A simple example using the gamlss packages

Flexible Regression and Smoothing

Bob Rigby[1] Mikis Stasinopoulos[1]

[1]STORM, London Metropolitan University

XXV SIMPOSIO INTERNACIONAL DE ESTADêSTICA, Armenia, Colombia, June 2015

The data

Data summary:

R data file: `film90` in package **gamlss.data** of dimensions 4015×14 but only two variables are used here.

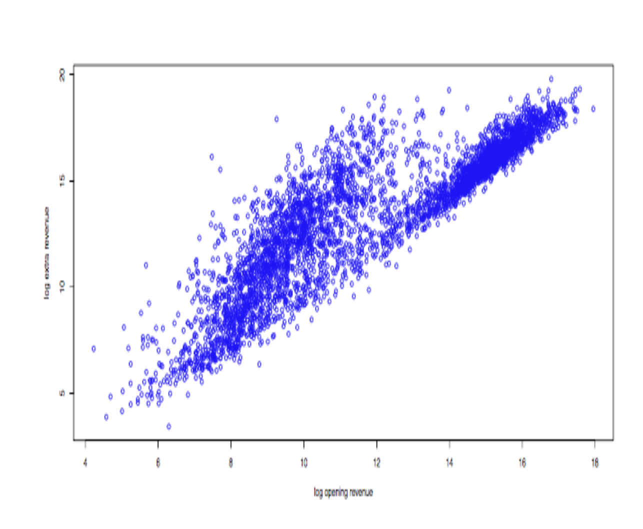
variables

`lborev1` : the log of box office revenues after the first week calculated in 1987 prices (the response variable)

`lboopen` : the log of box office opening week revenues calculated in 1987 prices

purpose: to demonstrate the fitting of a simple regression type model in the **gamlss** package.

Plot of the data



gamlss

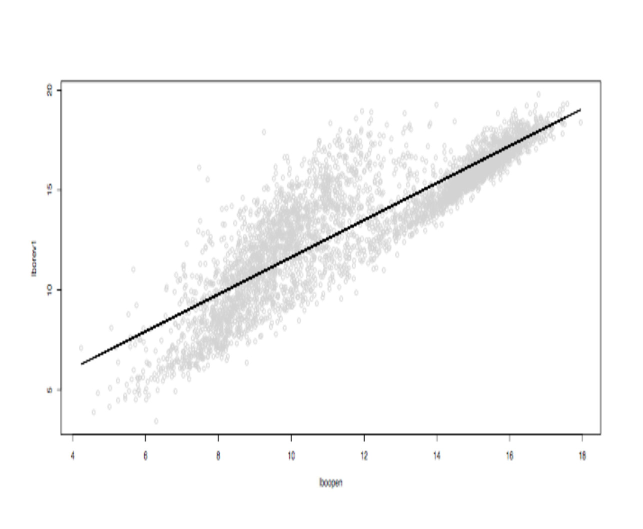
Fitting a parametric model

```
m0 <- gamlss(lborev1~lboopen, data=film90)

## GAMLSS-RS iteration 1: Global Deviance = 15078.88
## GAMLSS-RS iteration 2: Global Deviance = 15078.88

plot(lborev1~lboopen, data=film90,  col = "lightgray", lty=4)
lines(fitted(m0)~film90$lboopen)
```

```
m0 <- gamlss(lborev1~poly(lboopen,3), data=film90, family=NO)

## GAMLSS-RS iteration 1: Global Deviance = 14517.65
## GAMLSS-RS iteration 2: Global Deviance = 14517.65
```

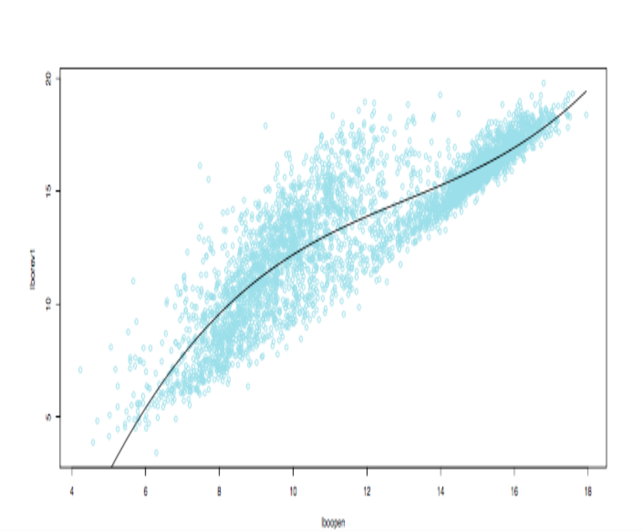
```
m00 <- gamlss(lborev1~lboopen+I(lboopen^2)+I(lboopen^3), data=film90,
                      family=NO)

## GAMLSS-RS iteration 1: Global Deviance = 14517.65
## GAMLSS-RS iteration 2: Global Deviance = 14517.65
```

The linear fit



gamlss

The polynomial fit



gamlss

The orthogonal polynomial summary

```
## Family:  c("NO", "Normal")
##
## Call:
## gamlss(formula = lborev1 ~ poly(lboopen, 3), family = NO, data = film90)
##
## Fitting method: RS()
##
## -------------------------------------------------------------------
## Mu link function:  identity
## Mu Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)         13.29257    0.02307  576.10   <2e-16 ***
## poly(lboopen, 3)1  180.61569    1.46494  123.29   <2e-16 ***
## poly(lboopen, 3)2  -29.66307    1.46494  -20.25   <2e-16 ***
## poly(lboopen, 3)3   20.30788    1.46494   13.86   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## -------------------------------------------------------------------
## Sigma link function:  log
## Sigma Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.38181    0.01114   34.28   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## -------------------------------------------------------------------
## No. of observations in the fit:  4031
## Degrees of Freedom for the fit:  5
##       Residual Deg. of Freedom:  4026
```

The polynomial summary

```
## ******************************************************************
## Family:  c("NO", "Normal")
##
## Call:  gamlss(formula = lborev1 ~ lboopen + I(lboopen^2) + I(lboopen^3),
##     family = NO, data = film90)
##
## Fitting method: RS()
##
## ------------------------------------------------------------------
## Mu link function:  identity
## Mu Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -2.244e+01  1.278e+00  -17.55   <2e-16 ***
## lboopen       7.174e+00  3.537e-01   20.28   <2e-16 ***
## I(lboopen^2) -4.985e-01  3.171e-02  -15.72   <2e-16 ***
## I(lboopen^3)  1.275e-02  9.195e-04   13.86   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ------------------------------------------------------------------
## Sigma link function:  log
## Sigma Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.38181    0.01114   34.28   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ------------------------------------------------------------------
## No. of observations in the fit:  4031
## Degrees of Freedom for the fit:  5
##       Residual Deg. of Freedom:  4026
##                       at cycle:  2
```

gamlss

The polynomial fitted model

The fitted model is given by $Y \sim NO(\hat{\mu}, \hat{\sigma})$ where $\hat{\mu} = \hat{\beta}_{01} + \hat{\beta}_{11}x + \hat{\beta}_{21}x^2 + \hat{\beta}_{31}x^3$, i.e.

$$\hat{\mu} = -22.437 + 7.174x - 0.499x^2 + 0.013x^3$$

and

$$\log(\hat{\sigma}) = \hat{\beta}_{02} = 0.3818$$

so $\hat{\sigma} = \exp(0.3818) = 1.465$ (since σ has a default log link function), where Y = `lborev1` and x = `lboopen`.

gamlss

The polynomial variance covariance

For robust standard erros use the argument `robust=TRUE`

```
# the variance-covariance
print(vcov(m00), digit=3)

##              (Intercept)   lboopen I(lboopen^2) I(lboopen^3) (Intercept)
## (Intercept)     1.63e+00 -4.49e-01     3.95e-02    -1.12e-03   -2.32e-11
## lboopen        -4.49e-01  1.25e-01    -1.11e-02     3.18e-04    6.38e-12
## I(lboopen^2)    3.95e-02 -1.11e-02     1.01e-03    -2.90e-05   -5.61e-13
## I(lboopen^3)   -1.12e-03  3.18e-04    -2.90e-05     8.46e-07    1.59e-14
## (Intercept)    -2.32e-11  6.38e-12    -5.61e-13     1.59e-14    1.24e-04

# the correlation matrix
print(vcov(m00, type="cor"), digit=3)

##              (Intercept)   lboopen I(lboopen^2) I(lboopen^3) (Intercept)
## (Intercept)     1.00e+00 -9.93e-01     9.74e-01    -9.49e-01   -1.63e-09
## lboopen        -9.93e-01  1.00e+00    -9.94e-01     9.79e-01    1.62e-09
## I(lboopen^2)    9.74e-01 -9.94e-01     1.00e+00    -9.95e-01   -1.59e-09
## I(lboopen^3)   -9.49e-01  9.79e-01    -9.95e-01     1.00e+00    1.55e-09
## (Intercept)    -1.63e-09  1.62e-09    -1.59e-09     1.55e-09    1.00e+00
```

gamlss

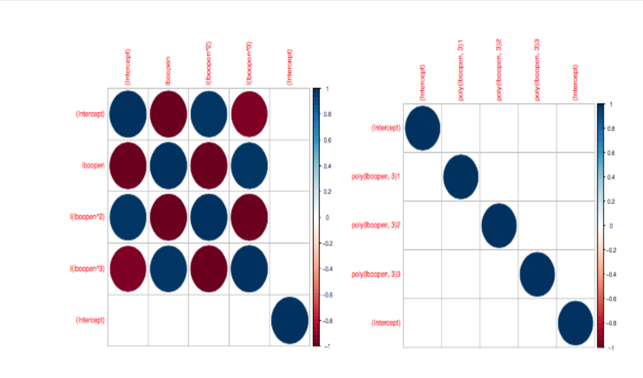
The polynomial standard errors

```
# standard errors
print(vcov(m00, type="se"), digits=2)

##  (Intercept)      lboopen I(lboopen^2) I(lboopen^3)  (Intercept)
##      1.27840      0.35369      0.03171      0.00092      0.01114

print(vcov(m00, type="se", robust=TRUE), digits=2)

##  (Intercept)      lboopen I(lboopen^2) I(lboopen^3)  (Intercept)
##       2.0171       0.5336       0.0455       0.0013       0.0135
```

The polynomial correlation matrix

```
library(corrplot)
op<-par(mfrow=c(1,2))
corrplot(vcov(m00, type="cor"))
corrplot(vcov(m0, type="cor"))
par(op)
```

The polynomial correlation matrix

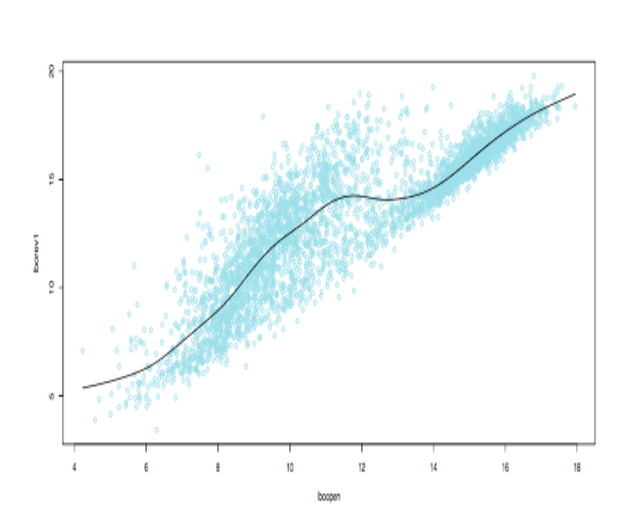


gamlss

Fitting a non-parametric smoothing model using pb()

```
m1<-gamlss(lborev1~pb(lboopen), data=film90, family=NO)

## GAMLSS-RS iteration 1: Global Deviance = 14085.78
## GAMLSS-RS iteration 2: Global Deviance = 14085.78
```

Plotting the fitted pb() model



gamlss

Summary of the fitted pb() model

```
## -------------------------------------------------------------------
## Mu link function:  identity
## Mu Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 2.352834   0.086899   27.08   <2e-16 ***
## pb(lboopen) 0.928404   0.007137  130.08   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## -------------------------------------------------------------------
## Sigma link function:  log
## Sigma Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.32824    0.01114   29.47   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## -------------------------------------------------------------------
## NOTE: Additive smoothing terms exist in the formulas:
##  i) Std. Error for smoothers are for the linear effect only.
## ii) Std. Error for the linear terms maybe are not accurate.
## -------------------------------------------------------------------
```

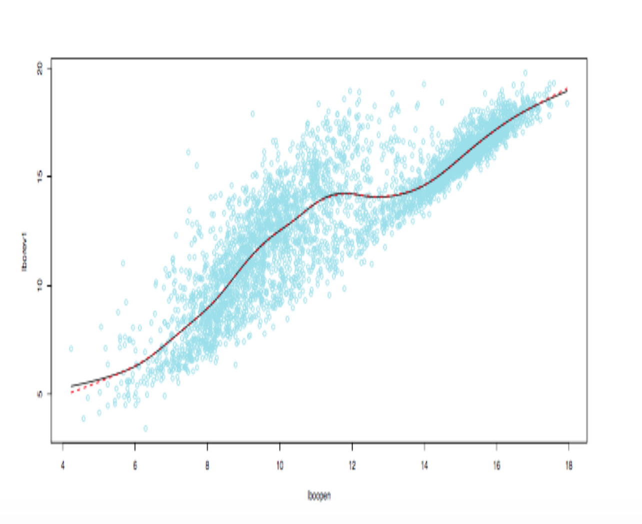
Fitting a non-parametric smoothing model using cs()

```
m2<-gamlss(lborev1~cs(lboopen,df=10), data=film90, family=NO)
```

```
## GAMLSS-RS iteration 1: Global Deviance = 14087.15
## . . .
## GAMLSS-RS iteration 2: Global Deviance = 14087.15
```

gamlss

Plotting the fitted pb() and cs() models



gamlss

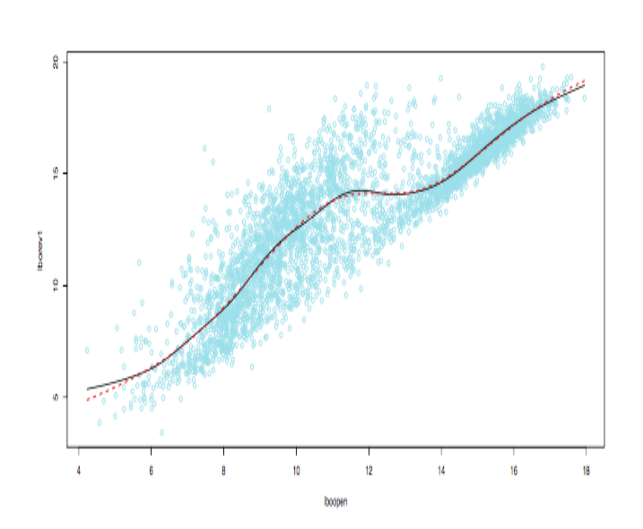
Fitting a non-parametric smoothing model using nn()

```
library(gamlss.add)
mnt <- gamlss(lborev1~nn(~lboopen,size=20, decay=0.1), data=film90, family=NO)
```

```
## GAMLSS-RS iteration 1: Global Deviance = 14166.07
## . . .
## GAMLSS-RS iteration 2: Global Deviance = 14108.85
```

gamlss

Plotting the fitted pb() and nn() models



gamlss

Extracting the fitted values for σ

```
fitted(m1,"sigma")[1]
```

```
##        1
## 1.388527
```

where [1] indicates the first value of the vector. The same values can be obtained using the more general function `predict()`:

```
predict(m1,what="sigma", type="response")[1]
```

```
##        1
## 1.388527
```

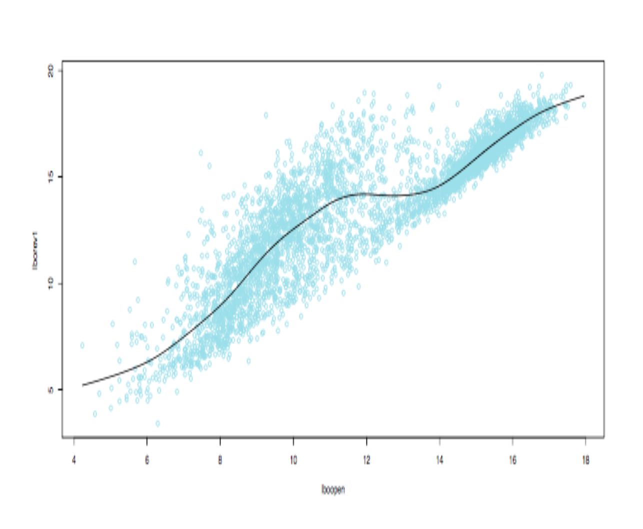
Modelling both μ and σ

```
m3 <- gamlss(lborev1~pb(lboopen),sigma.formula=~pb(lboopen),
data=film90, family=NO)
edfAll(m3)
```

```
## GAMLSS-RS iteration 1: Global Deviance = 12224
## . . .
## GAMLSS-RS iteration 4: Global Deviance = 12227
## $mu
## pb(lboopen)
##       12.41
##
## $sigma
## pb(lboopen)
##       10.96
```

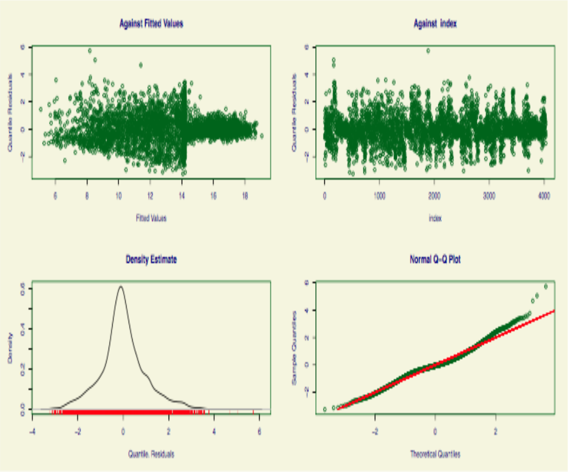
gamlss

Plot of the fitted μ and σ model

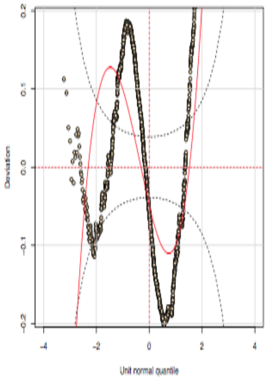


gamlss

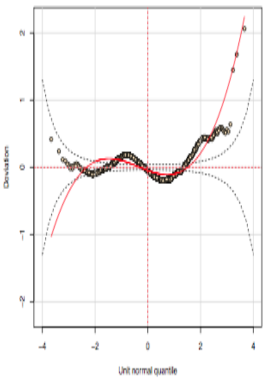
Diagnostic plots using plot()



Diagnostic plots using wp()



(a)



(b)

Fitting different distributions

```
m5 <-gamlss(lborev1~pb(lboopen), sigma.formula=~pb(lboopen),
            nu.formula=~pb(lboopen),
            data=film90, family=BCCG)
```

```
## GAMLSS-RS iteration 1: Global Deviance = 11845
## . . .
## GAMLSS-RS iteration 17: Global Deviance = 11769
```

To fit the Box-Cox Power Exponential (BCPE) a 4-parameter distribution) try:

```
m6 <-gamlss(lborev1~pb(lboopen), sigma.formula=~pb(lboopen),
            nu.formula=~pb(lboopen), tau.formula=~pb(lboopen),
            data=film90, start.from=m5, family=BCPE)
```

```
## GAMLSS-RS iteration 1: Global Deviance = 11699
## . . .
## GAMLSS-RS iteration 19: Global Deviance = 11696
```

Selection between models: AIC

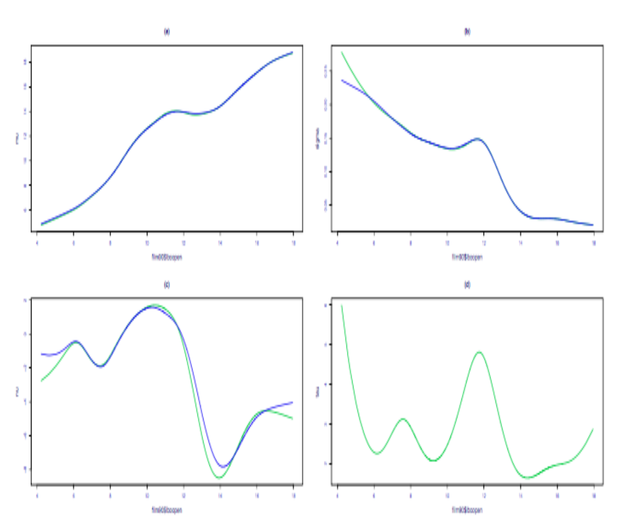
```
AIC(m0,m1,m2,m3,m4,m5,m6)

##              df      AIC
## m6  45.82525 11787.30
## m5  36.84390 11842.87
## m3  23.37377 12273.40
## m4  18.04177 12286.01
## m1  13.46241 14112.70
## m2  13.00200 14113.16
## m0   5.00000 14527.65
```

gamlss

Selection between models: SBC

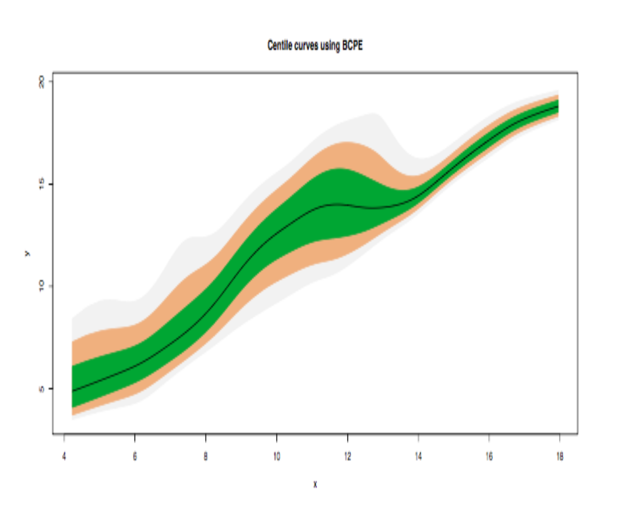
```
AIC(m0,m1,m2,m3,m4,m5,m6, k=log(4031))

##             df      AIC
## m5 36.84390 12075.05
## m6 45.82525 12076.08
## m4 18.04177 12399.71
## m3 23.37377 12420.69
## m2 13.00200 14195.09
## m1 13.46241 14197.54
## m0  5.00000 14559.15
```

Fitted values for m5 and m6 model

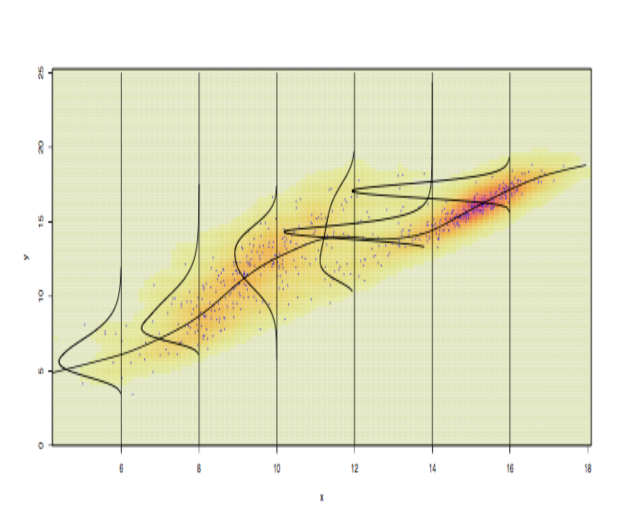


gamlss

Centile fan plot of model m6



Conditional distribution plot of model m6



gamlss

End

END

for more information see

www.gamlss.org

gamlss