GAM analysis of the weekly data

Willem Vervoort, Michaela Dolk & Floris van Ogtrop

2017-09-20

```
# 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(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. 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 only exception is model 6, which is dealt with in document 3B.

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
Log(Q) ~ trend + error	linear
Log(P) ~ trend + error	linear
Log(Q) ~ s(P) + trend + error	linear
Log (Q + 1) ~ s(P) + s(maxT	P) + trend + error
Log(Q + 1) ~ s(P) + s(maxT	P) + error
binomial log(P) ~ trend + error	40-year trend in weekly and daily rainfall occurence

Trend	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"	while the comparison with model 2 indicates the "amplification".
linear	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.
Mann-Kendall	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% {
# for (i in seq_along(Stations[,1])) {
#   i <- 1
  gamm.data <- subset(flow_rain_maxT_weekly,
```

```r
                      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= corCAR1(),
       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)

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	-0.0004162	0.1042	1981
RUTH	-0.0008201	1.807e-14	3874
CORA	-0.0003858	0.008591	6468
ELIZ	0.0001588	0.6635	3612
COCH	-0.000226	0.179	2352
COEN	0.0003334	0.3962	6630
SCOT	-0.0003297	0.2604	5093
HELL	-0.0001752	0.2486	2785
NIVE	-8.348e-05	0.7491	3697
MURR	-0.0002492	0.001871	827.6
SOUT	-0.0001727	0.007682	1943
YARR	-0.0003644	0.3426	4414
DOMB	0.1383	1.086e-28	4341

```r
rm(Store2)
```

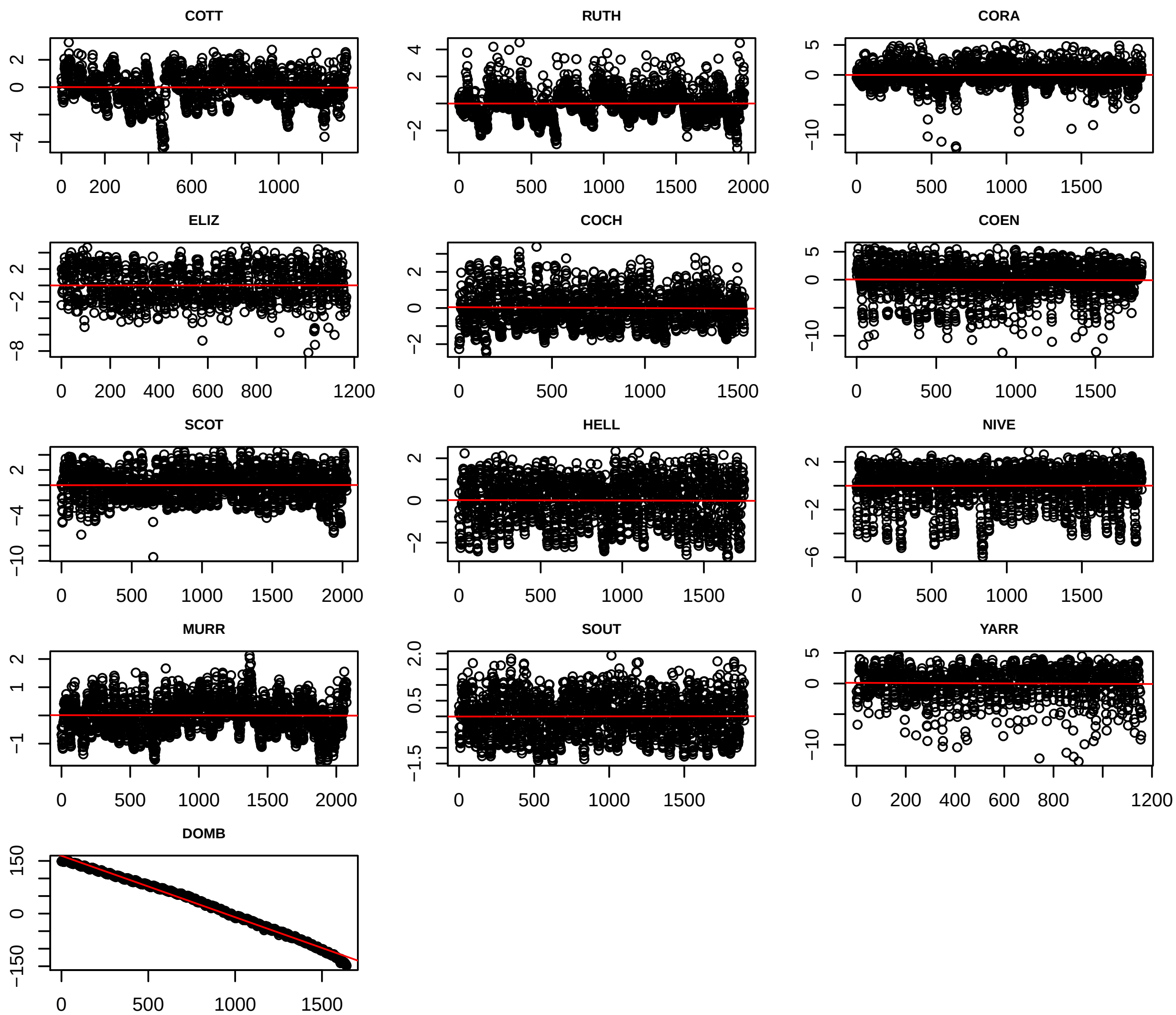

Figure 1: Residuals of linear mixed model analysis for trend in flow only

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$Station == Stations[i,1])
  gamm.data$Rain[gamm.data$Rain==0] <- NA
  gamm.data$trend <- 1:nrow(gamm.data)
  gam_TrendR <- gls(log(Rain)~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	-0.0001582	0.08666	3898
RUTH	-0.0001829	0.01233	6019
CORA	-0.0001492	0.03022	6153
ELIZ	-0.0002914	0.0248	3780

Station	Value	p.value	AIC
COCH	-0.0002699	0.02808	4703
COEN	-0.0005117	0.0007076	3969
SCOT	-5.494e-05	0.4506	6437
HELL	-2.483e-05	0.64	5218
NIVE	-0.0001866	0.001392	5378
MURR	-0.0001167	0.03638	6238
SOUT	-0.0001749	0.0001713	5869
YARR	-0.0001297	0.1612	6512
DOMB	-5.514e-05	0.5136	4534

```
rm(Store_Rain)
```

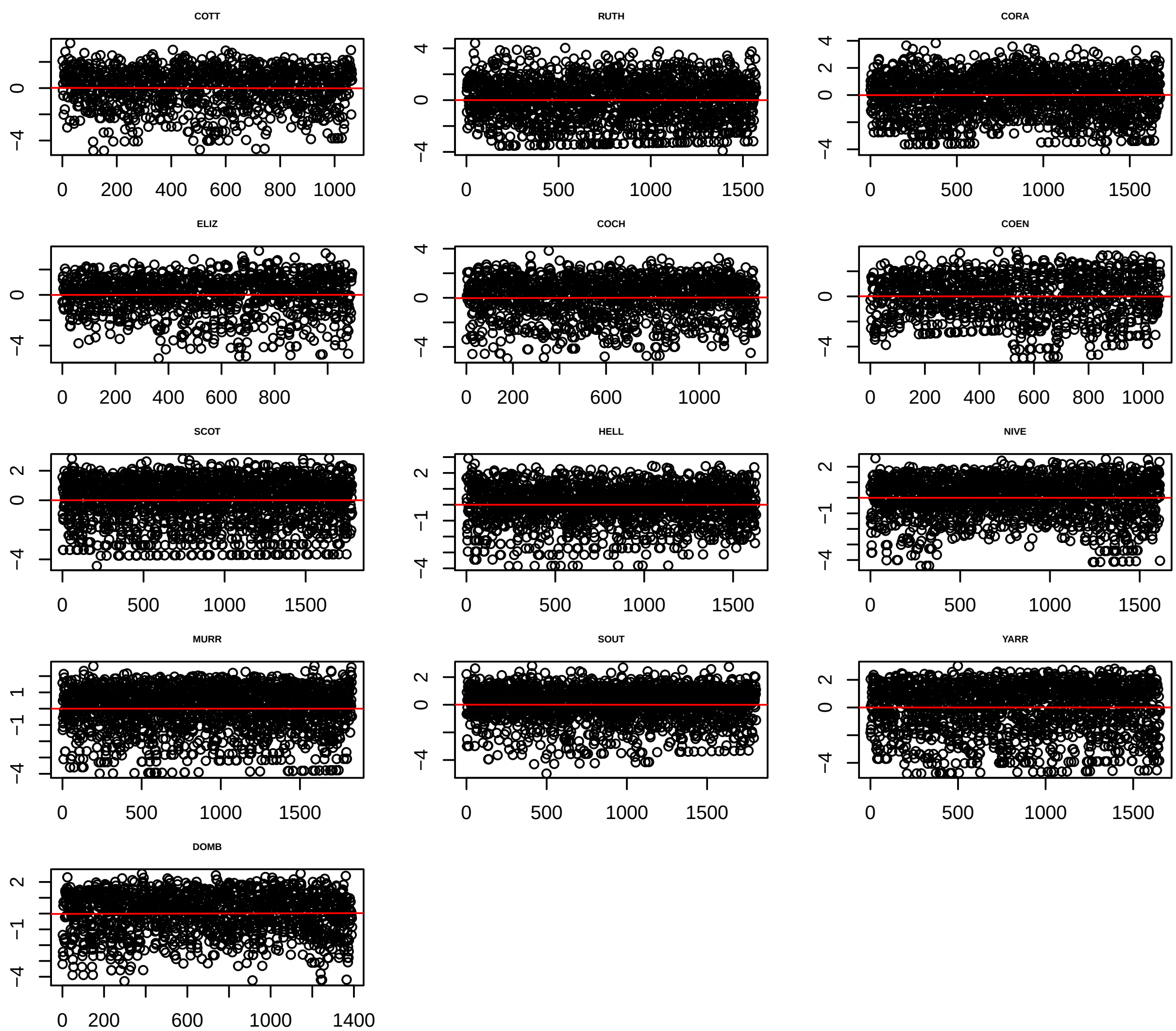

Figure 2: Residuals of linear mixed model analysis for trend in station rainfall data

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$Station == Stations[i,1])
  gamm.data$gridRain[gamm.data$gridRain==0] <- NA
  gamm.data$trend <- 1:nrow(gamm.data)
  gam_TrendGridR <- gls(log(gridRain)~trend, correlation= corCAR1(),
       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)


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	-0.0002285	0.07377	5525
RUTH	3.096e-05	0.7746	8862
CORA	-0.0001751	0.1139	9048
ELIZ	-0.001603	0.05815	11388
COCH	-0.0001978	0.4946	8921

Station	Value	p.value	AIC
COEN	-0.0001801	0.7148	11995
SCOT	0.0002808	0.1453	11359
HELL	-3.765e-05	0.5685	6004
NIVE	3.852e-05	0.5057	6478
MURR	-8.129e-05	0.4159	9662
SOUT	-2.797e-05	0.6771	7270
YARR	-4.22e-05	0.8494	11113
DOMB	-0.0003202	0.05589	8449

```
rm(Store_GridRain)
```

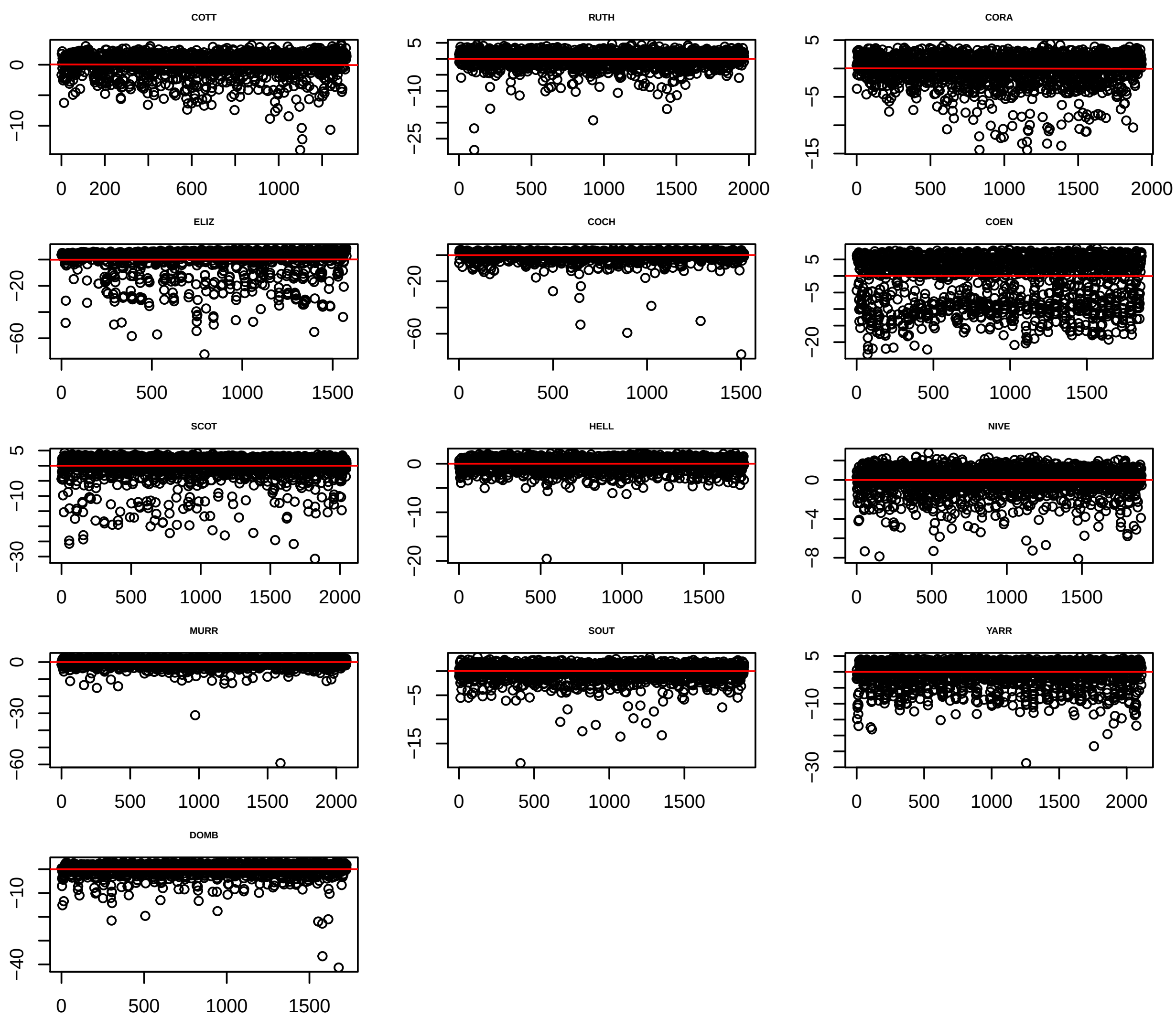

Figure 3: Residuals of linear mixed model analysis for trend in station rainfall data

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

```r
# 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$Station == Stations[i,1])
  gamm.data$Flow[gamm.data$Flow==0] <- NA
  gamm.data$trend <- 1:nrow(gamm.data)
  gam_TrendFlow_withR <- gamm(log(Flow)~s(Rain) + 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)

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	-0.0003577	0.1736	1528
RUTH	-0.0007711	4.075e-10	2368
CORA	-0.0003373	0.01406	5783
ELIZ	0.0001586	0.6017	3291
COCH	-0.0001283	0.4022	1580
COEN	0.0003541	0.3042	6236
SCOT	-0.0003011	0.2931	4249
HELL	-0.0001535	0.2896	2568
NIVE	-8.229e-05	0.7461	3683
MURR	-0.0002324	0.003982	94.89
SOUT	-0.0001372	0.03013	1255
YARR	-0.0004208	0.2431	4273
DOMB	-0.0005158	0.1562	3964

```r
rm(Store_FwR)
```

Gridded rainfall data

```r
#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$Station == Stations[i,1])
  gamm.data$Flow[gamm.data$Flow==0] <- NA
  gamm.data$trend <- 1:nrow(gamm.data)
  gam_TrendFlow_withGR <- gamm(log(Flow)~s(gridRain) + 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)

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")
}
```

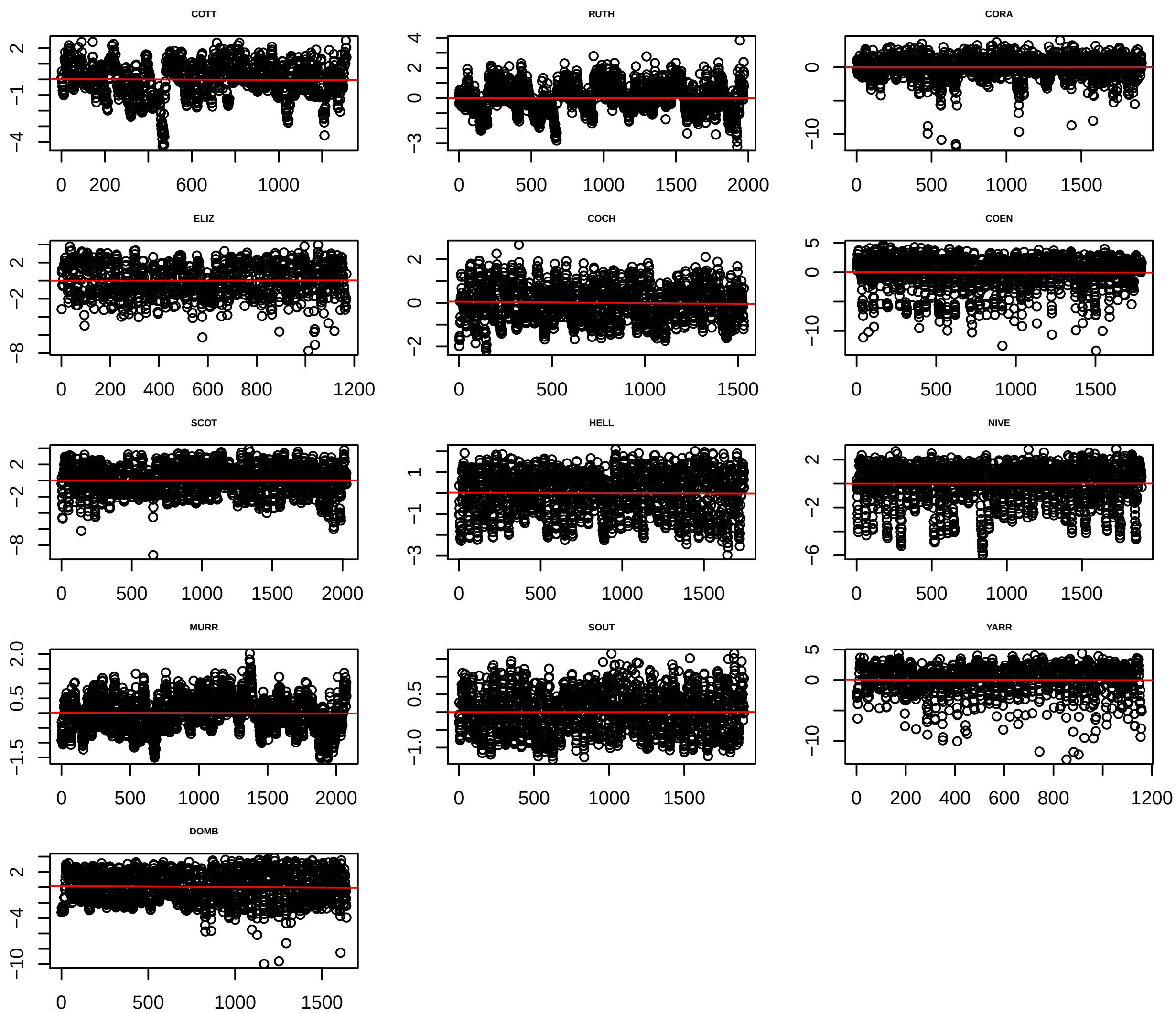

Figure 4: Residuals of GAMM analysis for trend in flow data taking into consideration station rainfall data

```
# 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 Gride
```

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

Station	Value	p.value	AIC
COTT	-0.000415	0.09513	1965
RUTH	-0.0008205	1.039e-14	3857
CORA	-0.0003858	0.008276	6453
ELIZ	0.0001613	0.6202	3583
COCH	-0.0002304	0.1558	2323
COEN	0.0003283	0.3817	6603
SCOT	-0.0003294	0.2482	5078
HELL	-0.0001752	0.2361	2768
NIVE	-8.332e-05	0.744	3685
MURR	-0.0002489	0.001578	807.1
SOUT	-0.000172	0.007211	1923
YARR	-0.0003994	0.2739	4390
DOMB	-0.000553	0.1545	4387

```
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,
```

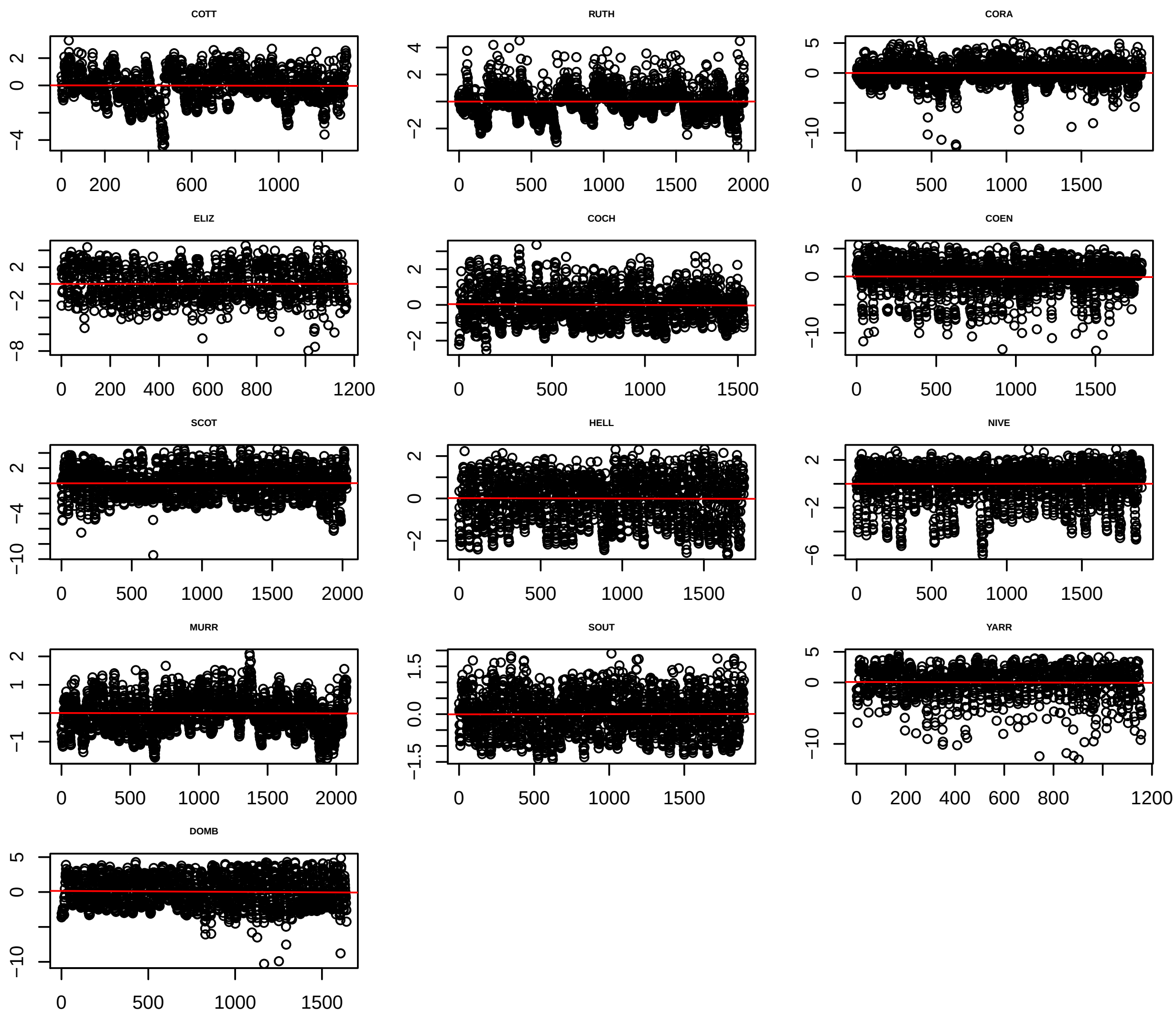

Figure 5: Residuals of GAMM analysis for trend in flow data taking into consideration gridded rainfall data

```
                      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)~s(Rain) + s(Rain, MaxT) +
                                       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)

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	-0.000308	0.2049	1479
RUTH	-0.0007593	1.634e-10	2330
CORA	-0.0003322	0.01196	5756
ELIZ	0.0001874	0.5427	3270
COCH	-0.0001309	0.3519	1516
COEN	0.000405	0.1687	6142
SCOT	-0.0002768	0.2074	4101
HELL	-0.0001414	0.1907	2395
NIVE	-6.918e-05	0.7618	3628
MURR	-0.0002198	0.002485	-122.2
SOUT	-0.0001262	0.01646	1183
YARR	-0.0006134	0.0445	4175
DOMB	-0.0005318	0.06405	3880

```r
rm(Store_FwRE)
```

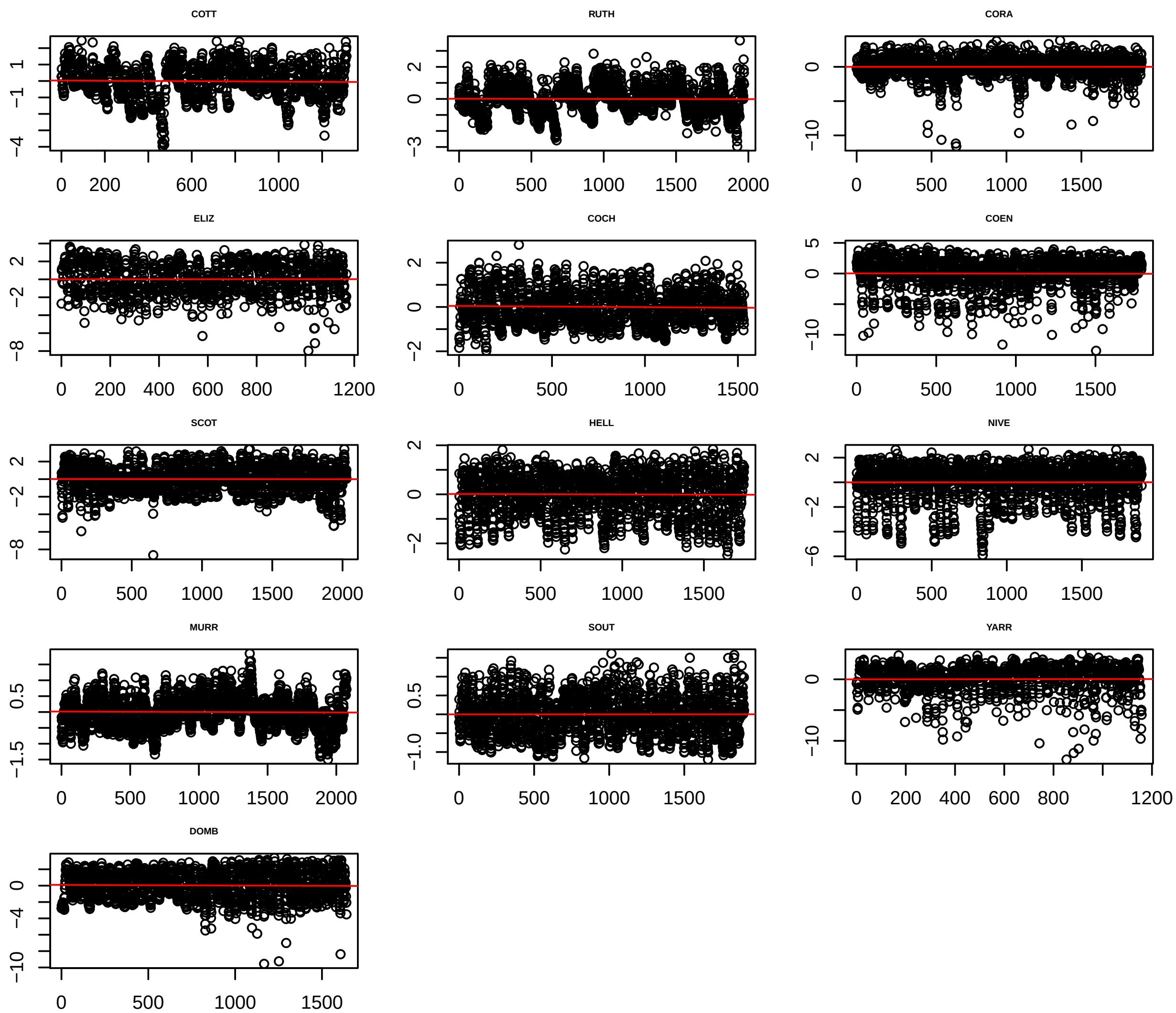

Figure 6: Residuals of GAMM analysis for trend in flow data removing station rainfall and evapotranspiration effects

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.

```r
# 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]),
```

```r
                  .packages="mgcv") %dopar% {
#  out <- list()
#for (i in 1:length(Stations[,1])) {
  gamm.data <- subset(weekGridRainAllDataout,
                      weekGridRainAllDataout$Station == Stations[i,1])
  gamm.data$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) ~
                                s(gridRain,k=10) +
                                s(gridRain,MaxT, k=10) +
                                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)

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 Gridd
```

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	-0.0003299	0.1358	1845
RUTH	-0.0007896	3.982e-14	3772
CORA	-0.0003629	0.00938	6379
ELIZ	0.000232	0.484	3454
COCH	-0.000208	0.1566	2144
COEN	0.0004239	0.148	6317

Station	Value	p.value	AIC
SCOT	-0.0002777	0.1159	4765
HELL	-0.000152	0.148	2518
NIVE	-7.085e-05	0.7572	3641
MURR	-0.0002313	0.0006513	580
SOUT	-0.000144	0.003405	1710
YARR	-0.000638	0.03162	4219
DOMB	-0.0005579	0.03056	4184

```
rm(Store_FwGRE)
```

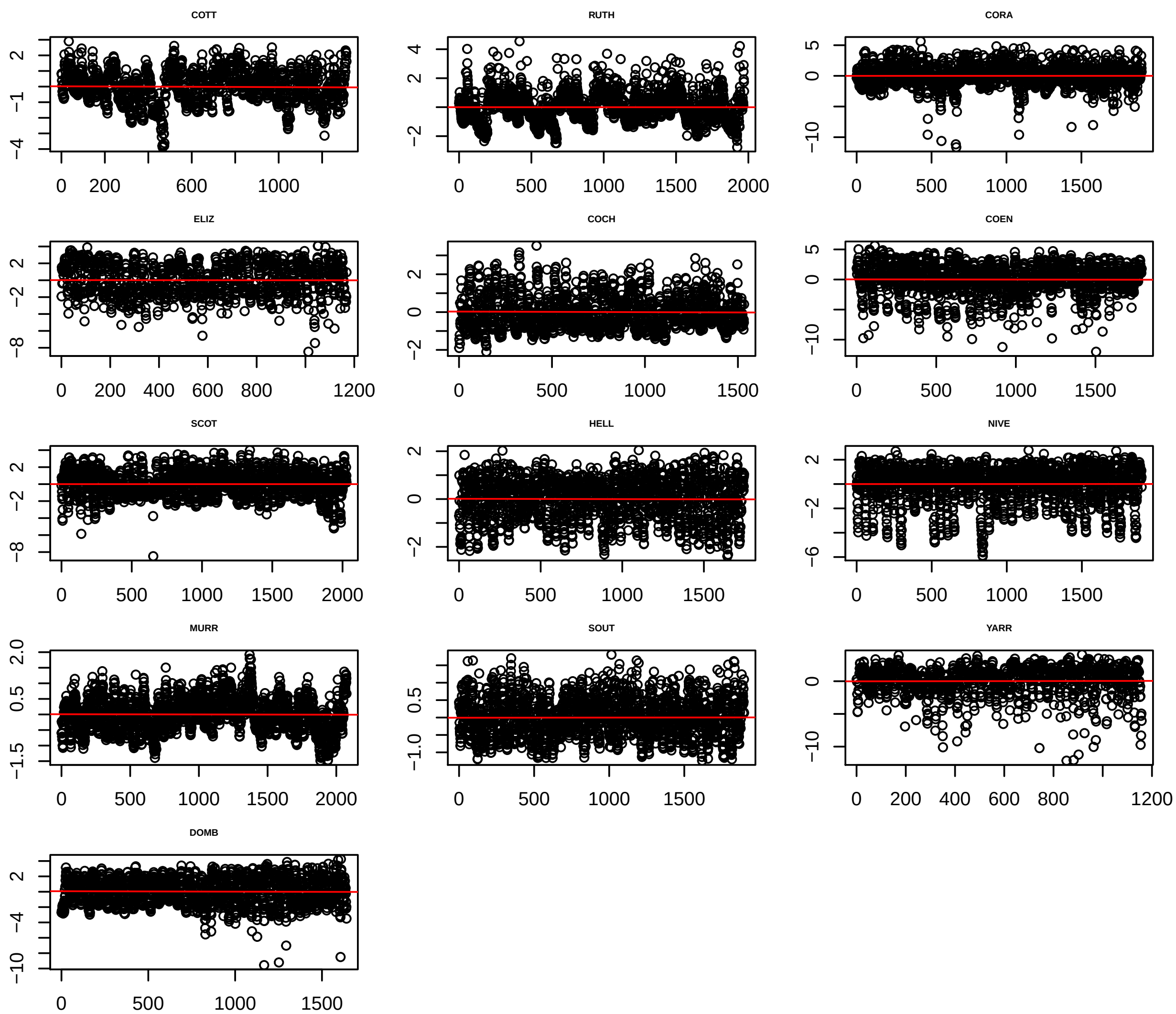

Figure 7: Residuals of GAMM analysis for trend in flow data removing gridded rainfall and evapotranspiration effects

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

```
# 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$Station == Stations[i,1])
  gamm.data$Flow[gamm.data$Flow==0] <- NA
  gam_Flow_withRandE <- gamm(log(Flow)~s(Rain) + s(Rain, MaxT) ,
                                  correlation= corCAR1(),
                             data=na.omit(gamm.data),
                                  control=list(niterEM=5))
  out <- list(model = gam_Flow_withRandE,
        results = data.frame(Station=Stations[i,1],
                            AIC=summary(gam_Flow_withRandE$lme)$AIC))
  out
}

stopCluster(cl)

# store results
save(Store_FwRE2,
     file=paste(storedir,
                "projectdata/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	1479
RUTH	2359
CORA	5761
ELIZ	3268
COCH	1515
COEN	6142
SCOT	4100
HELL	2395

Station	AIC
NIVE	3626
MURR	-115.8
SOUT	1187
YARR	4177
DOMB	3881

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)$Date))
}
resid_df <- do.call(merge.zoo,resid_list)
names(resid_df) <- Stations[,1]
# Bootstrap
set.seed(10)
# now run a loop over the number of years (create 41 different sets)
# do Mann Kendall test on each resonstituted series
# ---------------------------
#  -------------------------
resid_temp <- as.data.frame(resid_df)
resid_temp$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
 }
```

```r
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)
```

```r
save(hp, file="../projectdata/Figure8ResidGAM_MDPaper.Rdata")
save(real_df, file="../projectdata/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
```

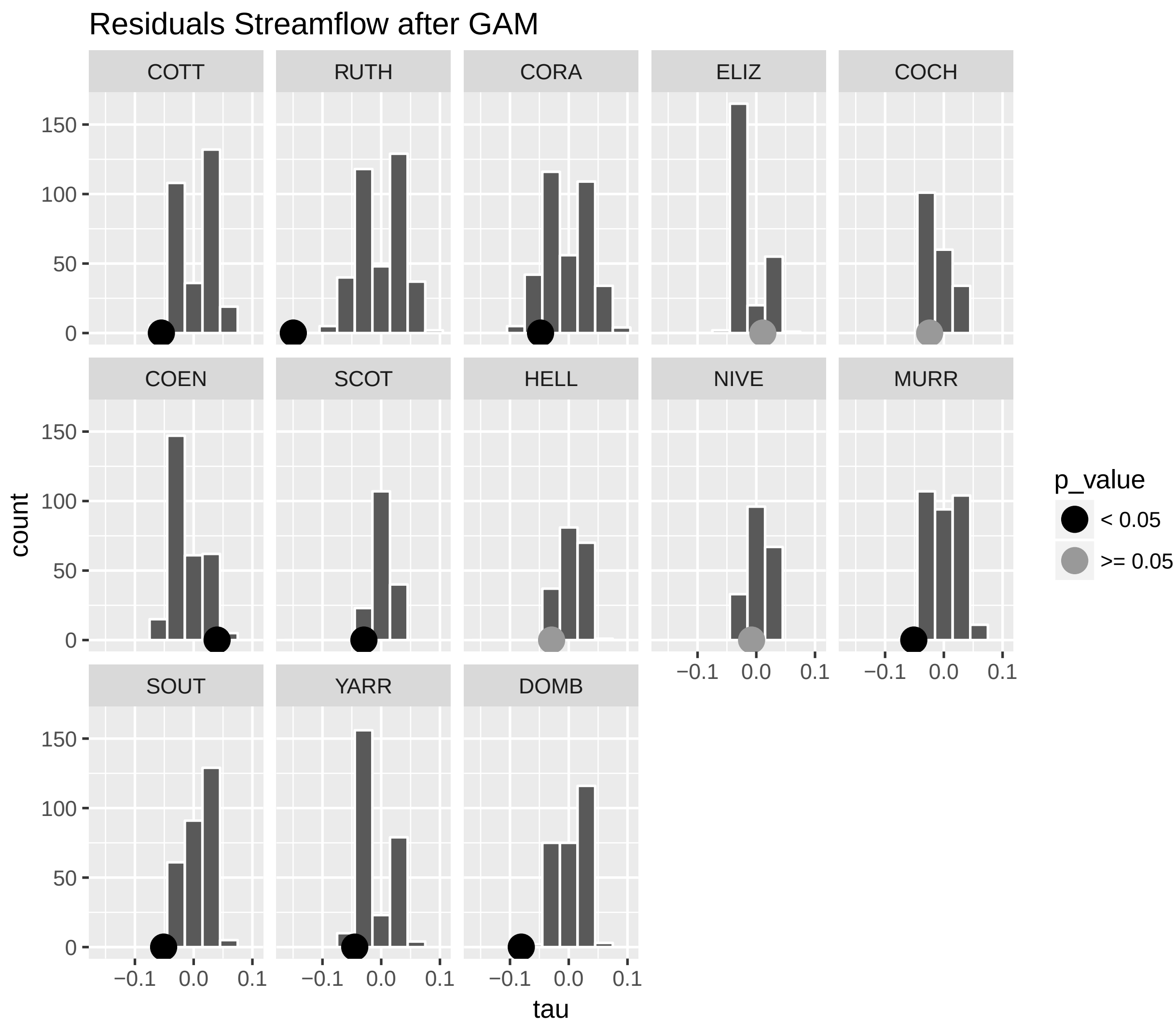

Figure 8: Mann Kendall analysis of the residuals of the streamflow after GAM model with rainfall and a Evapotranspiration

```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
0.002852	-0.05498	COTT	real
2.423e-23	-0.1496	RUTH	real
0.001699	-0.04802	CORA	real
0.5698	0.0111	ELIZ	real
0.1558	-0.0242	COCH	real
0.01142	0.0399	COEN	real
0.0469	-0.02944	SCOT	real
0.06769	-0.02918	HELL	real
0.5997	-0.008038	NIVE	real
0.0004808	-0.05115	MURR	real
0.0008428	-0.05113	SOUT	real
0.02162	-0.04507	YARR	real
9.382e-07	-0.0808	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$Station == Stations[i,1])
  gamm.data$Flow[gamm.data$Flow==0] <- NA

  gam_Flow_withGRE <- gamm(log(Flow) ~
                             s(gridRain, k=10) +
                             s(gridRain,MaxT, k=10),
                             correlation= corCAR1(),
                             data=na.omit(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
```

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

Station	AIC
COTT	1845
RUTH	3816
CORA	6384
ELIZ	3452
COCH	2144
COEN	6317
SCOT	4765
HELL	2518
NIVE	3639
MURR	588.8
SOUT	1717
YARR	4222
DOMB	4186

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(weekGridRainAllDataout,
                      weekGridRainAllDataout$Station == Stations[i,1])
  gamm.data$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)$Date))
}
resid_df <- do.call(merge.zoo,resid_list)
names(resid_df) <- Stations[,1]
# Bootstrap
# now run a loop over the number of years (create 41 different sets)
# do Mann Kendall test on each resonstituted series
# ---------------------------
#  -------------------------
resid_temp <- as.data.frame(resid_df)
resid_temp$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
```

```r
# 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")
```

```r
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)
```

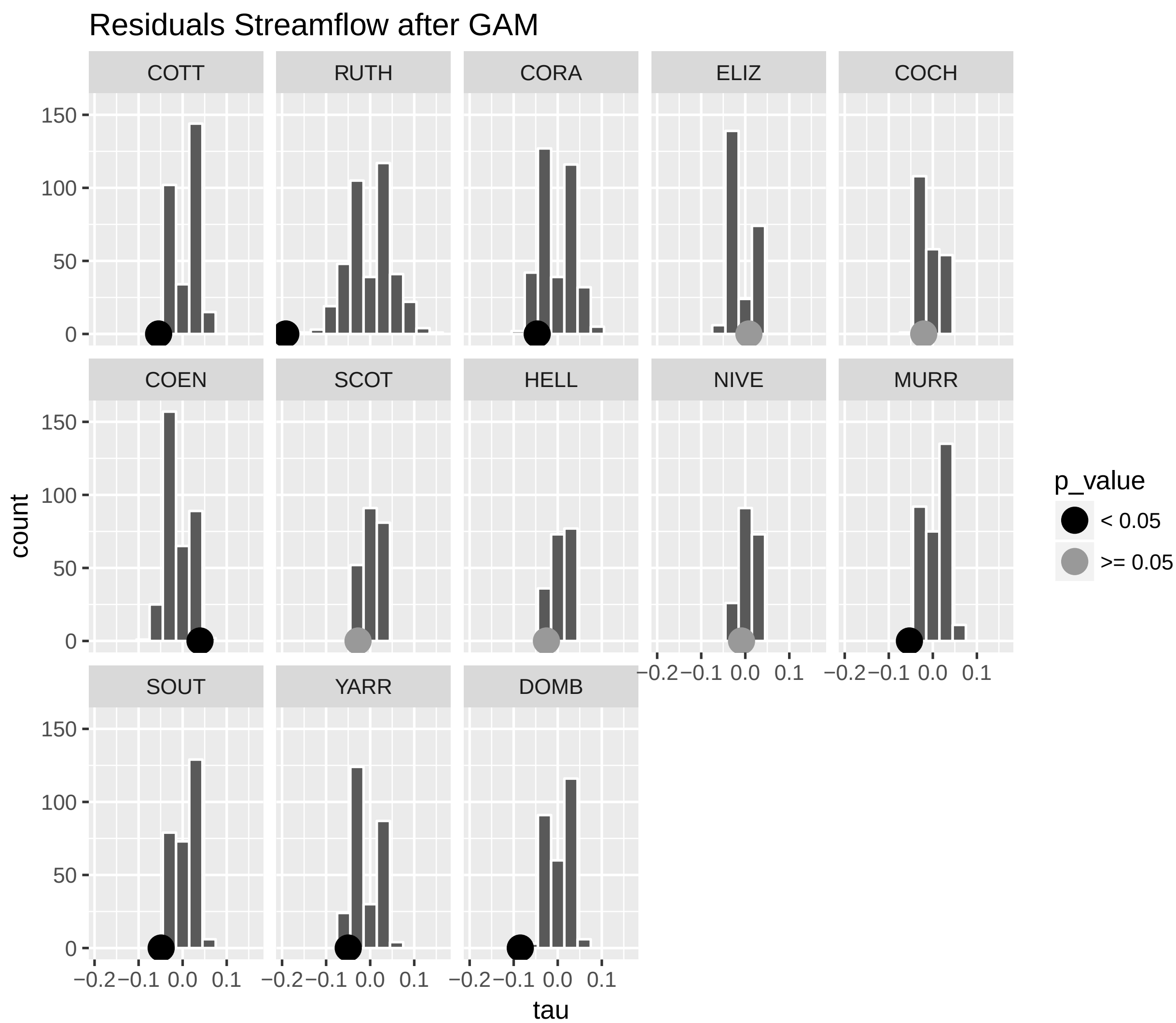

Figure 9: Mann Kendall analysis of the residuals of the streamflow after GAM model with gridded rainfall and a Evapotranspiration

```r
save(real_df, file="../projectdata/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
0.003079	-0.05455	COTT	real
0	-0.1915	RUTH	real

pvalue	tau	catch	type
0.002318	-0.04661	CORA	real
0.6736	0.008232	ELIZ	real
0.2262	-0.02063	COCH	real
0.0125	0.0394	COEN	real
0.06228	-0.02762	SCOT	real
0.1081	-0.02566	HELL	real
0.5771	-0.00854	NIVE	real
0.0002856	-0.05315	MURR	real
0.001451	-0.04877	SOUT	real
0.0111	-0.04983	YARR	real
2.469e-07	-0.08501	DOMB	real

```
rm(Store_FwGRE2)
```