GAM analysis of the weekly data

Willem Vervoort, Michaela Dolk & Floris van Ogtrop

2020-02-10

```
# root dir
knitr::opts_knit$set(root.dir =
                       "d:/cloudstor/Virtual Experiments/VirtExp")
knitr::opts_chunk$set(echo = TRUE)
# LOAD REQUIRED PACKAGES # #####
library(pander)
library(tidyverse)
library(xts)
library(zoo)
library(mgcv)
library(Kendall)
library(ggplot2)
library(doParallel)
library(foreach)

storedir <- "d:/cloudstor/virtual experiments"
```

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.).
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. In the analysis, the trend models are fitted to flow and rain data > 0, as this means the log transformation is better interpretable. In the end, the analysis is interested in changes in the positive flow values. A binary model could be fit to check if there is any change in the number of weeks or days with rainfall.
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/ClimCh_project_MD.Rdata")
```

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_delim("documents/Table2Models.csv", delim=";")
```

```
## Parsed with column specification:
## cols(
##   No = col_double(),
##   Model = col_character(),
##   Trend = col_character(),
##   Analysis = col_character()
## )
```

```
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
1	Log(Q) ~ trend + error	linear
2	Log(P) ~ trend + error	linear
3	Log(Q) ~ s(P) + trend + error	linear
4	Log (Q + 1) ~ s(P) + s(maxT, P) + trend + error	linear
5	Log(Q + 1) ~ s(P) + s(maxT, P) + error	Mann-Kendall

Analysis
40-year trend in streamflow to compare with Mann Kendall analysis.
40-year trend in rainfall to compare with Mann Kendall analysis.
The trend in this model relative to model 1 indicates the importance of other processes.
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

```r
# run the gls model on flowtrend only
#for (i in seq_along(Stations[,1])) {
cl <- makeCluster(6) # create a cluster with 6 cores
registerDoParallel(cl) # register the cluster
# use a foreach loop to calibrate
Store2 <- foreach(i = 1:length(Stations[,1]),
                  .packages="mgcv") %dopar% {
# for (i in seq_along(Stations[,1])) {
#   i <- 1
  gamm.data <- subset(flow_rain_maxT_weekly,
                      flow_rain_maxT_weekly$Station == Stations[i,1])
  gamm.data$Flow[gamm.data$Flow==0] <- NA
  gamm.data$trend <- 1:nrow(gamm.data)
  gam_TrendOnly <- gls(log(Flow)~trend, correlation= corAR1(),
       data=na.omit(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)

# store results
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	-0.0003546	0.01402	3704
RUTH	-0.000799	3.137e-15	4720
CORA	-0.0003226	0.01401	7208
ELIZ	0.00045	0.1192	4684
COCH	-1.016e-05	0.9262	4002
COEN	0.0002589	0.4821	7510
SCOT	-0.0002845	0.3101	5690
HELL	-0.0001685	0.1736	4020
NIVE	-6.304e-05	0.7877	4471
MURR	-0.0002631	4.74e-06	2583
SOUT	-0.0001576	0.004749	3227
YARR	-0.0002415	0.5285	4543
DOMB	-0.0002252	0.5674	5257

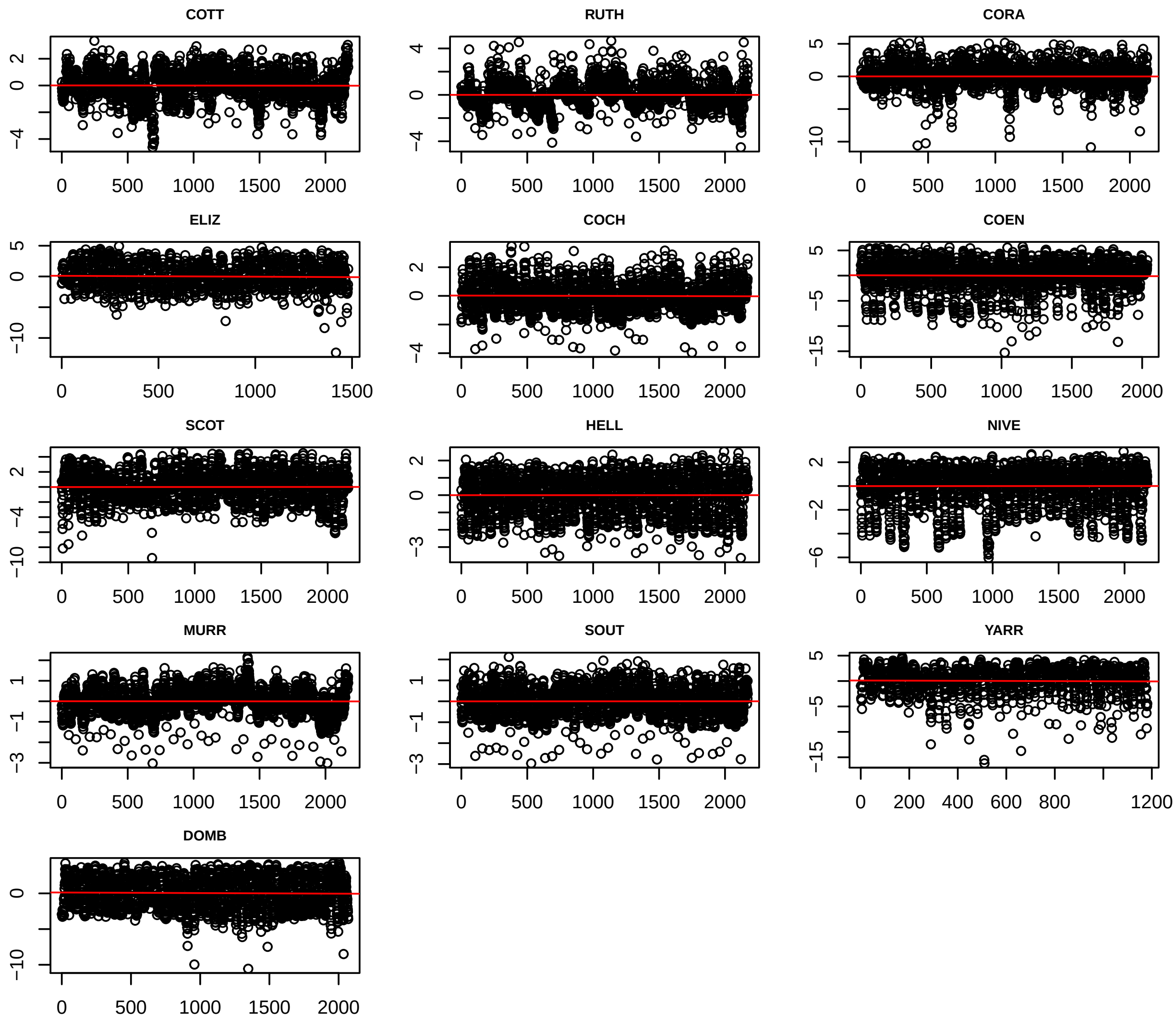

Figure 1: Residuals of linear mixed model analysis for trend in flow only

```
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")
}
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

```
# create an empty list
# and an empty dataframe to store results
# run the gls model on flowtrend only
cl <- makeCluster(6) # create a cluster with 6 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$Station == Stations[i,1])
  gamm.data$Rain[gamm.data$Rain==0] <- NA
  gamm.data$trend <- 1:nrow(gamm.data)
  gam_TrendR <- gls(log(Rain)~trend, correlation= corAR1(),
       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)


# 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	5.108e-05	0.4865	5185
RUTH	-0.0001815	0.009602	6573
CORA	-0.0001515	0.01549	6815
ELIZ	-0.0002335	0.05193	4130
COCH	0.0002211	0.003489	6562
COEN	-0.0004279	0.001529	4514
SCOT	-3.649e-05	0.6008	6710
HELL	3.56e-05	0.4334	6414
NIVE	-0.000183	0.0004455	6097
MURR	-0.0001045	0.04252	6324
SOUT	-0.0001966	1.669e-05	6501

Station	Value	p.value	AIC
YARR	-7.085e-05	0.4359	6421
DOMB	6.066e-05	0.3966	5649

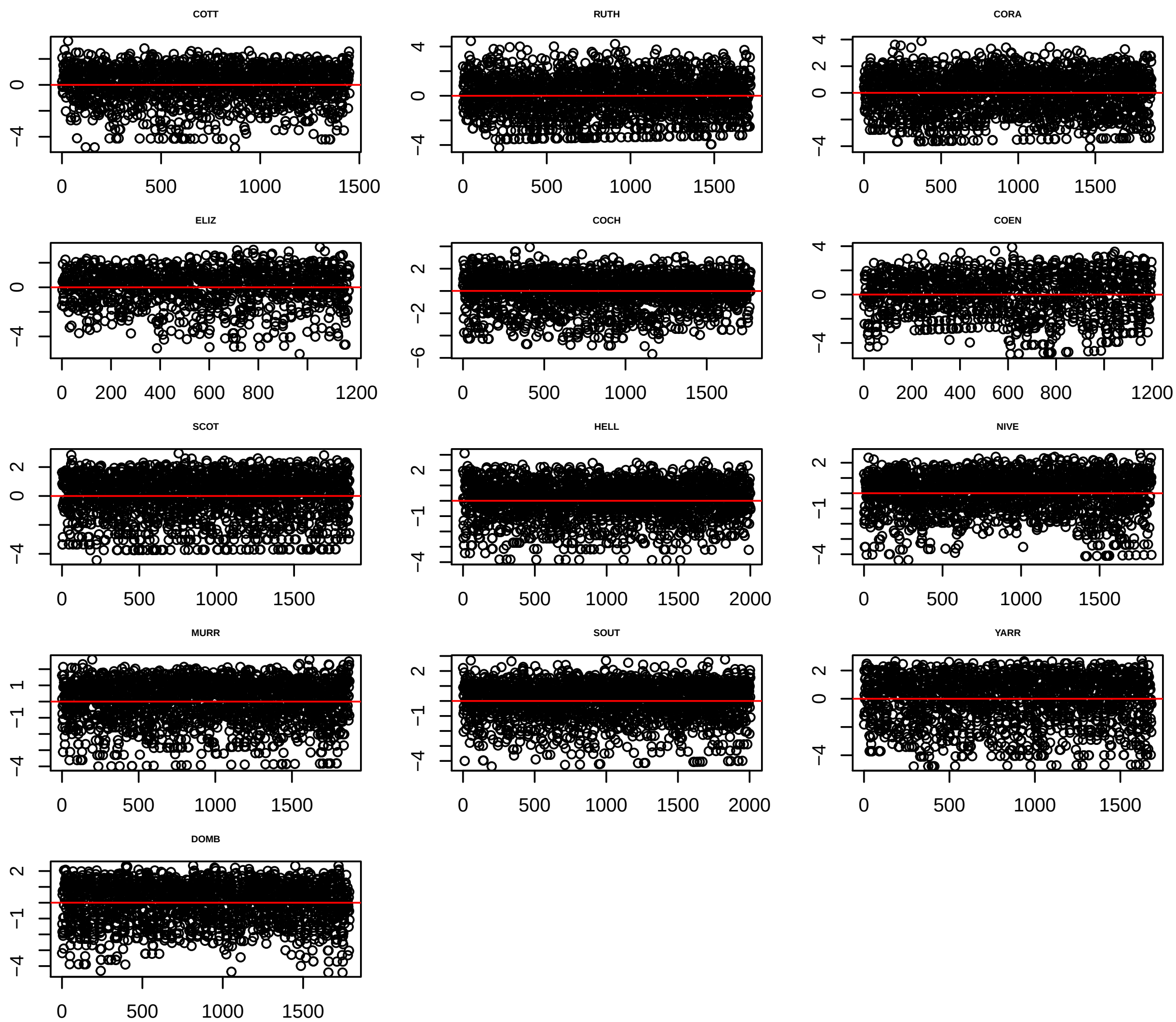

Figure 2: Residuals of linear mixed model analysis for trend in station rainfall data

```
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")
}
rm(Store_Rain)
```

Rainfall gridded data analysis

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

```r
# create an empty list
# and an empty dataframe to store results
# run the gls model on gridraintrend only
cl <- makeCluster(6) # create a cluster with 6 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(flow_rain_maxT_weekly,
                      flow_rain_maxT_weekly$Station == Stations[i,1])
  gamm.data$gridRain[gamm.data$gridRain==0] <- NA
  gamm.data$trend <- 1:nrow(gamm.data)
  gam_TrendGridR <- gls(log(gridRain)~trend, correlation= corAR1(),
       data=na.omit(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)

# 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	-0.0001669	0.003116	6950
RUTH	-0.0001857	0.007627	7266
CORA	-0.0001504	0.02484	6775
ELIZ	8.306e-05	0.4373	4704
COCH	-7.317e-05	0.4043	7071
COEN	-0.0002984	0.05627	4792
SCOT	-5.27e-05	0.4394	6724
HELL	-4.105e-05	0.4827	7295
NIVE	-0.000108	0.03926	7217
MURR	-9.361e-05	0.1143	7078
SOUT	-0.0001672	0.00117	7250
YARR	-1.924e-05	0.8348	6119
DOMB	-0.0001051	0.2424	7232

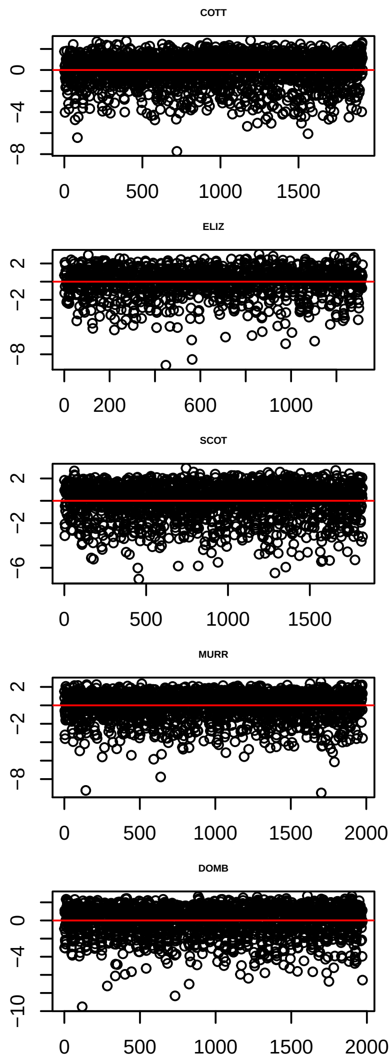

Figure 3: Residuals of linear mixed model analysis for trend in gridded rainfall data

```
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")
}
rm(Store_GridRain)
```

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).

For the GAMM models, the flexibility of the smooths is constrained to be k = 5 (**???**). This should be flexible enough for general purposes and not lead to overfitting. However, we needed to use k = 5 rather than a stiffer k = 3 to allow convergence for some of the models.

Station rainfall data

```
# Gamm model with flow and rain
cl <- makeCluster(6) # create a cluster with 6 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$Station == Stations[i,1])
  #gamm.data$Flow[gamm.data$Flow==0] <- NA
  gamm.data$trend <- 1:nrow(gamm.data)
  gam_TrendFlow_withR <- gamm(log(Flow+1)~s(Rain, k=5) + trend,
                              correlation= corCAR1(), data=na.omit(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)

# 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	-0.0002565	0.01313	1630
RUTH	-0.0005318	5.668e-12	2098
CORA	-0.0001701	0.001904	3820
ELIZ	-4.057e-05	0.8521	3189
COCH	4.103e-06	0.9654	2542
COEN	-1.684e-05	0.9298	2923
SCOT	-7.084e-05	0.4252	1490
HELL	-0.0001295	0.23	3118
NIVE	-6.483e-05	0.6832	3503

Station	Value	p.value	AIC
MURR	-0.0002062	3.506e-05	900.9
SOUT	-0.0001138	0.02167	2183
YARR	-0.0001077	0.01987	-506.7
DOMB	-0.0001219	0.5047	2104

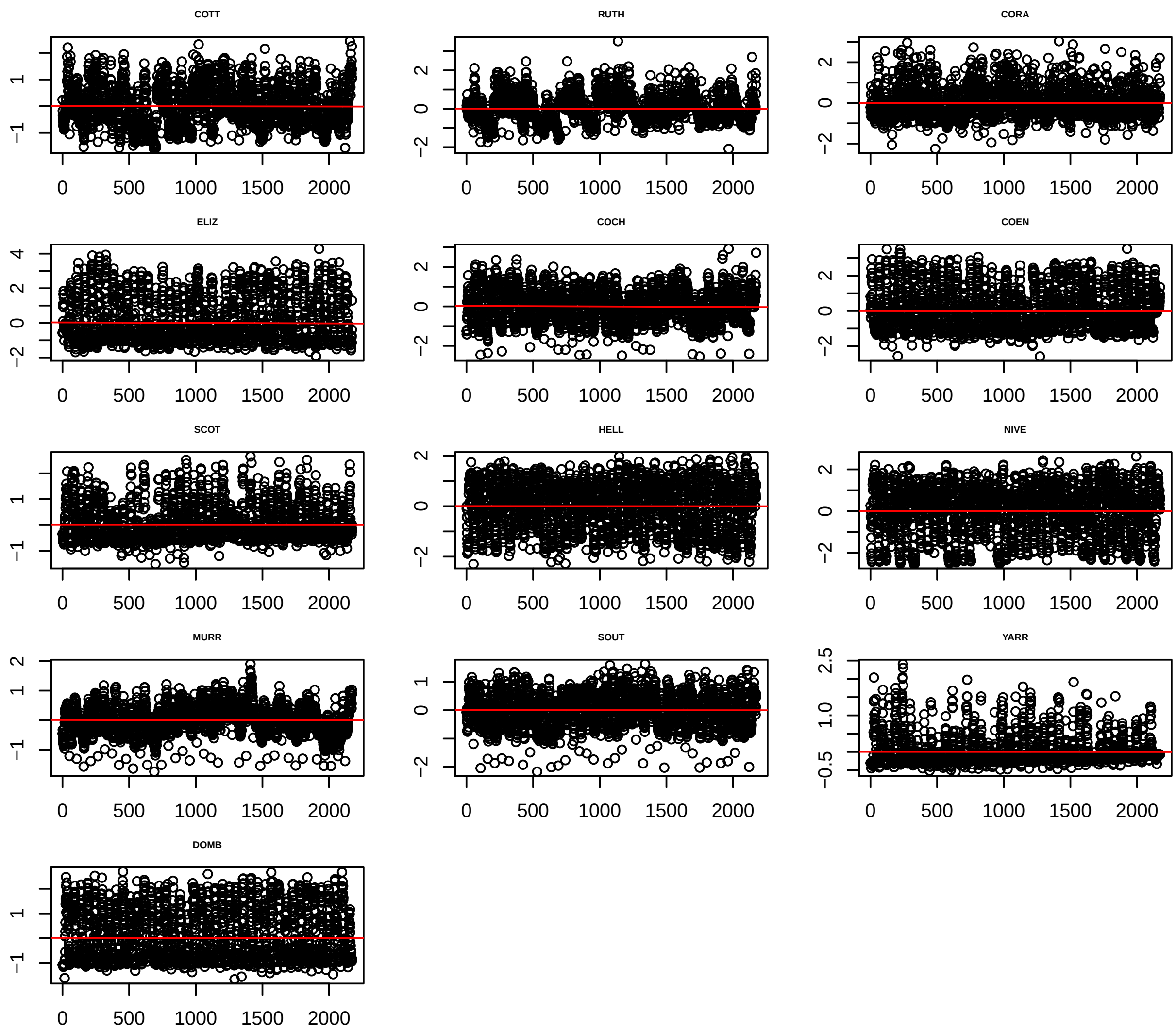

Figure 4: Residuals of GAMM analysis for trend in flow data taking into consideration station rainfall data

```
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")
}
rm(Store_FwR)
```

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(flow_rain_maxT_weekly,
                      flow_rain_maxT_weekly$Station == Stations[i,1])
  #gamm.data$Flow[gamm.data$Flow==0] <- NA
  gamm.data$trend <- 1:nrow(gamm.data)
  gam_TrendFlow_withGR <- gamm(log(Flow+1)~s(gridRain, k=3) + trend,
                               correlation= corCAR1(),
                            data=na.omit(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)

# 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 Gridd
```

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

Station	Value	p.value	AIC
COTT	-0.0002306	0.0313	1347
RUTH	-0.0005375	5.441e-12	1957
CORA	-0.0001632	0.002504	3801
ELIZ	-3.02e-05	0.8857	2737
COCH	-1.3e-06	0.989	2157
COEN	-1.277e-05	0.9453	2791
SCOT	-6.444e-05	0.4501	1561
HELL	-0.0001372	0.1585	2902
NIVE	-6.139e-05	0.6897	3395
MURR	-0.0002025	3.964e-05	774.9
SOUT	-0.0001143	0.01588	2240
YARR	-0.0001101	0.01643	-512.9
DOMB	-0.0001137	0.5228	2128

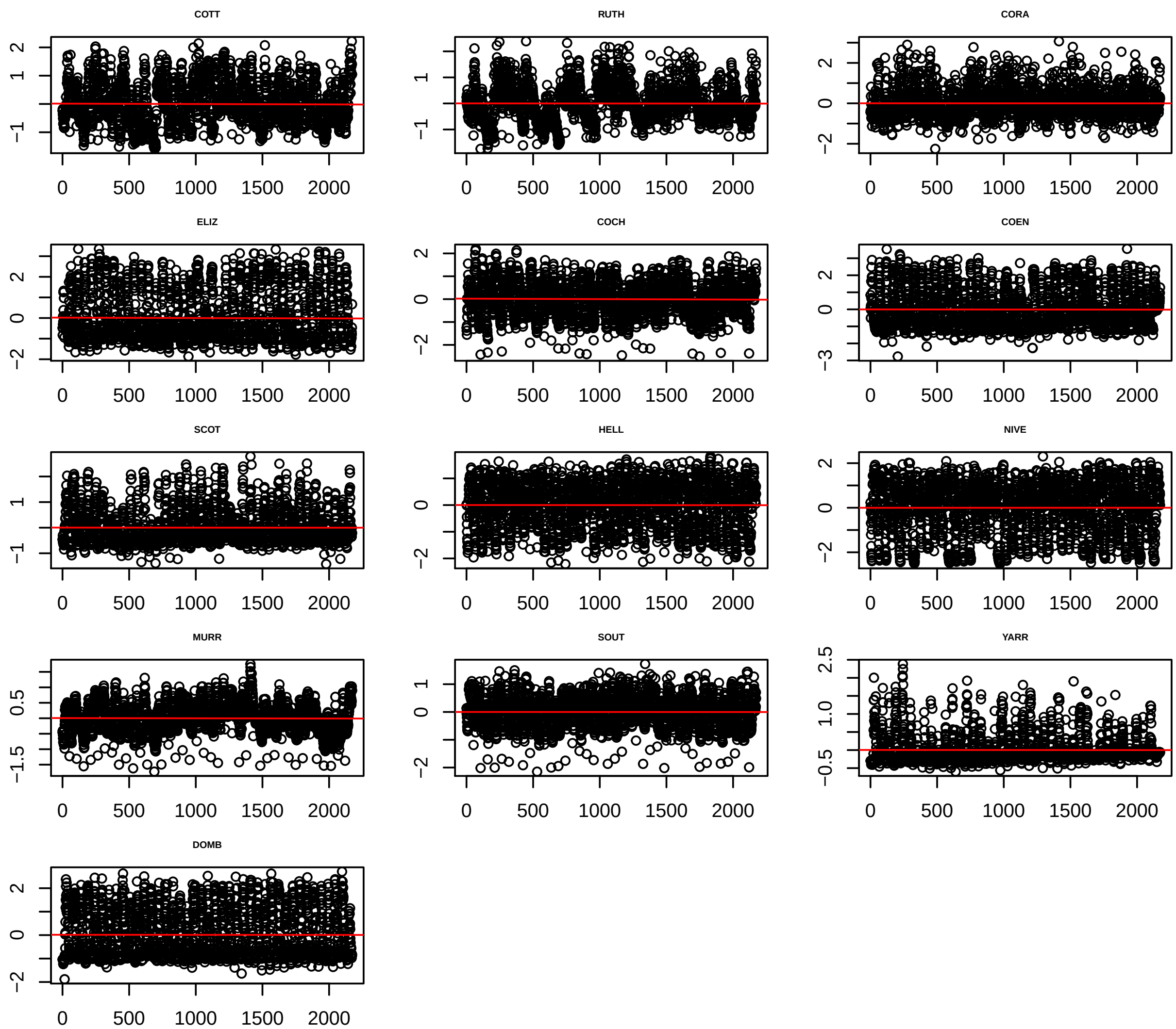

Figure 5: Residuals of GAMM analysis for trend in flow data taking into consideration gridded rainfall data

```
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")
}
rm(Store_FwGR)
```

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$Station == Stations[i,1])
  #gamm.data$Flow[gamm.data$Flow==0] <- NA
  gamm.data$trend <- 1:nrow(gamm.data)
  gam_TrendFlow_withRandE <- gamm(log(Flow+1)~s(Rain, k=3) + s(Rain, MaxT, k=3) +
                                      trend, correlation= corCAR1(),
                                      data=na.omit(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)

# 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 Evapotranspiration

Station	Value	p.value	AIC
COTT	-0.0002628	0.02516	1326
RUTH	-0.0005444	7.67e-10	1960
CORA	-0.0001674	0.001817	3861
ELIZ	-4.086e-05	0.8563	3114
COCH	-3.187e-06	0.9773	2187
COEN	-2.193e-05	0.9095	2916
SCOT	-7.285e-05	0.3837	1554
HELL	-0.0001359	0.2603	3065

Station	Value	p.value	AIC
NIVE	-7.237e-05	0.6789	3420
MURR	-0.0002052	0.002626	-201.2
SOUT	-0.0001213	0.05517	1380
YARR	-0.0001075	0.01846	-509
DOMB	-0.0001218	0.4992	2120

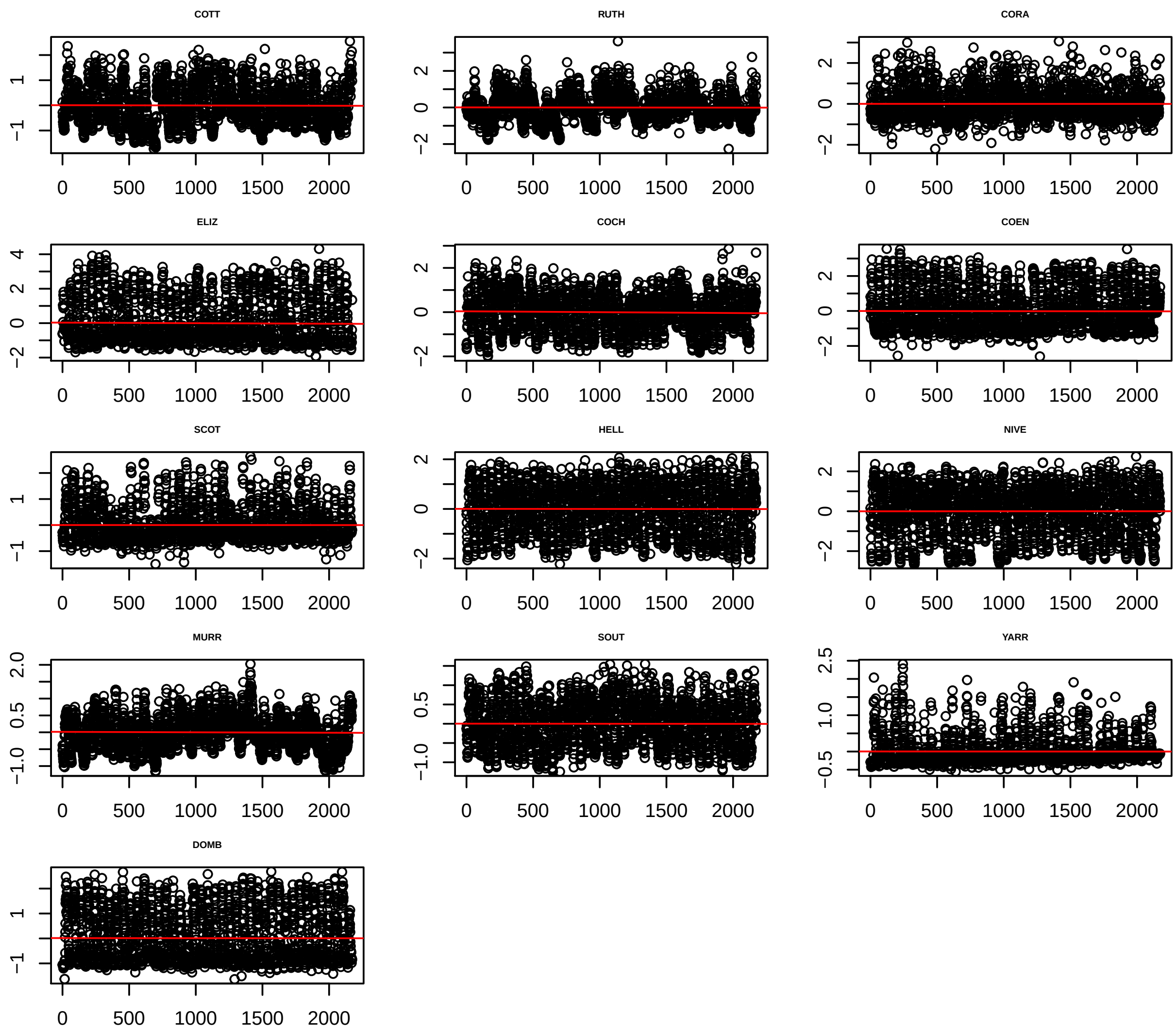

Figure 6: Residuals of GAMM analysis for trend in flow data removing station rainfall and evapotranspiration effects

```
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")
}
rm(Store_FwRE)
```

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 = 3 (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(6) # create a cluster with 6 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(flow_rain_maxT_weekly,
                      flow_rain_maxT_weekly$Station == Stations[i,1])
  #gamm.data$Flow[gamm.data$Flow==0] <- NA
  gamm.data$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=3) +
                                s(gridRain,MaxT, k=3) +
                                trend,
                                correlation= corCAR1(),
                                data=na.omit(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)

# 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	-0.0002363	0.0591	997.8
RUTH	-0.0005485	6.455e-10	1746
CORA	-0.0001627	0.002392	3804
ELIZ	-3.019e-05	0.8896	2676
COCH	-8.489e-06	0.9418	1693
COEN	-1.569e-05	0.9327	2781

Station	Value	p.value	AIC
SCOT	-6.624e-05	0.4324	1556
HELL	-0.0001438	0.1991	2815
NIVE	-6.999e-05	0.6893	3291
MURR	-0.000202	0.003849	-404.6
SOUT	-0.0001251	0.03497	1460
YARR	-0.0001098	0.01567	-514.3
DOMB	-0.0001138	0.521	2131

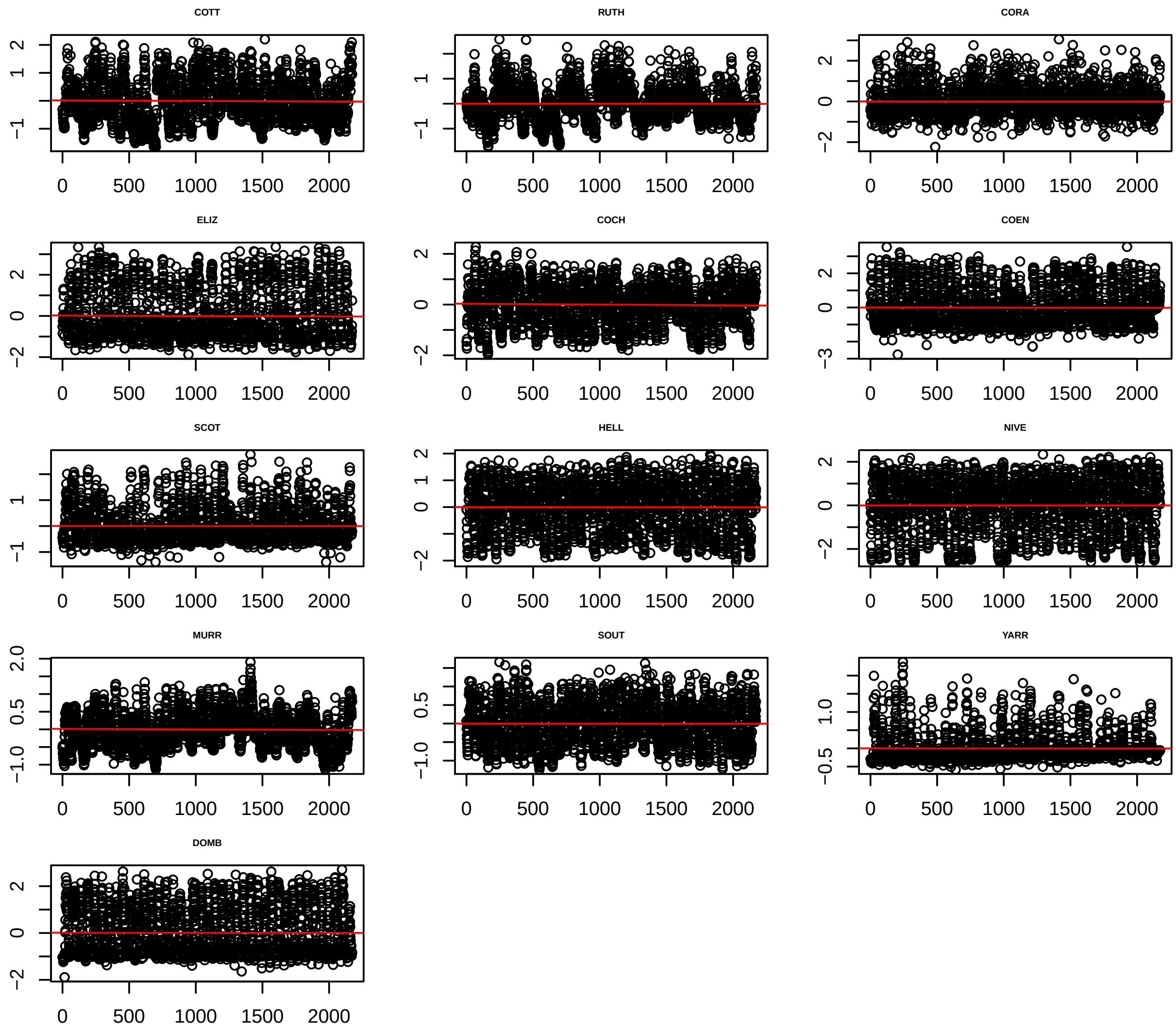

Figure 7: Residuals of GAMM analysis for trend in flow data removing gridded rainfall and evapotranspiration effects

```
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")
}
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. The bootstrap Mann Kendall is retained for completeness

Station rainfall data

```
# run the gamm model on rain, maxT and flow
cl <- makeCluster(6) # create a cluster with 6 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$Station == Stations[i,1])
  #gamm.data$Flow[gamm.data$Flow==0] <- NA
  gam_Flow_withRandE <- gamm(log(Flow+1)~s(Rain, k=3) + s(Rain, MaxT, k=3) ,
                              correlation= corCAR1(),
                         data=na.omit(gamm.data))#,
                              #control=list(niterEM=5))
  out <- list(model = gam_Flow_withRandE,
        results = data.frame(Station=Stations[i,1],
                         AIC=summary(gam_Flow_withRandE$lme)$AIC))
  out
}

stopCluster(cl)

# store results
save(Store_FwRE2,
     file=paste(storedir,
               "projectdata/StoreFwRE2_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	1329
RUTH	1989
CORA	3869
ELIZ	3112
COCH	2185
COEN	2914
SCOT	1552

Station	AIC
HELL	3065
NIVE	3418
MURR	-194.8
SOUT	1381
YARR	-505.6
DOMB	2119

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])) {
  gamm.data <- subset(flow_rain_maxT_weekly,
                      flow_rain_maxT_weekly$Station == Stations[i,1])
  #gamm.data$Flow[gamm.data$Flow==0] <- NA

  resid_list[[i]] <- zoo(residuals(Store_FwRE2[[i]]$model$lme,
                 type="normalized"),
                 order.by=as.Date(na.omit(gamm.data)$Date))
}
resid_df <- do.call(merge.zoo,resid_list)
names(resid_df) <- Stations[,1]

# write this out to run MK_LTP on HPC
save(resid_df,file="data/GAMMmodel5Resid.rdata")
# 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(6) # create a cluster with 6 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)
```

```
save(hp, file="../projectdata/Figure7ResidGAM_MDPaper.Rdata")
save(real_df, file="../projectdata/MKResidGAM_MDPaper.Rdata")
# # production quality tiff, this is Figure 7 in the manuscript
tiff("../manuscript/Figure7_ResidGAM_MDPaper.tif",res=600,compression="lzw",
     width=10*480,height=10*480)
print(hp)
dev.off()
```

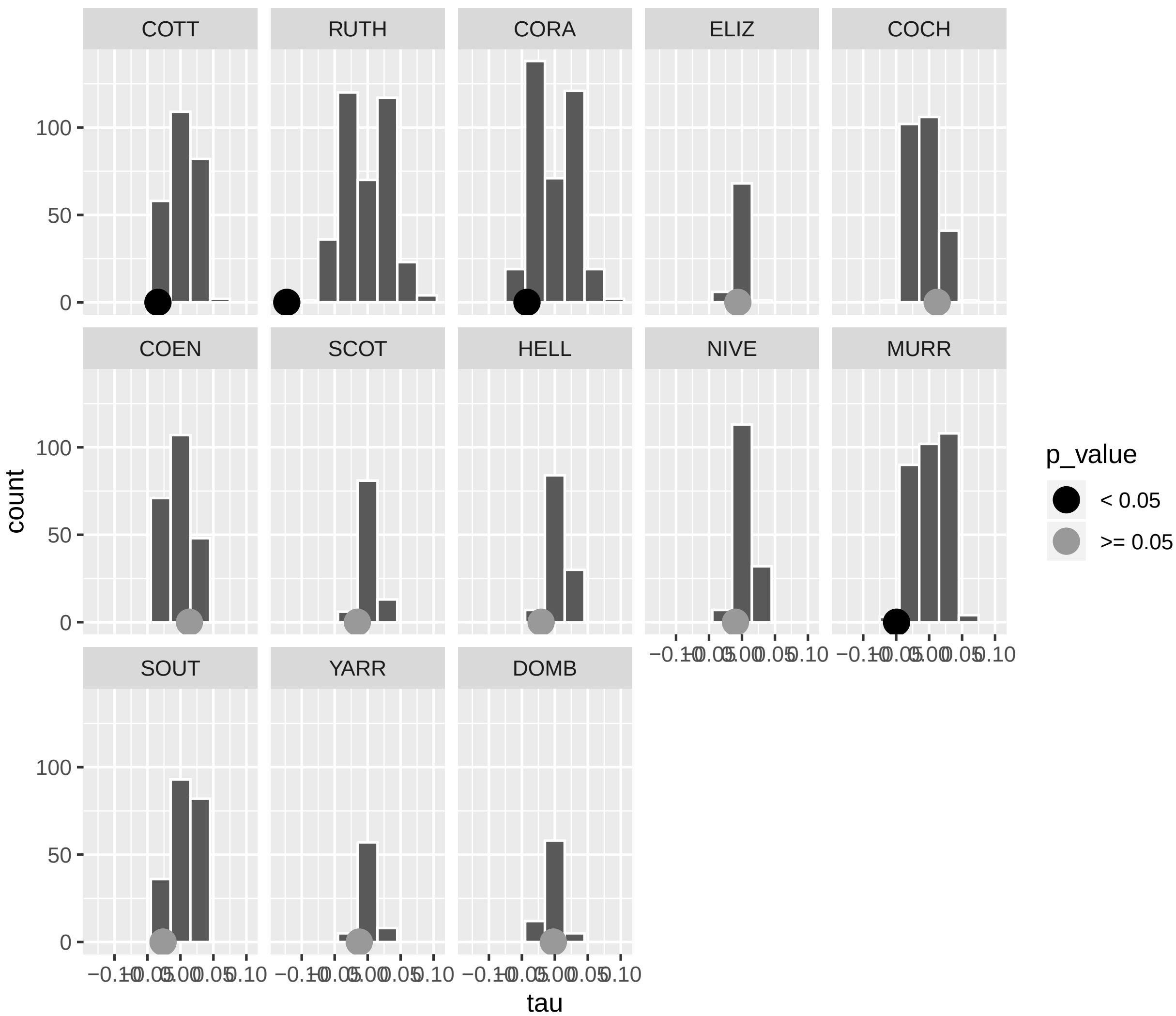

Figure 8: Mann Kendall analysis of the residuals of the streamflow after GAM model with rainfall and a Evapotranspiration

```
## 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
0.01704	-0.03416	COTT	real
1.072e-17	-0.1226	RUTH	real
0.003204	-0.0422	CORA	real
0.6624	-0.006251	ELIZ	real
0.3978	0.01211	COCH	real
0.3411	0.01363	COEN	real
0.2769	-0.01557	SCOT	real
0.1466	-0.02078	HELL	real
0.4908	-0.009866	NIVE	real
0.0005525	-0.04945	MURR	real
0.06524	-0.02639	SOUT	real
0.3738	-0.01273	YARR	real
0.8758	-0.002239	DOMB	real

```r
rm(Store_FwRE2)
```

Read in the MKLTP results on the GAM residuals for the station data

The Mann-Kendall analysis was moved to the HPC in January 2019 to be able to use the MKLTP analysis. Here the results of that analysis is read into the document. The script for the MKLTP analysis is called: "ResidualsMannkendallLTP.R" and this is in the VirtExp/Rcode/HPC folder on github. The bootstrap is retained for completeness

```r
load(paste(storedir,
                "projectdata/HPCresults/Store_Res_Mod5.Rdata", sep="/"))

MK_standard <- lapply(Store_resid, "[[","Mann_Kendall")
Hsignif <- lapply(Store_resid, "[[","Significance_of_H")
MK_LTP <- lapply(Store_resid, "[[","Mann_Kendall_LTP")

ResMod5_MKLTP <- cbind(do.call(rbind, MK_standard),do.call(rbind, Hsignif),
                    do.call(rbind, MK_LTP))
colnames(ResMod5_MKLTP)[c(6,8,10)] <- paste(c("MK","Hest","MKLTP"),
                                          "2_sided_pvalue",sep="_")

weeklyTable_Mod5 <- data.frame(Catchment = Stations[,1],
                  `tau MK` = ResMod5_MKLTP[,1],
                  `p-value MK` = ResMod5_MKLTP[,6],
                  `Hurst p-value` = ResMod5_MKLTP[,8],
                  `MK LTP p-value` = ResMod5_MKLTP[,10])
save(weeklyTable_Mod5, file="../projectdata/MKResidGAM_MDPaper.Rdata")
pander(weeklyTable_Mod5,
```

```
      caption="Mann-Kendall test (Hamed, 2008)
     results on the residuals of model 5 (Station data).
      p-values are considered significant at the 5% level.")
```

Table 12: Mann-Kendall test (Hamed, 2008) results on the residuals of model 5 (Station data). p-values are considered significant at the 5% level.

Catchment	tau.MK	p.value.MK	Hurst.p.value	MK.LTP.p.value
COTT	-0.03416	0.01704	6.066e-05	0.002952
RUTH	-0.1226	1.072e-17	0.04105	3.842e-21
CORA	-0.0422	0.003204	0.6936	0.003426
ELIZ	-0.00625	0.6624	0.1472	0.6997
COCH	0.01211	0.3978	5.344e-09	0.238
COEN	0.01363	0.3411	2.695e-05	0.23
SCOT	-0.01557	0.2769	4.648e-19	0.06725
HELL	-0.02078	0.1466	4.824e-06	0.06107
NIVE	-0.009866	0.4908	2.605e-10	0.6551
MURR	-0.04945	0.0005525	1.019e-13	1.06e-07
SOUT	-0.02639	0.06524	1.271e-06	0.01554
YARR	-0.01273	0.3738	4.097e-21	0.1228
DOMB	-0.002239	0.8758	0.0002911	0.8492

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, had to bump up to 5 to allow convergence.

```
# run the gamm model on rain, maxT and flow
cl <- makeCluster(6) # create a cluster with 6 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(flow_rain_maxT_weekly,
                      flow_rain_maxT_weekly$Station == Stations[i,1])
  #gamm.data$Flow[gamm.data$Flow==0] <- NA

  gam_Flow_withGRE <- gamm(log(Flow+1) ~
                             s(gridRain, k=3) +
                             s(gridRain,MaxT, k=3),
                             correlation= corCAR1(),
                             data=gamm.data)
  out <- list(model = gam_Flow_withGRE,
        results = data.frame(Station=Stations[i,1],
                            AIC=summary(gam_Flow_withGRE$lme)$AIC))
  out
}

stopCluster(cl)
```

```r
# 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 Gridd
```

Table 13: Mixed model results for analysis of trend in flow data taking into account Gridded Rainfall and Evapotranspiration

Station	AIC
COTT	1008
RUTH	1777
CORA	3812
ELIZ	2675
COCH	1695
COEN	2781
SCOT	1555
HELL	2824
NIVE	3297
MURR	-391.7
SOUT	1468
YARR	-511.7
DOMB	2130

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])) {
  gamm.data <- subset(flow_rain_maxT_weekly,
                      flow_rain_maxT_weekly$Station == Stations[i,1])
  #gamm.data$Flow[gamm.data$Flow==0] <- NA

  resid_list[[i]] <- zoo(residuals(Store_FwGRE2[[i]]$model$lme,
                 type="normalized"),
                 order.by=as.Date(na.omit(gamm.data)$Date))
}
resid_df <- do.call(merge.zoo,resid_list)
names(resid_df) <- Stations[,1]

# write this out to run MK_LTP on HPC
save(resid_df,file="data/GAMMmodel5Resid_grid.rdata")


# Bootstrap
# now run a loop over the number of years (create 41 different sets)
# do Mann Kendall test on each resonstituted series
# ---------------------------
#  -------------------------
```

```r
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(6) # create a cluster with 6 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
```

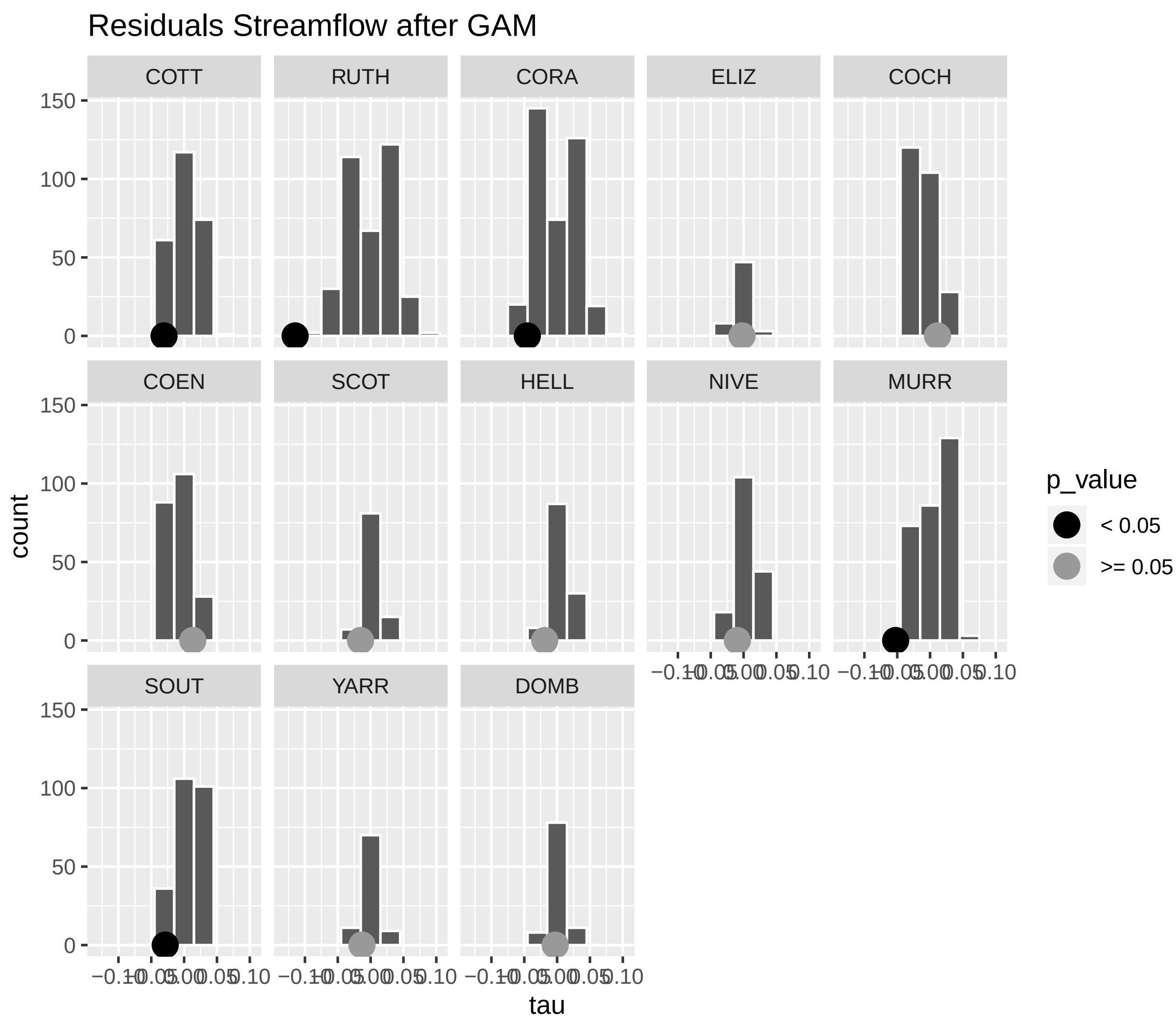

Figure 9: Mann Kendall analysis of the residuals of the streamflow after GAM model with gridded rainfall and a Evapotranspiration

```
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)
```

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

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

Table 14: Mann Kendall results for the GAMM residuals with gridded rainfall

pvalue	tau	catch	type
0.03265	-0.03058	COTT	real
9.926e-16	-0.1149	RUTH	real
0.001545	-0.04533	CORA	real
0.8694	-0.002354	ELIZ	real
0.428	0.01135	COCH	real
0.3568	0.01319	COEN	real
0.2824	-0.01539	SCOT	real
0.1878	-0.01886	HELL	real
0.5077	-0.009483	NIVE	real
0.0002532	-0.05238	MURR	real
0.04419	-0.02881	SOUT	real
0.3578	-0.01317	YARR	real
0.8399	-0.002892	DOMB	real

```r
rm(Store_FwGRE2)
```

Read in the MKLTP results on the GAM residuals for the gridded data

The Mann-Kendall analysis was moved to the HPC in January 2019 to be able to use the MKLTP analysis. Here the results of that analysis is read into the document. The script for the MKLTP analysis is called: "ResidualsMannkendallLTP.R" and this is in the VirtExp/Rcode/HPC folder on github.

```r
load(paste(storedir,
                "projectdata/HPCresults/Store_Res_Mod5_grid.Rdata", sep="/"))

MK_standard <- lapply(Store_resid_grid, "[[","Mann_Kendall")
Hsignif <- lapply(Store_resid_grid, "[[","Significance_of_H")
MK_LTP <- lapply(Store_resid_grid, "[[","Mann_Kendall_LTP")

ResMod5_grid_MKLTP <- cbind(do.call(rbind, MK_standard),do.call(rbind, Hsignif),
                    do.call(rbind, MK_LTP))
colnames(ResMod5_grid_MKLTP)[c(6,8,10)] <- paste(c("MK","Hest","MKLTP"),
                                          "2_sided_pvalue",sep="_")

weeklyTable_Mod5_grid <- data.frame(Catchment = Stations[,1],
                  `tau_MK` = ResMod5_grid_MKLTP[,1],
                  `p-value MK` = ResMod5_grid_MKLTP[,6],
                  `Hurst p-value` = ResMod5_grid_MKLTP[,8],
                  `MK LTP p-value` = ResMod5_grid_MKLTP[,10])

save(weeklyTable_Mod5_grid, file="../projectdata/GrMKResidGAM_MDPaper.Rdata")
pander(weeklyTable_Mod5_grid,
```

```
 caption="Mann-Kendall test (Hamed, 2008)
results on the residuals of model 5 (Gridded data).
 p-values are considered significant at the 5% level.")
```

Table 15: Mann-Kendall test (Hamed, 2008) results on the residuals of model 5 (Gridded data). p-values are considered significant at the 5% level.

Catchment	tau_MK	p.value.MK	Hurst.p.value	MK.LTP.p.value
COTT	-0.03058	0.03265	0.00869	0.01465
RUTH	-0.1149	9.926e-16	5.389e-05	1.241e-23
CORA	-0.04533	0.001545	0.7481	0.001752
ELIZ	-0.002354	0.8694	0.08395	0.8865
COCH	0.01135	0.428	3.596e-08	0.2784
COEN	0.01319	0.3568	0.005884	0.2884
SCOT	-0.01539	0.2824	4.325e-19	0.07028
HELL	-0.01886	0.1878	5.535e-08	0.07301
NIVE	-0.009483	0.5077	1.128e-08	0.6553
MURR	-0.05238	0.0002532	9.954e-14	1.772e-08
SOUT	-0.02881	0.04419	2.581e-06	0.008855
YARR	-0.01317	0.3578	1.711e-19	0.1192
DOMB	-0.002892	0.8399	1.505e-05	0.7974