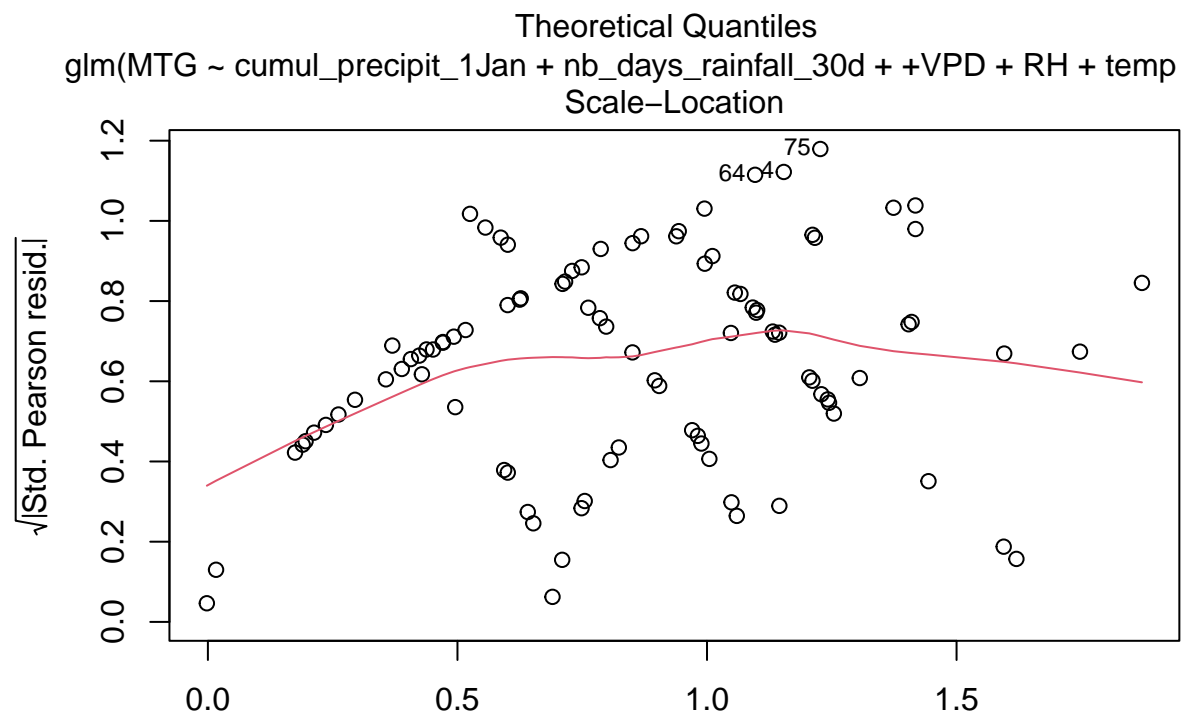
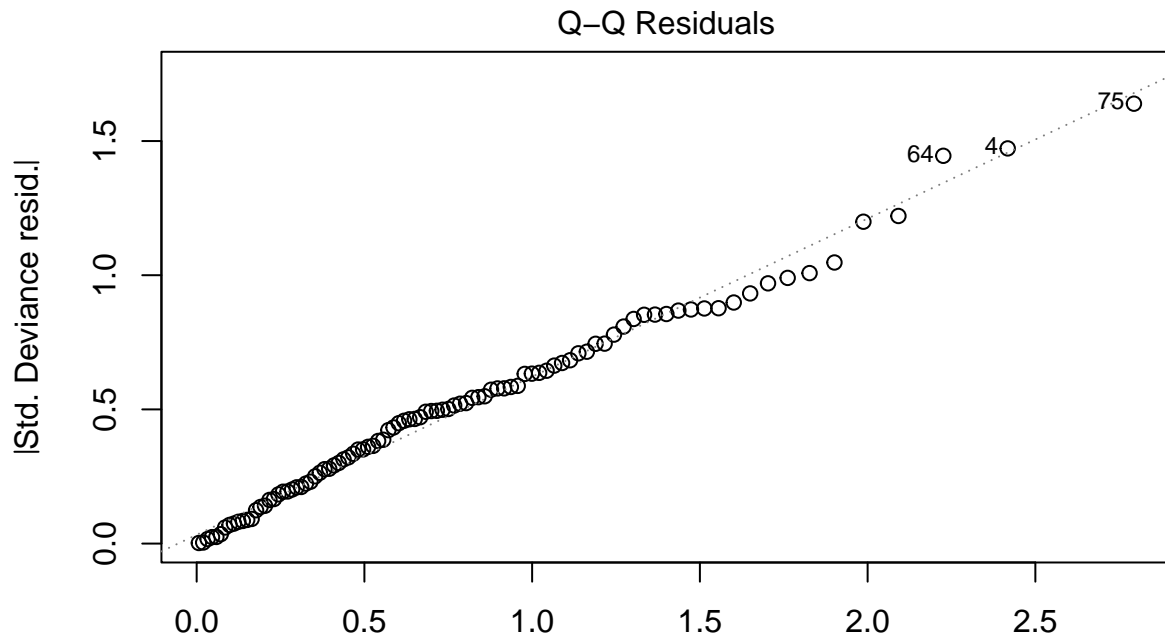
## DATASET BBCH 0:12

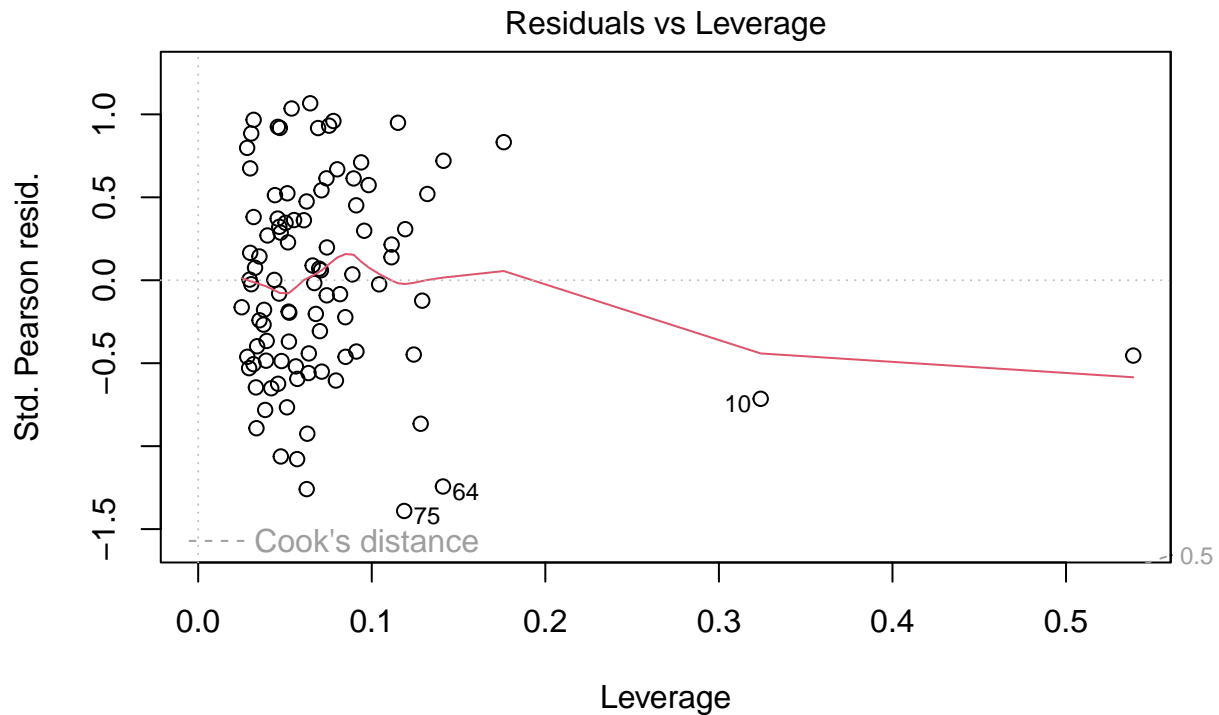
```
df <- df_all %>% filter(df_all$BBCH < 13)
```

```
model_MGT(df)
```

## Histogram of df\$MTG

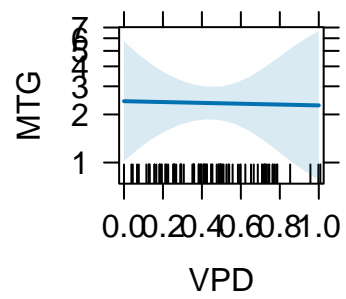
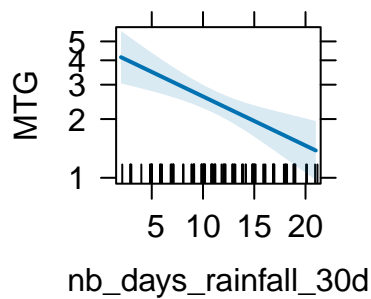
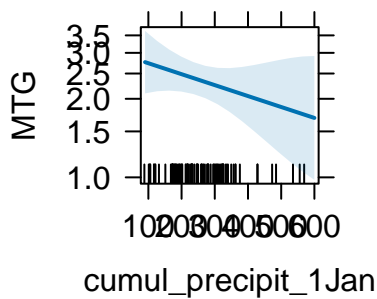






glm(MTG ~ cumul\_precipit\_1Jan + nb\_days\_rainfall\_30d + VPD + RH + temp + TDD)

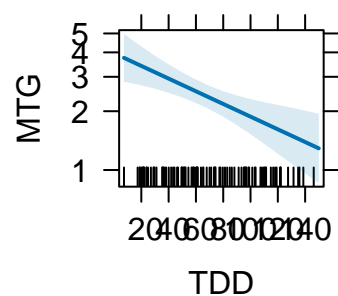
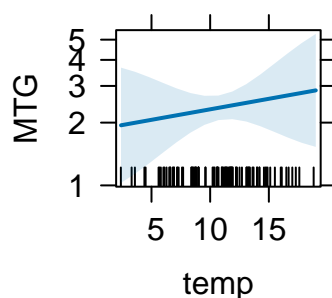
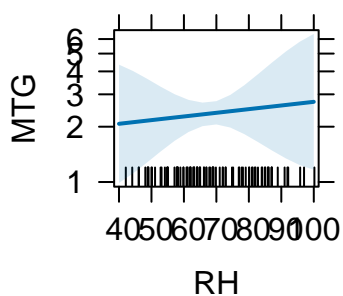
cumul\_precipit\_1Jan effect plot nb\_days\_rainfall\_30d effect plot VPD effect plot



**RH effect plot**

**temp effect plot**

**TDD effect plot**



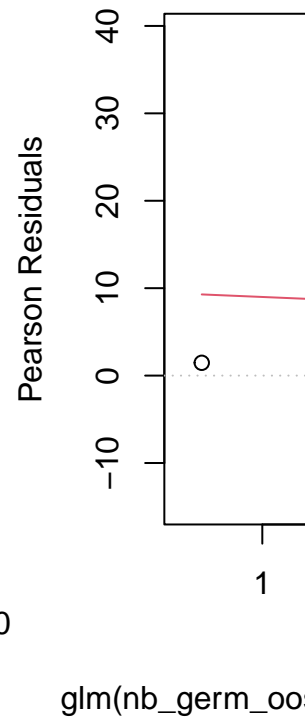
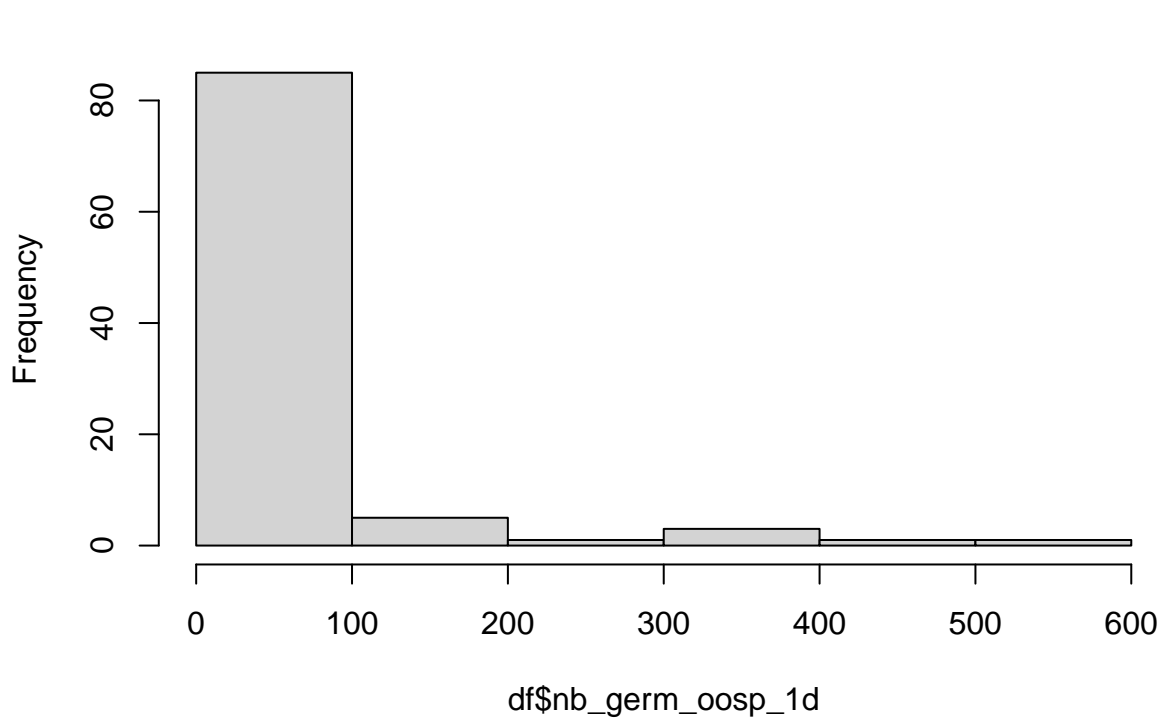
```
## # A tibble: 7 x 5
##   term                estimate std.error statistic  p.value
##   <chr>              <dbl>    <dbl>     <dbl>    <dbl>
## 1 (Intercept)        1.77      1.10      1.62    0.106
```

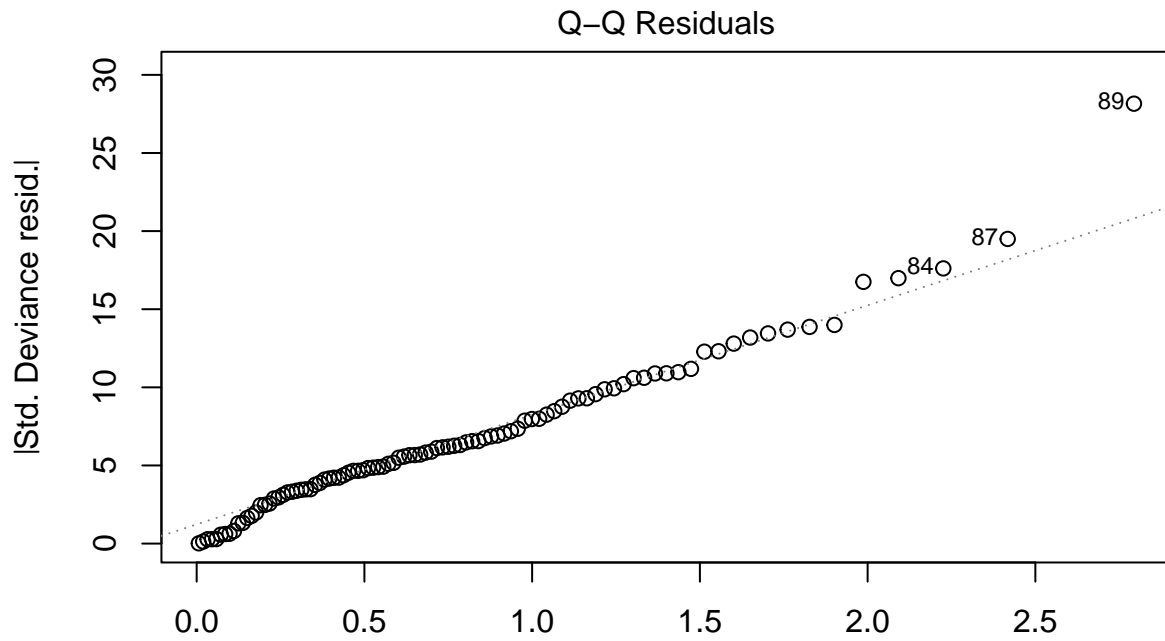


```
## 2 cumul_precipit_1Jan -0.000966 0.000772 -1.25 0.211
## 3 nb_days_rainfall_30d -0.0579 0.0162 -3.58 0.000348
## 4 VPD -0.0607 0.980 -0.0620 0.951
## 5 RH 0.00459 0.0134 0.344 0.731
## 6 temp 0.0232 0.0379 0.613 0.540
## 7 TDD -0.00747 0.00227 -3.29 0.00101
```

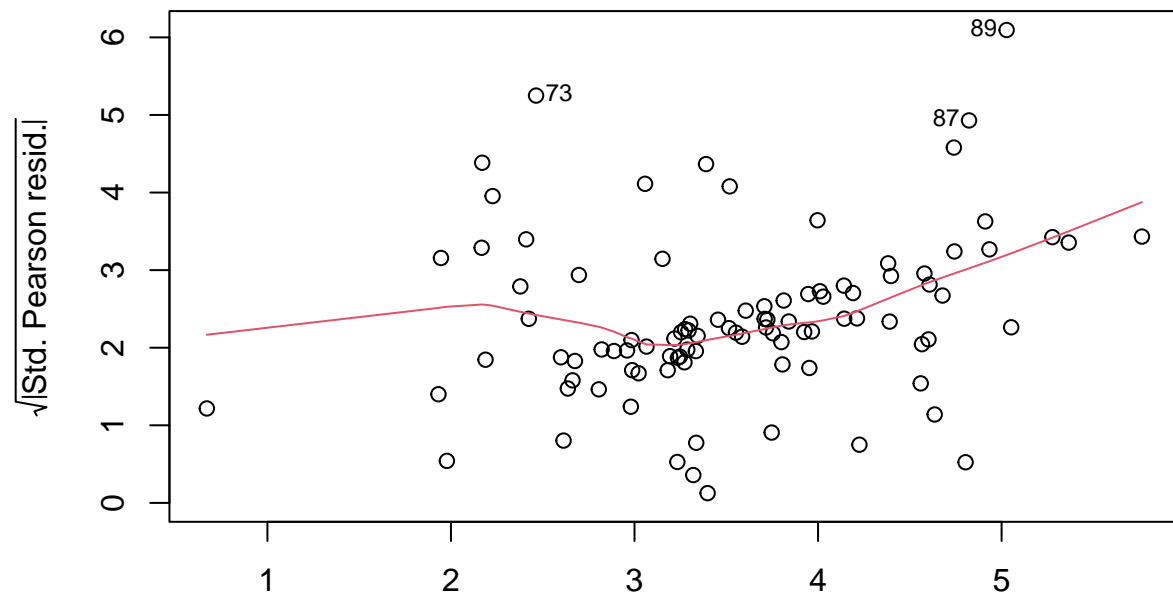
```
model_Nspores1d(df)
```

**Histogram of df\$nb\_germ\_oosp\_1d**

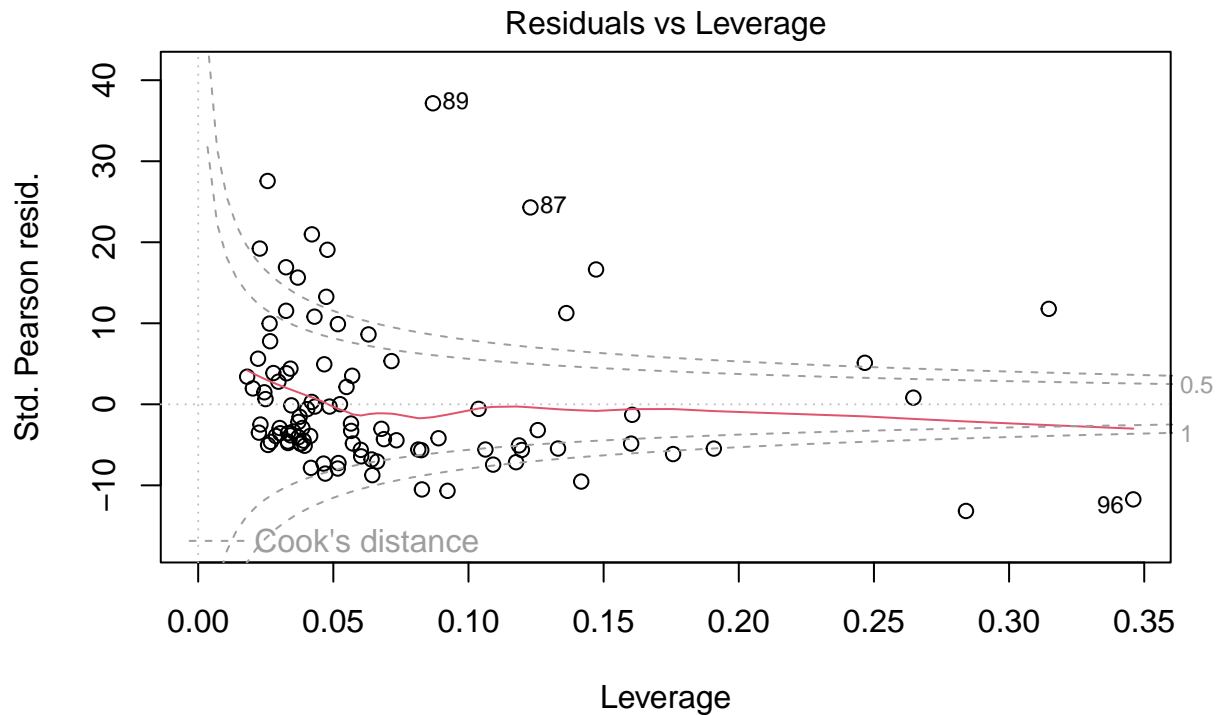




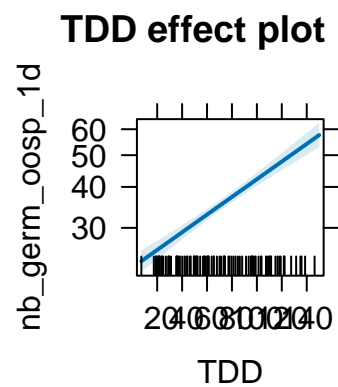
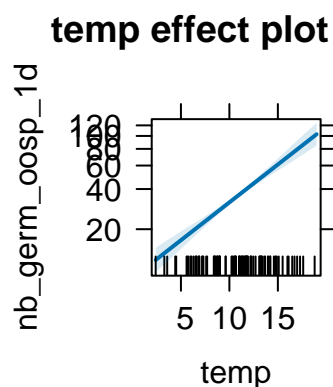
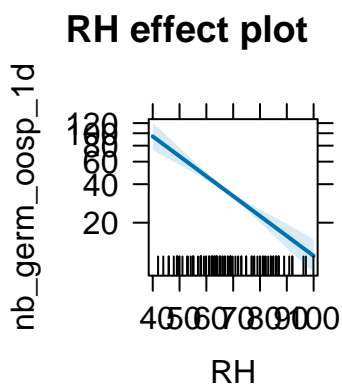
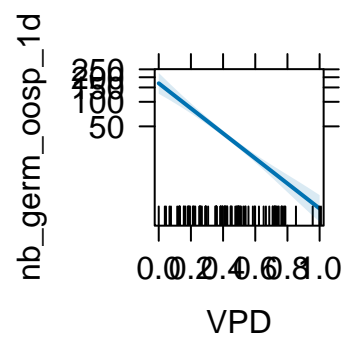
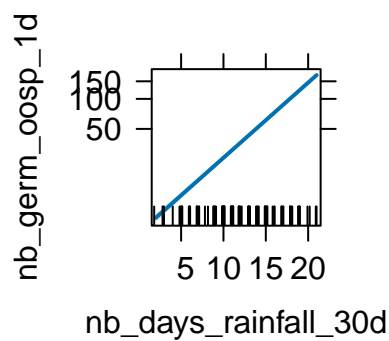
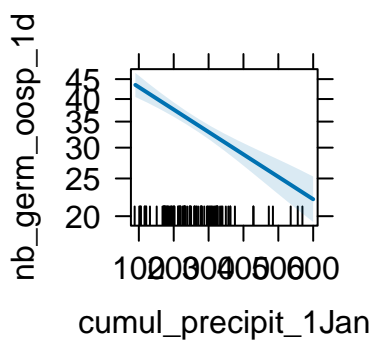
glm(nb\_germ\_oosp\_1d ~ cumul\_precipit\_1Jan + nb\_days\_rainfall\_30d + VPD + RH)  
Scale-Location



glm(nb\_germ\_oosp\_1d ~ cumul\_precipit\_1Jan + nb\_days\_rainfall\_30d + VPD + RH)



glm(nb\_germ\_oosp\_1d ~ cumul\_precipit\_1Jan + nb\_days\_rainfall\_30d + VPD + RH +  
 cumul\_precipit\_1Jan effect plot nb\_days\_rainfall\_30d effect plot VPD effect plot

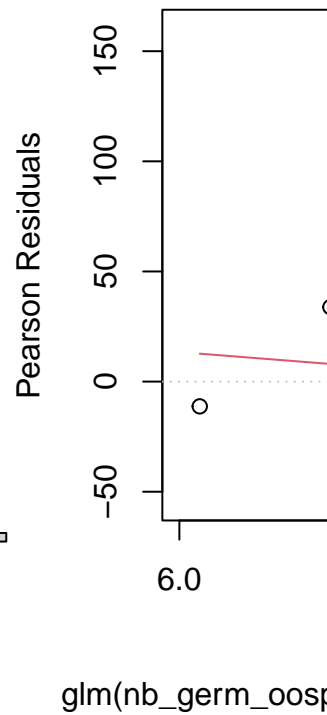
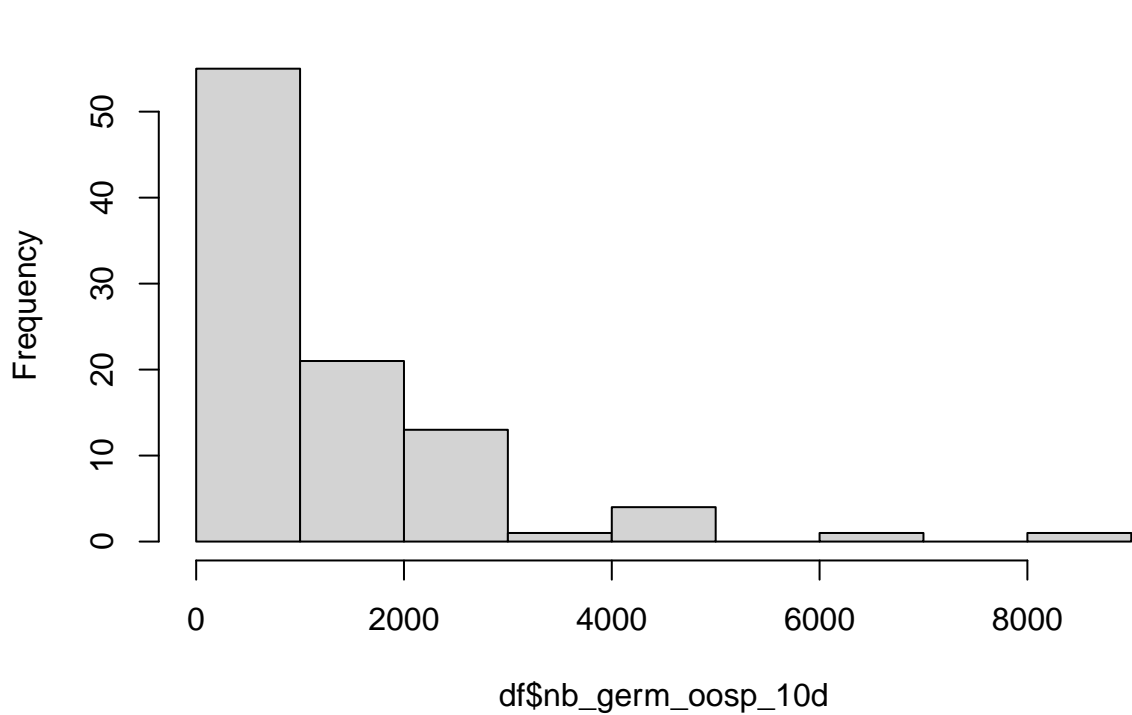


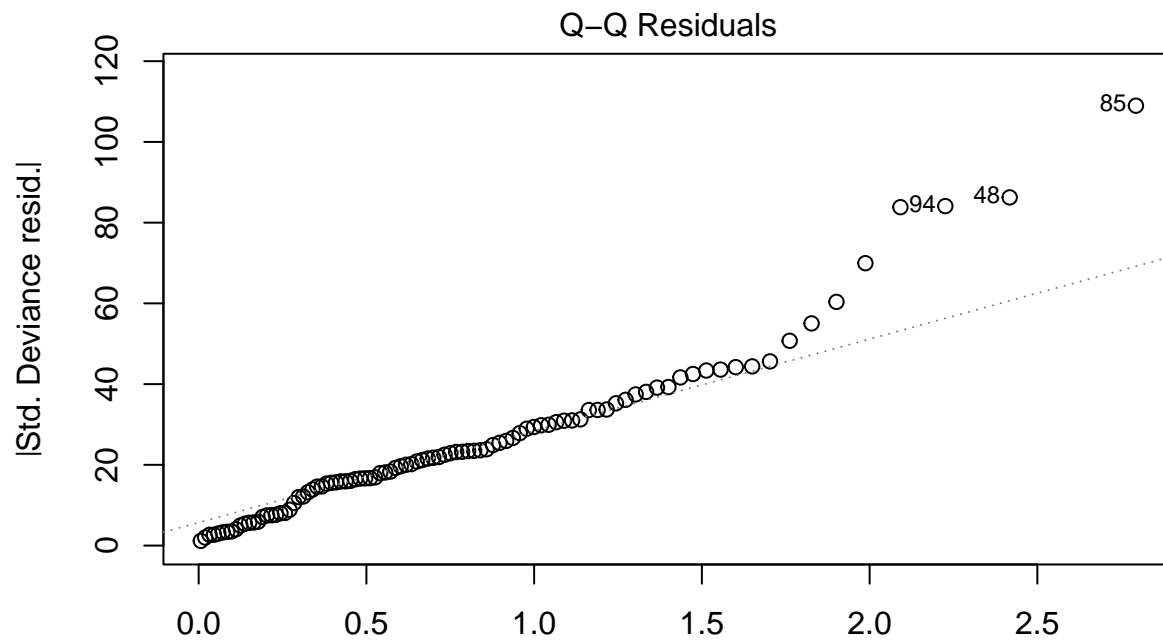
```
## # A tibble: 7 x 5
##   term                estimate std.error statistic  p.value
##   <chr>              <dbl>    <dbl>    <dbl>    <dbl>
## 1 (Intercept)        4.02      0.358     11.2 2.46e-29
```

```
## 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
```

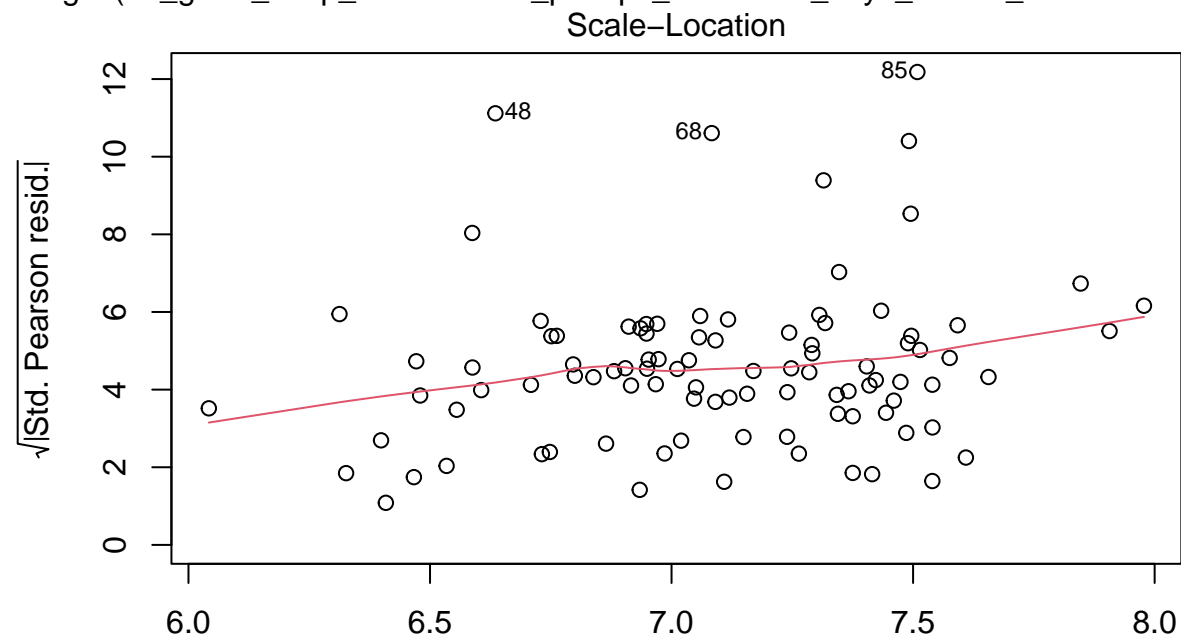
```
model_Nspores10d(df)
```

**Histogram of df\$nb\_germ\_oosp\_10d**

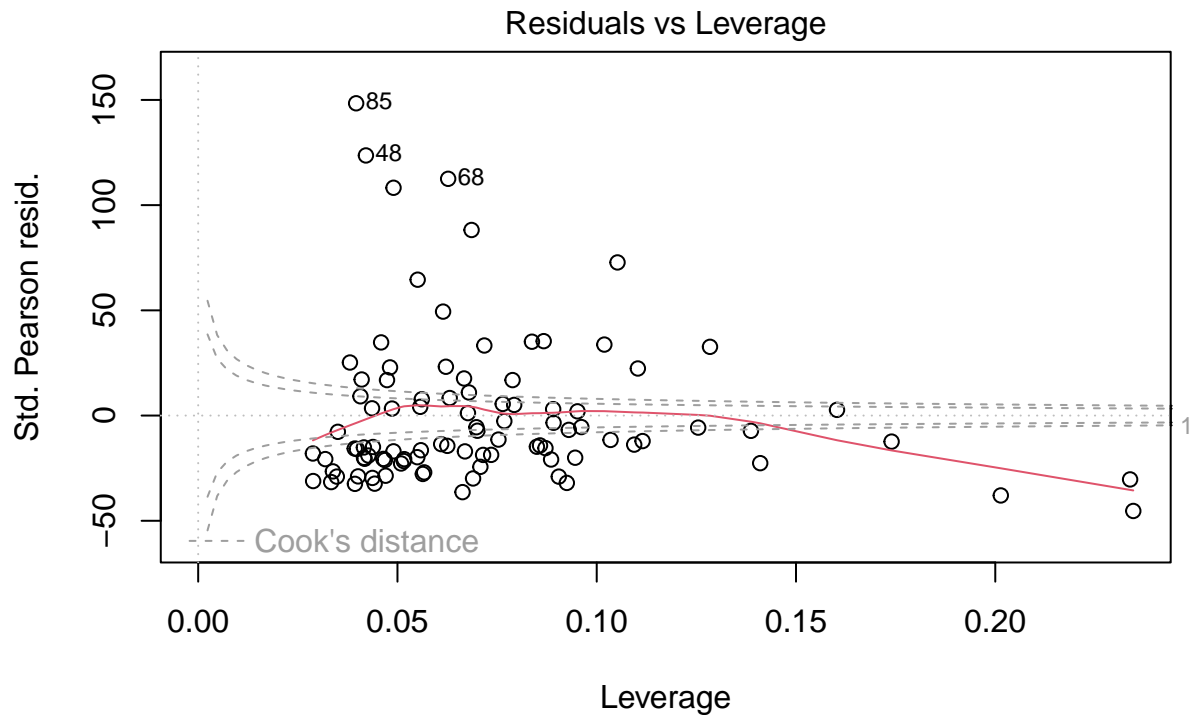




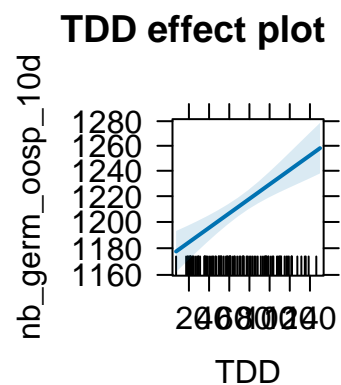
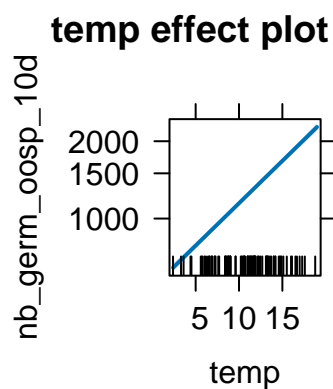
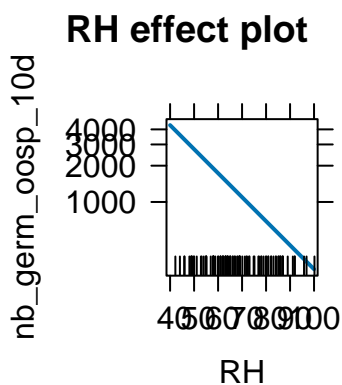
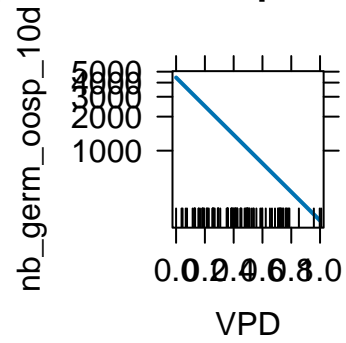
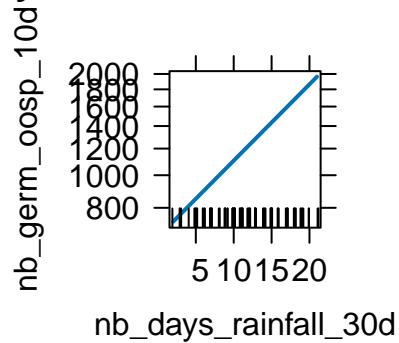
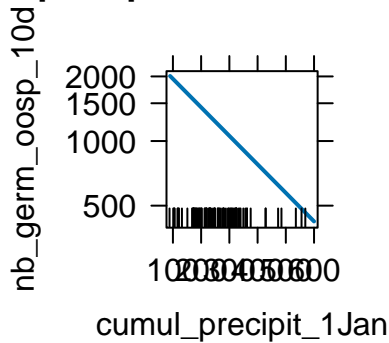
glm(nb\_germ\_oosp\_10d ~ cumul\_precipit\_1Jan + nb\_days\_rainfall\_30d + VPD + R .



glm(nb\_germ\_oosp\_10d ~ cumul\_precipit\_1Jan + nb\_days\_rainfall\_30d + VPD + R .



glm(nb\_germ\_oosp\_10d ~ cumul\_precipit\_1Jan + nb\_days\_rainfall\_30d + VPD + R .  
 cumul\_precipit\_1Jan effect plot nb\_days\_rainfall\_30d effect plot VPD effect plot



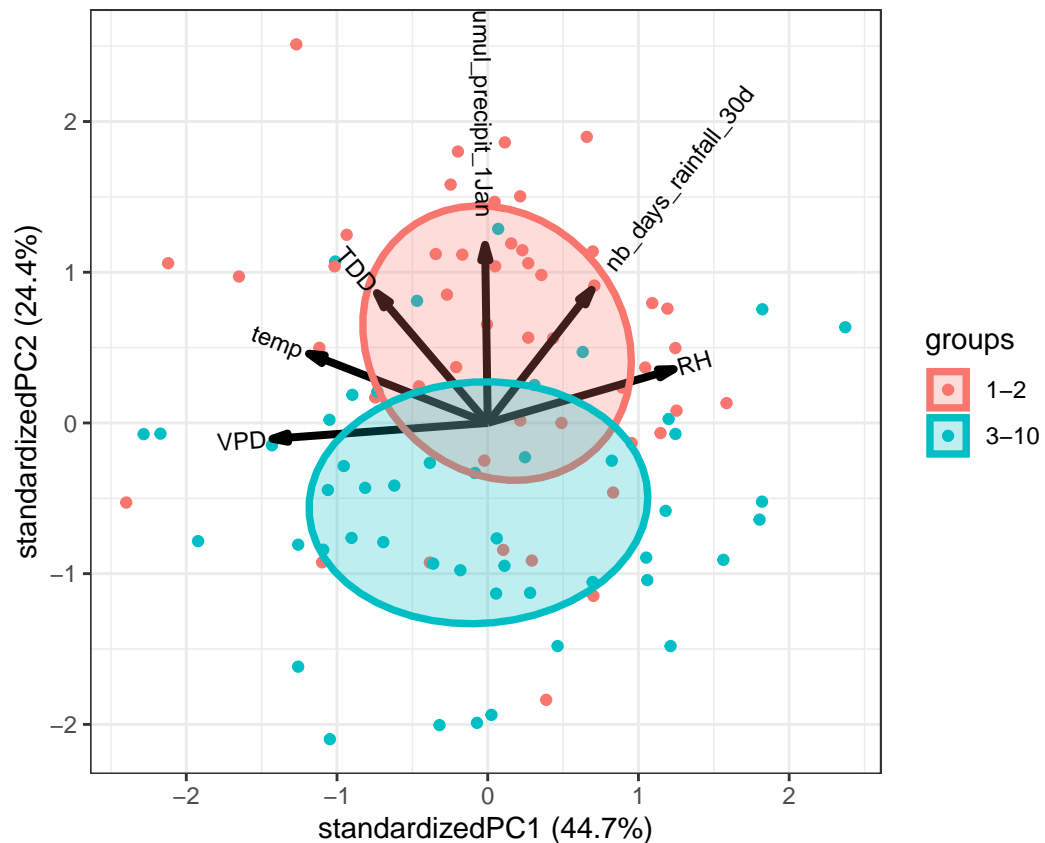
```
## # A tibble: 7 x 5
##   term          estimate std.error statistic    p.value
##   <chr>          <dbl>    <dbl>    <dbl>    <dbl>
## 1 (Intercept)    10.8      0.0623    174.      0
```

```
## 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)
```

```
## Importance of components:
```

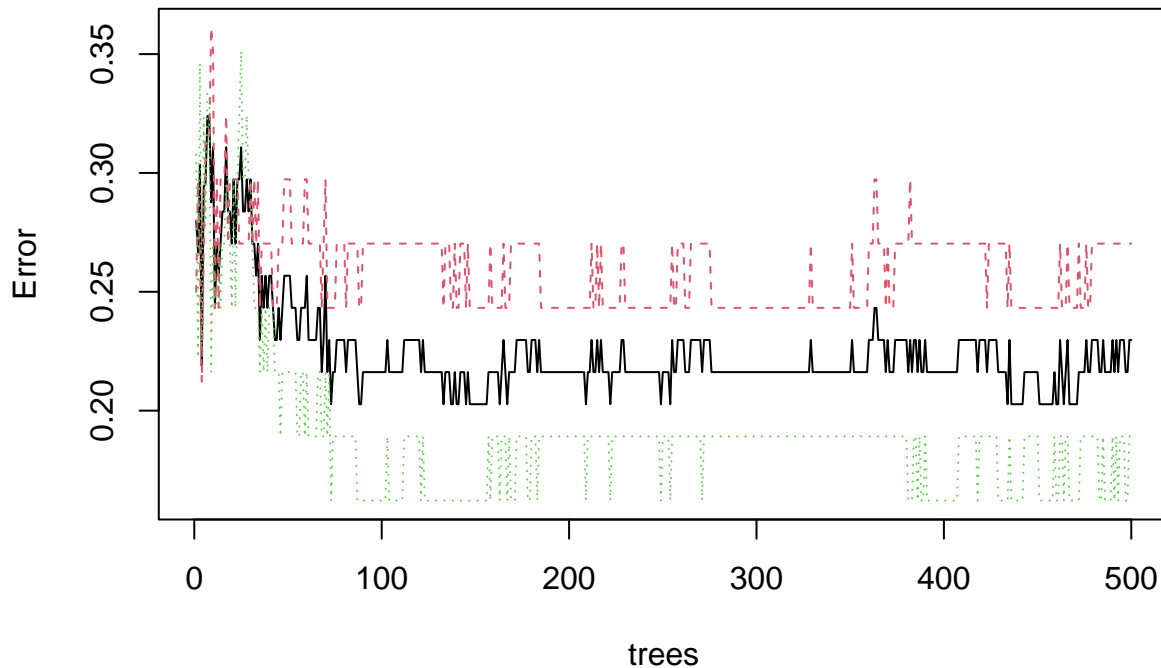
```
##          PC1    PC2    PC3    PC4    PC5    PC6
## Standard deviation  1.6379 1.2106 0.8839 0.7777 0.65931 0.17559
## Proportion of Variance 0.4471 0.2442 0.1302 0.1008 0.07245 0.00514
## Cumulative Proportion 0.4471 0.6914 0.8216 0.9224 0.99486 1.00000
##          PC1    PC2    PC3    PC4
## cumul_precipit_1Jan -0.007127124 0.65356359 -0.3992252 0.63981644
## nb_days_rainfall_30d 0.281886996 0.48853867 -0.3831484 -0.72696270
## VPD -0.583691549 -0.05809163 -0.2658824 -0.10922479
## RH 0.504439125 0.19741174 0.4599211 0.05363645
## temp -0.484689275 0.25422327 0.2046934 -0.21041246
## TDD -0.300683514 0.47666456 0.6080266 -0.05549627
##          PC5    PC6
## cumul_precipit_1Jan -0.06329066 0.007230953
## nb_days_rainfall_30d 0.07860160 -0.020346757
## VPD -0.04752153 -0.755677022
## RH -0.43959307 -0.546738898
## temp -0.69767659 0.357101434
## TDD 0.55458371 -0.045178590
```



```
random_forest(df)
```

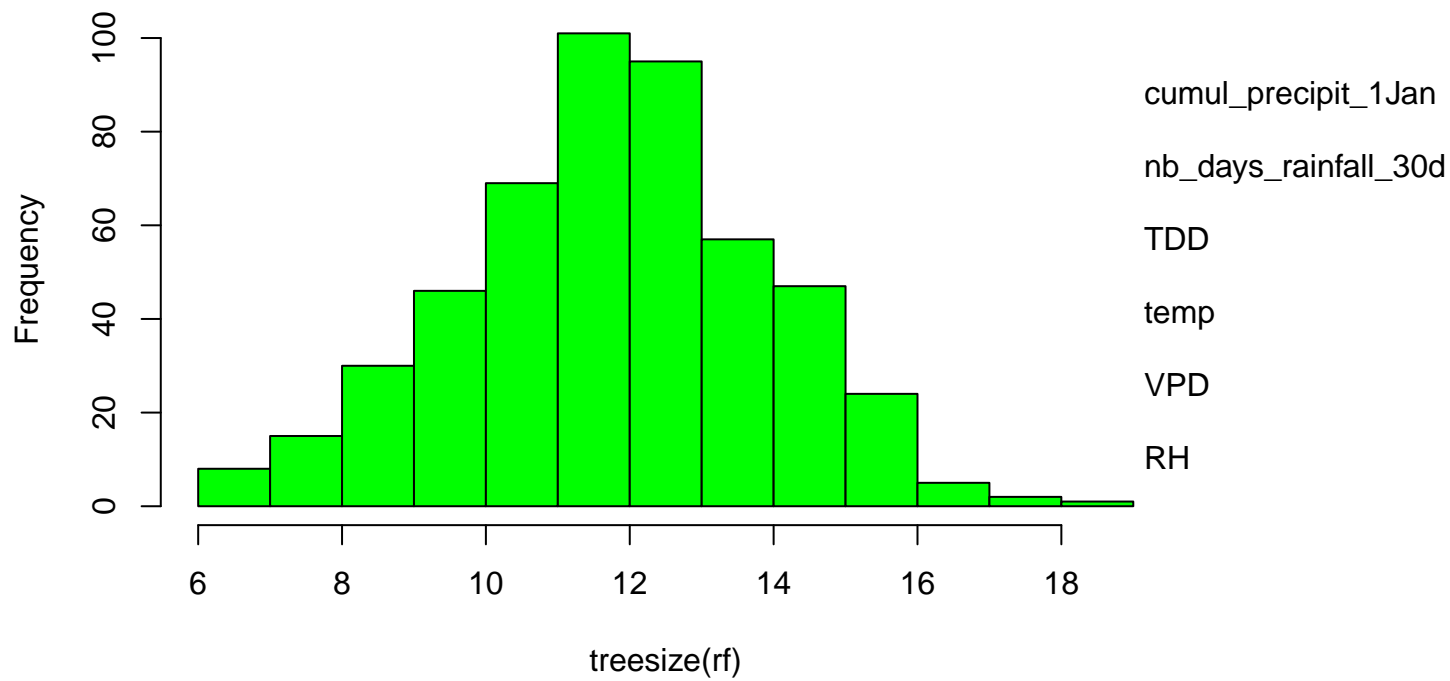
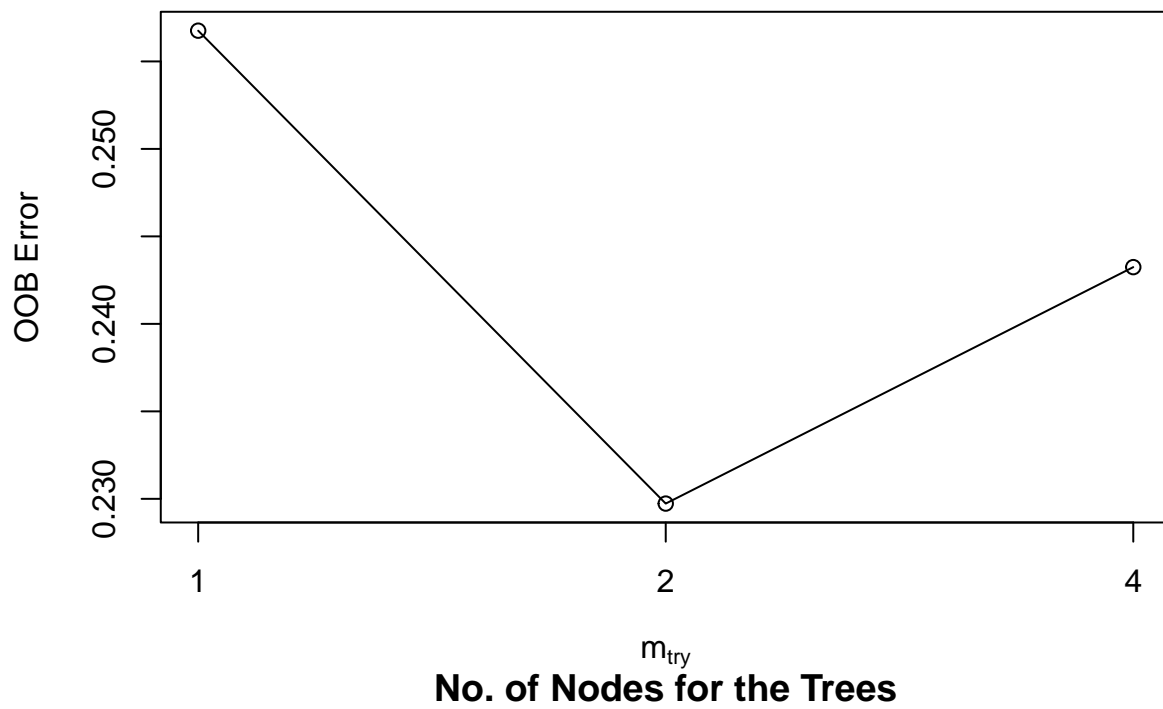
```
##
## Call:
## randomForest(formula = MTG_cat ~ ., data = train, proximity = TRUE,      mtry = 3)
##           Type of random forest: classification
##           Number of trees: 500
## No. of variables tried at each split: 3
##
##           OOB estimate of  error rate: 22.97%
## Confusion matrix:
##      1-2 3-10 class.error
## 1-2   27   10  0.2702703
## 3-10   7   30  0.1891892
```

**rf**



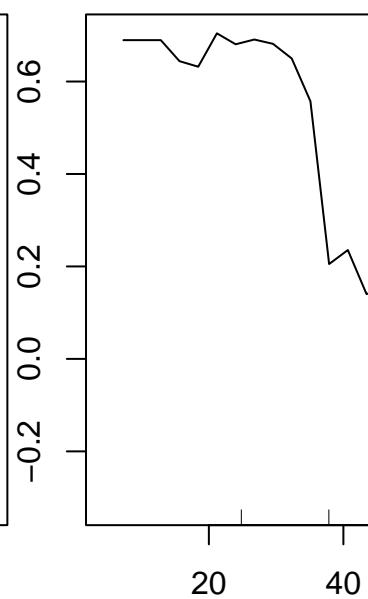
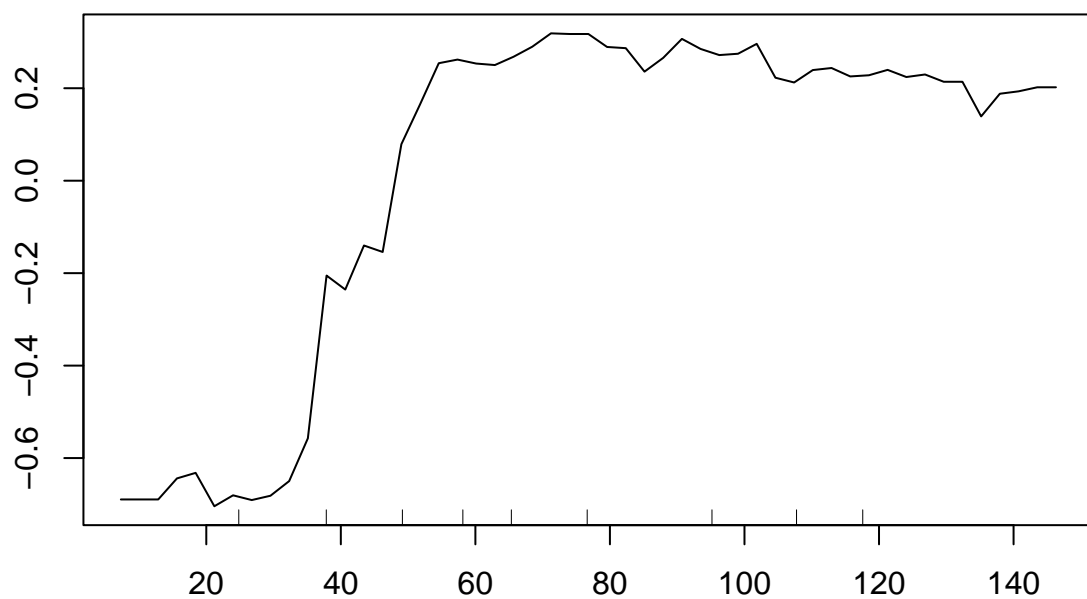
```
## mtry = 2  OOB error = 22.97%
## Searching left ...
## mtry = 4    OOB error = 24.32%
## -0.05882353 0.05
## Searching right ...
## mtry = 1    OOB error = 25.68%
## -0.1176471 0.05
```



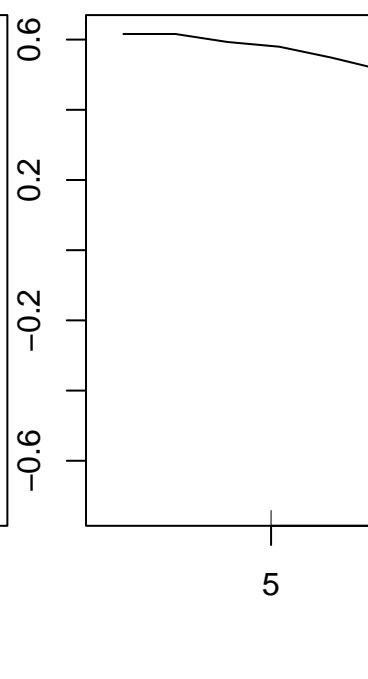
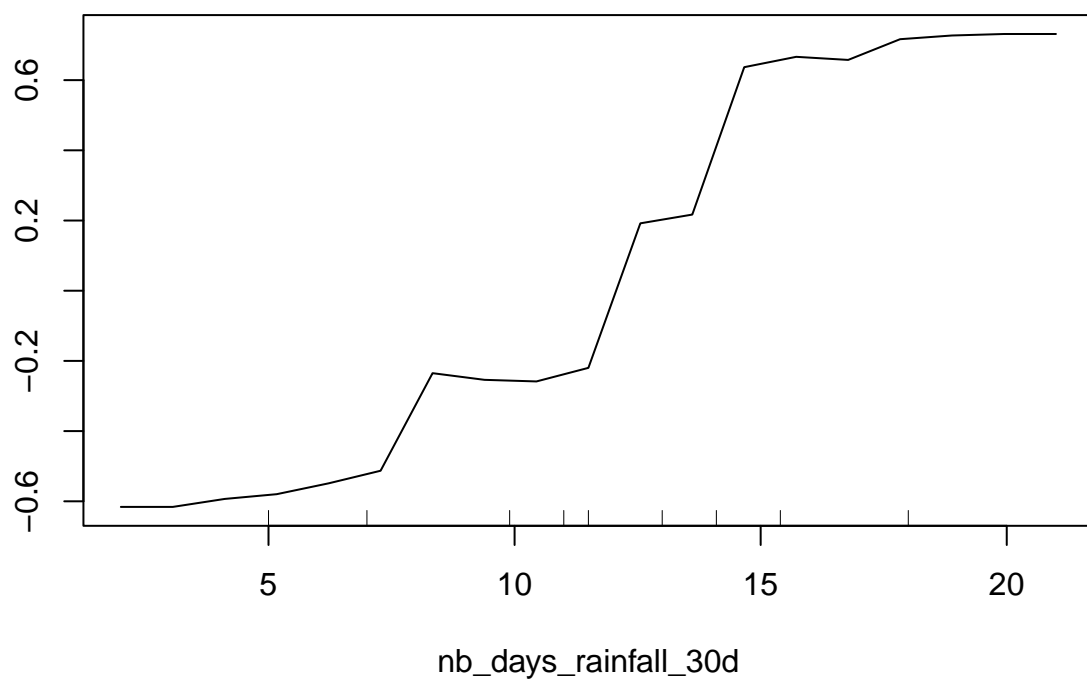


```
##          MeanDecreaseGini
## cumul_precipit_1Jan      9.462054
## nb_days_rainfall_30d     8.262184
## VPD                      4.045020
## RH                       2.868870
## temp                     4.685147
## TDD                      7.209482
```

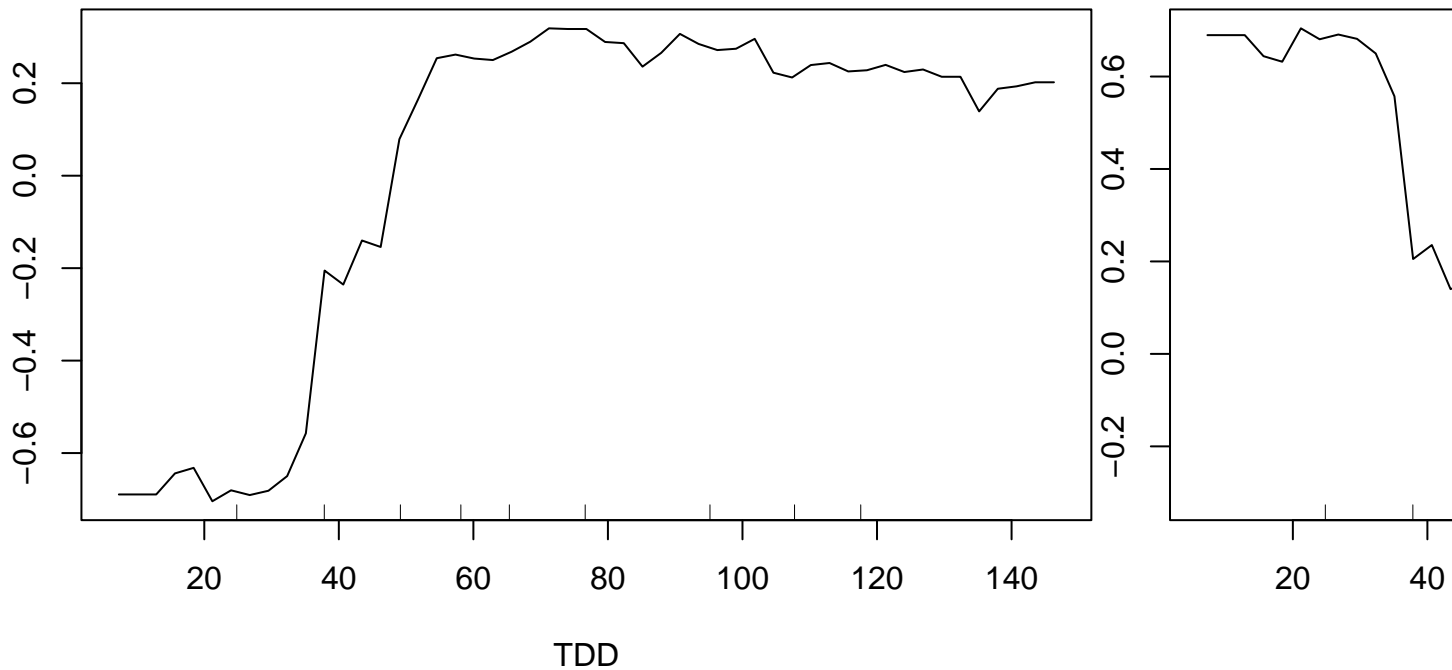
**Partial Dependence on TDD**



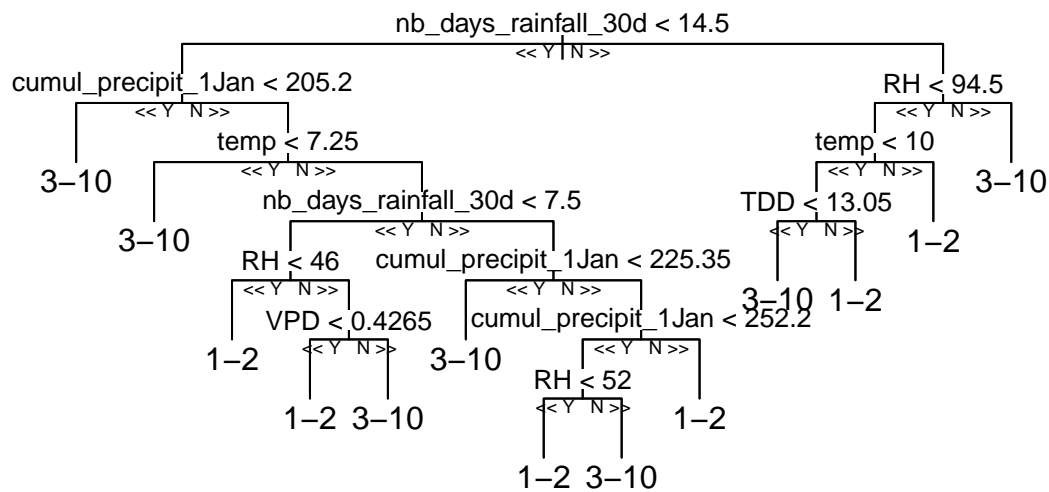
**Partial Dependence on nb\_days\_rainfall\_30d**



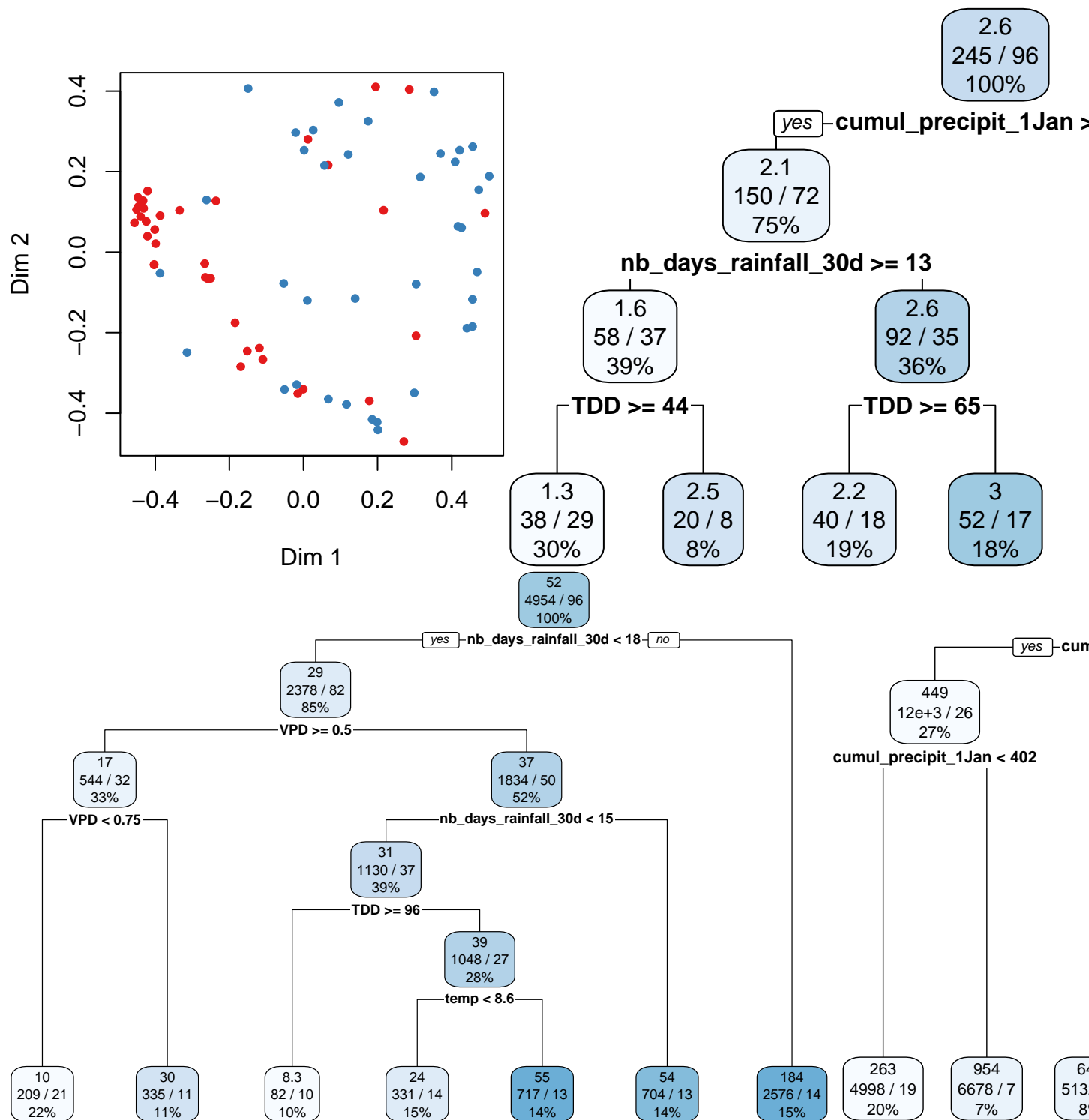
## Partial Dependence on TDD



```
## Skipping install of 'reptree' from a github remote, the SHA1 (7ebb9ff7) has not changed since last :
## Use `force = TRUE` to force installation
```

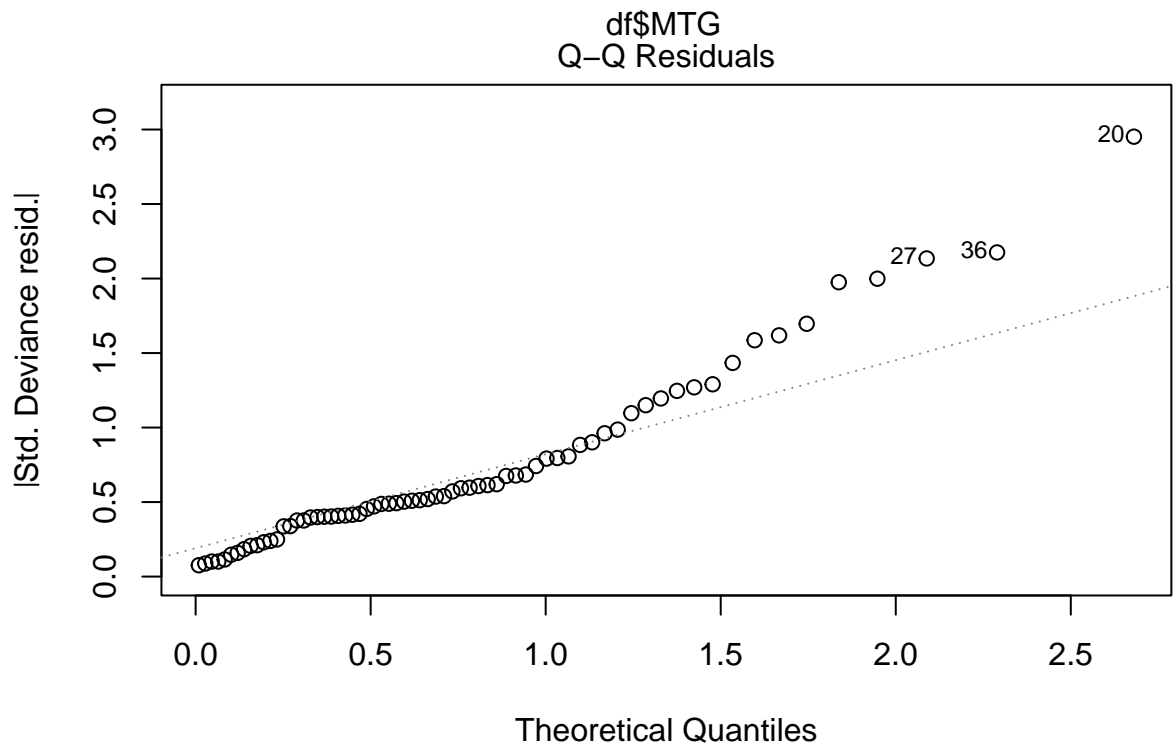
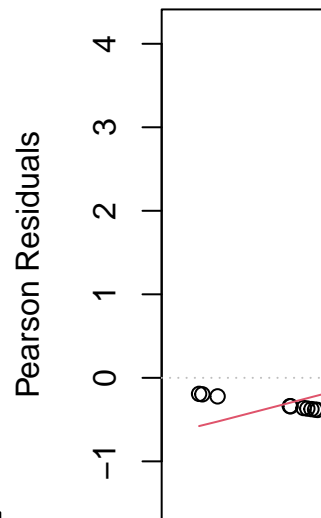
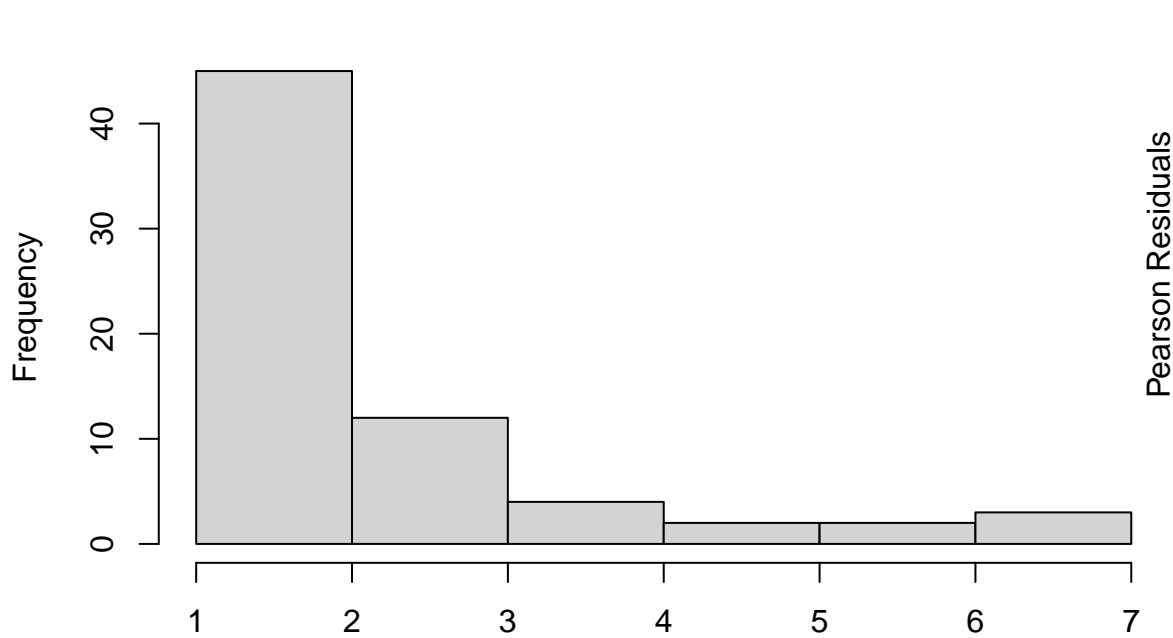


```
## Warning in RColorBrewer::brewer.pal(nlevs, "Set1"): minimal value for n is 3, returning requested palette
```



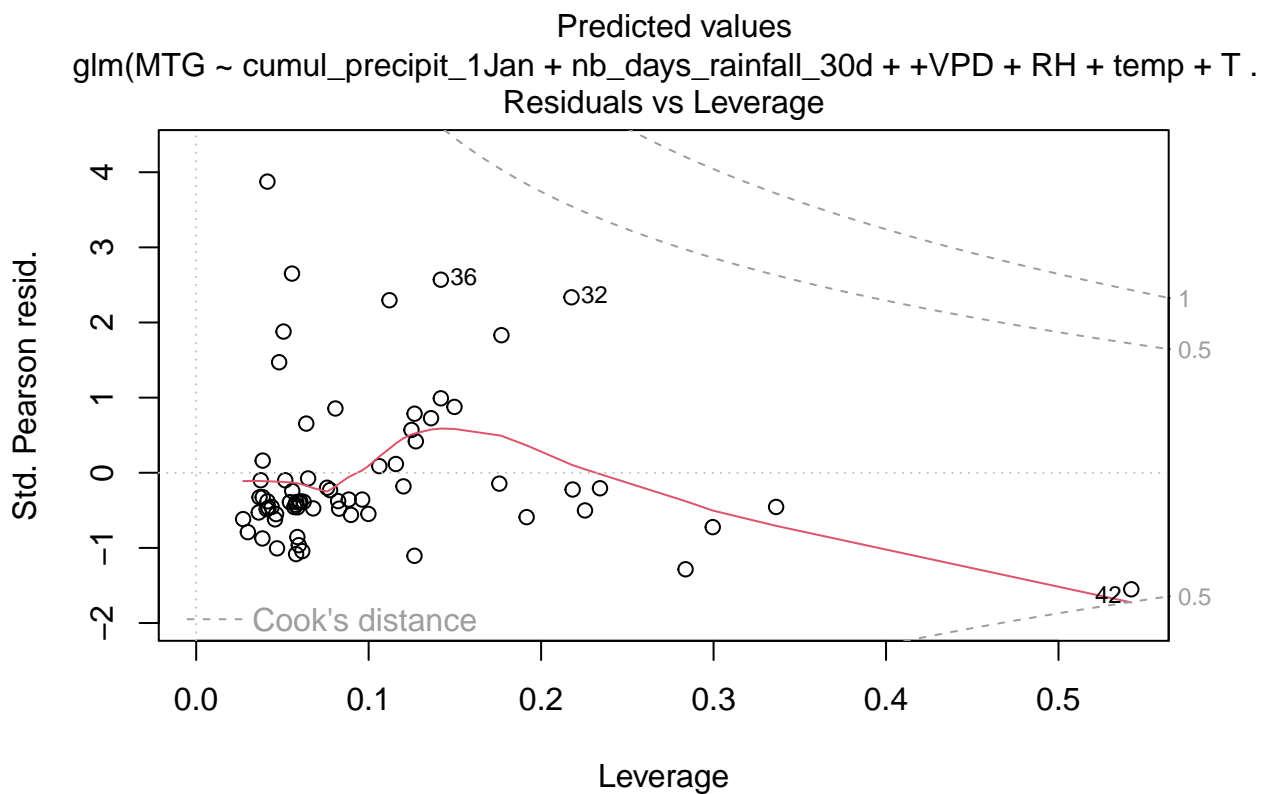
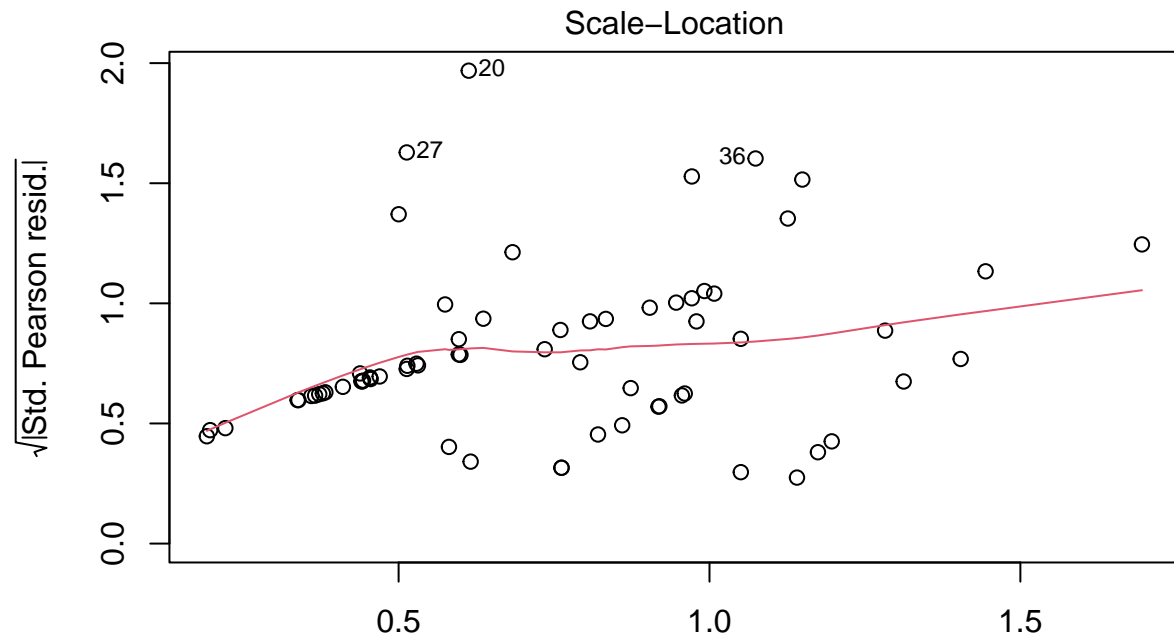
```
## DATASET BBCH 13:+
df <- df_all %>% filter(df_all$BBCH >= 13)
model_MGT(df)
```

# Histogram of df\$MTG

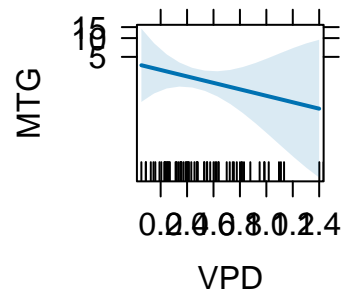
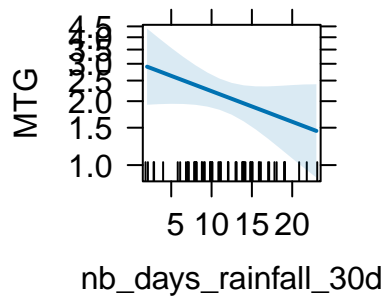
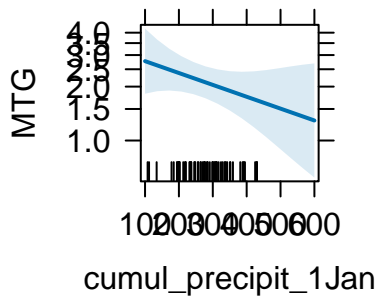


glm(MTG ~ cumul

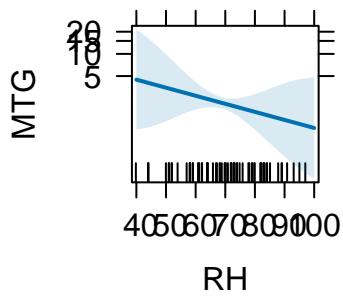
Theoretical Quantiles  
glm(MTG ~ cumul\_precipit\_1Jan + nb\_days\_rainfall\_30d + +VPD + RH + temp + T .



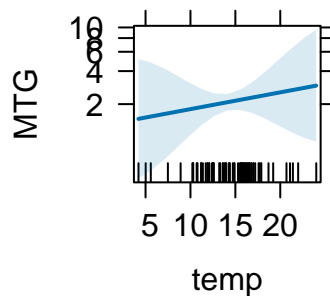
ul\_precipit\_1Jan effect plot    nb\_days\_rainfall\_30d effect plot    VPD effect plot



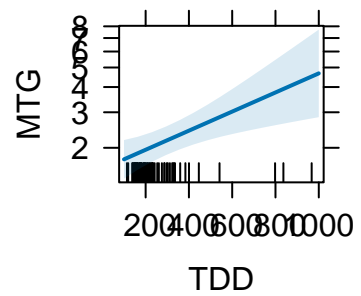
**RH effect plot**



**temp effect plot**



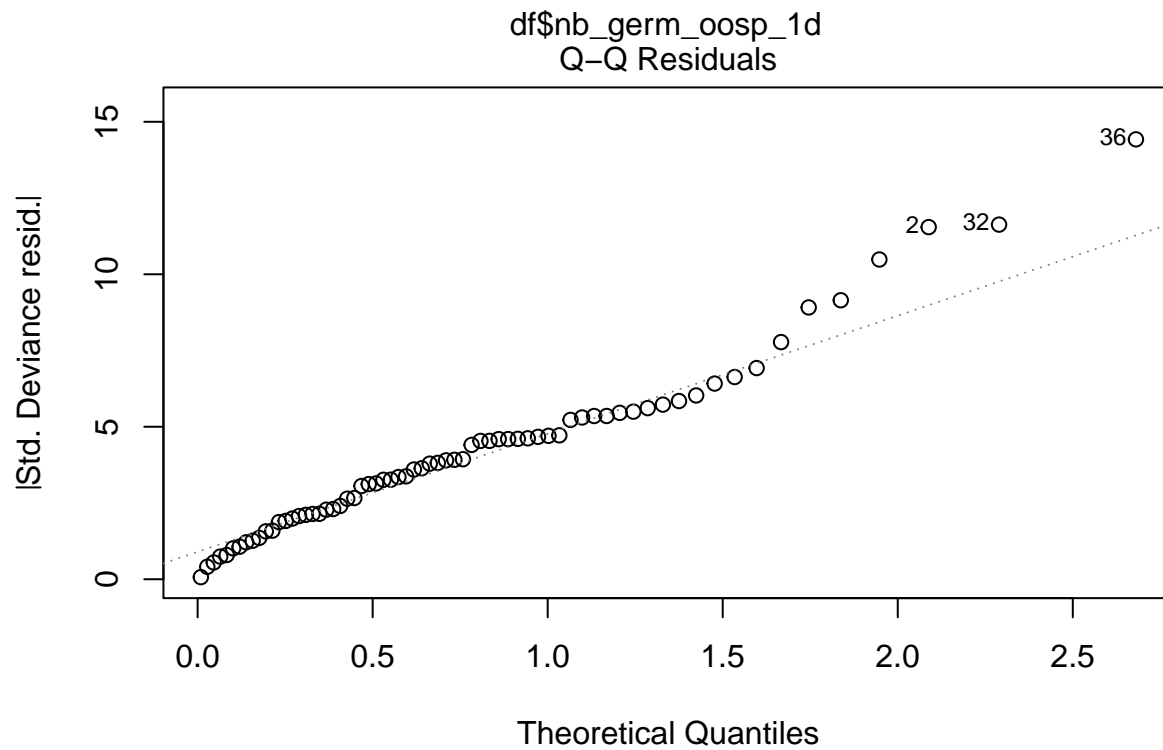
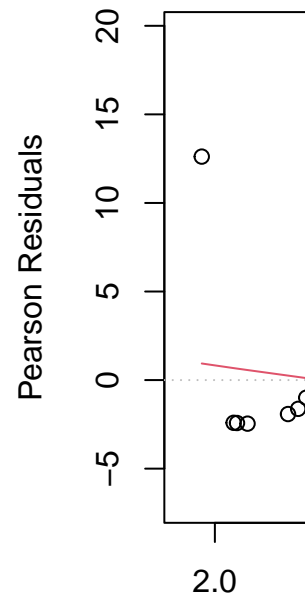
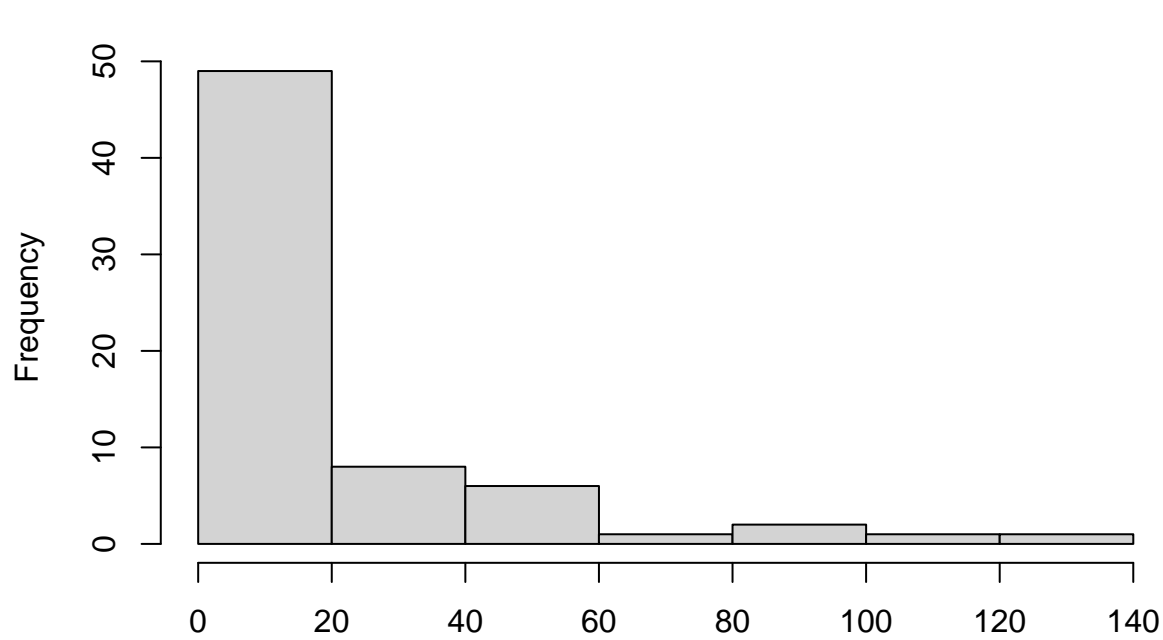
**TDD effect plot**



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

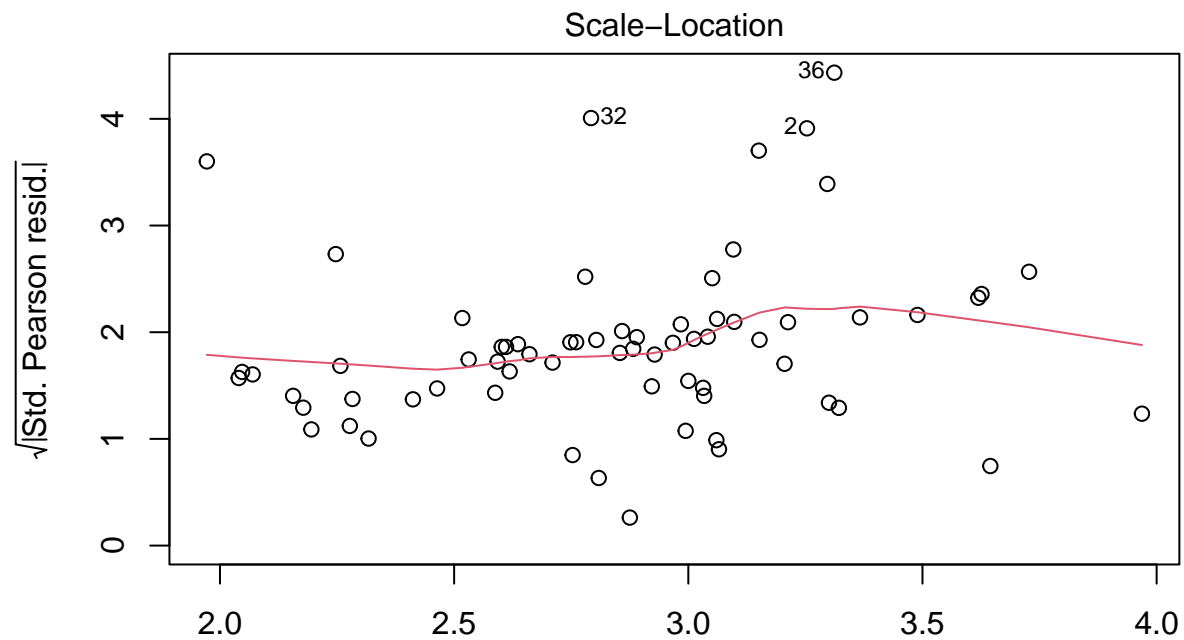
Histogram of df\$nb\_germ\_oosp\_1d



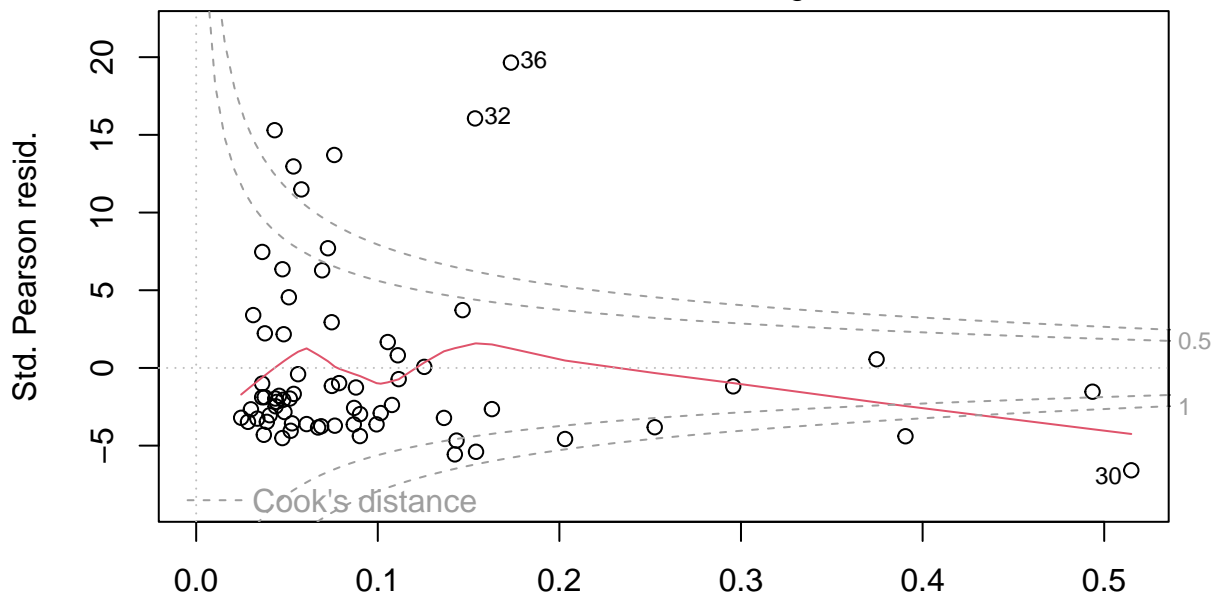
glm(nb\_germ\_oosp\_1d ~ cumul\_precipit\_1Jan + nb\_days\_rainfall\_30d + VPD + RH ,

Theoretical Quantiles  
glm(nb\_germ\_oosp\_1d ~ cumul\_precipit\_1Jan + nb\_days\_rainfall\_30d + VPD + RH ,



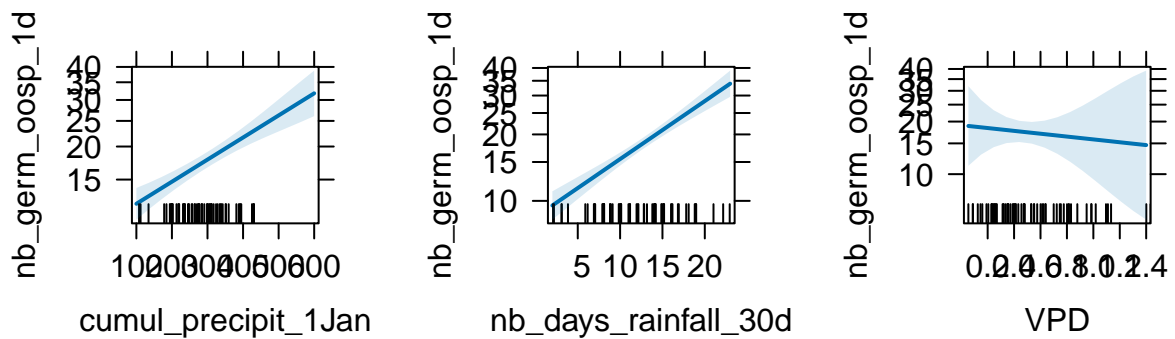


glm(nb\_germ\_oosp\_1d ~ cumul\_precipit\_1Jan + nb\_days\_rainfall\_30d + VPD + RH .  
Residuals vs Leverage

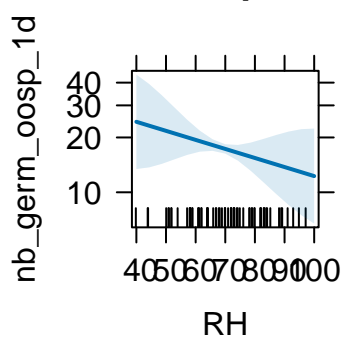


Leverage  
glm(nb\_germ\_oosp\_1d ~ cumul\_precipit\_1Jan + nb\_days\_rainfall\_30d + VPD + RH .

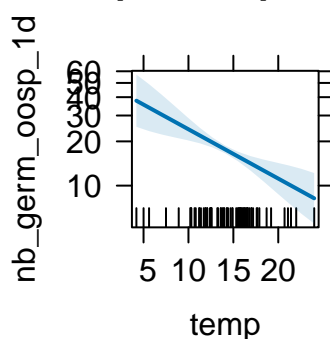
# 



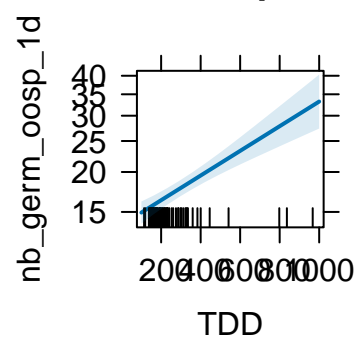
## RH effect plot



## temp effect plot



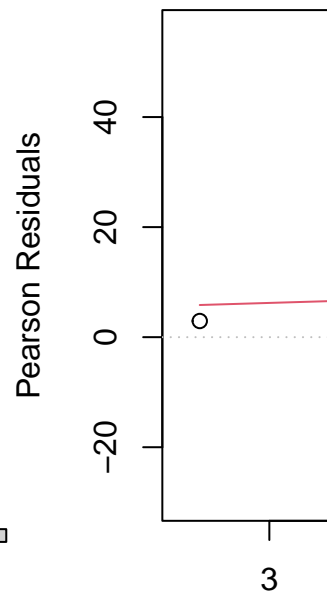
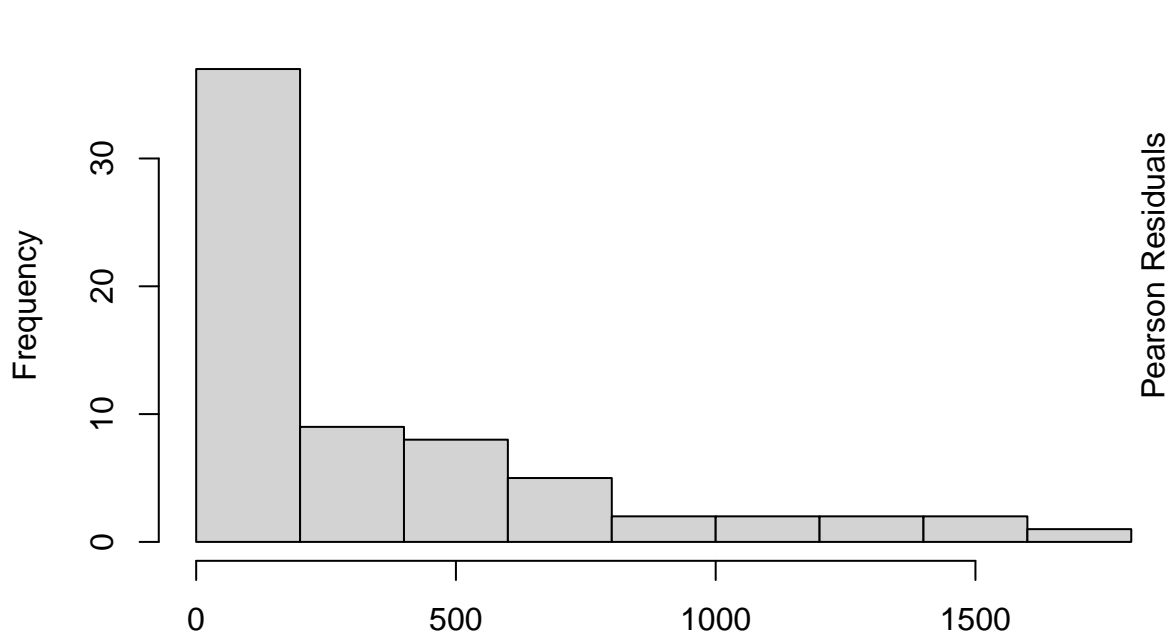
## TDD effect plot



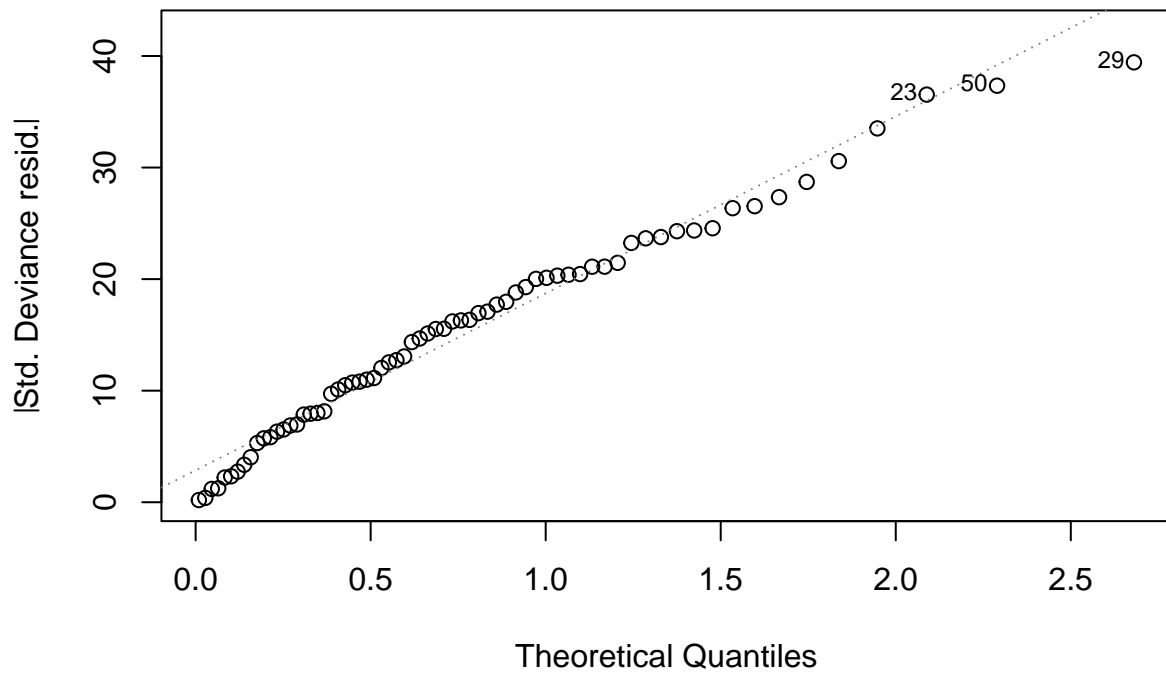
```
## # A tibble: 7 x 5
##   term                estimate std.error statistic  p.value
##   <chr>              <dbl>    <dbl>    <dbl>    <dbl>
## 1 (Intercept)        3.36      0.763      4.41 1.05e- 5
## 2 cumul_precipit_1Jan 0.00192 0.000321    6.00 1.97e- 9
## 3 nb_days_rainfall_30d 0.0605  0.00632    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

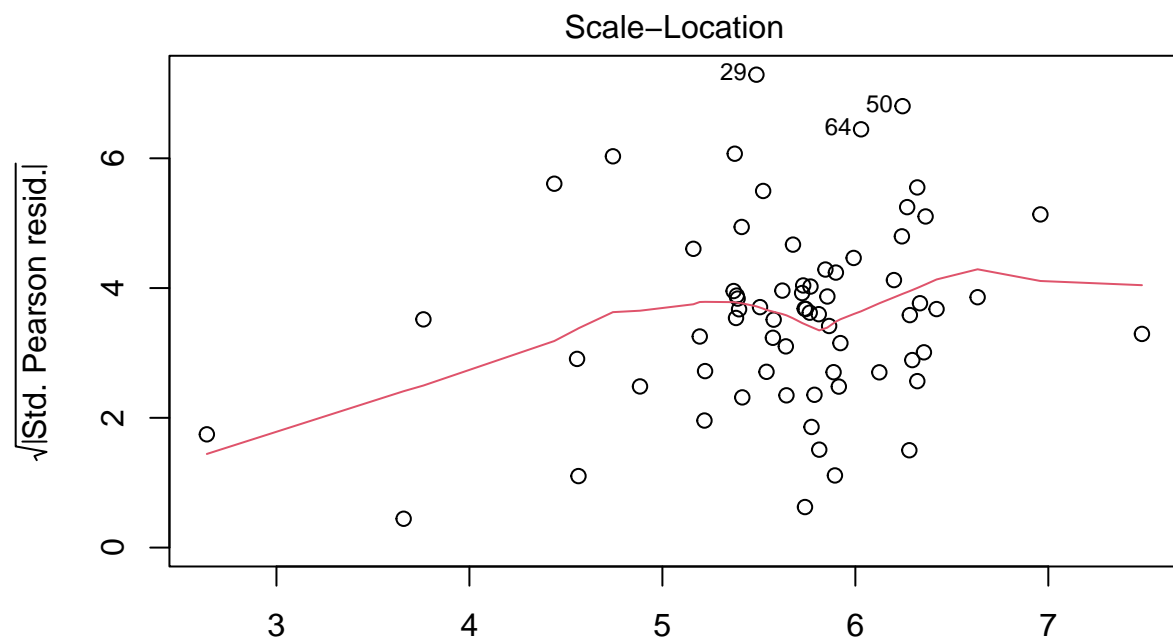


df\$nb\_germ\_oosp\_10d  
Q-Q Residuals

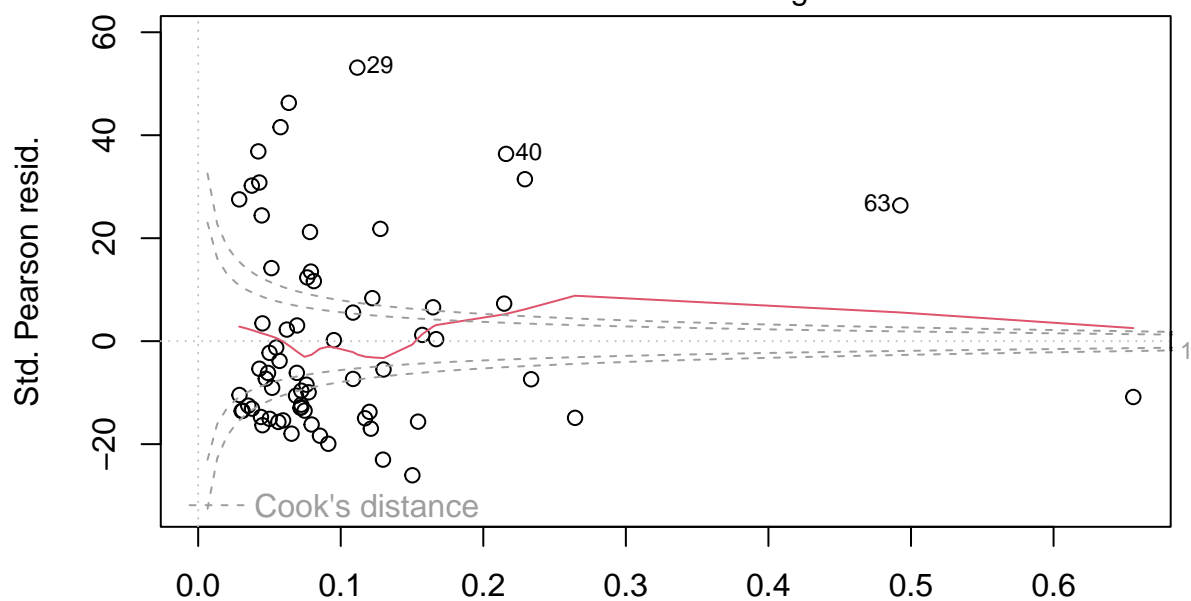


glm(nb\_germ\_oosp\_10d ~ cumul\_precipit\_1Jan + nb\_days\_rainfall\_30d + VPD + R .

glm(nb\_germ\_oosp\_10d ~ cumul\_precipit\_1Jan + nb\_days\_rainfall\_30d + VPD + R .

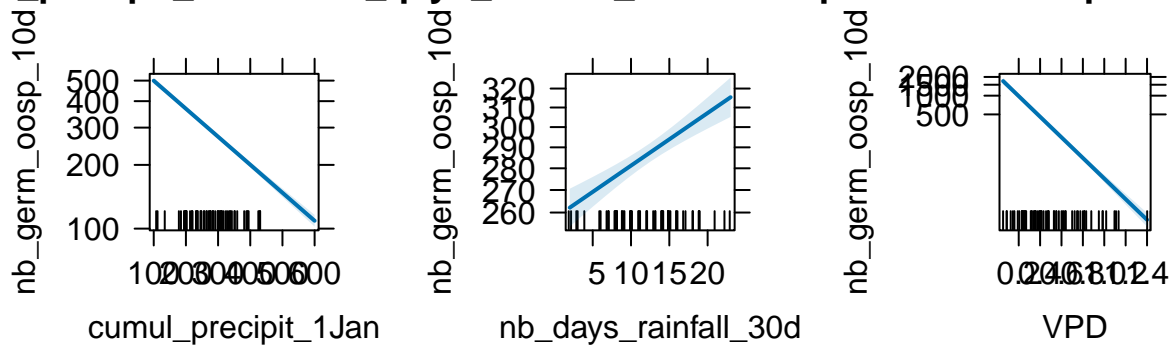


glm(nb\_germ\_oosp\_10d ~ cumul\_precipit\_1Jan + nb\_days\_rainfall\_30d + VPD + R .  
Residuals vs Leverage

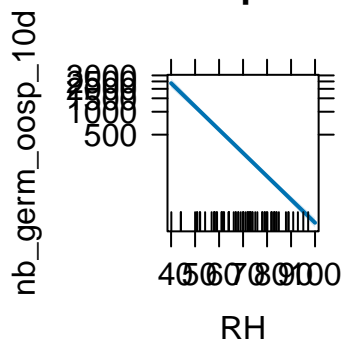


Leverage  
glm(nb\_germ\_oosp\_10d ~ cumul\_precipit\_1Jan + nb\_days\_rainfall\_30d + VPD + R .

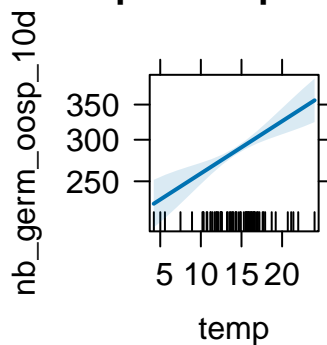
## ul\_precipit\_1Jan effect plot, nb\_days\_rainfall\_30d effect plot, VPD effect plot



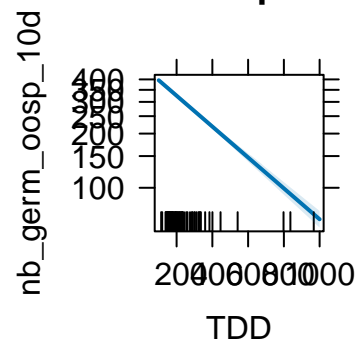
## RH effect plot



## temp effect plot



## TDD effect plot

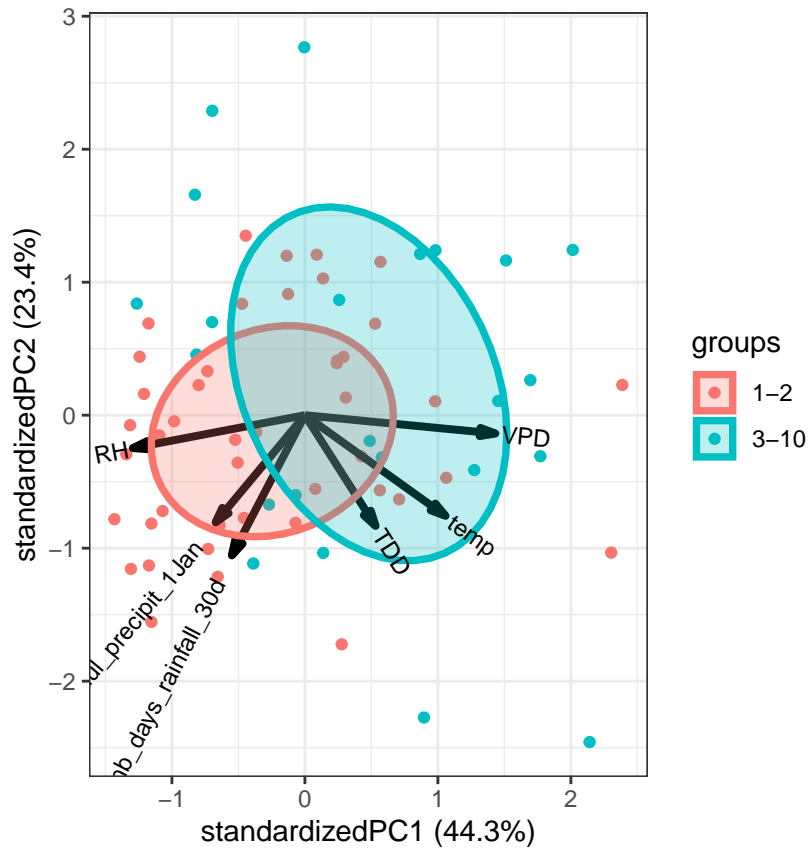


```
## # A tibble: 7 x 5
##   term                estimate std.error statistic    p.value
##   <chr>              <dbl>    <dbl>    <dbl>    <dbl>
## 1 (Intercept)       13.5      0.178      76.2      0
## 2 cumul_precipit_1Jan -0.00305 0.0000947  -32.2  8.25e-227
## 3 nb_days_rainfall_30d  0.00881 0.00138     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)
```

```
## Importance of components:
##               PC1    PC2    PC3    PC4    PC5    PC6
## Standard deviation  1.6302 1.1851 0.9099 0.8070 0.6632 0.13772
## Proportion of Variance 0.4429 0.2341 0.1380 0.1085 0.0733 0.00316
## Cumulative Proportion 0.4429 0.6770 0.8150 0.9235 0.9968 1.00000
##               PC1    PC2    PC3    PC4
## cumul_precipit_1Jan -0.2769867 -0.45713611  0.46125715  0.707974231
## nb_days_rainfall_30d -0.2265899 -0.59800837  0.22838069 -0.619343897
## VPD                  0.5907516 -0.07807785  0.24636952  0.003750625
## RH                   -0.5335250 -0.14030990 -0.30935410 -0.116177435
## temp                 0.4362037 -0.42943050  0.07957991 -0.158379491
## TDD                  0.2191470 -0.47246057 -0.75654031  0.276749523
##               PC5    PC6
## cumul_precipit_1Jan -0.007710035  0.016119706
## nb_days_rainfall_30d  0.394070569  0.002604559
## VPD                   0.079285329  0.760209346
```

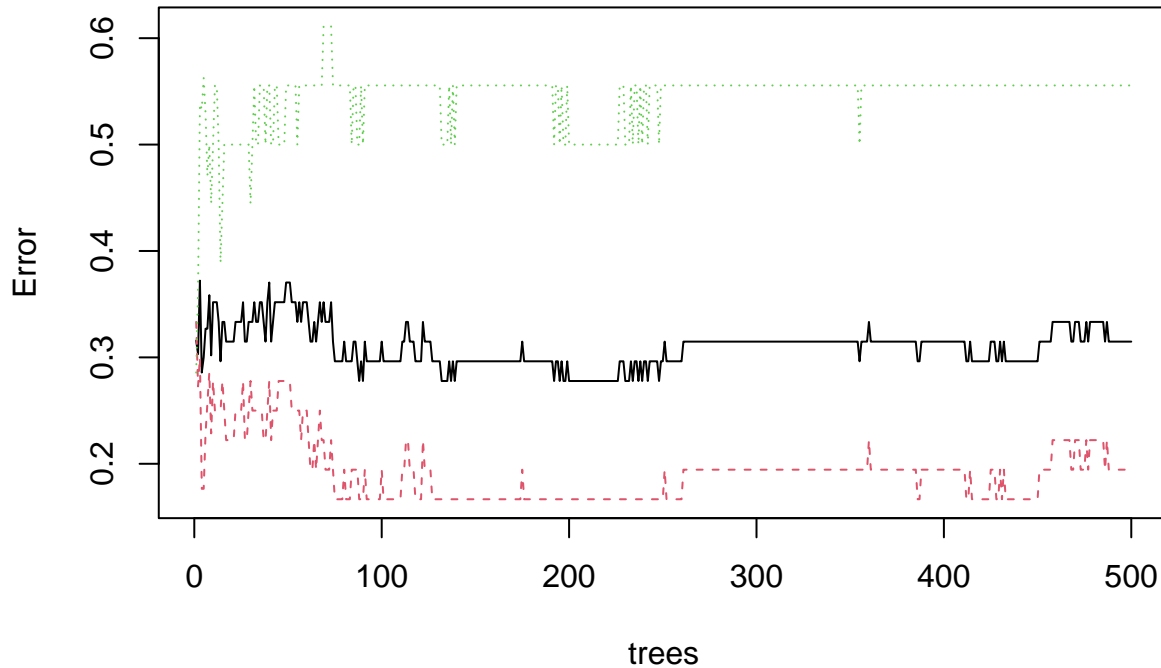
```
## RH -0.526680142 0.555945257
## temp -0.693670445 -0.335738053
## TDD 0.282474991 -0.004466881
```



```
random_forest(df)
```

```
##
## Call:
## randomForest(formula = MTG_cat ~ ., data = train, proximity = TRUE,      mtry = 3)
##           Type of random forest: classification
##           Number of trees: 500
## No. of variables tried at each split: 3
##
##           OOB estimate of  error rate: 31.48%
## Confusion matrix:
##      1-2 3-10 class.error
## 1-2   29   7  0.1944444
## 3-10  10   8  0.5555556
```

rf



```
## mtry = 2  OOB error = 33.33%
## Searching left ...
## mtry = 4    OOB error = 31.48%
## 0.05555556 0.05

## Warning in randomForest.default(x, y, mtry = mtryCur, ntree = ntreeTry, :
## invalid mtry: reset to within valid range

## mtry = 8    OOB error = 29.63%
## 0.05882353 0.05

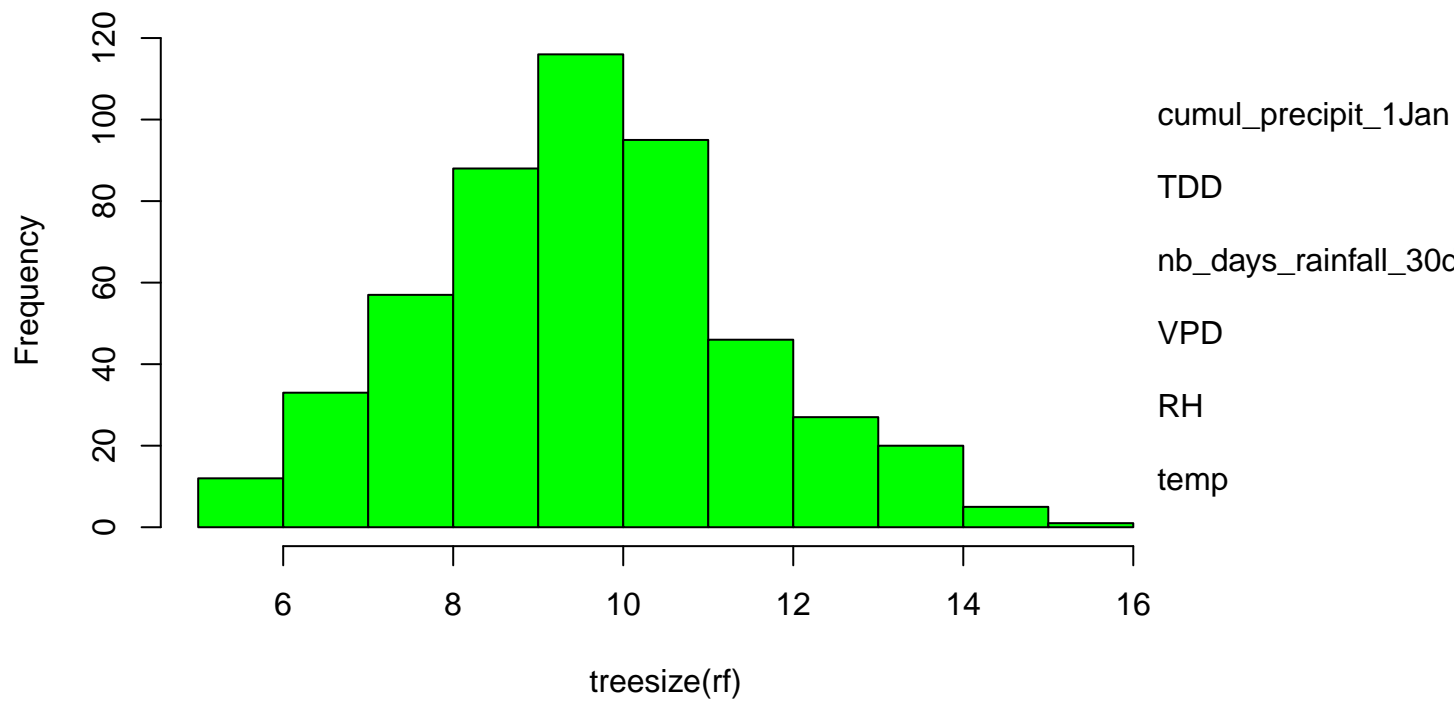
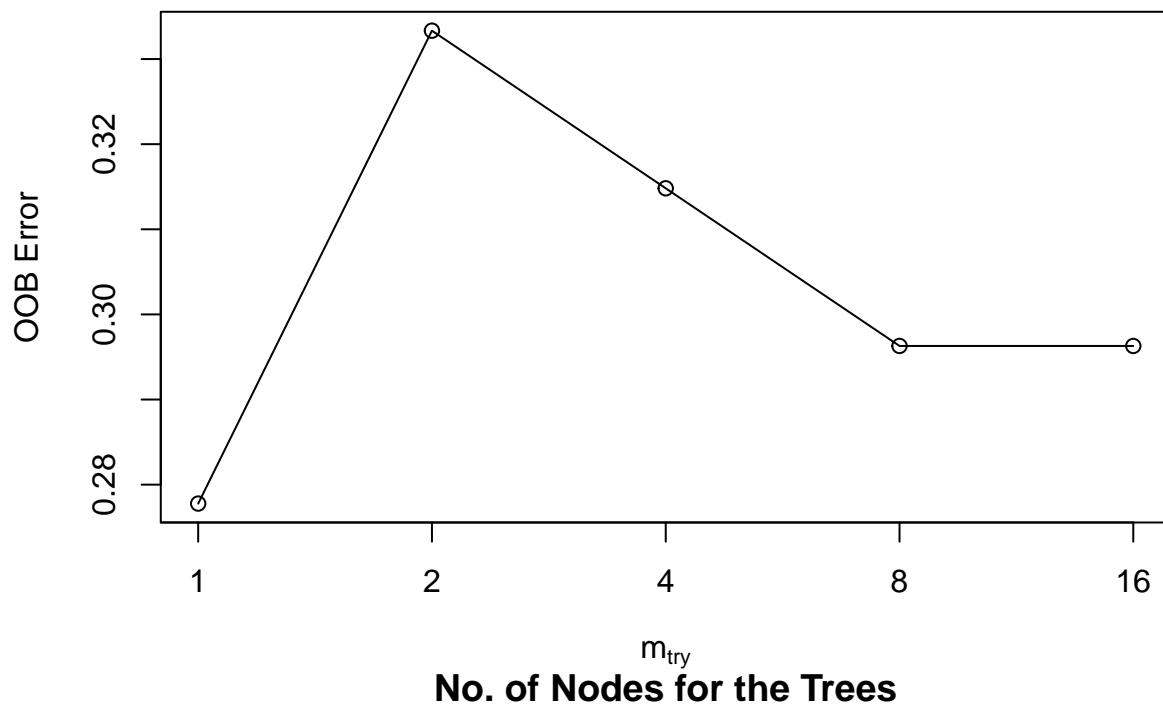
## Warning in randomForest.default(x, y, mtry = mtryCur, ntree = ntreeTry, :
## invalid mtry: reset to within valid range

## mtry = 16   OOB error = 29.63%
## 0 0.05
## Searching right ...
## mtry = 1    OOB error = 27.78%
## 0.0625 0.05

## Warning in randomForest.default(x, y, mtry = mtryCur, ntree = ntreeTry, :
## invalid mtry: reset to within valid range

## mtry = 0    OOB error = 31.48%
## -0.1333333 0.05

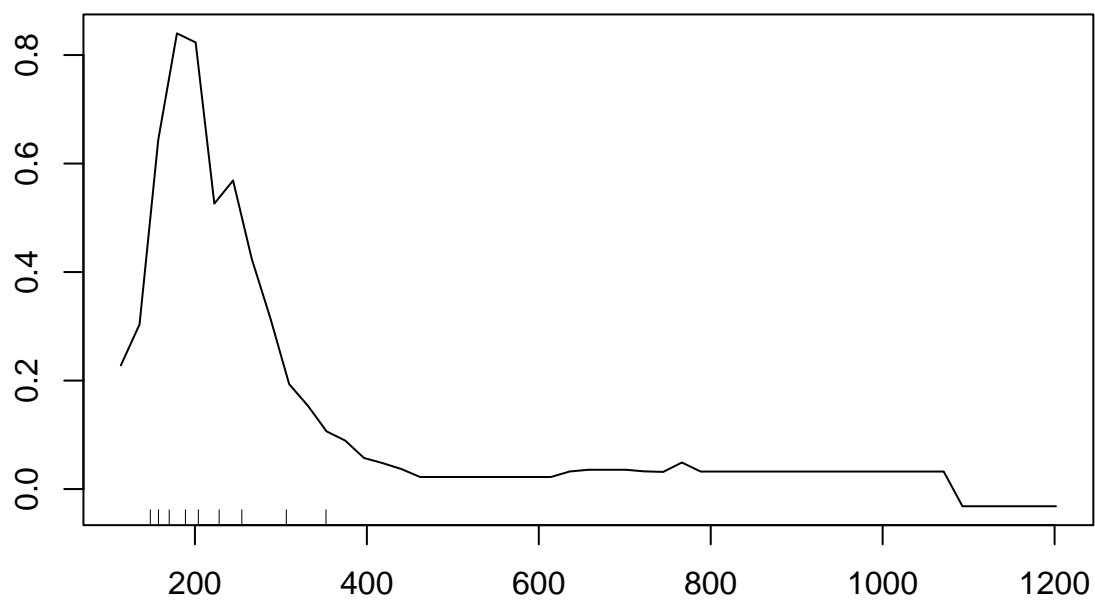
## Warning in xy.coords(x, y, xlabel, ylabel, log): 1 x value <= 0 omitted from
## logarithmic plot
```



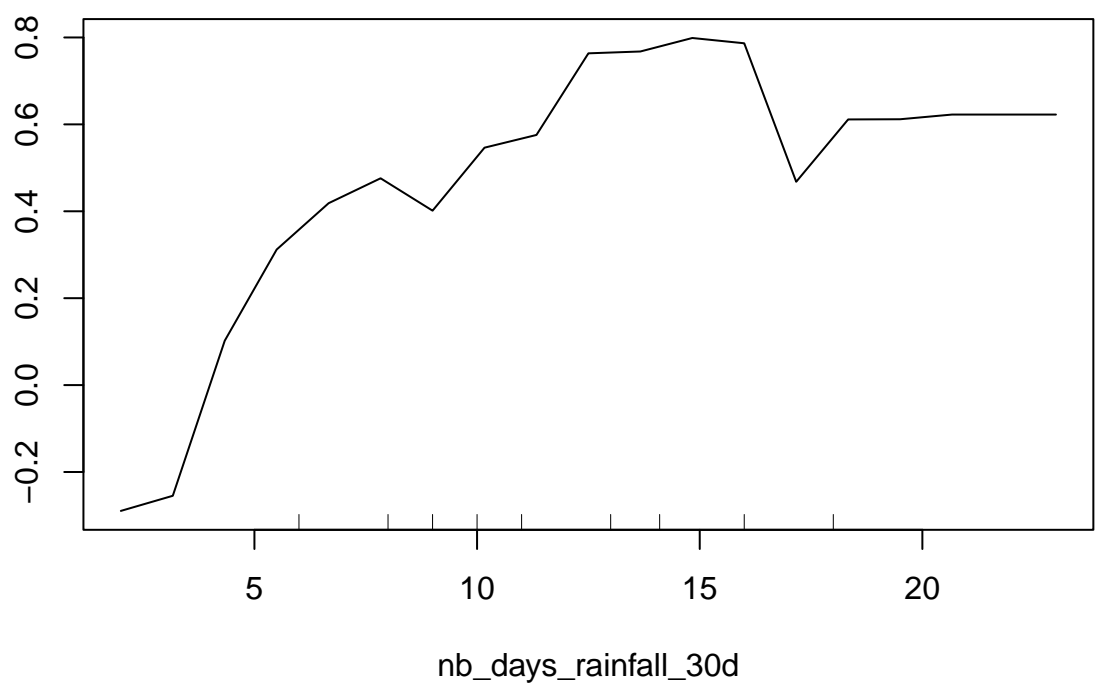
```
##               MeanDecreaseGini
## cumul_precipit_1Jan          6.462931
## nb_days_rainfall_30d          3.679121
## VPD                           3.482929
## RH                            3.351965
## temp                          2.971933
## TDD                           3.705640
```



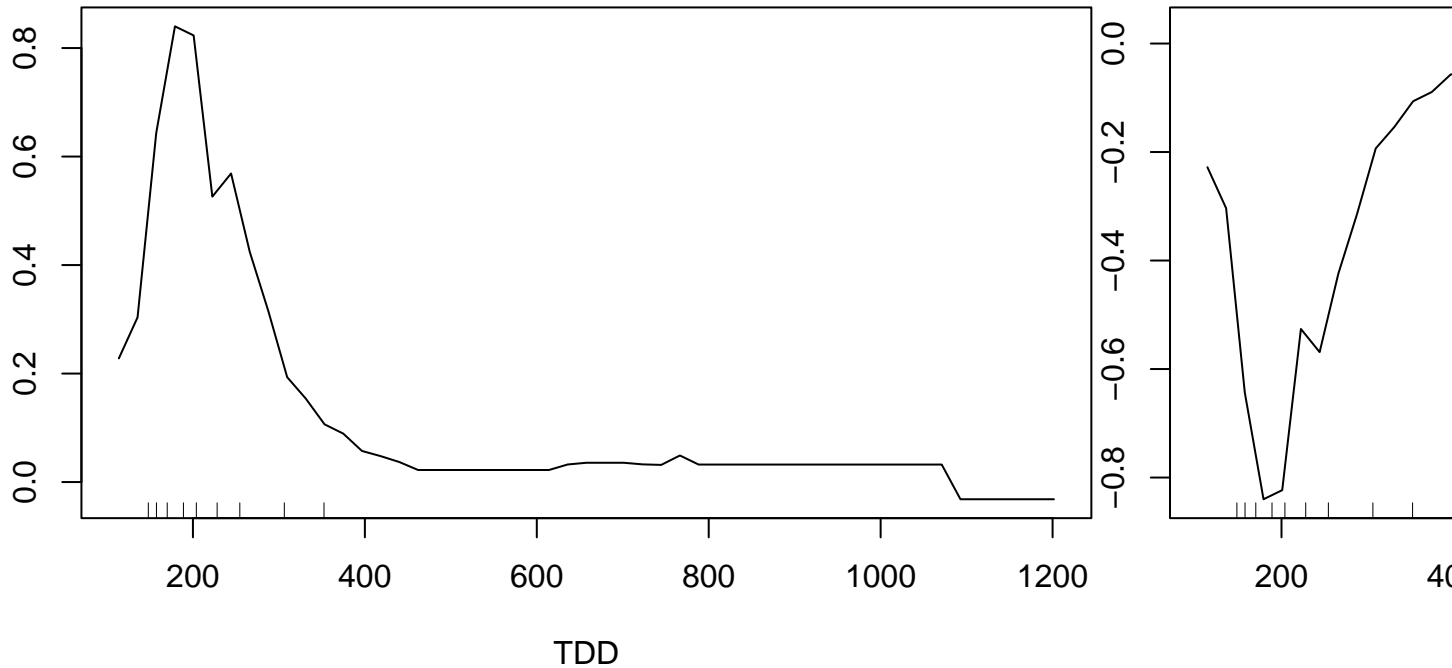
**Partial Dependence on TDD**



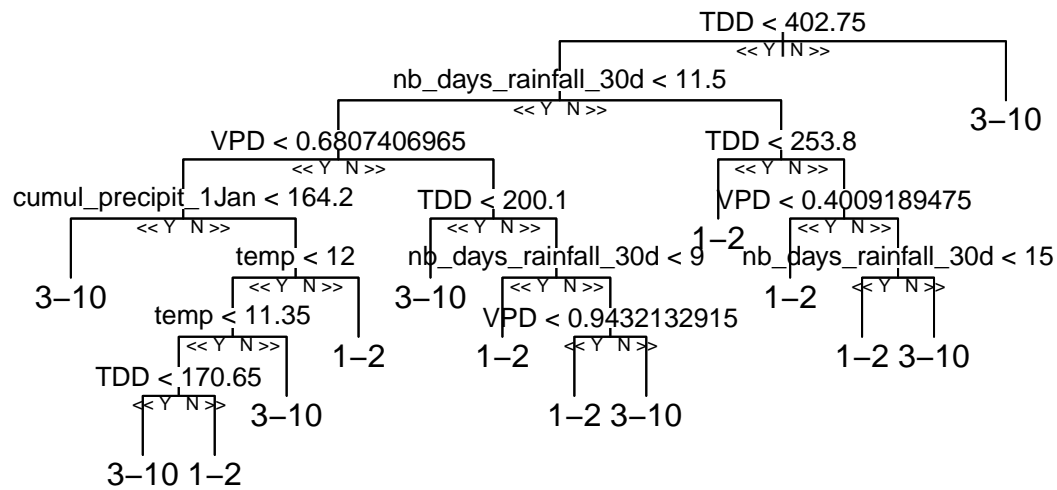
**Partial Dependence on nb\_days\_rainfall\_30d**



## Partial Dependence on TDD



```
## Skipping install of 'reptree' from a github remote, the SHA1 (7ebb9ff7) has not changed since last
## Use `force = TRUE` to force installation
```



```
## Warning in RColorBrewer::brewer.pal(nlevs, "Set1"): minimal value for n is 3, returning requested palette
```

