Multivariate Extremes

Whitney Huang

November 05, 2024

Contents

Motivating (simulated) example

```
library(MASS)
# Simulate bivariate normal with corr = 0.75
rho = 0.75
set.seed(123)
sim <- mvrnorm(n = 1000, mu = rep(0, 2),
               Sigma = matrix(c(1, rho, rho, 1), 2, 2))
cor(sim)
```

```
##           [,1]      [,2]
## [1,] 1.0000000 0.7432828
## [2,] 0.7432828 1.0000000
```

```
library(evd)
# Sample size
n <- 1000
set.seed(123)
# Simulate bivariate extreme random vector with
# unit Frechet and logistic copula (alpha = 0.5)
x <- rmvevd(n, dep = .5, model = "log", mar = c(0, 1, 0))
cor(x)
```

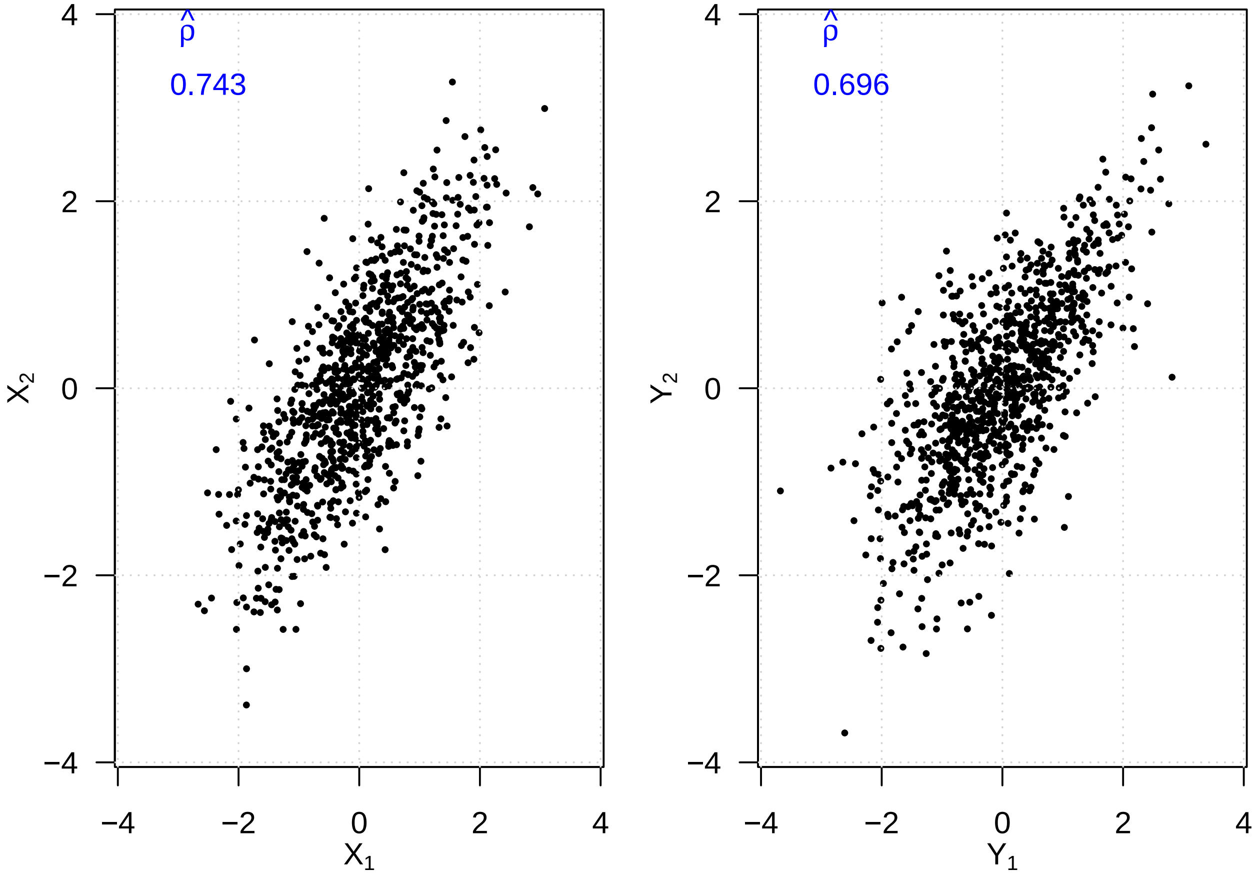
```
##             [,1]      [,2]
## [1,] 1.0000000 0.7351708
## [2,] 0.7351708 1.0000000
```

```
# transfer the marginals to standard normal
u1 <- pgev(x[, 1]); n1 <- qnorm(u1)
u2 <- pgev(x[, 2]); n2 <- qnorm(u2)
cor(cbind(n1, n2))
```

```
##          n1       n2
## n1 1.000000 0.696237
## n2 0.696237 1.000000
```

Plot the data

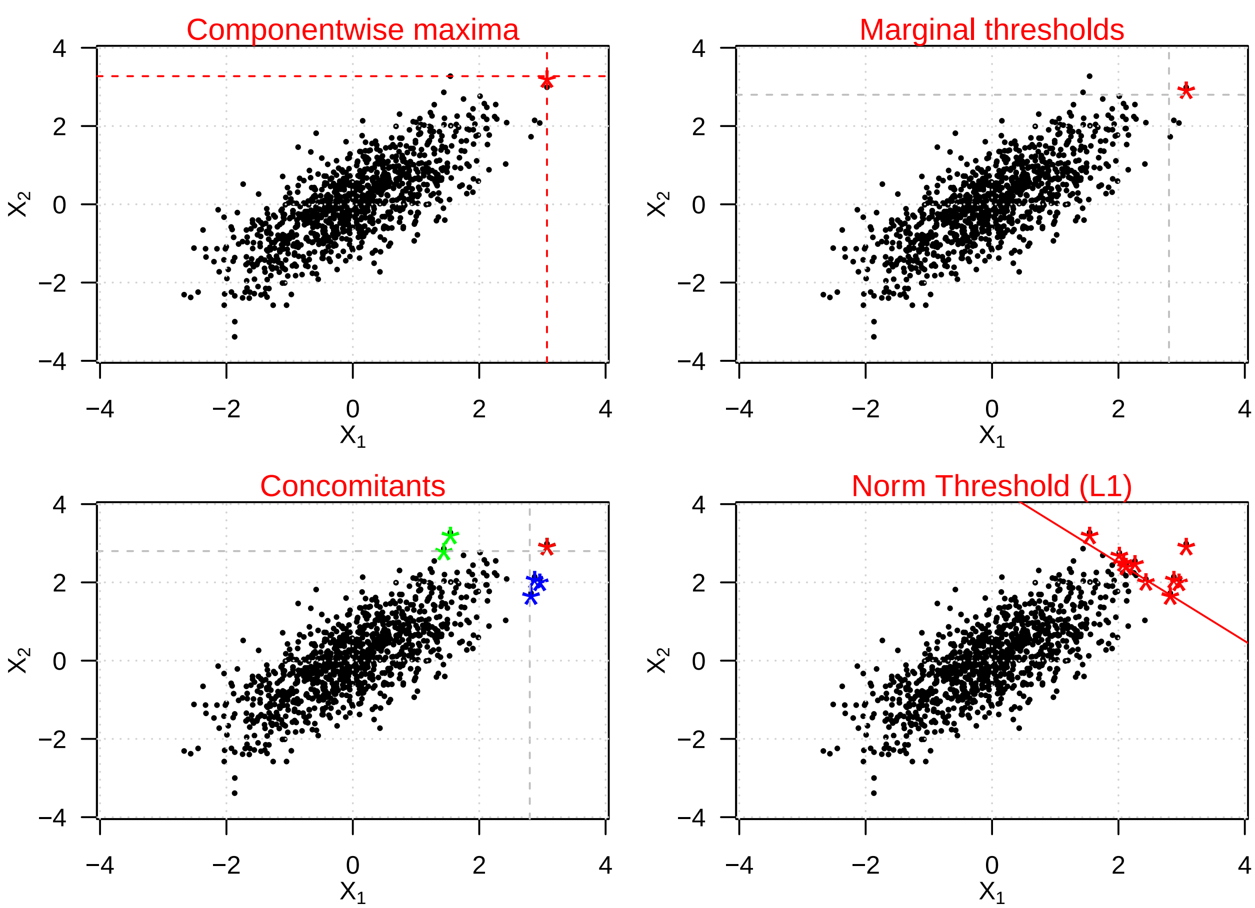
```
par(mfrow = c(1, 2), las = 1, mgp = c(2, 1, 0),
    mar = c(3.6, 3.6, 0.8, 0.6))
plot(sim, pch = 16, cex = 0.5,
     xlab = expression(X[1]),
     ylab = expression(X[2]), xlim = c(-3.75, 3.75),
     ylim = c(-3.75, 3.75))
grid()
legend("topleft", legend = round(cor(sim)[1, 2], 3),
       title = expression(hat(rho)), text.col = "blue",
       bty = "n")
plot(cbind(n1, n2), pch = 16, cex = 0.5,
     xlab = expression(Y[1]),
     ylab = expression(Y[2]), xlim = c(-3.75, 3.75),
     ylim = c(-3.75, 3.75))
grid()
legend("topleft", legend = round(cor(n1, n2), 3),
       title = expression(hat(rho)), text.col = "blue",
       bty = "n")
```



Multivariate extremes are not uniquely defined

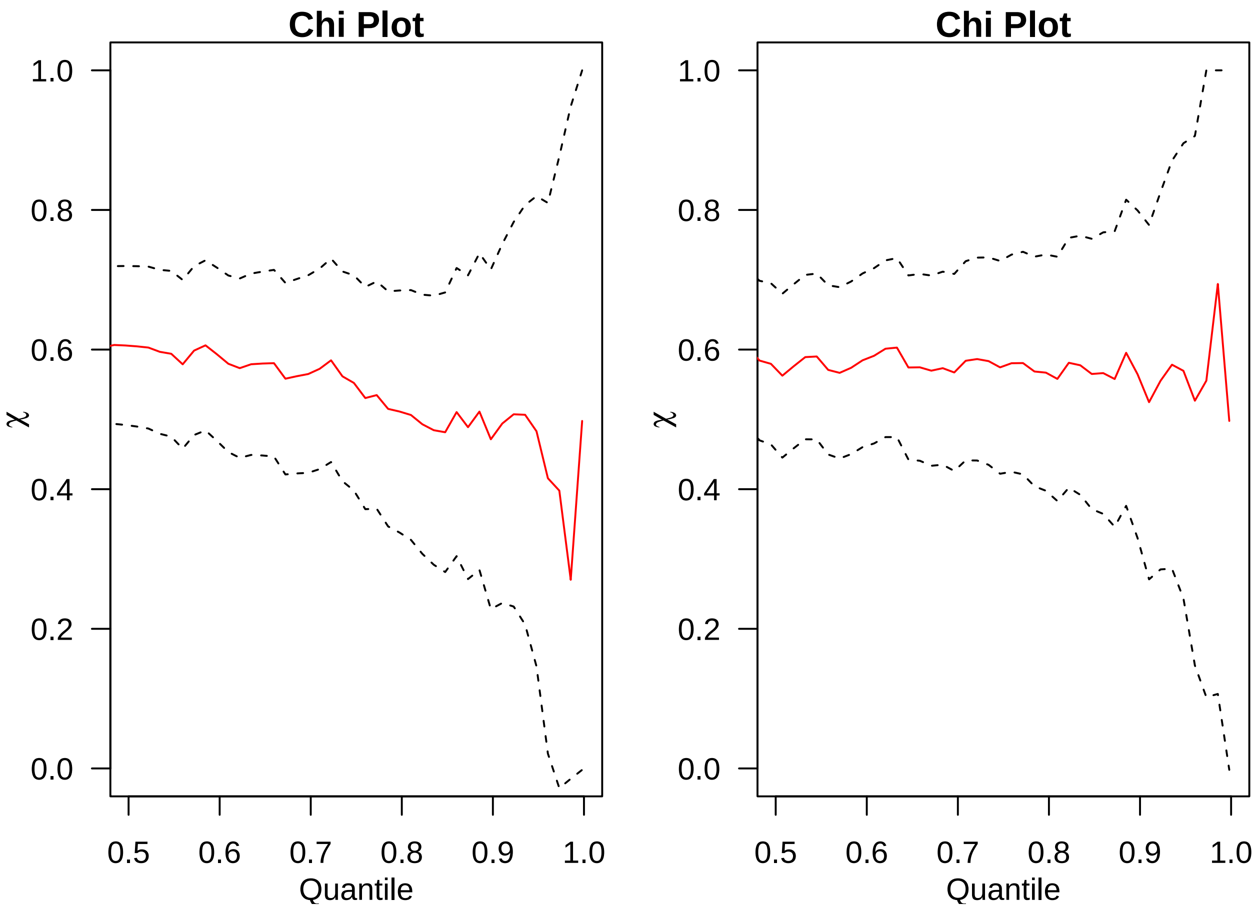
```
par(mfrow = c(2, 2), las = 1, mgp = c(2, 1, 0),
    mar = c(3.6, 3.6, 1, 0.6))
plot(sim, pch = 16, cex = 0.5,
     xlab = expression(X[1]),
     ylab = expression(X[2]), xlim = c(-3.75, 3.75),
     ylim = c(-3.75, 3.75))
points(max(sim[, 1]), max(sim[, 2]), pch = "*",
       cex = 2, col = "red")
abline(h = max(sim[, 2]), lty = 2, col = "red")
abline(v = max(sim[, 1]), lty = 2, col = "red")
grid()
mtext("Componentwise maxima", col = "red")
plot(sim, pch = 16, cex = 0.5,
     xlab = expression(X[1]),
     ylab = expression(X[2]), xlim = c(-3.75, 3.75),
     ylim = c(-3.75, 3.75))
ex1 <- which(sim[, 1] >= 2.8)
ex2 <- which(sim[, 2] >= 2.8)
ex <- intersect(ex1, ex2)
points(sim[ex, 1], sim[ex, 2], pch = "*", cex = 2, col = "red")
abline(h = 2.8, lty = 2, col = "gray")
abline(v = 2.8, lty = 2, col = "gray")
```

```r
grid()
mtext("Marginal thresholds", col = "red")
plot(sim, pch = 16, cex = 0.5,
     xlab = expression(X[1]),
     ylab = expression(X[2]), xlim = c(-3.75, 3.75),
     ylim = c(-3.75, 3.75))
points(sim[ex1, 1], sim[ex1, 2], pch = "*", cex = 2, col = "blue")
points(sim[ex2, 1], sim[ex2, 2], pch = "*", cex = 2, col = "green")
points(sim[ex, 1], sim[ex, 2], pch = "*", cex = 2, col = "red")
abline(h = 2.8, lty = 2, col = "gray")
abline(v = 2.8, lty = 2, col = "gray")
grid()
mtext("Concomitants", col = "red")
plot(sim, pch = 16, cex = 0.5,
     xlab = expression(X[1]),
     ylab = expression(X[2]), xlim = c(-3.75, 3.75),
     ylim = c(-3.75, 3.75))
l1 <- which(sim[, 1] + sim[, 2] >=4.5)
points(sim[l1, 1], sim[l1, 2], pch = "*", cex = 2, col = "red")
abline(a = 4.5, b = -1, col = "red")
grid()
mtext("Norm Threshold (L1)", col = "red")
```

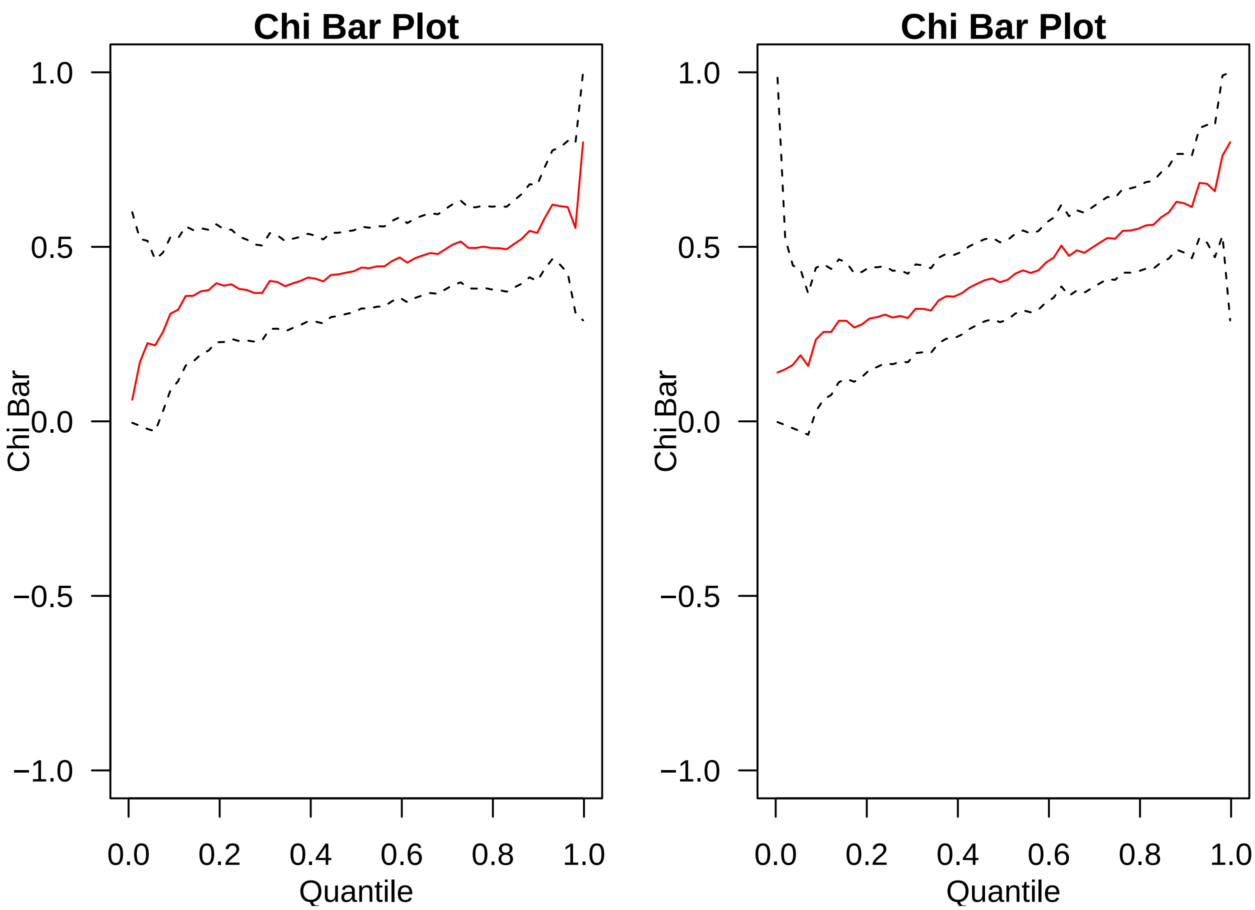


Plot $\chi(u)$

```
par(mfrow = c(1, 2), las = 1, mgp = c(2, 1, 0),
    mar = c(3.6, 3.6, 1, 0.6))
chiplot(sim, which = 1, ylim1 = c(0, 1), nq = 80, col = "red",
        xlim = c(0.5, 1), ylab1 = expression(chi))
chiplot(cbind(n1, n2), which = 1, ylim1 = c(0, 1), nq = 80, col = "red", xlim = c(0.5, 1), , ylab1 = exp
```




```
# we can change which = 1 to which =2
par(mfrow = c(1, 2), las = 1, mgp = c(2, 1, 0),
    mar = c(3.6, 3.6, 1, 0.6))
chiplot(sim, which = 2, ylim1 = c(0, 1), nq = 60, col = "red")
chiplot(cbind(n1, n2), which = 2, ylim1 = c(0, 1), nq = 60, col = "red")
```



Annual Maximum Wind Speeds at Albany and Hartford

```r
library(ismev)
data(wind); head(wind)
```

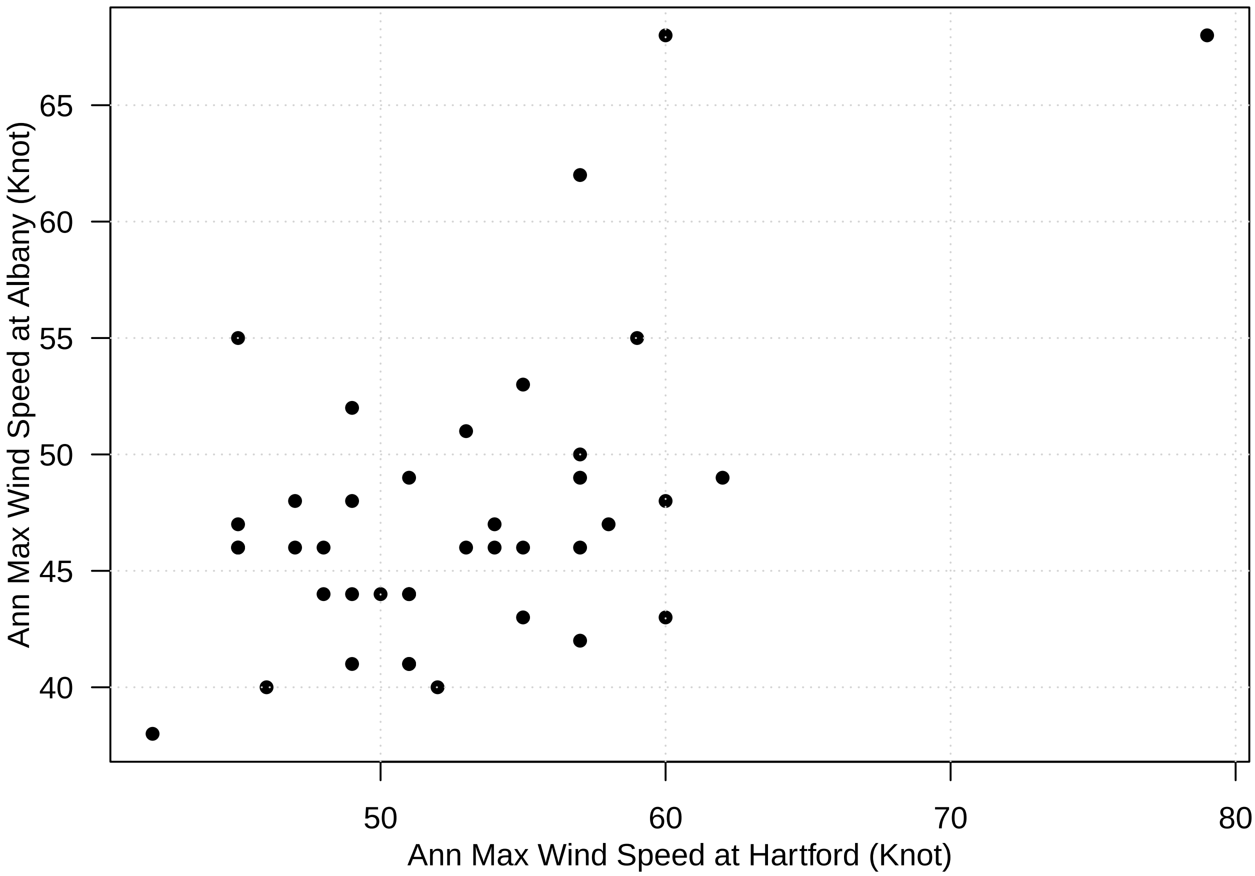
```
##    Year Hartford Albany
## 33 1944       49     52
## 34 1945       54     46
## 35 1946       60     48
## 36 1947       49     44
## 37 1948       57     42
## 38 1949       51     41
```

```r
hartford <- wind[, 2]; albany <- wind[, 3]
blockmax <- cbind(hartford, albany)
```

Plot the componentwise maxima "data"

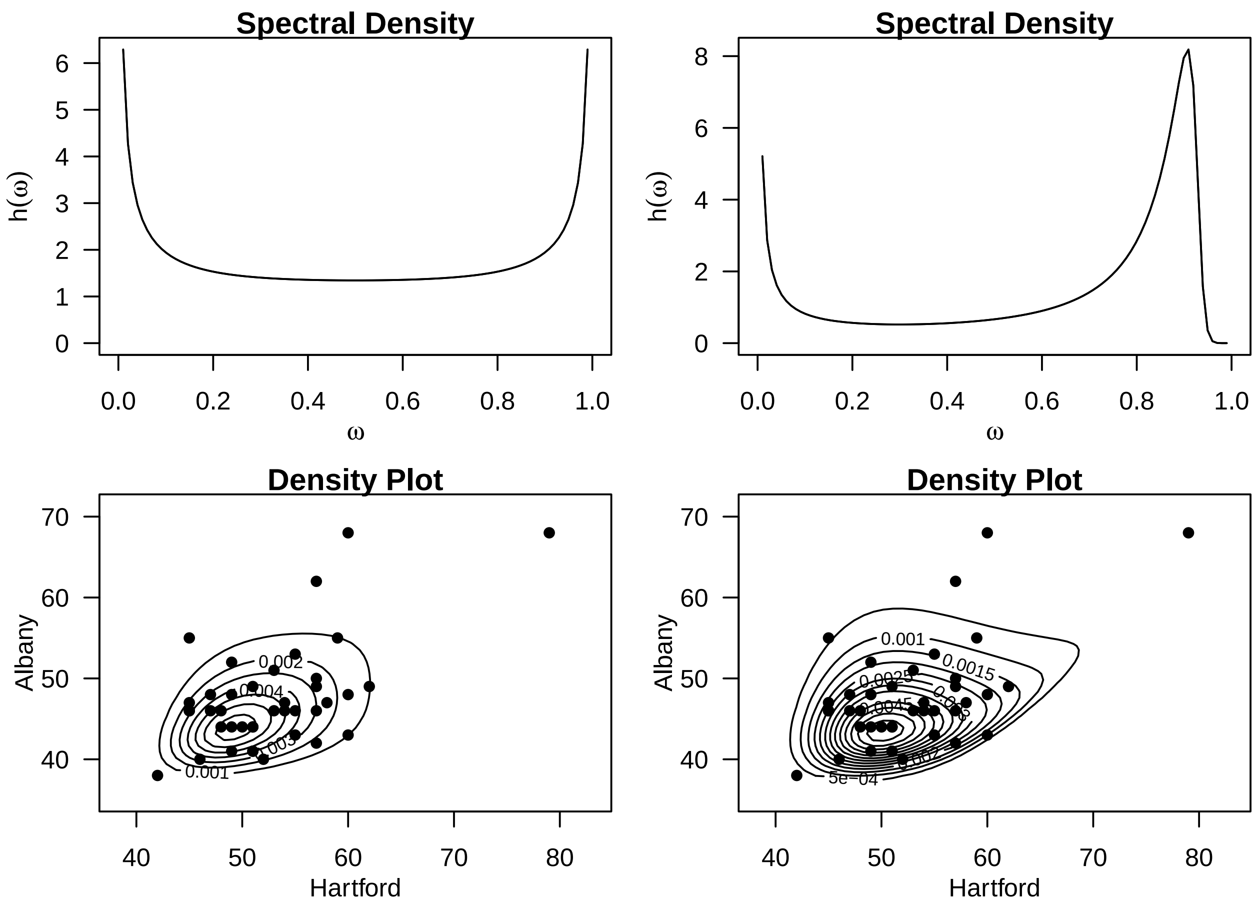
```r
par(las = 1, mgp = c(2, 1, 0), mar = c(3.6, 3.6, 1, 0.6))
plot(blockmax, pch = 16, cex = 1,
```

```
     xlab = "Ann Max Wind Speed at Hartford (Knot)",
     ylab = "Ann Max Wind Speed at Albany (Knot)")
grid()
```

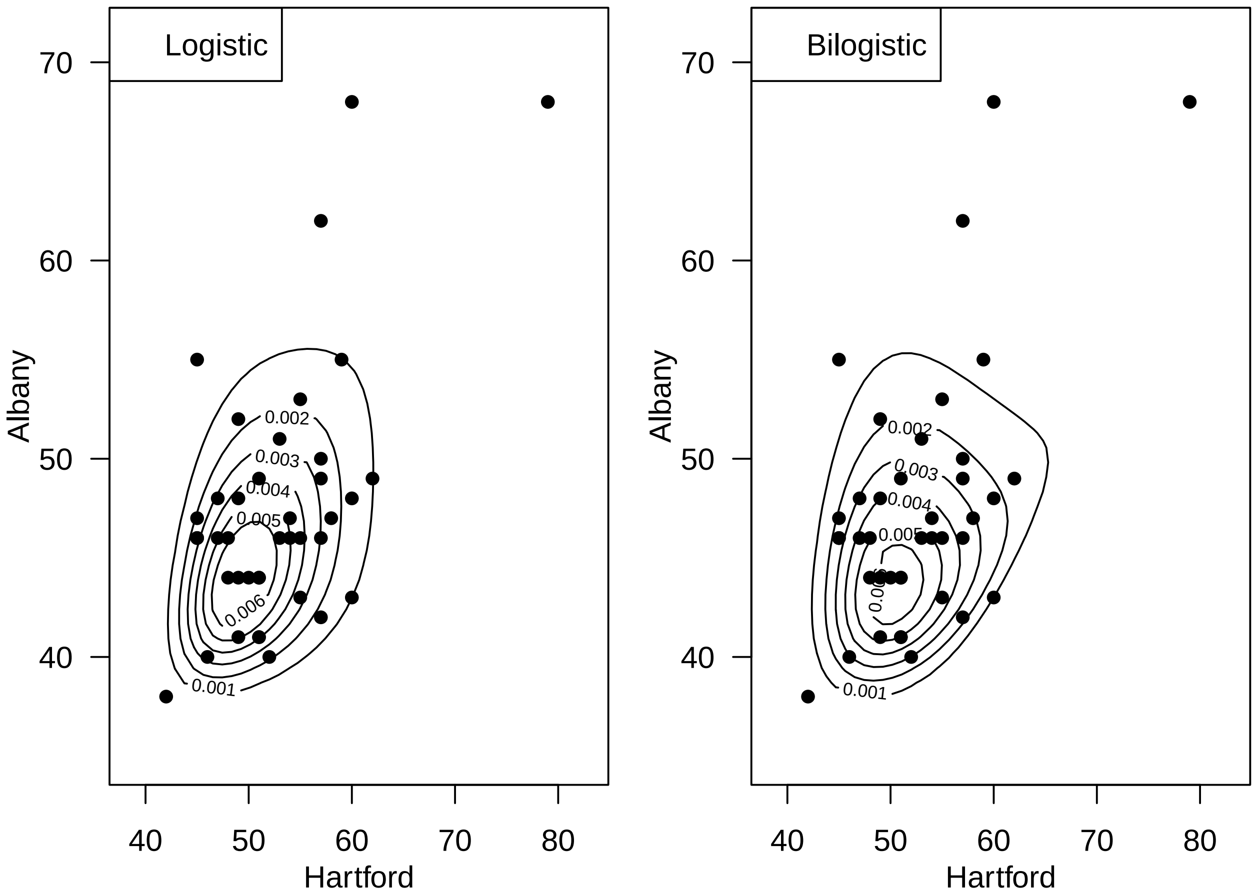


Fit bivariate GEV

```
fit1 <- fbvevd(blockmax, model = "log")
fit2 <- fbvevd(blockmax, model = "bilog")

par(las = 1, mgp = c(2, 1, 0), mar = c(3.6, 3.6, 1, 0.6), mfrow = c(2, 2))
plot(fit1, which = 6,
     xlab = expression(omega), ylab = expression(h(omega)))
plot(fit2, which = 6,
     xlab = expression(omega), ylab = expression(h(omega)))
plot(fit1, which = 3, pch = 16, cex = 0.35, xlab = "Hartford", ylab = "Albany")
plot(fit2, which = 3, pch = 16, cex = 0.35, xlab = "Hartford", ylab = "Albany")
```




```r
par(las = 1, mgp = c(2, 1, 0), mar = c(3.6, 3.6, 0.2, 0.3), mfrow = c(1, 2))
plot(fit1, which = 3, pch = 16, cex = 0.15, xlab = "Hartford", ylab = "Albany", main = "",
     levels = seq(0.001, 0.006, 0.001))
legend("topleft", legend = "Logistic")
plot(fit2, which = 3, pch = 16, cex = 0.15, xlab = "Hartford", ylab = "Albany", main = "",
     levels = seq(0.001, 0.006, 0.001))
legend("topleft", legend = "Bilogistic")
```



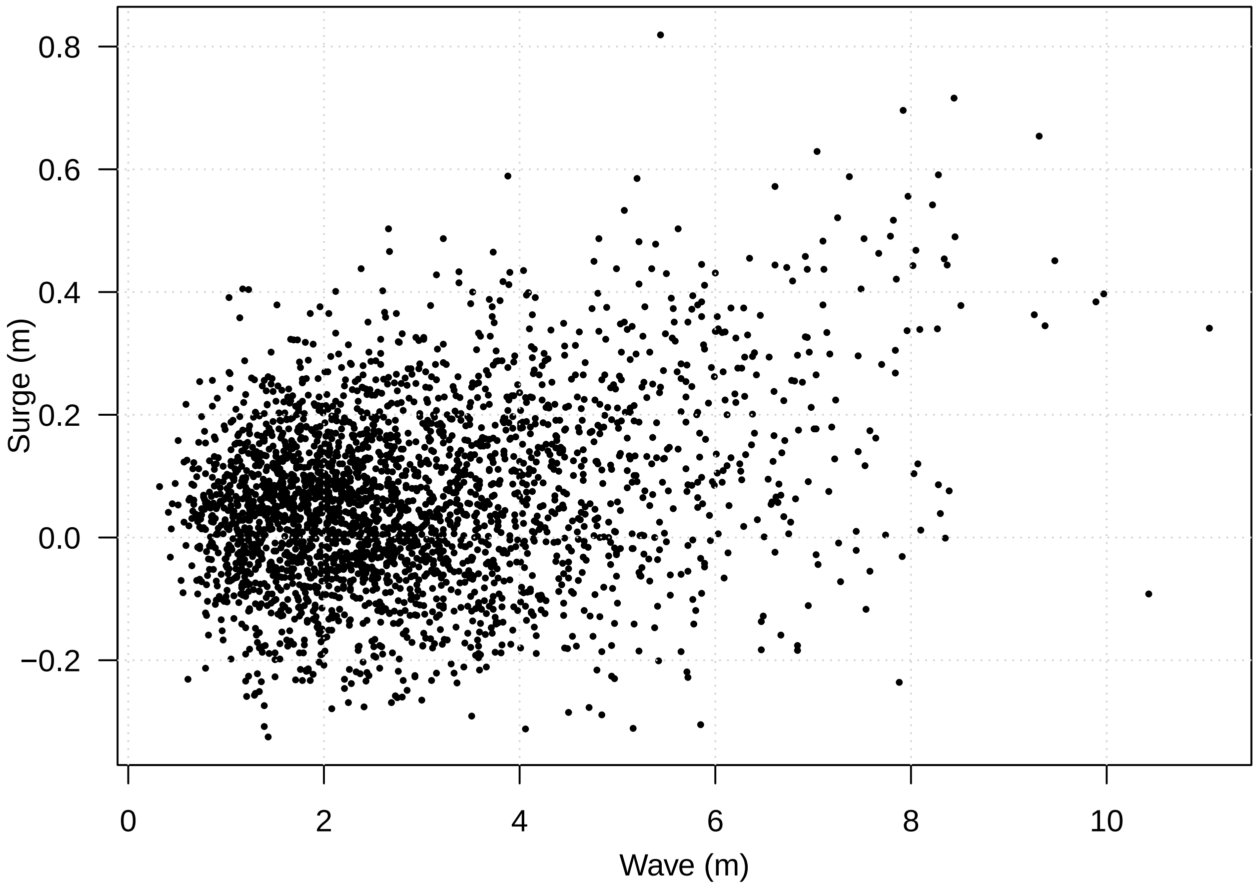
Wave and Surge Heights in South-West England

Let's plot the data as well as the $\chi(u)$ plot

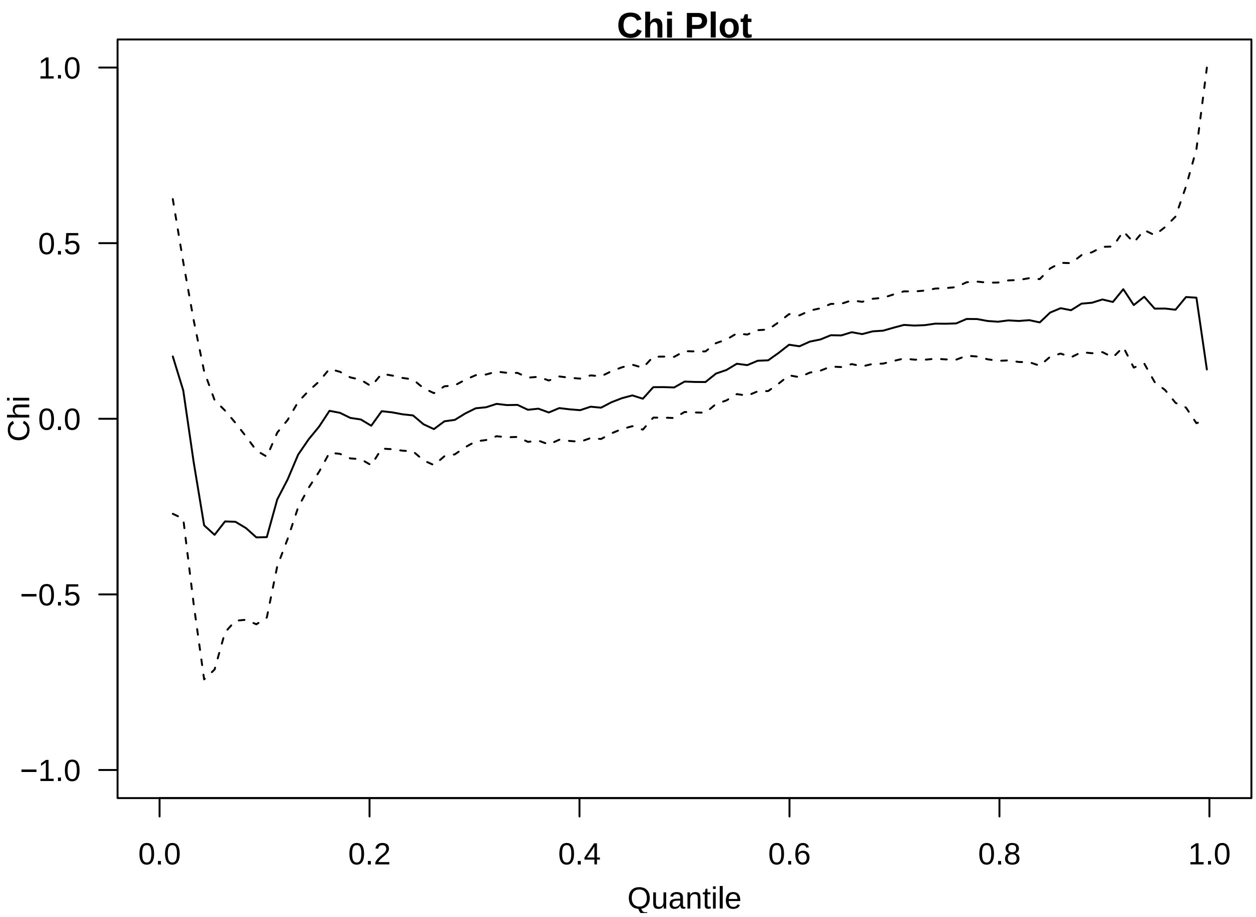
```
data(wavesurge); head(wavesurge)
```

```
##   wave  surge
## 1 1.50 -0.009
## 2 1.83 -0.053
## 3 2.44 -0.024
## 4 1.68  0.000
## 5 1.49  0.079
## 6 1.20  0.068
```

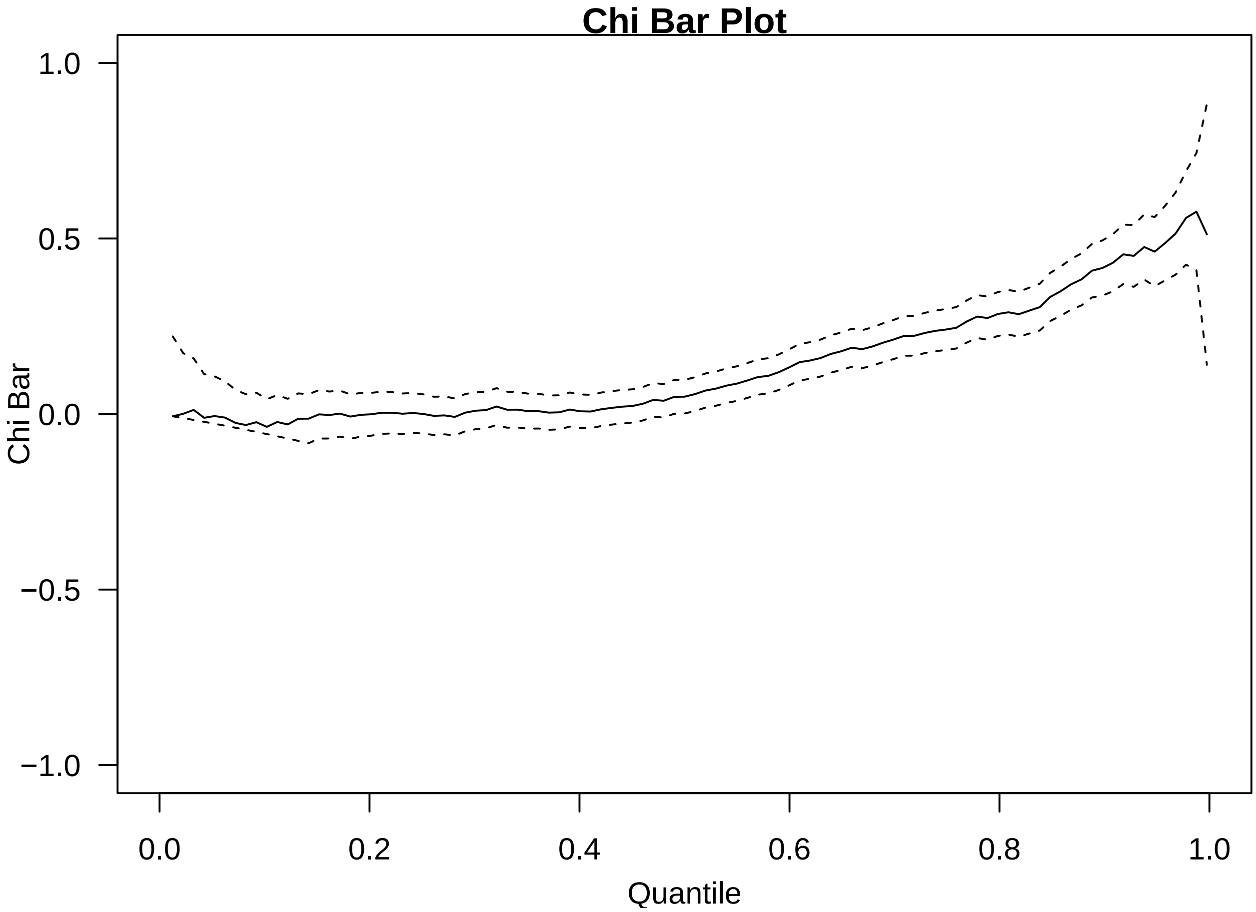
```
wave <- wavesurge[, 1]; surge <- wavesurge[, 2]

par(las = 1, mgp = c(2.2, 1, 0), mar = c(3.6, 3.6, 0.8, 0.6))
plot(wave, surge, pch = 16, cex = 0.5,
     xlab = "Wave (m)", ylab = "Surge (m)")
grid()
```




```
chiplot(wavesurge)
```

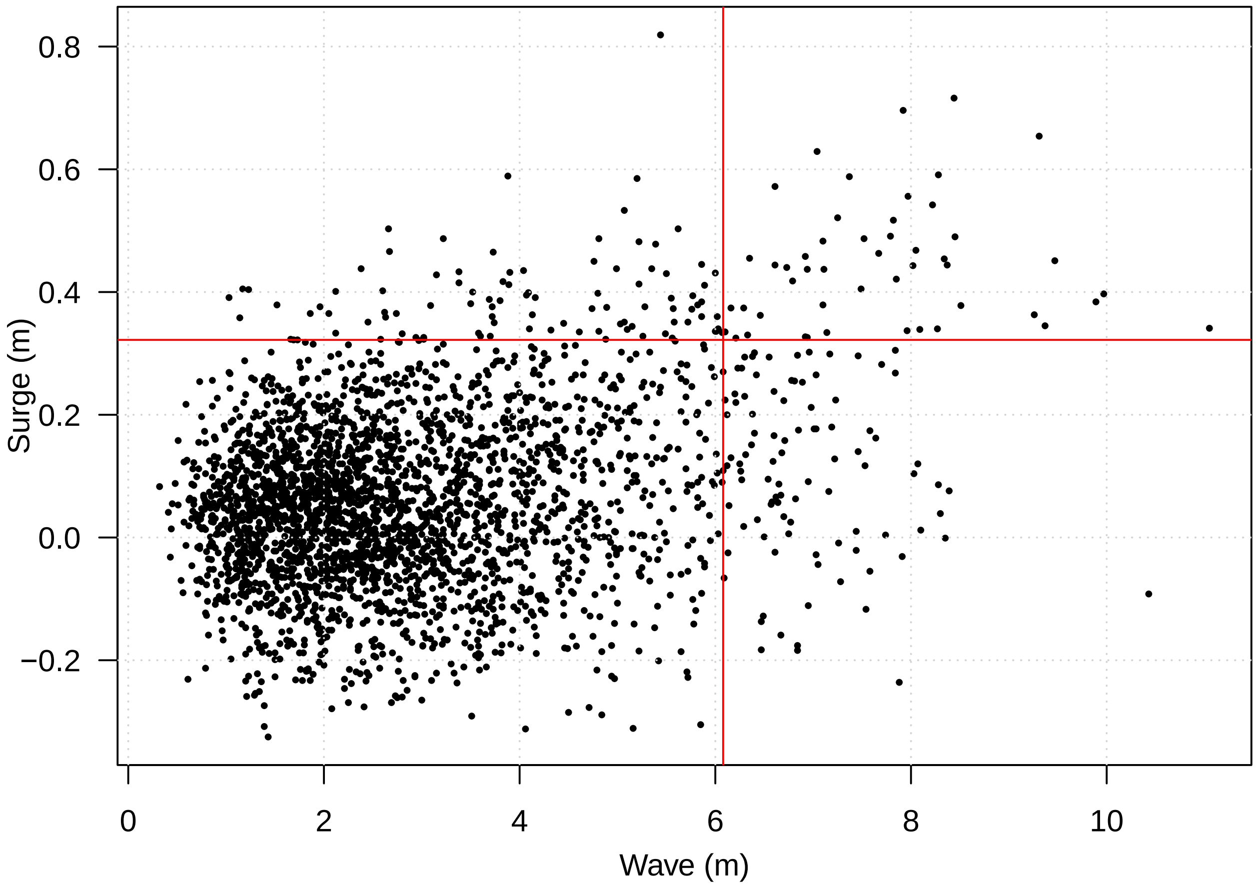


Chi Plot
1.0
0.5
0.0
−0.5
−1.0
Chi
0.0
0.2
0.4
0.6
0.8
1.0
Quantile

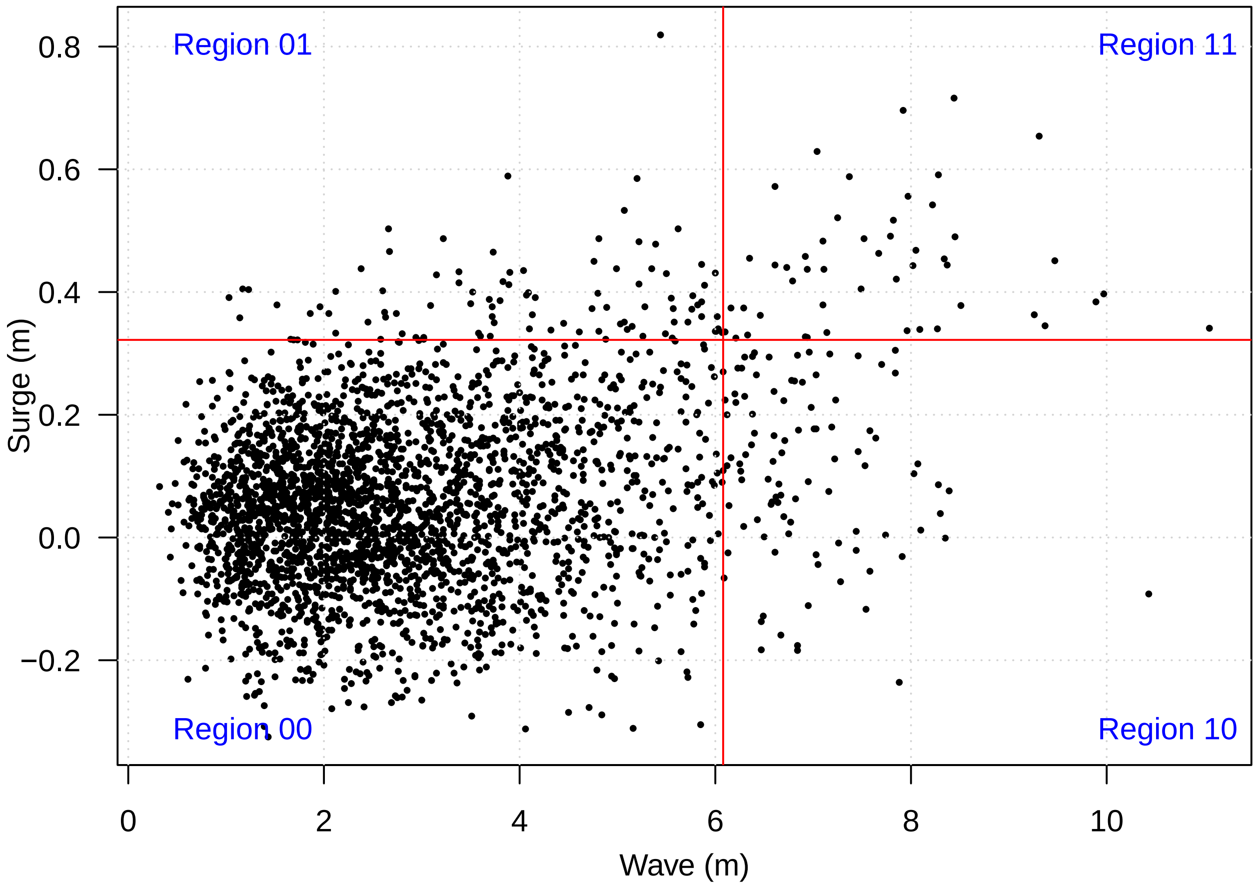


Fit bivariate threshold excess models

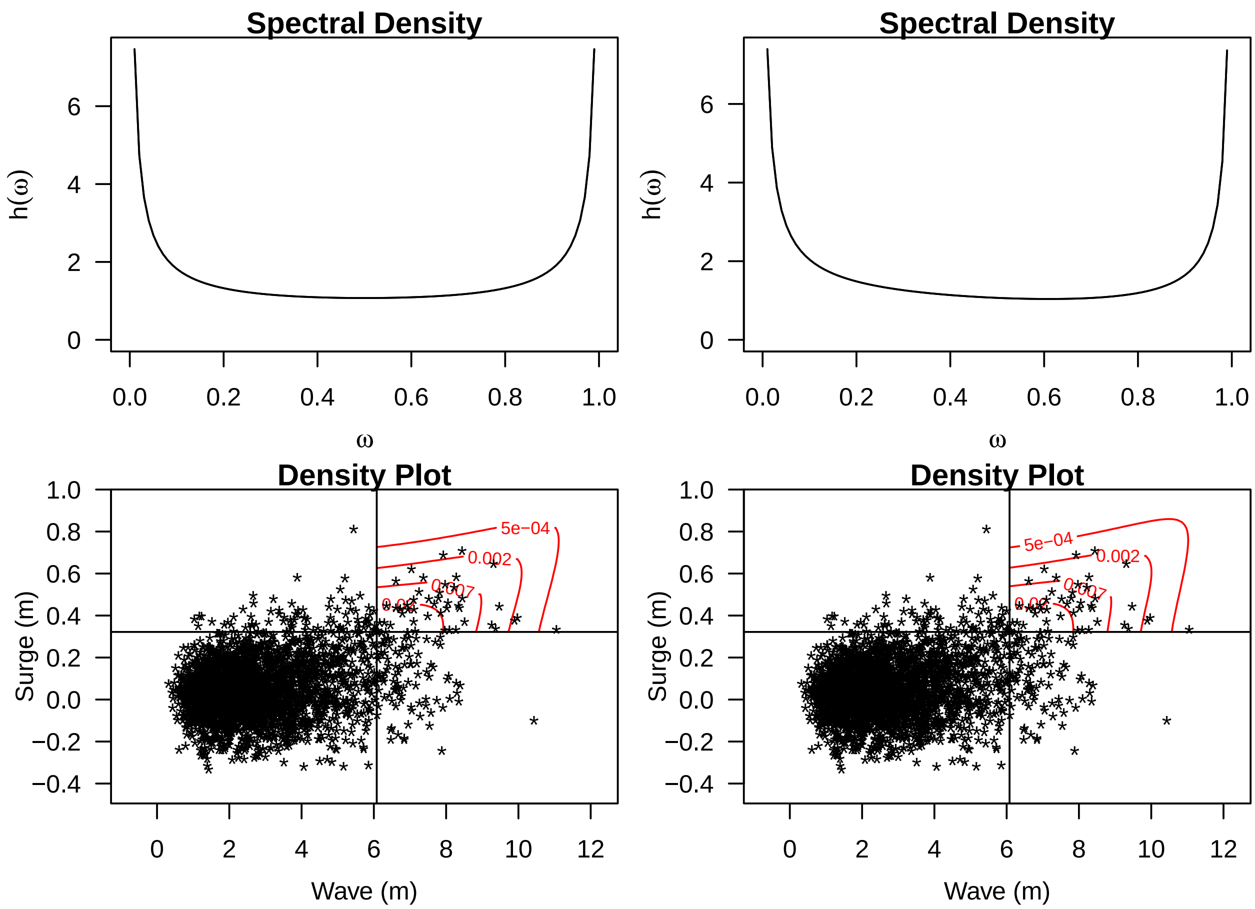
```
thresh <- c(quantile(wave, 0.95), quantile(surge, 0.95))
par(las = 1, mgp = c(2.2, 1, 0), mar = c(3.6, 3.6, 0.8, 0.6))
plot(wave, surge, pch = 16, cex = 0.5,
     xlab = "Wave (m)", ylab = "Surge (m)")
grid()
abline(v = thresh[1], col = "red")
abline(h = thresh[2], col = "red")
```



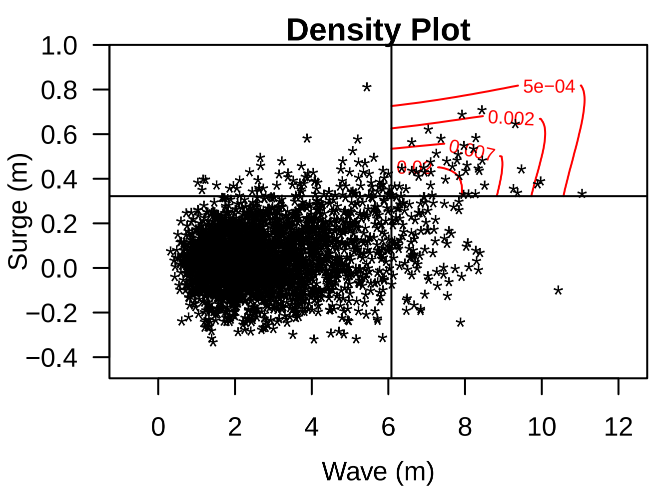

```
par(las = 1, mgp = c(2.2, 1, 0), mar = c(3.6, 3.6, 0.8, 0.6))
plot(wave, surge, pch = 16, cex = 0.5,
     xlab = "Wave (m)", ylab = "Surge (m)")
grid()
abline(v = thresh[1], col = "red")
abline(h = thresh[2], col = "red")
legend("topleft", legend = "Region 01", bty = "n", text.col = "blue")
legend("topright", legend = "Region 11", bty = "n", text.col = "blue")
legend("bottomleft", legend = "Region 00", bty = "n", text.col = "blue")
legend("bottomright", legend = "Region 10", bty = "n", text.col = "blue")
```




```
fitlogistic <- fbvpot(wavesurge, thresh)
fitbilogistic <- fbvpot(wavesurge, thresh, model = "bilog")

par(las = 1, mgp = c(2.4, 1, 0), mar = c(3.6, 3.6, 1, 0.6), mfrow = c(2, 2))
plot(fitlogistic, which = 4,
     xlab = expression(omega), ylab = expression(h(omega)))
plot(fitbilogistic, which = 4,
     xlab = expression(omega), ylab = expression(h(omega)))
plot(fitlogistic, which = 1, pch = "*", xlab = "Wave (m)", ylab = "Surge (m)",
     col = "red")
plot(fitbilogistic, which = 1, pch = "*", xlab = "Wave (m)", ylab = "Surge (m)", col = "red")
```




```
par(las = 1, mgp = c(2.4, 1, 0), mar = c(3.6, 3.6, 1, 0.6))
plot(fitlogistic, which = 1, pch = "*", xlab = "Wave (m)", ylab = "Surge (m)", col = "red")
```



Conditional extreme values modelling [Heffernan and Tawn, 2004]

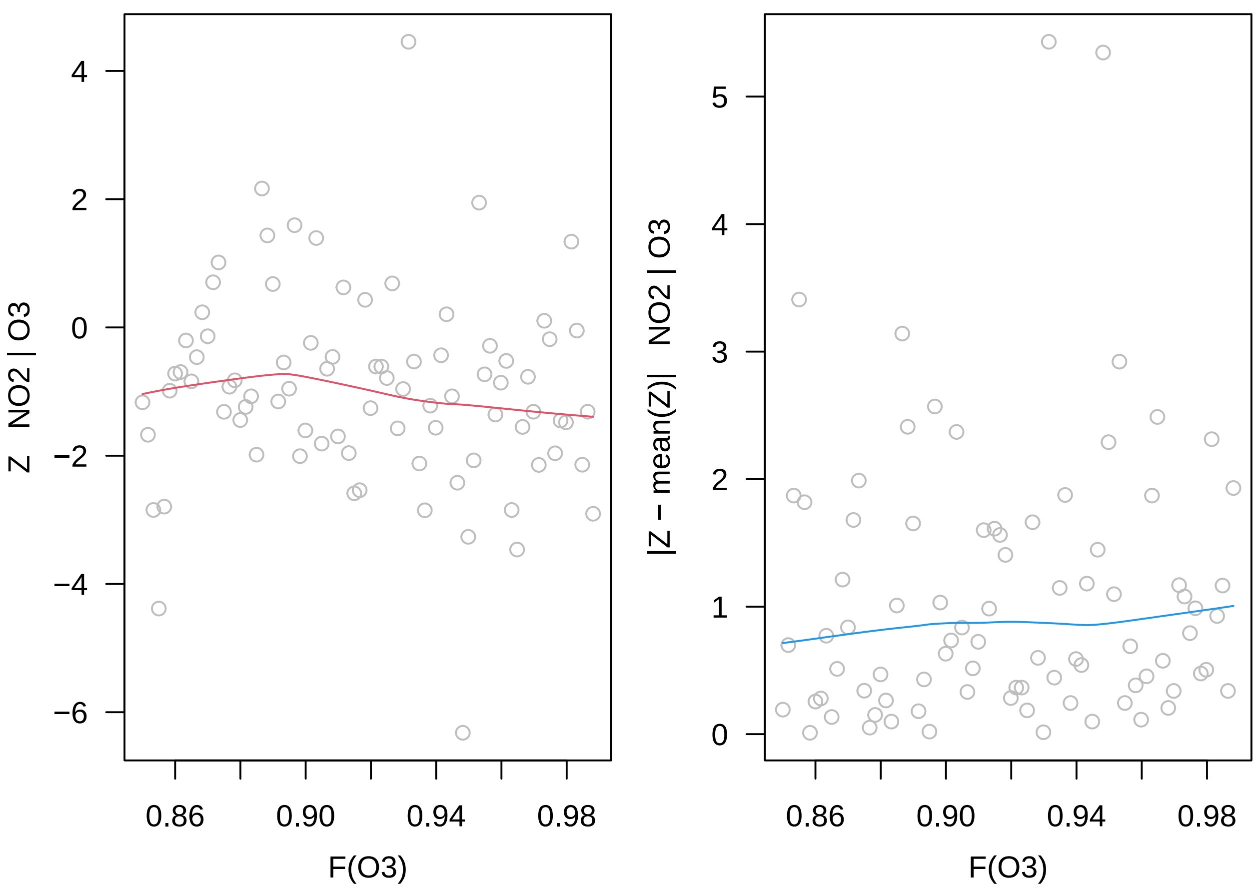
Air pollution data from Leeds (U.K.) city centre, collected from 1994 to 1998. The summer data set corresponds to the months of April to July inclusive. The winter data set corresponds to the months of

November to February inclusive. Some outliers have been removed, as discussed by Heffernan and Tawn, 2004. Here we only consider NO_2 and O_3 during summer to illustrate the CEV modeling framework.

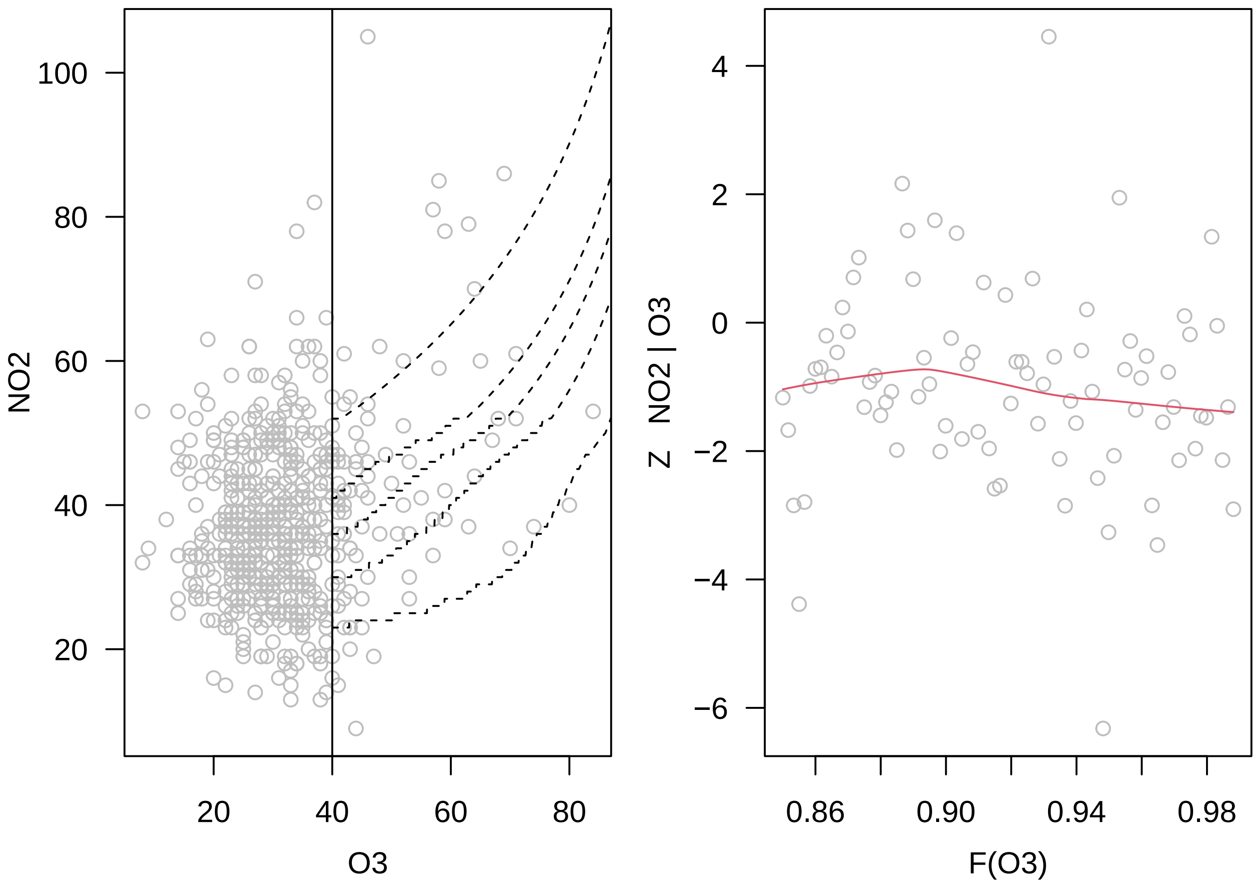
```r
library(texmex)
data(winter); data(summer)
(CEV_Fit <- mex(summer[, 1:2], mqu = .9, dqu = 0.85, which = "O3"))
```

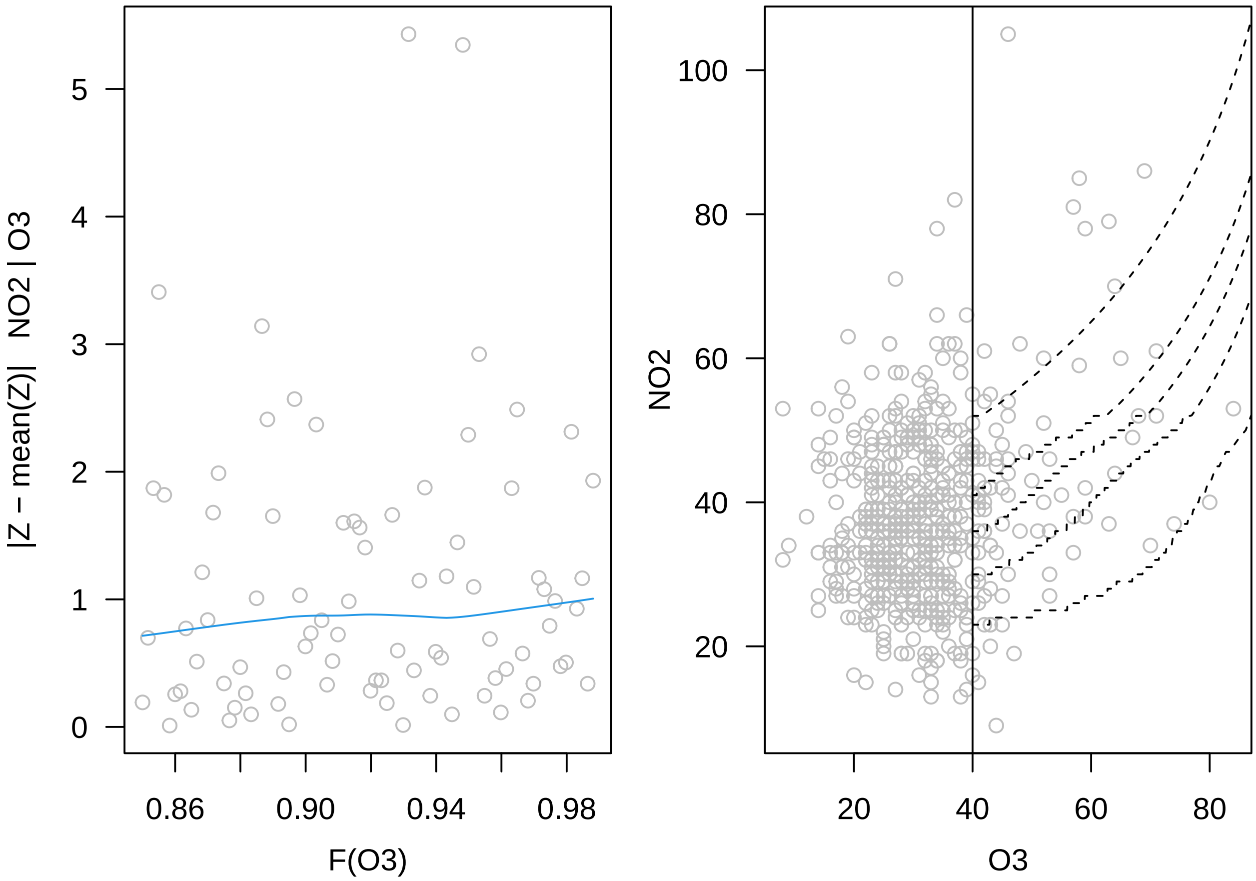
```
## mex(data = summer[, 1:2], which = "O3", mqu = 0.9, dqu = 0.85)
##
##
## Marginal models:
##
## Dependence model:
##
## Conditioning on O3 variable.
## Thresholding quantiles for transformed data: dqu = 0.85
## Using laplace margins for dependence estimation.
## Constrained estimation of dependence parameters using v = 10 .
## Log-likelihood = -164.8171
##
## Dependence structure parameter estimates:
##        a      b
## 0.7730 0.1878
```

```r
par(las = 1, mgp = c(2.4, 1, 0), mar = c(3.6, 3.6, 1, 0.6), mfrow = c(1, 2))
plot(CEV_Fit)
```



```
par(las = 1, mgp = c(2.4, 1, 0), mar = c(3.6, 3.6, 1, 0.6))
plot(CEV_Fit)
```





|Z − mean(Z)| NO2 | O3
F(O3)
0.86
0.90
0.94
0.98
0
1
2
3
4
5
NO2
O3
20
40
60
80
100