GAM analysis of the weekly data

Willem Vervoort, Michaela Dolk & Floris van Ogtrop

2017-08-29

```
# root dir
knitr::opts_knit$set(root.dir =
                       "C:/Users/rver4657/ownCloud/Virtual Experiments/VirtExp")
knitr::opts_chunk$set(echo = TRUE)
# LOAD REQUIRED PACKAGES # #####
library(pander)
library(tidyr)
library(xts)
library(zoo)
library(mgcv)
library(oz)
library(Kendall)
library(ggplot2)
library(doParallel)
library(foreach)
```

This rmarkdown document and the resulting pdf are stored on github. All directories (apart from the root working directory) refer to the directories in this repository

Introduction

This document is related to the manuscript "Disentangling climate change trends in Australian streamflow" (vervoort et al.), submitted to Journal of Hydrology.
This part of the series covers the analysis of the streamflow data using Generalised Additive models (GAM) testing for a trend in the data, or testing for a trend in the residuals. In particular, this extends the Mann Kendall analysis, as the Mann Kendall tau only indicates a strength, significance and direction of the trend, but does not quantify the magnitude of the trend.
The different models are outlined in **Table 2** in the main manuscript. The methods describe in detail how the GAM are developed with reference to the underlying theory.
This document follows **Table 2** in the series of models, so the headings (and model numbers) relate to this table.

The data

Using the datasets that were developed earlier, we can load in the daily data for streamflow, rainfall and temperature.

```
load("data/DailyDataIncludingGridded.Rdata")
load("data/ClimCh_project_MD.Rdata")
# correct the column name of maxT in GridRainAllDataout
colnames(GridRainAllDataout)[5] <- "MaxT"
```

The models (from Table 2 in the manuscript)

Table 2 in the manuscript (reproduced below) outlines the different models that were analysed using the statistical general additive models.

```
table2 <- read.csv("documents/Table2Models.csv")
pander(table2,caption = "Model structures used in the Generalised additive modelling analysis")
```

Table 1: Model structures used in the Generalised additive modelling analysis (continued below)

No	Model	Trend
Log(Q + 1) ~ trend + error	linear	40-year trend in streamflow to compare with Mann Kendall analysis
Log(P + 1) ~ trend + error	linear	40-year trend in rainfall to compare with Mann Kendall analysis
Log(Q + 1) ~ s(P) + trend + error	linear	The trend in this model relative to model 1 indicates the importance of "other processes"
Log (Q + 1) ~ s(P) + s(maxT	P) + trend + error	linear
Log(Q + 1) ~ s(P) + s(maxT	P) + error	Mann-Kendall

Analysis
while the comparison with model 2 indicates the "amplification".
Difference between model 3 and 4 is the effect of evapotranspiration on the trend. The remaining trend is related to changes over time in the rainfall runoff response.
`Check if linear trend assumption is biased.`

Model 1 Only flow and trend

The first 2 models are actually not generalised additive mixed models (GAM) as the models only analyse a linear trend. To match the GAM analysis, we used generalised least squares (`gls()`) in R. This still allows correlated errors to be analysed

```
# run the gls model on flowtrend only
#for (i in seq_along(Stations[,1])) {
cl <- makeCluster(4) # create a cluster with 4 cores
registerDoParallel(cl) # register the cluster
# use a foreach loop to calibrate
Store2 <- foreach(i = 1:length(Stations[,1]),
                  .packages="mgcv") %dopar% {
```

```r
# for (i in seq_along(Stations[,1])) {
#   i <- 1
  gamm.data <- subset(flow_rain_maxT_weekly,
                      flow_rain_maxT_weekly$Station == Stations[i,1])
  gamm.data$trend <- 1:nrow(gamm.data)
  gam_TrendOnly <- gls(log(Flow +1)~trend, correlation= corCAR1(),
       data=gamm.data)
  out <- list(model = gam_TrendOnly,
              results = data.frame(Station=Stations[i,1],
              t(summary(gam_TrendOnly)$tTable[2,c(1,4)]),
                   AIC=summary(gam_TrendOnly)$AIC))
  out
}
stopCluster(cl)

par(mfrow=c(5,3),mar=c(2,2,2,2))
for (i in seq_along(Stations[,1])) {
  res <- residuals(Store2[[i]]$model)
  plot(res, main=Stations[i,1], cex.main=0.7,
       ylab="normalised residuals",xlab="")
  n <- length(res)
  abline(lsfit(1:n, res), col="red")
}

# store results
storedir <- "c:/users/rver4657/owncloud/virtual experiments"
save(Store2,file=paste(storedir,
                "projectdata/Store2_TrendOnlyAnalysis.RData",
                sep="/"))
output <- do.call(rbind, lapply(1:length(Store2), function(i) rbind(Store2[[i]][[2]])))
pander(output, caption="Mixed model results for analysis of trend in flow only")
```

Table 3: Mixed model results for analysis of trend in flow only

Station	Value	p.value	AIC
COTT	-0.0002592	0.0285	1673
RUTH	-0.0005892	3.149e-14	3476
CORA	-0.0002107	0.00178	5085
ELIZ	4.887e-06	0.9848	3702
COCH	-5.77e-06	0.9613	2957
COEN	-2.977e-05	0.8874	4056
SCOT	-8.134e-05	0.3733	2795
HELL	-0.0001375	0.2707	2935
NIVE	-6.446e-05	0.718	3101
MURR	-0.0002164	0.001545	250.2
SOUT	-0.0001543	0.01232	1928
YARR	-0.0001143	0.02487	-139.8
DOMB	-0.0001237	0.5306	2784

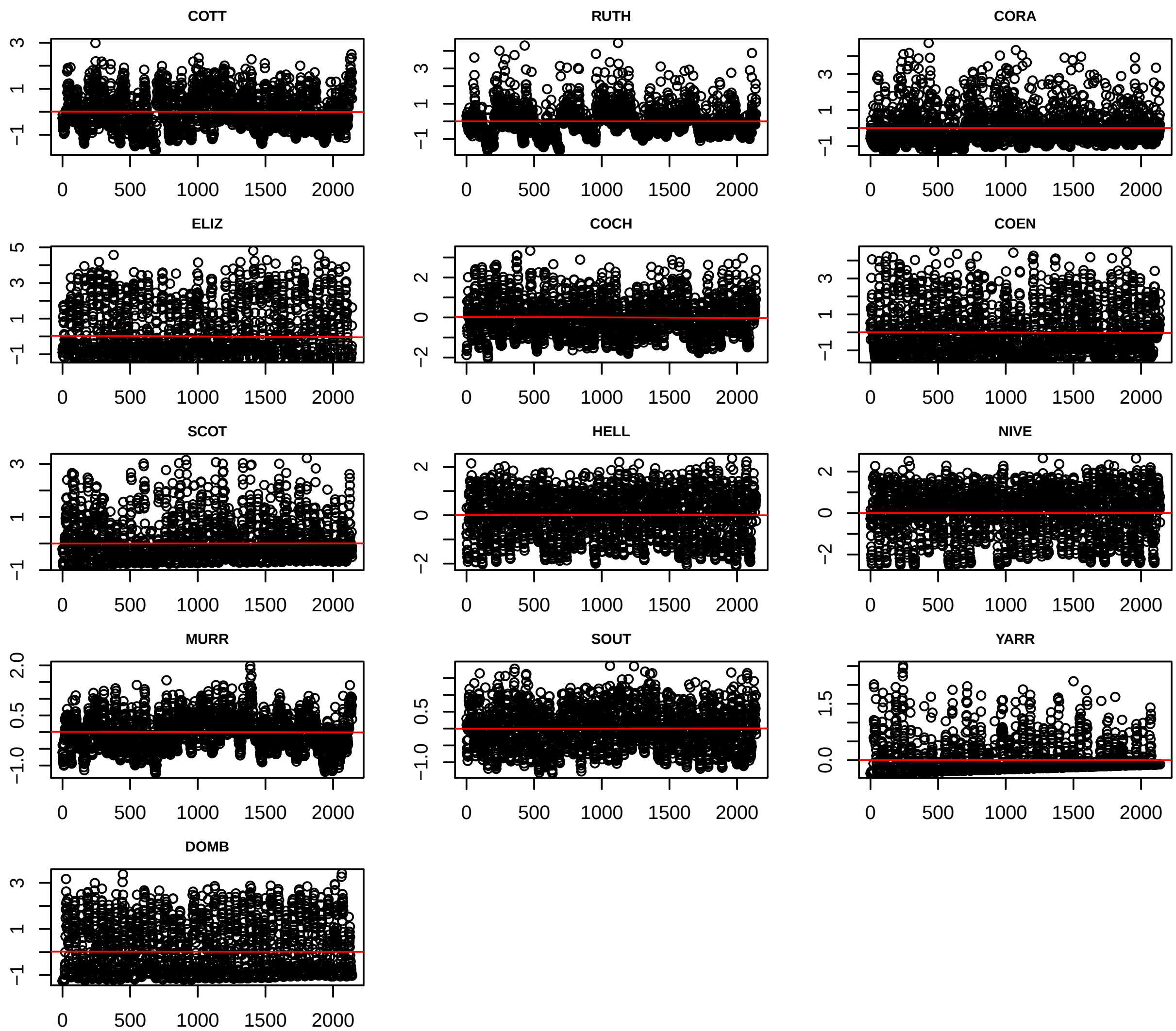

Figure 1: Residuals of linear mixed model analysis for trend in flow only

```r
rm(Store2)
```

Model 2 trend in rain

Rainfall Station measured data

Similar to the flow data, this analysis uses `gls()` to run the linear mixed model to test for a trend in the data and compare to the Mann-Kendall results

```r
# create an empty list
# and an empty dataframe to store results
# run the gls model on flowtrend only
cl <- makeCluster(4) # create a cluster with 4 cores
registerDoParallel(cl) # register the cluster
# use a foreach loop to calibrate
Store_Rain <- foreach(i = 1:length(Stations[,1]),
                 .packages="mgcv") %dopar% {
  gamm.data <- subset(flow_rain_maxT_weekly,
                      flow_rain_maxT_weekly$Station == Stations[i,1])
  gamm.data$trend <- 1:nrow(gamm.data)
  gam_TrendR <- gls(log(Rain + 1)~trend, correlation= corCAR1(),
       data=na.omit(gamm.data))
  out <- list(model = gam_TrendR,
              results = data.frame(Station=Stations[i,1],
                     t(summary(gam_TrendR)$tTable[2,c(1,4)]),
                                  AIC=summary(gam_TrendR)$AIC))
  out
}
stopCluster(cl)


par(mfrow=c(5,3),mar=c(2,2,2,2))
for (i in seq_along(Stations[,1])) {
  res <- residuals(Store_Rain[[i]]$model)
  plot(res, main=Stations[i,1], cex.main=0.5,
       ylab="normalised residuals",xlab="")
  n <- length(res)
  abline(lsfit(1:n, res), col="red")
}

# store results
save(Store_Rain,file=paste(storedir,
                  "projectdata/StoreRain_TrendAnalysis.RData",
                    sep="/"))
output <- do.call(rbind, lapply(1:length(Store_Rain),
                     function(i) rbind(Store_Rain[[i]][[2]])))
pander(output, caption="Mixed model results for analysis of trend in Station Rainfall")
```

Table 4: Mixed model results for analysis of trend in Station Rainfall

Station	Value	p.value	AIC
COTT	-0.0003816	8.77e-06	4748

Station	Value	p.value	AIC
RUTH	-0.0001305	0.03456	6883
CORA	-4.979e-05	0.3874	6621
ELIZ	-0.0003745	0.06944	6183
COCH	-0.0006523	1.916e-06	5896
COEN	-5.208e-05	0.7176	6742
SCOT	-7.987e-05	0.2414	6768
HELL	-0.000191	9.632e-05	5332
NIVE	-7.42e-05	0.2209	6667
MURR	-9.83e-05	0.0713	6858
SOUT	-0.0001819	4.055e-05	5937
YARR	-0.0001191	0.2118	7593
DOMB	-0.0002795	0.003782	5895

```
rm(Store_Rain)
```

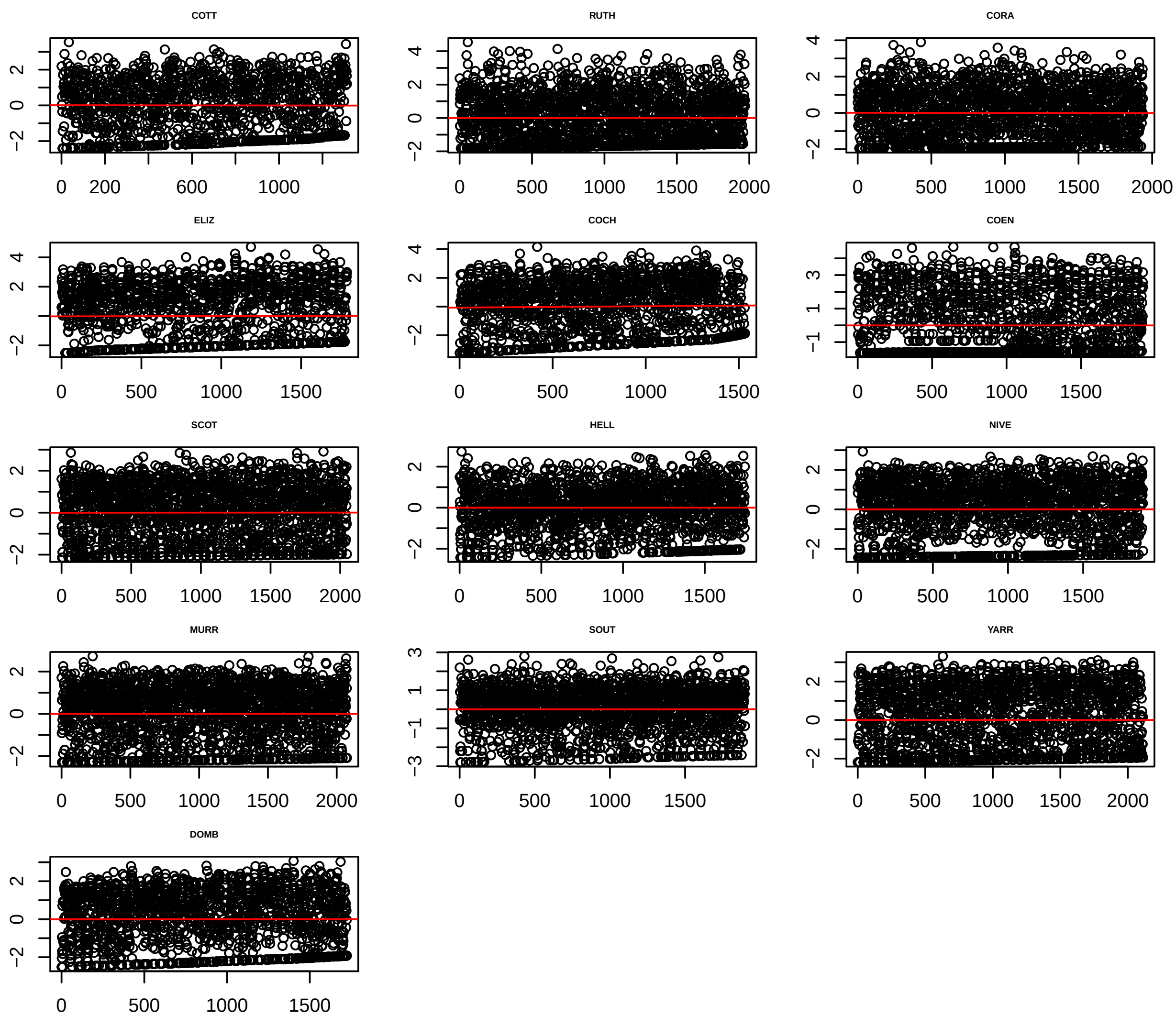

Figure 2: Residuals of linear mixed model analysis for trend in station rainfall data

Rainfall gridded data analysis

Repeat the rainfall analysis for the gridded data to compare station data to gridded data.

```
# create an empty list
# and an empty dataframe to store results
# run the gls model on gridraintrend only
cl <- makeCluster(4) # create a cluster with 4 cores
registerDoParallel(cl) # register the cluster
# use a foreach loop to calibrate
Store_GridRain <- foreach(i = 1:length(Stations[,1]),
                 .packages="mgcv") %dopar% {
  gamm.data <- subset(weekGridRainAllDataout,
                      weekGridRainAllDataout$Station == Stations[i,1])
  gamm.data$trend <- 1:nrow(gamm.data)
  gam_TrendGridR <- gls(log(gridRain +1)~trend, correlation= corCAR1(),
       data=gamm.data)
  out <- list(model = gam_TrendGridR,
              results = data.frame(Station=Stations[i,1],
                         t(summary(gam_TrendGridR)$tTable[2,c(1,4)]),
                              AIC=summary(gam_TrendGridR)$AIC))
  out
    }
stopCluster(cl)


par(mfrow=c(5,3),mar=c(2,2,2,2))
for (i in seq_along(Stations[,1])) {
  res <- residuals(Store_GridRain[[i]]$model)
  plot(res, main=Stations[i,1], cex.main=0.5,
       ylab="normalised residuals",xlab="")
  n <- length(res)
  abline(lsfit(1:n, res), col="red")
}

# store results
save(Store_GridRain,
     file=paste(storedir,
                "projectdata/StoreGridRain_TrendAnalysis.RData",
                sep="/"))
output <- do.call(rbind, lapply(1:length(Store_GridRain),
              function(i) rbind(Store_GridRain[[i]][[2]])))
pander(output, caption="Mixed model results for analysis of trend in Gridded Rainfall")
```

Table 5: Mixed model results for analysis of trend in Gridded Rainfall

Station	Value	p.value	AIC
COTT	-9.652e-05	0.07725	7093
RUTH	1.331e-06	0.981	7024
CORA	-3.491e-05	0.5401	7266
ELIZ	-8.524e-06	0.9642	6715
COCH	8.23e-05	0.3916	7827
COEN	-9.398e-06	0.9478	7213

Station	Value	p.value	AIC
SCOT	1.005e-05	0.878	7006
HELL	-1.731e-05	0.7256	6360
NIVE	3.02e-05	0.4828	6185
MURR	-2.165e-05	0.6871	6875
SOUT	-1.499e-05	0.7395	6513
YARR	8.52e-07	0.9929	7314
DOMB	-4.553e-05	0.5818	6689

```
rm(Store_GridRain)
```

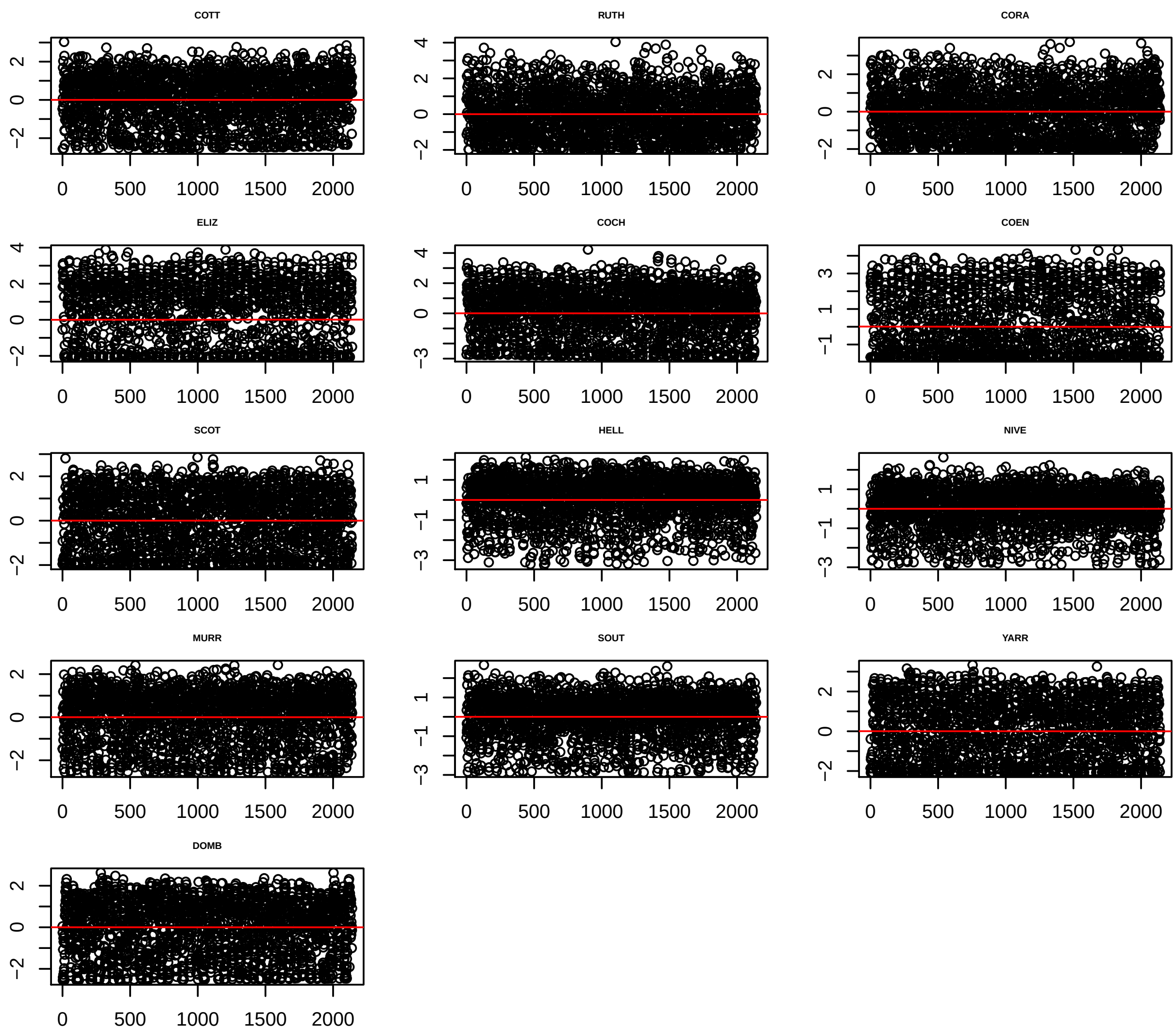

Figure 3: Residuals of linear mixed model analysis for trend in station rainfall data

Model 3 GAMM with rainfall

This model analyses flow as a function of rainfall only. This is therefore an analysis of the rainfall runoff coefficient, taking into account a possible time trend in the data. If the trend in this analysis is significant, then this is a measure of how the rainfall runoff coefficient has changed over time.
Again the analysis is run twice, once with station rainfall data (model 3a) and once with gridded rainfall data (model 3b).

Station rainfall data

```
# Gamm model with flow and rain
cl <- makeCluster(4) # create a cluster with 4 cores
registerDoParallel(cl) # register the cluster
# use a foreach loop to calibrate
Store_FwR <- foreach(i = 1:length(Stations[,1]),
                 .packages="mgcv") %dopar% {
  gamm.data <- subset(flow_rain_maxT_weekly,
                      flow_rain_maxT_weekly$Station == Stations[i,1])
  gamm.data$trend <- 1:nrow(gamm.data)
  gam_TrendFlow_withR <- gamm(log(Flow +1)~s(Rain) + trend,
                              correlation= corCAR1(), data=gamm.data)
  out <- list(model = gam_TrendFlow_withR,
              results = data.frame(Station=Stations[i,1],
                          t(summary(gam_TrendFlow_withR$lme)$tTable[2,c(1,5)]),
                              AIC=summary(gam_TrendFlow_withR$lme)$AIC))
  out
   }
stopCluster(cl)

par(mfrow=c(5,3), mar=c(2,2,2,2))
for (i in seq_along(Stations[,1])) {
  res <- residuals(Store_FwR[[i]]$model$lme)
  plot(res, main=Stations[i,1], cex.main=0.5,
       ylab="normalised residuals",xlab="")
  n <- length(res)
  abline(lsfit(1:n, res), col="red")
}

# store results
save(Store_FwR,
     file=paste(storedir,
              "projectdata/StoreFwR_TrendAnalysis.RData",
              sep="/"))
output <- do.call(rbind, lapply(1:length(Store_FwR),
                                function(i) rbind(Store_FwR[[i]][[2]])))
pander(output, caption="Mixed model results for analysis of trend in flow data taking into account Rain
```

Table 6: Mixed model results for analysis of trend in flow data taking into account Rainfall

Station	Value	p.value	AIC
COTT	-0.0002635	0.1214	807.7

Station	Value	p.value	AIC
RUTH	-0.0005527	6.471e-11	1519
CORA	-0.0001906	0.0007155	3700
ELIZ	-0.000197	0.4233	2813
COCH	-0.0001328	0.3417	1439
COEN	-2.843e-05	0.8813	2859
SCOT	-7.227e-05	0.4154	1603
HELL	-0.0001427	0.2644	2361
NIVE	-8.442e-05	0.6261	3014
MURR	-0.0001986	0.004321	-472.7
SOUT	-0.0001322	0.0259	1082
YARR	-0.000118	0.01033	-417.6
DOMB	-0.0002598	0.1395	1926

```
rm(Store_FwR)
```

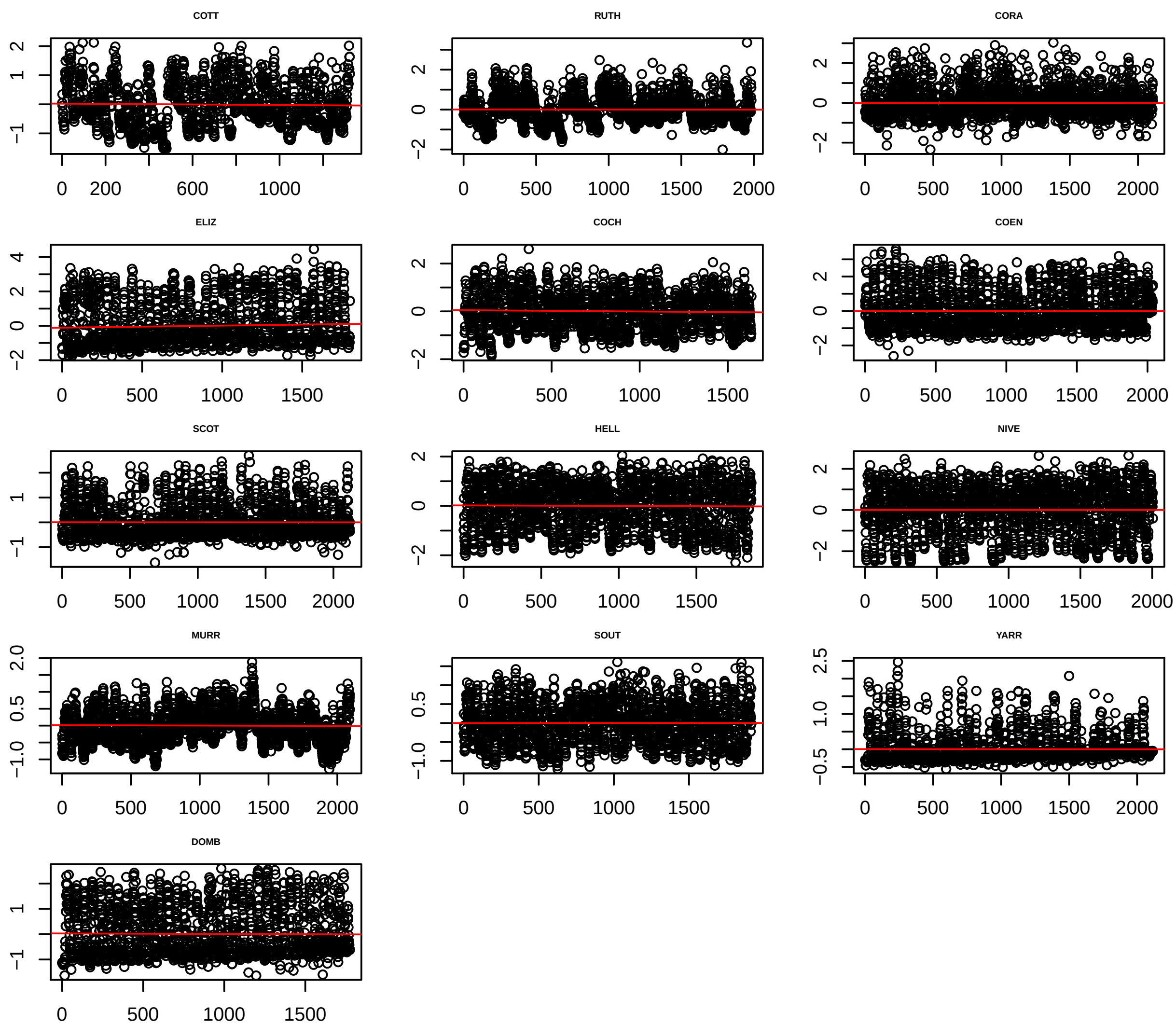

Figure 4: Residuals of GAMM analysis for trend in flow data taking into consideration station rainfall data

Gridded rainfall data

```
#gam model with flow and gridded rainfall
cl <- makeCluster(4) # create a cluster with 4 cores
registerDoParallel(cl) # register the cluster
# use a foreach loop to calibrate
Store_FwGR <- foreach(i = 1:length(Stations[,1]),
                .packages="mgcv") %dopar% {
  gamm.data <- subset(weekGridRainAllDataout,
                      weekGridRainAllDataout$Station == Stations[i,1])
  gamm.data$trend <- 1:nrow(gamm.data)
  gam_TrendFlow_withGR <- gamm(log(Flow +1)~s(gridRain) + trend,
                                  correlation= corCAR1(), data=gamm.data)
  out <- list(model = gam_TrendFlow_withGR,
              results = data.frame(Station=Stations[i,1],
                            t(summary(gam_TrendFlow_withGR$lme)$tTable[2,c(1,5)]),
                                  AIC=summary(gam_TrendFlow_withGR$lme)$AIC))
  out
   }
stopCluster(cl)

par(mfrow=c(5,3), mar=c(2,2,2,2))
for (i in seq_along(Stations[,1])) {
  res <- residuals(Store_FwGR[[i]]$model$lme)
  plot(res, main=Stations[i,1], cex.main=0.5,
       ylab="normalised residuals",xlab="")
  n <- length(res)
  abline(lsfit(1:n, res), col="red")
}

# store results
save(Store_FwGR,
     file=paste(storedir,
                "projectdata/StoreFwGR_TrendAnalysis.RData",
                sep = "/"))
output <- do.call(rbind, lapply(1:length(Store_FwGR),
                                function(i) rbind(Store_FwGR[[i]][[2]])))
pander(output, caption="Mixed model results for analysis of trend in flow data taking into account Grid
```

Table 7: Mixed model results for analysis of trend in flow data taking into account Gridded Rainfall

Station	Value	p.value	AIC
COTT	-0.0002591	0.02594	1657
RUTH	-0.0005896	1.902e-14	3456
CORA	-0.0002107	0.001708	5067
ELIZ	3.483e-06	0.9888	3687
COCH	-7.97e-06	0.9452	2929
COEN	-3.002e-05	0.8833	4037
SCOT	-8.132e-05	0.3642	2775
HELL	-0.0001371	0.2639	2919
NIVE	-6.429e-05	0.7136	3087
MURR	-0.0002163	0.001295	229.7

Station	Value	p.value	AIC
SOUT	-0.000154	0.01168	1909
YARR	-0.0001139	0.02261	-160.9
DOMB	-0.000124	0.5217	2770

```r
rm(Store_FwGR)
```

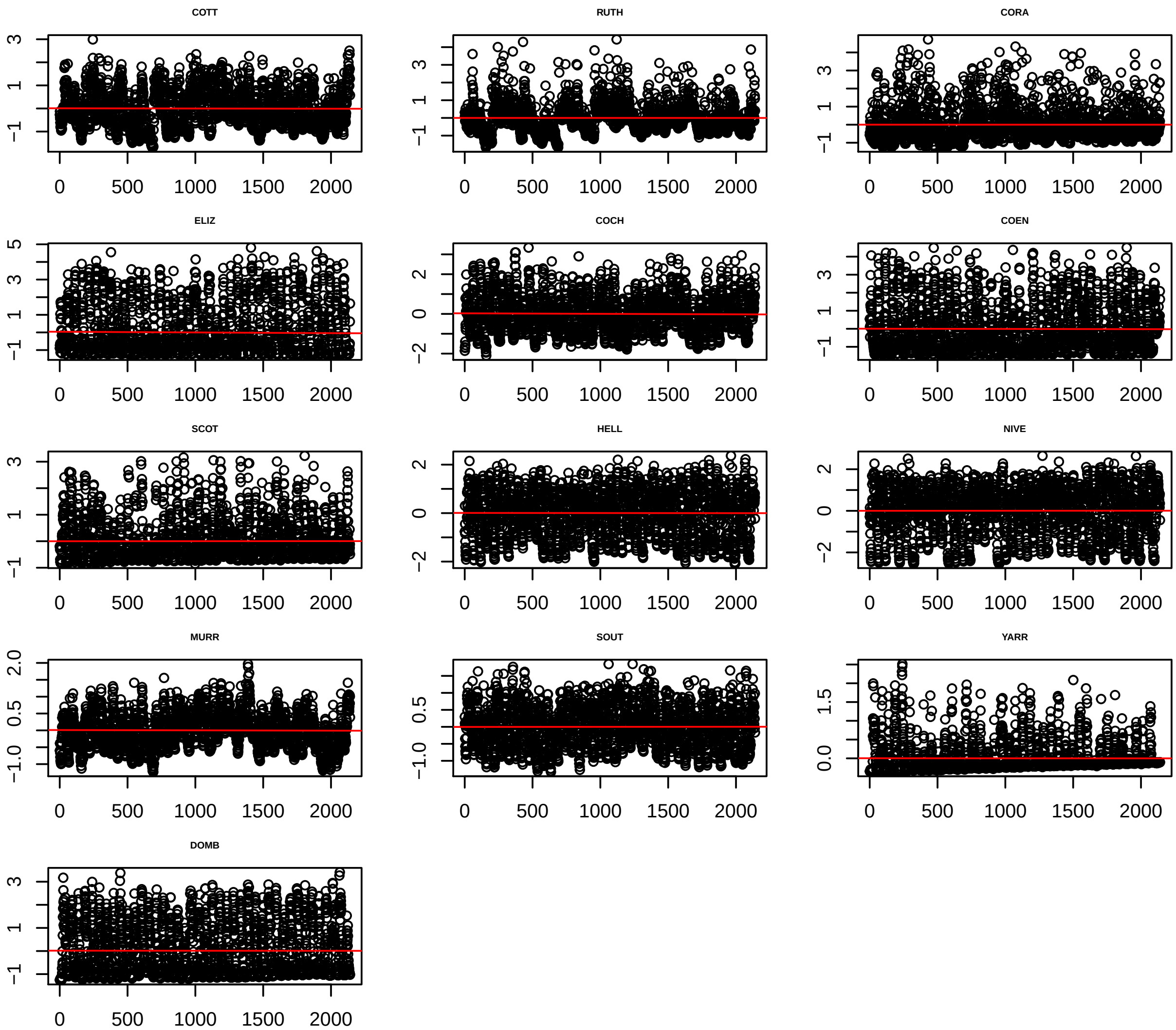

Figure 5: Residuals of GAMM analysis for trend in flow data taking into consideration gridded rainfall data

Model 4. GAMM with rain & s(rain,MaxT) and trend

This model analyses flow as a function of rainfall and the interaction between rainfall and maximum temperature, which is conceptualised as the actual evapotranspiration. This is therefore an analysis of the rainfall runoff coefficient, taking into account the changes in evapotranspiration and a possible time trend in the data. If the trend in this analysis is significant, then this is a measure of how the rainfall runoff coefficient has changed over time.

Again the analysis is run twice, once with station rainfall data (model 3a) and once with gridded rainfall data (model 3b).

Station rainfall data

```
# run the gamm model on rain, maxT and flow
cl <- makeCluster(4) # create a cluster with 4 cores
registerDoParallel(cl) # register the cluster
# use a foreach loop to calibrate
Store_FwRE <- foreach(i = 1:length(Stations[,1]),
                 .packages="mgcv") %dopar% {
  gamm.data <- subset(flow_rain_maxT_weekly,
                      flow_rain_maxT_weekly$Station == Stations[i,1])
  gamm.data$trend <- 1:nrow(gamm.data)
  gam_TrendFlow_withRandE <- gamm(log(Flow +1)~s(Rain) + s(Rain, MaxT) +
                                   trend, correlation= corCAR1(),
                                   data=gamm.data, control=list(niterEM=0))
  out <- list(model = gam_TrendFlow_withRandE,
        results = data.frame(Station=Stations[i,1],
                      t(summary(gam_TrendFlow_withRandE$lme)$tTable[2,c(1,5)]),
                            AIC=summary(gam_TrendFlow_withRandE$lme)$AIC))
  out
 }
stopCluster(cl)

par(mfrow=c(5,3), mar=c(2,2,2,2))
for (i in seq_along(Stations[,1])) {
  res <- residuals(Store_FwRE[[i]]$model$lme)
  plot(res, main=Stations[i,1], cex.main=0.5,
       ylab="normalised residuals",xlab="")
  n <- length(res)
  abline(lsfit(1:n, res), col="red")
}

# store results
save(Store_FwRE,
     file=paste(storedir,
                "projectdata/StoreFwRE_TrendAnalysis.RData",
                sep = "/"))
output <- do.call(rbind, lapply(1:length(Store_FwRE),
                                function(i) rbind(Store_FwRE[[i]][[2]])))
pander(output, caption="Mixed model results for the analysis of trend in flow data taking into account R
```

Table 8: Mixed model results for the analysis of trend in flow data taking into account Rainfall and Evapotranspiratiion

Station	Value	p.value	AIC
COTT	-0.0002282	0.154	749.1
RUTH	-0.0005415	4.384e-11	1475
CORA	-0.000195	0.0009571	3347
ELIZ	-0.0001815	0.4428	2726
COCH	-0.000131	0.3165	1362

Station	Value	p.value	AIC
COEN	-4.817e-06	0.9785	2709
SCOT	-8.012e-05	0.2356	1356
HELL	-0.0001206	0.1994	2113
NIVE	-6.754e-05	0.6476	2922
MURR	-0.0001848	0.003078	-670.3
SOUT	-0.0001189	0.01581	1013
YARR	-0.0001128	0.00571	-491.7
DOMB	-0.0002406	0.07439	1811

```r
rm(Store_FwRE)
```

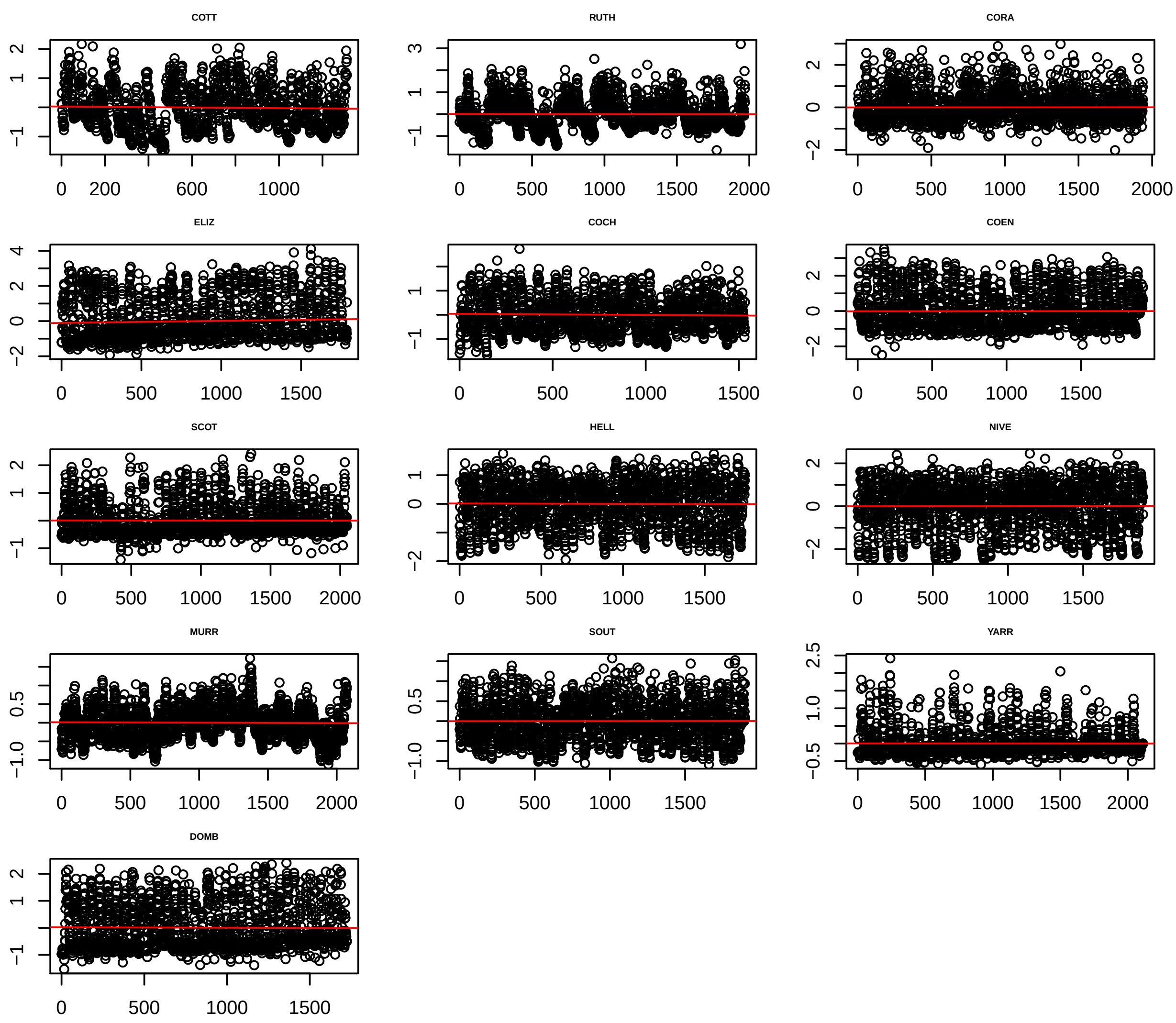

Figure 6: Residuals of GAMM analysis for trend in flow data removing station rainfall and evapotranspiration effects

Gridded rainfall data

The gridded rainfall data for some reason is slightly more complex and creates minor problems with fitting using GAMM. The difficulty is in setting k, which determines the flexibility of the smooths. The default value is k = -1 which allows the optimisation to set the smooths, but for R > 3.4.0 and updated package mgcv (20/05/2017) this fails. Setting k = 10 (which should be more than flexible enough) allows the models to run. This should not influence the results.

```
# run the gamm model on gridded rain, maxT and flow
cl <- makeCluster(4) # create a cluster with 4 cores
registerDoParallel(cl) # register the cluster
# use a foreach loop to calibrate
Store_FwGRE <- foreach(i = 1:length(Stations[,1]),
                 .packages="mgcv") %dopar% {
#  out <- list()
#for (i in 1:length(Stations[,1])) {
  gamm.data <- subset(weekGridRainAllDataout,
                      weekGridRainAllDataout$Station == Stations[i,1])
  gamm.data$trend <- 1:nrow(gamm.data)
  # need to set k is 10 (large enough), as it cannot be default k=-1
  # due to missing data. This is for R > 3.4.0
  # unclear, runs with k= -1 in R 3.3.0
  gam_TrendFlow_withGRE <- gamm(log(Flow +1) ~
                                s(gridRain,k=10) +
                                s(gridRain,MaxT, k=10) +
                                trend,
                                correlation= corCAR1(),
                                data=gamm.data,
                                control=list(niterEM=5))
  out <- list(model = gam_TrendFlow_withGRE,
        results = data.frame(Station=Stations[i,1],
       t(summary(gam_TrendFlow_withGRE$lme)$tTable[2,c(1,5)]),
       AIC=summary(gam_TrendFlow_withGRE$lme)$AIC))
  out
}
stopCluster(cl)

par(mfrow=c(5,3),mar=c(2,2,2,2))
for (i in seq_along(Stations[,1])) {
  res <- residuals(Store_FwGRE[[i]]$model$lme)
  plot(res, main=Stations[i,1], cex.main=0.5,
       ylab="normalised residuals",xlab="")
  n <- length(res)
  abline(lsfit(1:n, res), col="red")
}

# store results
save(Store_FwGRE,
     file=paste(storedir,
                "projectdata/StoreFwGRE_TrendAnalysis.RData",
                sep="/"))
output <- do.call(rbind, lapply(1:length(Store_FwGRE),
                                function(i) rbind(Store_FwGRE[[i]][[2]])))
pander(output, caption="Mixed model results for analysis of trend in flow data taking into account Grid
```

Table 9: Mixed model results for analysis of trend in flow data taking into account Gridded Rainfall and Evapotranspiration

Station	Value	p.value	AIC
COTT	-0.0002296	0.03122	1473
RUTH	-0.0005616	1.505e-13	3371
CORA	-0.0002075	0.003263	4578
ELIZ	2.366e-05	0.9237	3442
COCH	1.793e-05	0.8752	2548
COEN	3.763e-05	0.8352	3546
SCOT	-7.685e-05	0.1452	2397
HELL	-0.0001146	0.1792	2558
NIVE	-5.128e-05	0.7348	3019
MURR	-0.0001994	0.0005311	20.41
SOUT	-0.0001292	0.006673	1680
YARR	-0.0001097	0.009742	-218.3
DOMB	-0.0001214	0.3872	2551

```r
rm(Store_FwGRE)
```

Model 5, same as model 4, but no trend and Mann Kendall on the residuals

This last model is to check the trend with GAMM analysis with the analysis using Mann-Kendall. So rather than incorporating a trend in the model, we analyse the residuals using Mann-Kendall for a trend. In this case we drop the plotting of the residuals.

Station rainfall data

```r
# run the gamm model on rain, maxT and flow
cl <- makeCluster(4) # create a cluster with 4 cores
registerDoParallel(cl) # register the cluster
# use a foreach loop to calibrate
Store_FwRE2 <- foreach(i = 1:length(Stations[,1]),
                 .packages="mgcv") %dopar% {
  gamm.data <- subset(flow_rain_maxT_weekly,
                      flow_rain_maxT_weekly$Station == Stations[i,1])
  gam_Flow_withRandE <- gamm(log(Flow +1)~s(Rain) + s(Rain, MaxT) ,
                                  correlation= corCAR1(), data=gamm.data,
                                  control=list(niterEM=5))
  out <- list(model = gam_Flow_withRandE,
        results = data.frame(Station=Stations[i,1],
                      AIC=summary(gam_Flow_withRandE$lme)$AIC))
  out
}

stopCluster(cl)

# store results
```

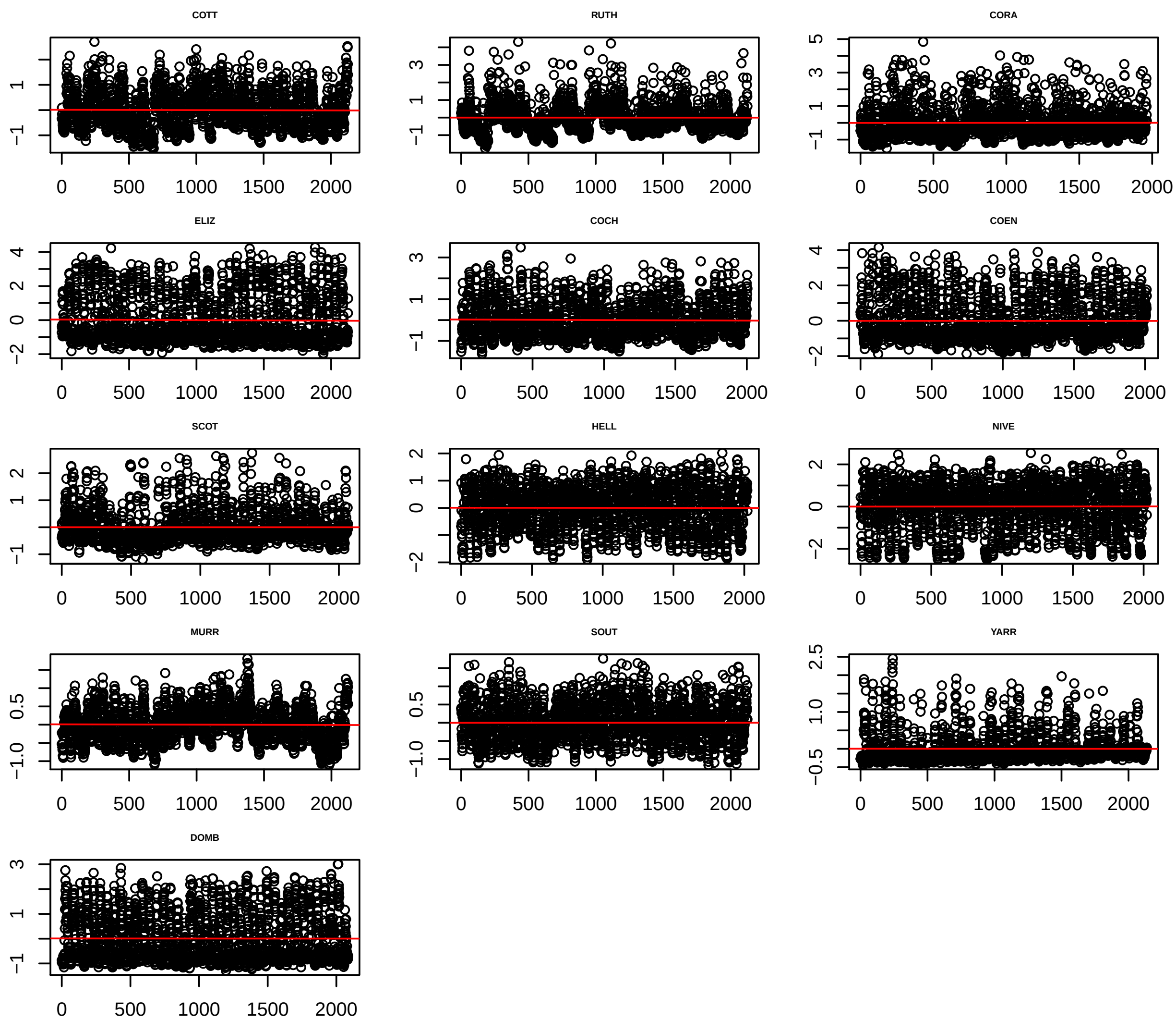

Figure 7: Residuals of GAMM analysis for trend in flow data removing gridded rainfall and evapotranspiration effects

```r
save(Store_FwRE2,
     file=paste(storedir,
                "projectdata/StoreFwRE_Analysis.RData",
                sep="/"))
output <- do.call(rbind, lapply(1:length(Store_FwRE2),
                                function(i) rbind(Store_FwRE2[[i]][[2]])))
pander(output, caption="Mixed model results for the analysis of flow data taking into account Rainfall a
```

Table 10: Mixed model results for the analysis of flow data taking into account Rainfall and Evapotranspiration

Station	AIC
COTT	749.1
RUTH	1507
CORA	3355
ELIZ	2724
COCH	1361
COEN	2707
SCOT	1355
HELL	2113
NIVE	2920
MURR	-664.1
SOUT	1017
YARR	-486.4
DOMB	1812

Now do the Mann-Kendall analysis on the residuals

```r
# do mann kendall on the residuals
resid_list <- vector("list", length=length(Stations[,1]))
for (i in seq_along(Stations[,1])) {
  resid_list[[i]] <- zoo(residuals(Store_FwRE2[[i]]$model$lme,
                type="normalized"),
                order.by=as.Date(na.omit(subset(flow_rain_maxT_weekly,
}
resid_df <- do.call(merge.zoo,resid_list)
names(resid_df) <- Stations[,1]
# Bootstrap
set.seed(10)
# now run a loop over the number of years (create 41 different sets)
# do Mann Kendall test on each resonstituted series
# ---------------------------
#  -------------------------
resid_temp <- as.data.frame(resid_df)
resid_temp$years <- format(time(resid_df),"%Y")
split_resid <- split(resid_temp[,1:13],resid_temp$years)


cl <- makeCluster(4) # create a cluster with 4 cores
registerDoParallel(cl) # register the cluster
# use a foreach loop to calibrate
MK_list <- foreach(i = 1:500,
                 .packages=c("Kendall", "xts")) %dopar% {
```

```
  # reorganise the list elements
  series <- sample(1:nyears(resid_df),nyears(resid_df))
  for (j in 1:length(series)) {
    if (j==1) {
      new_df <- as.data.frame(split_resid[[series[j]]])
    } else {
      new_df <- rbind(new_df,as.data.frame(split_resid[[series[j]]]))
    }
  }
  # run mann kendall on the columns and store the results
  mk_r <- apply(new_df,2,MannKendall)

  out <- do.call(cbind,mk_r)
  out
 }
stopCluster(cl)


MK_df <- do.call(rbind,MK_list)

pvalues <- subset(MK_df, rownames(MK_df)=="sl")
tau <- subset(MK_df, rownames(MK_df)=="tau")

sig_set <- list()

for (i in 1:ncol(pvalues)) {
  set <- data.frame(pvalue=as.numeric(pvalues[,i]),
                    tau=as.numeric(tau[,i]),catch=rep(colnames(MK_df)[i],nrow(tau)))
  sig_set[[i]] <- set[set$pvalue < 0.5,]
}

sig_set_a <- do.call(rbind,sig_set)
sig_set_a$type <- rep("bootstrap",nrow(sig_set_a))


MK_resid <- do.call(rbind,lapply(resid_list,MannKendall))


real_df <- data.frame(pvalue = as.numeric(MK_resid[,2]),
                      tau = as.numeric(MK_resid[,1]),
                      catch=Stations[,1],
                      type=rep("real",nrow(MK_resid)))
# A histogram of taus

hp <- ggplot(sig_set_a, aes(x=tau)) + geom_histogram(binwidth=0.03,colour="white")
# Histogram of significant tau's, divided by catch
# With panels that have the same scaling, but different range
# (and therefore different physical sizes)
hp <- hp + facet_wrap(~ catch,ncol=5)
# add a red point for the real slope from the data
p_value <- ifelse(real_df$pvalue<0.05,"< 0.05",">= 0.05")
hp <- hp + geom_point(data=real_df,aes(x=tau, y=0,colour=p_value),
                shape=16,size=5) +
```

```
  facet_wrap(~ catch,ncol=5)+ ggtitle("Residuals Streamflow after GAM") #+
hp <- hp + scale_colour_grey(start = 0, end = 0.6)
print(hp)
```

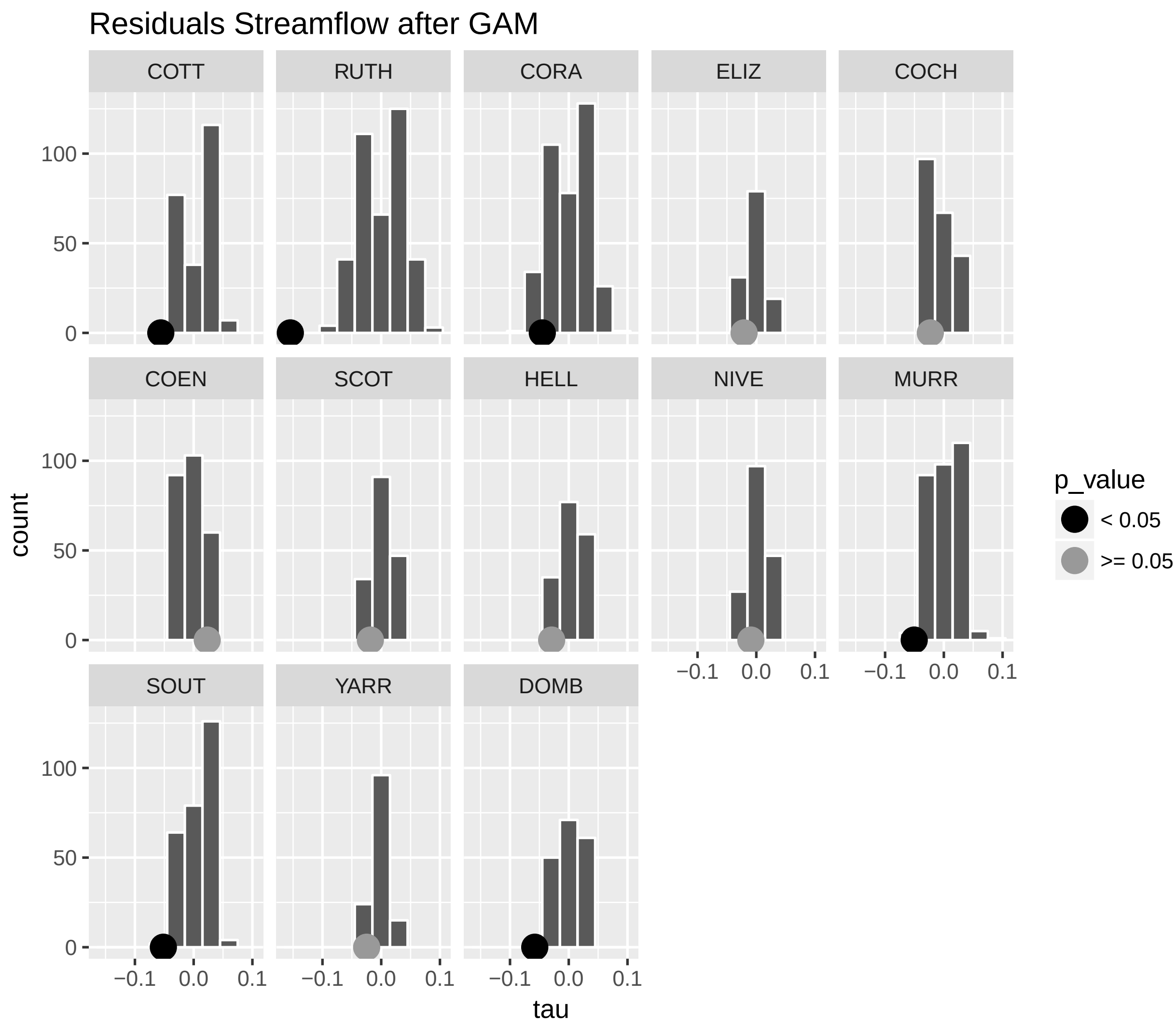

Figure 8: Mann Kendall analysis of the residuals of the streamflow after GAM model with rainfall and a Evapotranspiration

```
save(hp, file="../Figure8ResidGAM_MDPaper.Rdata")
save(real_df, file="../MKResidGAM_MDPaper.Rdata")
# # production quality tiff, this is Figure 8 in the manuscript
tiff("../manuscript/Figure8_ResidGAM_MDPaper.tif",res=600,compression="lzw",
     width=10*480,height=10*480)
print(hp)
dev.off()
```

```
## pdf
##   2
```

```r
pander(real_df, caption="Mann Kendall results for the GAMM residuals, ref Figure 8")
```

Table 11: Mann Kendall results for the GAMM residuals, ref Figure 8

pvalue	tau	catch	type
0.002394	-0.05596	COTT	real
0	-0.1547	RUTH	real
0.002975	-0.04502	CORA	real
0.1877	-0.0208	ELIZ	real
0.1764	-0.02305	COCH	real
0.1292	0.02312	COEN	real
0.2107	-0.01845	SCOT	real
0.06842	-0.0291	HELL	real
0.5471	-0.009223	NIVE	real
0.0005784	-0.05042	MURR	real
0.000755	-0.0516	SOUT	real
0.08815	-0.02476	YARR	real
0.0003182	-0.05786	DOMB	real

```r
rm(Store_FwRE2)
```

Gridded rainfall data

Do the same for the gridded rainfall data See the comments with model 4b with regard to setting the parameter k. Kept this consistent between model 4b and 5b.

```r
# run the gamm model on rain, maxT and flow
cl <- makeCluster(4) # create a cluster with 4 cores
registerDoParallel(cl) # register the cluster
# use a foreach loop to calibrate
Store_FwGRE2 <- foreach(i = 1:length(Stations[,1]),
                 .packages="mgcv") %dopar% {
  gamm.data <- subset(weekGridRainAllDataout,
                      weekGridRainAllDataout$Station == Stations[i,1])
  gam_Flow_withGRE <- gamm(log(Flow +1) ~
                             s(gridRain, k=10) +
                             s(gridRain,MaxT, k=10),
                             correlation= corCAR1(),
                             data=gamm.data,
                             control=list(niterEM=5))
  out <- list(model = gam_Flow_withGRE,
        results = data.frame(Station=Stations[i,1],
                            AIC=summary(gam_Flow_withGRE$lme)$AIC))
  out
}

stopCluster(cl)

# store results
save(Store_FwGRE2,
     file=paste(storedir,
```

```
                "projectdata/StoreFwGRE2_TrendAnalysis.RData",
                sep="/"))
output <- do.call(rbind, lapply(1:length(Store_FwGRE2),
                               function(i) rbind(Store_FwGRE2[[i]][[2]])))
pander(output, caption="Mixed model results for analysis of trend in flow data taking into account Grid
```

Table 12: Mixed model results for analysis of trend in flow data taking into account Gridded Rainfall and Evapotranspiration

Station	AIC
COTT	1476
RUTH	3415
CORA	4584
ELIZ	3440
COCH	2546
COEN	3544
SCOT	2397
HELL	2558
NIVE	3017
MURR	29.58
SOUT	1685
YARR	-213.9
DOMB	2550

Now do the Mann-Kendall analysis on the residuals

```
# do mann kendall on the residuals
resid_list <- vector("list", length=length(Stations[,1]))
for (i in seq_along(Stations[,1])) {
  resid_list[[i]] <- zoo(residuals(Store_FwGRE2[[i]]$model$lme,
                 type="normalized"),
                 order.by=as.Date(na.omit(subset(flow_rain_maxT_weekly,
}
resid_df <- do.call(merge.zoo,resid_list)
names(resid_df) <- Stations[,1]
# Bootstrap
# now run a loop over the number of years (create 41 different sets)
# do Mann Kendall test on each resonstituted series
# ---------------------------
#  -------------------------
resid_temp <- as.data.frame(resid_df)
resid_temp$years <- format(time(resid_df),"%Y")
split_resid <- split(resid_temp[,1:13],resid_temp$years)


cl <- makeCluster(4) # create a cluster with 4 cores
registerDoParallel(cl) # register the cluster
# use a foreach loop to calibrate
MK_list <- foreach(i = 1:500,
                 .packages=c("Kendall","xts")) %dopar% {
  # reorganise the list elements
  series <- sample(1:nyears(resid_df),nyears(resid_df))
  for (j in 1:length(series)) {
```

```
    if (j==1) {
      new_df <- as.data.frame(split_resid[[series[j]]])
    } else {
      new_df <- rbind(new_df,as.data.frame(split_resid[[series[j]]]))
    }
  }
  # run mann kendall on the columns and store the results
  mk_r <- apply(new_df,2,MannKendall)

  out <- do.call(cbind,mk_r)
  out
 }
stopCluster(cl)

MK_df <- do.call(rbind,MK_list)

pvalues <- subset(MK_df, rownames(MK_df)=="sl")
tau <- subset(MK_df, rownames(MK_df)=="tau")

sig_set <- list()

for (i in 1:ncol(pvalues)) {
  set <- data.frame(pvalue=as.numeric(pvalues[,i]),
                  tau=as.numeric(tau[,i]),catch=rep(colnames(MK_df)[i],nrow(tau)))
  sig_set[[i]] <- set[set$pvalue < 0.5,]
}

sig_set_a <- do.call(rbind,sig_set)
sig_set_a$type <- rep("bootstrap",nrow(sig_set_a))


MK_resid <- do.call(rbind,lapply(resid_list,MannKendall))


real_df <- data.frame(pvalue = as.numeric(MK_resid[,2]),
                      tau = as.numeric(MK_resid[,1]),
                      catch=Stations[,1],
                      type=rep("real",nrow(MK_resid)))
# A histogram of taus

hp <- ggplot(sig_set_a, aes(x=tau)) + geom_histogram(binwidth=0.03,colour="white")
# Histogram of significant tau's, divided by catch
# With panels that have the same scaling, but different range
# (and therefore different physical sizes)
hp <- hp + facet_wrap(~ catch,ncol=5)
# add a red point for the real slope from the data
p_value <- ifelse(real_df$pvalue<0.05,"< 0.05",">= 0.05")
hp <- hp + geom_point(data=real_df,aes(x=tau, y=0,colour=p_value),
                shape=16,size=5) +
  facet_wrap(~ catch,ncol=5)+ ggtitle("Residuals Streamflow after GAM") #+
hp <- hp + scale_colour_grey(start = 0, end = 0.6)
print(hp)
```

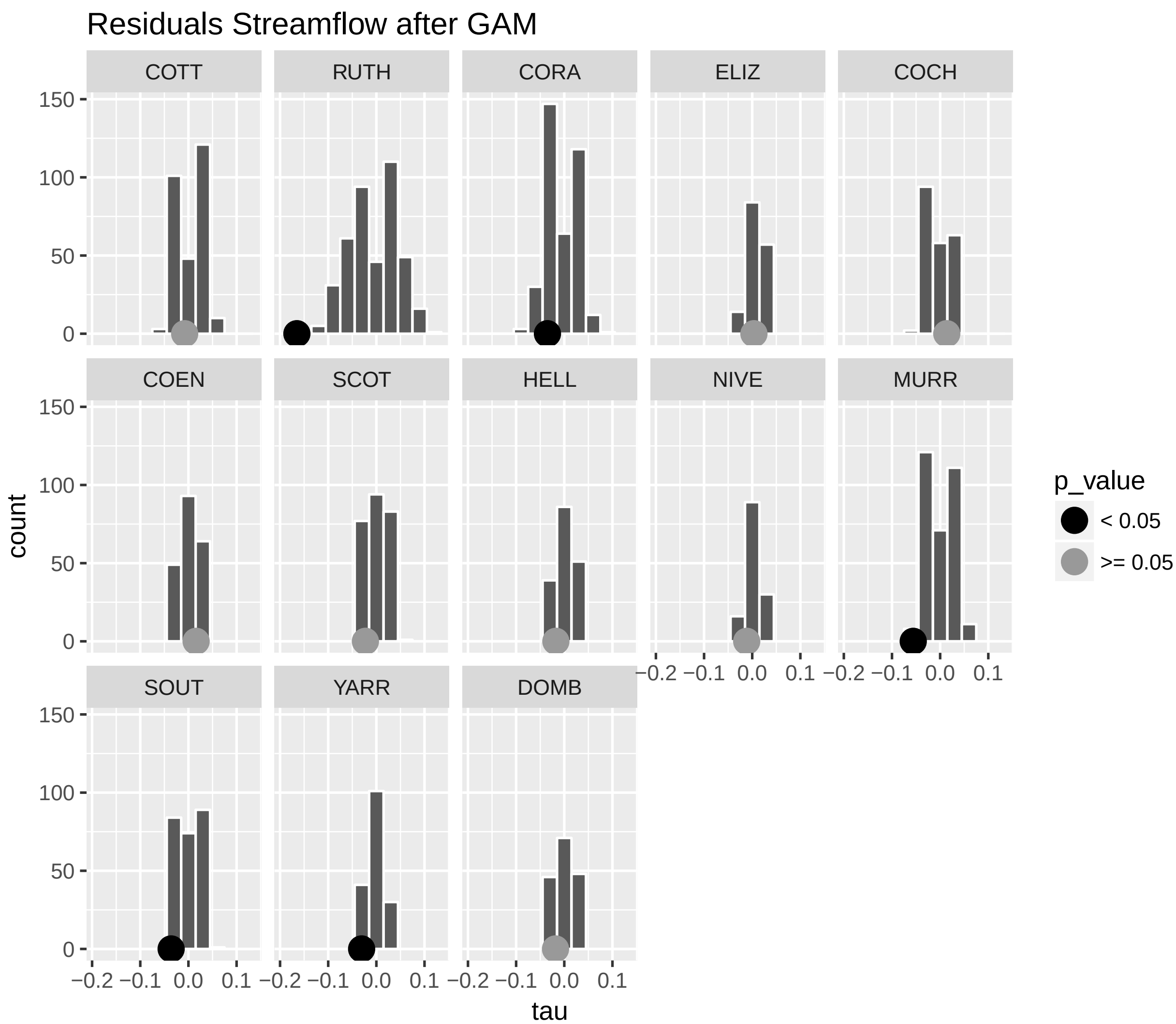

Figure 9: Mann Kendall analysis of the residuals of the streamflow after GAM model with gridded rainfall and a Evapotranspiration

```r
save(real_df, file="../GrMKResidGAM_MDPaper.Rdata")

pander(real_df, caption="Mann Kendall results for the GAMM residuals with gridded rainfall")
```

Table 13: Mann Kendall results for the GAMM residuals with gridded rainfall

pvalue	tau	catch	type
0.6709	-0.007832	COTT	real
0	-0.1651	RUTH	real
0.02107	-0.03496	CORA	real
0.8209	0.003574	ELIZ	real
0.4201	0.01374	COCH	real
0.2947	0.01597	COEN	real
0.1228	-0.02275	SCOT	real
0.2729	-0.01751	HELL	real
0.4546	-0.01145	NIVE	real
0.000136	-0.0559	MURR	real
0.01897	-0.03594	SOUT	real
0.03498	-0.03061	YARR	real
0.2565	-0.01824	DOMB	real

```r
rm(Store_FwGRE2)
```