MultiHazard: Excercises

Robert Jane

December 14, 2019

R Markdown

Impute missing values at Well G_580A using readings at Well G-860.

```
#Reading in the record at G_3356
G_580A<-read.table("C:\\Users\\ro327497\\Documents\\SFWMD\\SFWMD Data\\GH-580A_GWStageFormated.txt",hea
G_860<-read.csv("C:\\Users\\ro327497\\Documents\\SFWMD\\SFWMD Data\\G_860.csv")[-16508,c(3,4)]
#Converting Date column to "Date"" object
G_580A$Date<-seq(as.Date("1985-06-01"), as.Date("2019-02-03"), by="day")
G_860$Date<-seq(as.Date("1973-10-01"), as.Date("2018-12-10"), by="day")
#Converting column containing the readings to a "numeric"" object
G_580A$Value<-as.numeric(as.character(G_580A$Value))
```

```
## Warning: NAs introduced by coercion
```

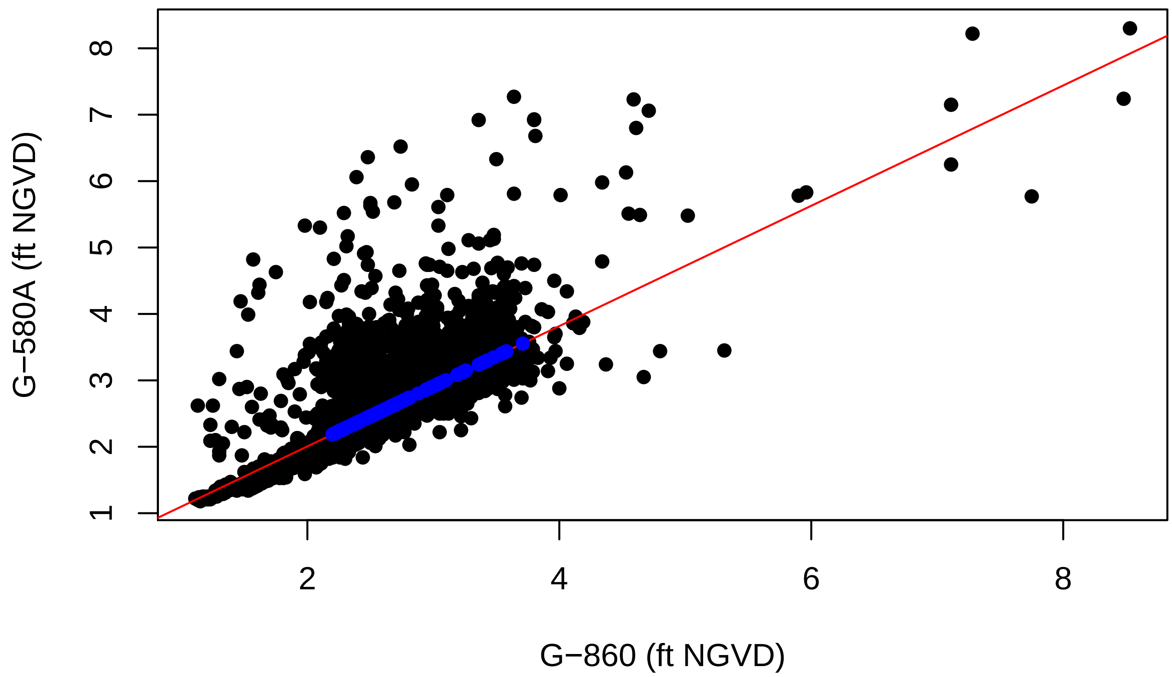
```
G_860$Value<-as.numeric(as.character(G_860$Value))
```

Warning message confirms there are NAs in the record at Well G_580A. Before carrying out the imputation the two data frames need to be merged.

```
#Merge the two data frames by date -  only keeping dates of the series at G_580A
GW_S22<-left_join(G_580A,G_860,by="Date")
colnames(GW_S22)<-c("Date","G580A","G860")
```

The imputation can now be carried out:

```
#Carrying out imputation
Imp<-Imputation(Data=GW_S22,Variable="G580A",x_lab="G-860 (ft NGVD)", y_lab="G-580A (ft NGVD)")
```



Show the resulting completed time series and summary of fitted regression model:

```r
head(Imp$Data)
```

```
##         Date G580A G860 ValuesFilled
## 1 1985-06-01  2.15 2.27         2.15
## 2 1985-06-02  2.12 2.25         2.12
## 3 1985-06-03  2.10 2.21         2.10
## 4 1985-06-04  2.07 2.18         2.07
## 5 1985-06-05  2.05 2.16         2.05
## 6 1985-06-06  2.02 2.12         2.02
```

```r
Imp$Model
```

```
##
## Call:
## lm(formula = data[, variable] ~ data[, Other.variable])
##
## Residuals:
##     Min       1Q  Median      3Q     Max
## -1.5536 -0.1135 -0.0554  0.0144  3.9195
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)            0.194313   0.016002   12.14   <2e-16 ***
## data[, Other.variable] 0.905710   0.005905  153.38   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

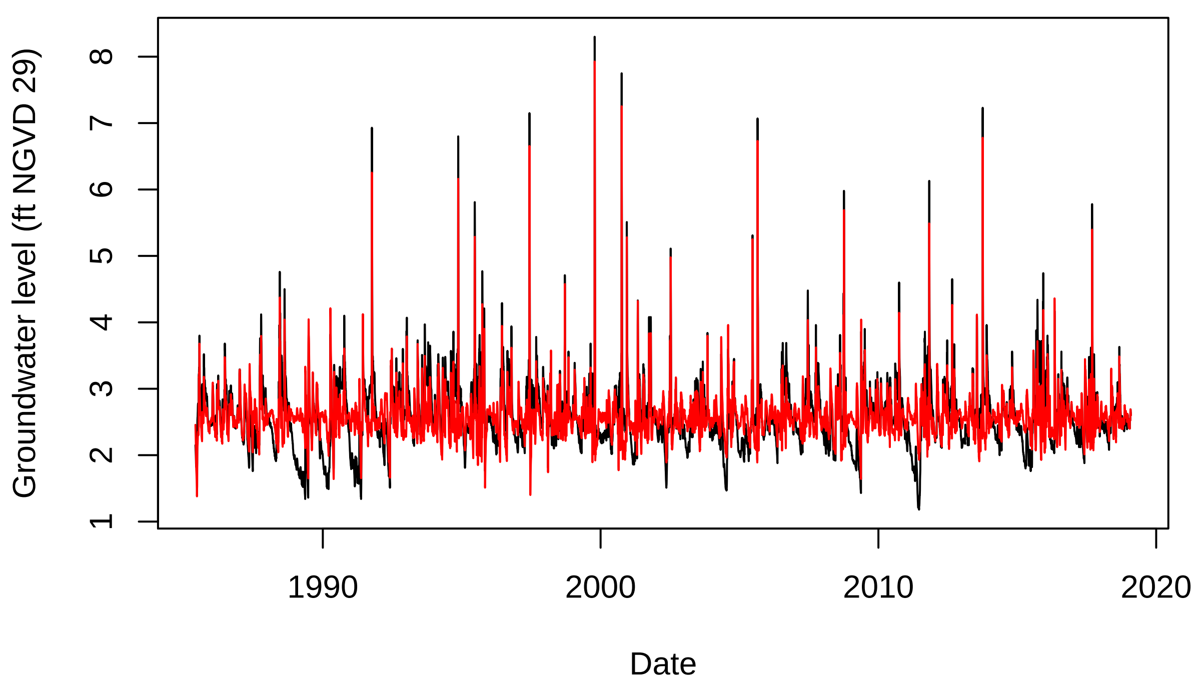
```
##
## Residual standard error: 0.2964 on 11878 degrees of freedom
##   (421 observations deleted due to missingness)
## Multiple R-squared:  0.6645, Adjusted R-squared:  0.6645
## F-statistic: 2.352e+04 on 1 and 11878 DF,  p-value: < 2.2e-16
```

Are any values still NA?

```r
G_580A_ValueFilled_NA<-which(is.na(Imp$Data$ValuesFilled)==TRUE)
length(G_580A_ValueFilled_NA)
```

```
## [1] 0
```

```r
#Creating a data frame with the imputed series alongside the corresponding dates
G_580A_Imp<-data.frame(Imp$Data$Date,Imp$Data$ValuesFilled)
colnames(G_580A_Imp)<-c("Date","ValuesFilled")
#Detrending
G_580A_Detrend<-Detrend_3Month(Data=G_580A_Imp,PLOT=TRUE,x_lab="Date",
                               y_lab="Groundwater level (ft NGVD 29)")
```



Print the first few values of the detrended series:

```r
head(G_580A_Detrend)
```

```
## [1] 2.456600 2.403339 2.358774 2.302469 2.255296 2.200513
```

Create a data frame containing the detrended groundwater series at site S_22 i.e. G_580A_Detrend and corresponding dates in column 1.

```
S22.Groundwater.Detrend.df<-data.frame(as.Date(GW_S22$Date),G_580A_Detrend)
colnames(S22.Groundwater.Detrend.df)<-c("Date","Groundwater")
```

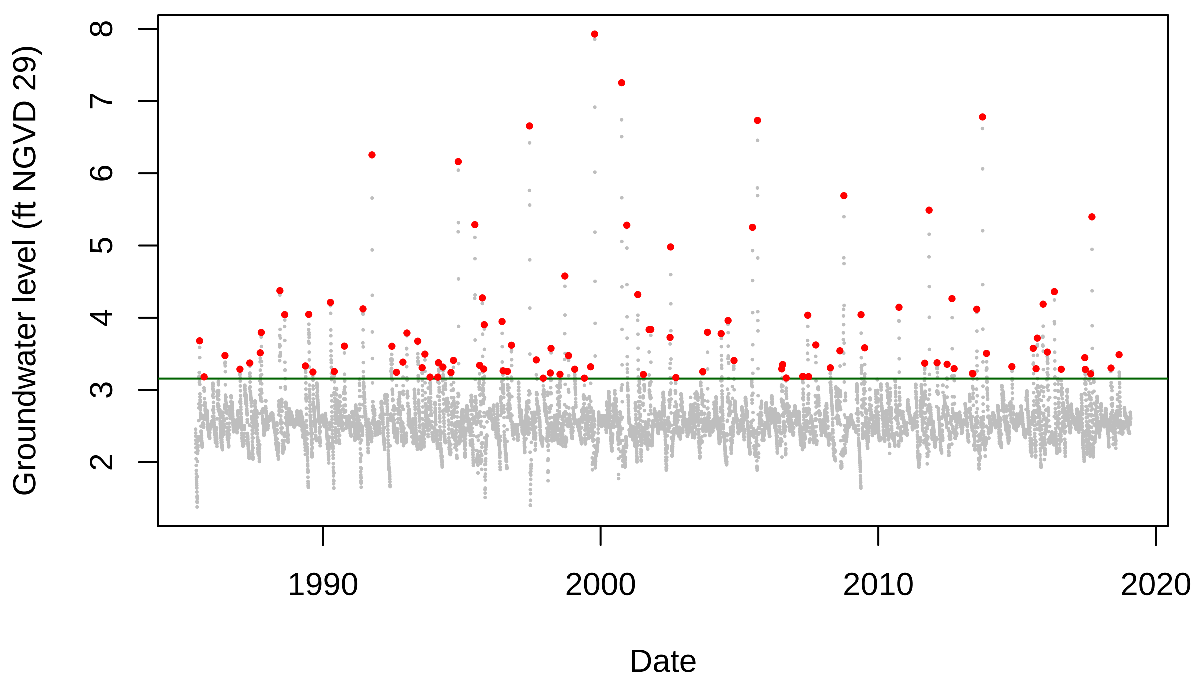
Declustering

Use the `Decluster()` to decluster the using a threshold of 0.95 and separation criterion of 3 days

```
G_580A.Declustered<-Decluster(Data=G_580A_Detrend,u=0.95,SepCrit=3,mu=365.25)
```

Plot showing the completed, detrended record at Well G-3356 (grey circles) along with cluster maxima (red circles) identified using a 95% threshold (green line) and three day separation criterion. Declustered series is obtained by applying the `$Declustered` suffix to an `Decluster` object.

```
G_580A_Imp$Detrend<-G_580A_Detrend
plot(as.Date(G_580A_Imp$Date),G_580A_Imp$Detrend,col="Grey",pch=16,
     cex=0.25,xlab="Date",ylab="Groundwater level (ft NGVD 29)")
abline(h=G_580A.Declustered$Threshold,col="Dark Green")
points(as.Date(G_580A_Imp$Date[G_580A.Declustered$EventsMax]),
       G_580A.Declustered$Declustered[G_580A.Declustered$EventsMax],col="Red",pch=16,cex=0.5)
```



Print the other outputs from the `Decluster()` function

```
G_580A.Declustered$Threshold
```

```
## [1] 3.157355
```

```
G_580A.Declustered$EventsPerYear
```

```
## [1] 3.147427
```

In preparation for later work, assign the detrended and declustered groundwater series at site S22 a name.

```
S22.Groundwater.Detrend.Declustered<-G_580A.Declustered$Declustered
```

Preparation for multivariate analysis

Reading in the other rainfall and O-sWL series at Site S22

```
S22.Rainfall.df<-read.csv("C:\\Users\\ro327497\\Documents\\SFWMD\\SFWMD Data\\S22\\Miami_Airport_df.csv
S22.Rainfall.df$Date<-as.Date(S22.Rainfall.df$Date)
S22.OsWL.df<-read.csv("C:\\Users\\ro327497\\Documents\\SFWMD\\SFWMD Data\\S22\\S22_T_MAX_Daily_Complete
S22.OsWL.df$Date<-as.Date(S22.OsWL.df$Date)
```

Detrend and O-sWL series

```
S22.OsWL.Detrend<-Detrend_3Month(Data=S22.OsWL.df,PLOT=FALSE,x_lab="Date",
                                 y_lab="O-sWL (ft NGVD 29)")
```

Create a dataframe with the date alongside the detrended OsWL series

```
S22.OsWL.Detrend.df<-data.frame(as.Date(S22.OsWL.df$Date),S22.OsWL.Detrend)
colnames(S22.OsWL.Detrend.df)<-c("Date","OsWL")
```

Declustering rainfall and O-sWL series at site S20,

```
#Setting missing values to zero
S22.Rainfall.Declustered<-Decluster(Data=S22.Rainfall.df$Value,u=0.95,SepCrit=3)$Declustered

## Warning in x.exceed.max.position[i] <- (x.exceed.lower.bound) +
## which(Data[(x.exceed.lower.bound + : number of items to replace is not a
## multiple of replacement length

## Warning in x.exceed.max.position[i] <- (x.exceed.lower.bound) +
## which(Data[(x.exceed.lower.bound + : number of items to replace is not a
## multiple of replacement length

#S20.O-sWL does not contain any missing values and so can be declustered without
S22.OsWL.Detrend.Declustered<-Decluster(Data=S22.OsWL.Detrend,u=0.95,SepCrit=3,mu=365.25)$Declustered
```

Creating data frames with the date alongside declustered series

```
S22.OsWL.Detrend.Declustered.df<-data.frame(S22.OsWL.df$Date,S22.OsWL.Detrend.Declustered)
colnames(S22.OsWL.Detrend.Declustered.df)<-c("Date","OsWL")
S22.Rainfall.Declustered.df<-data.frame(S22.Rainfall.df$Date,S22.Rainfall.Declustered)
colnames(S22.Rainfall.Declustered.df)<-c("Date","Rainfall")
S22.Groundwater.Detrend.Declustered.df<-data.frame(G_580A$Date,S22.Groundwater.Detrend.Declustered)
colnames(S22.Groundwater.Detrend.Declustered.df)<-c("Date","OsWL")
```

Use the Dataframe_Combine function to create data frames containing all observations of the original, detrended if necessary, and declustered time series. On dates where not all variables are observed, missing values are assigned NA.

```
S22.Detrend.df<-Dataframe_Combine(data.1<-S22.Rainfall.df,
                                  data.2<-S22.OsWL.Detrend.df,
                                  data.3<-S22.Groundwater.Detrend.df,
                                  names=c("Rainfall","OsWL","Groundwater"))
S22.Detrend.Declustered.df<-Dataframe_Combine(data.1<-S22.Rainfall.Declustered.df,
                              data.2<-S22.OsWL.Detrend.Declustered.df,
                              data.3<-S22.Groundwater.Detrend.Declustered.df,
                              names=c("Rainfall","OsWL","Groundwater"))
```

Trivariate analysis

The three higher dimensional (>3) approaches in the pakage to model the joint distribution of rainfall, O-sWL and groundwater level are:

Standard (trivariate) copula Pair Copula Construction *Heffernan and Tawn (2004)

Fit multiple independent GPD to declustered excesses

```
S22.Migpd<-Migpd_Fit(Data=S22.Detrend.Declustered.df[,-1],mqu=c(0.99,0.99,0.95))
```

Fit a Gaussian copula

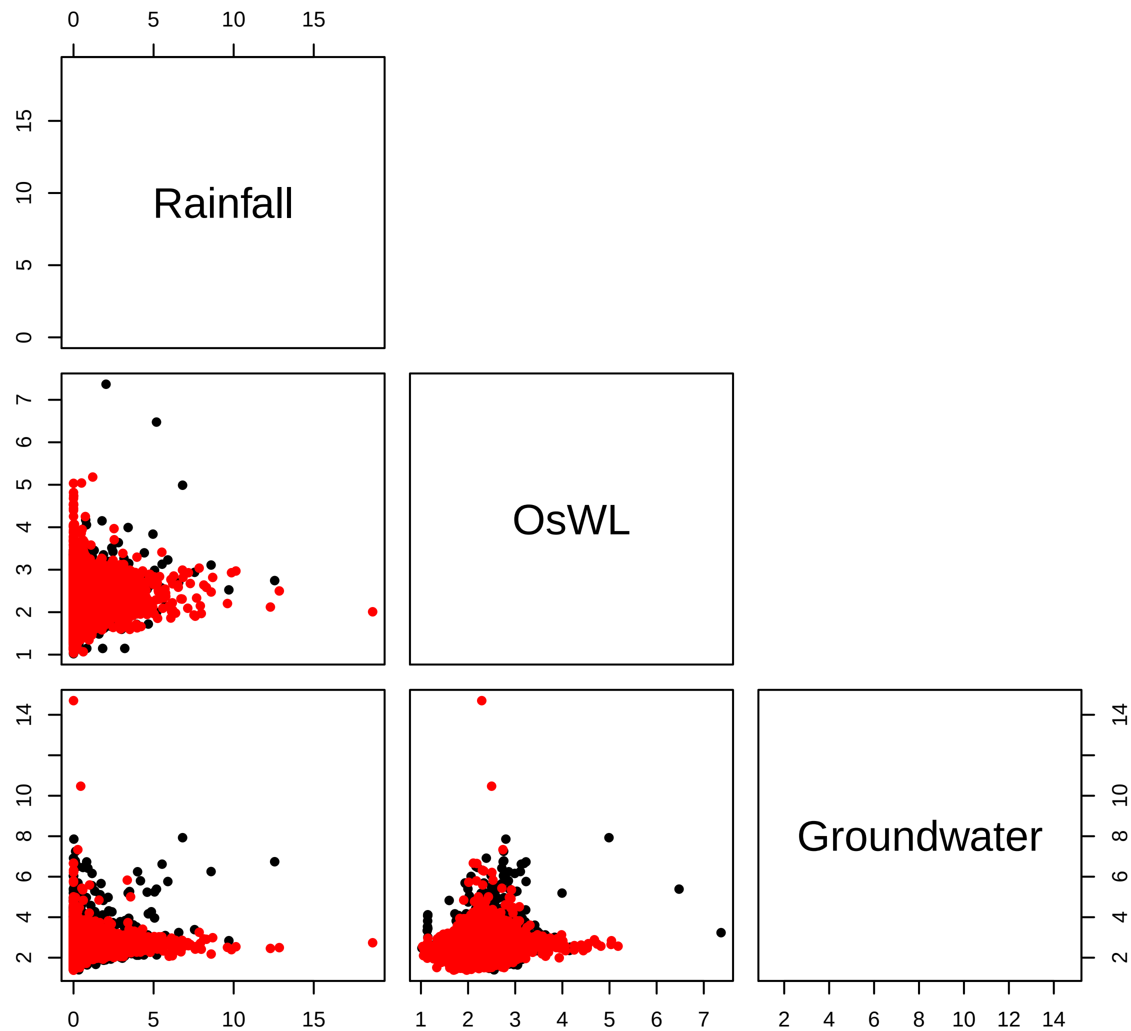
```
S22.Gaussian<-Standard_Copula_Fit(Data=S22.Detrend.df,Copula_Type="Gaussian")
```

Simulate a 100 years worth of events from the Gaussian copula

```
S22.Gaussian.Sim<-Standard_Copula_Sim(Data=S22.Detrend.df,Marginals=S22.Migpd,
                                      Copula=S22.Gaussian,N=100)
```

Plotting the observationed and simulated values

```
S22.Pairs.Plot.Data<-data.frame(rbind(na.omit(S22.Detrend.df[,-1]),S22.Gaussian.Sim$x.Sim),
                c(rep("Observation",nrow(na.omit(S22.Detrend.df))),rep("Simulation",nrow(S20.Gaussian.Si
colnames(S22.Pairs.Plot.Data)<-c(names(S22.Detrend.df)[-1],"Type")
pairs(S22.Pairs.Plot.Data[,1:3],col=ifelse(S22.Pairs.Plot.Data$Type=="Observation","Black","Red"),upper
```



Use the Standard_Copula_Sel command to deduce the best fitting standard trivariate copula:

```
Standard_Copula_Sel(Data=S22.Detrend.df)
```

```
## Warning in var.mpl(copula, u): the covariance matrix of the parameter
## estimates is computed as if 'df.fixed = TRUE' with df = 15.6721505967246
```

```
##     Copula        AIC
## 1 Gaussian -206.87828
## 2    t-cop -260.40591
## 3   Gumbel -240.43433
## 4  Clayton  -63.09137
## 5    Frank -142.32563
```

Fit a vine copula:

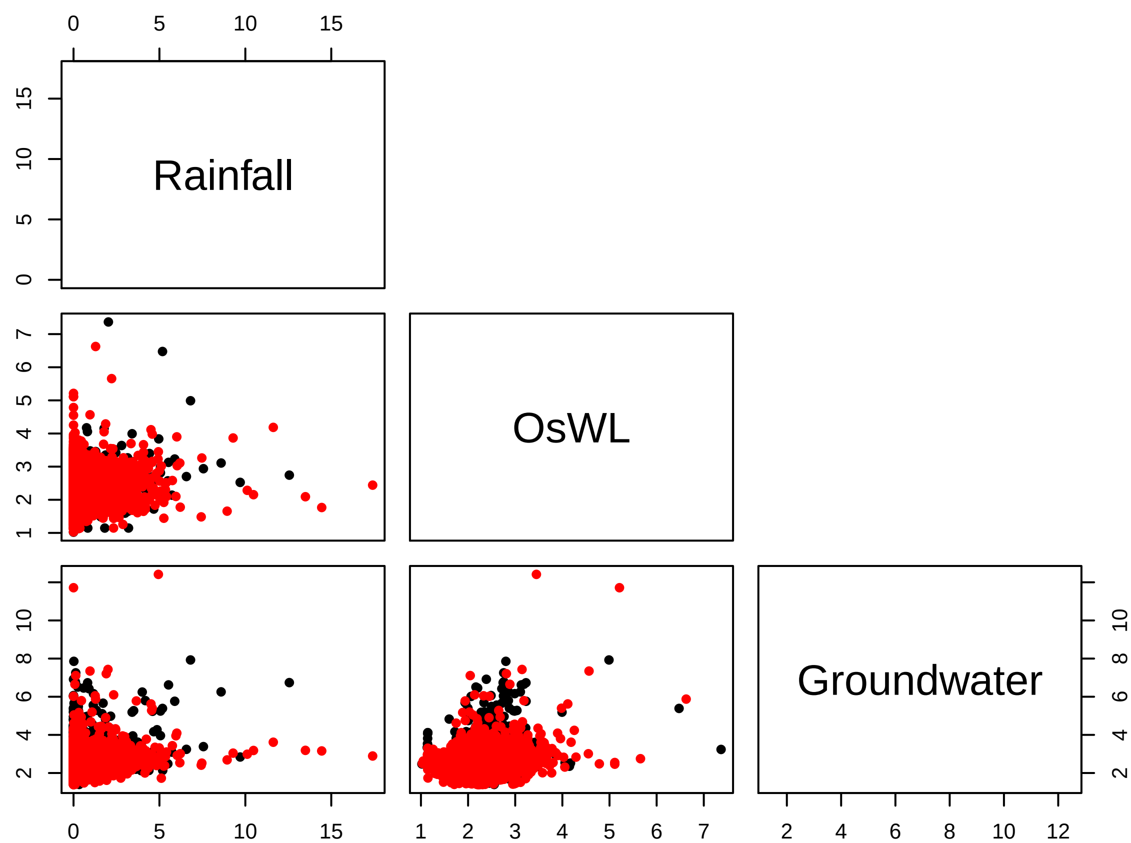
```
S22.Vine<-Vine_Copula_Fit(Data=S22.Detrend.df)
```

Simulate a 100 years worth of events from the Gaussian copula:

```
S22.Vine.Sim<-Vine_Copula_Sim(Data=S22.Detrend.df,Marginals=S22.Migpd,
                              Vine_family=S22.Vine$Family,Vine_par=S22.Vine$Par,
                              Vine_par2=S22.Vine$Par2,N=100)
```

Plot the observationed and simulated values

```
S22.Pairs.Plot.Data<-data.frame(rbind(na.omit(S22.Detrend.df[,-1]),S22.Vine.Sim$x.Sim),c(rep("Observati
colnames(S22.Pairs.Plot.Data)<-c(names(S22.Detrend.df)[-1],"Type")
pairs(S22.Pairs.Plot.Data[,1:3],col=ifelse(S22.Pairs.Plot.Data$Type=="Observation","Black","Red"),upper
```



Use the `HT04` command to fit the HT04 model and simulate 100 years worth of events

```
S22.HT04<-HT04(data_Detrend_Dependence_df=S22.Detrend.df,
               data_Detrend_Declustered_df=S22.Detrend.Declustered.df,u_Dependence=0.995,
               Migpd=S22.Migpd,mu=365.25,N=100)
```

Print the fitted model when rainfall is the conditioning variable

```
S22.HT04$Model$Rainfall
```

```
## mexDependence(x = Migpd, which = colnames(data_Detrend_Dependence_df)[1],
##     dqu = u_Dependence, margins = Margins, constrain = FALSE,
##     v = V, maxit = Maxit)
##
##
```

```
## Marginal models:
##
## Dependence model:
##
## Conditioning on Rainfall variable.
## Thresholding quantiles for transformed data: dqu = 0.995
## Using gumbel margins for dependence estimation.
## Log-likelihood = -136.9253 -145.8972
##
## Dependence structure parameter estimates:
##     OsWL Groundwater
## a 1.0000      1.0000
## b 0.4054     -0.1943
```

Assign the name `S22.HT04.Sim` to the simulations that have been transformed back to the origional scale

```
S22.HT04.Sim<-S22.HT04$x.sim
```

Plot the simulated data:

```
S22.Pairs.Plot.Data<-data.frame(rbind(na.omit(S22.Detrend.df[,-1]),S22.HT04.Sim),
                                c(rep("Observation",nrow(na.omit(S22.Detrend.df))),rep("Simulation",nrow(S2
colnames(S22.Pairs.Plot.Data)<-c(names(S22.Detrend.df)[-1],"Type")
pairs(S22.Pairs.Plot.Data[,1:3],col=ifelse(S22.Pairs.Plot.Data$Type=="Observation","Black","Red"),upper
```