Beautiful graphs with npde 3.0

Février 2021

Emmanuelle Comets, Romain Leroux

Contributors to npde: Karl Brendel, Marc Cerou, Romain Leroux, Thi Huyen Tram Nguyen, France Mentré

INSERM, IAME UMR 1137, Paris, France

Université de Paris, Paris, France.

Npde website:`www.npde.biostat.fr`

Contents

1 Running npde

```
> # virload20
> yvir20 <- autonpde(virload20,simvirload,iid=1,ix=2,iy=3, icens=4, boolsave=
    FALSE)
> # theopp
> theofit1 <- autonpde(theopp,simtheopp,1,3,4,boolsave=FALSE, units=list(x="hr"
    ,y="mg/L"))
> # warfarine
> wbase<-autonpde(namobs=warfarin,namsim=simwarfarinBase, iid=1,ix=2,iy=4,icov=
    c(3,6:8),namsav="warfBase", units=list(x="hr",y="ug/L", covariates=c("mg","
    kg","-","yr")))
> wcov<-autonpde(namobs=warfarin,namsim=simwarfarinCov, iid=1,ix=2,iy=4,icov=c
    (3,6:8),namsav="warfCov", units=list(x="hr",y="ug/L", covariates=c("mg","kg
    ","-","yr")))
```

```
> yvir20
Object of class NpdeObject
------------------------------------------
----        Component data            ----
------------------------------------------
Object of class NpdeData
    Structured data: Log_VL ~ Time | ID
This object has the following components:
     data: data
     with 50 subjects
      300 observations
The data has the following components
     X: Time ()
     Y: Log_VL ()
     individual model predictions: ipred
     missing data: mdv  (1=missing)
     censored data: cens  (1=censored)
      LOQ:     1.3
------------------------------------------
----        Component results         ----
------------------------------------------
Object of class NpdeRes
  containing the following elements:
    predictions (ypred)
    prediction discrepancies (pd)
    normalised prediction distribution errors (npde)
    completed responses (ycomp) for censored data
    decorrelated responses (ydobs)
    ploq: probability of being <LOQ for each observation
  the dataframe has  300 non-missing observations .
First 10 lines of results, removing missing observations:
```

```
       ypred ycomp    pd        npde       ydobs     tnpde
1  4.4099809  3.95 0.307 -0.50437199 -0.50389104 1.4356698
2  2.4975561  1.91 0.229 -0.55923698 -0.57968771 1.3609386
3  1.9229180  1.89 0.493  1.11232137  1.03971370 3.6377559
4  1.6290177  1.69 0.538  0.37454350  0.39208593 2.6328341
5  1.3697648  1.63 0.640  0.48736457  0.47523309 2.7865068
6  0.8601327  1.55 0.821  1.19011804  0.95569947 3.7437222
7  4.3566026  4.42 0.534  0.08532879  0.06968412 2.2388969
8  2.4636655  2.55 0.579  0.06521854  0.09265785 2.2115048
9  1.9063313  1.48 0.265 -1.35317415 -1.37472078 0.2795226
10 1.6029737  1.48 0.424  0.78236516  0.66459383 3.1883250
> theofit1
Object of class NpdeObject
------------------------------------------
----        Component data            ----
------------------------------------------
Object of class NpdeData
    Structured data: Conc ~ Time | ID
This object has the following components:
     data: data
     with 12 subjects
      120 observations
The data has the following components
     X: Time (hr)
     Y: Conc (mg/L)
     missing data: mdv  (1=missing)
------------------------------------------
----        Component results         ----
------------------------------------------
Object of class NpdeRes
  containing the following elements:
    predictions (ypred)
    prediction discrepancies (pd)
    normalised prediction distribution errors (npde)
    completed responses (ycomp) for censored data
    decorrelated responses (ydobs)
  the dataframe has  120 non-missing observations and 132 lines.
First 10 lines of results, removing missing observations:
      ypred ycomp    pd       ydobs      npde     tnpde
2  2.923864  2.84 0.550 -0.05124648 0.1256613  5.403817
3  4.682299  6.57 0.850  1.96398150 2.0537489 11.302178
4  6.264357 10.50 0.990  2.56602650 2.3263479 12.136107
5  6.986255  9.66 0.980  0.41616411 0.5244005  6.623631
6  6.511039  8.58 0.930  0.28430866 0.2533471  5.794430
7  5.895675  8.36 0.960  0.54879386 0.6744898  7.082780
8  5.064736  7.47 0.970  1.79335938 1.6448536 10.051295
9  4.302909  6.89 0.990  0.80506269 0.7721932  7.381673
10 3.294020  5.94 0.995  1.91537662 1.7506861 10.375055
```

```
11 1.168743  3.28 0.995  3.25535923 2.5758293 12.899315
> wbase
Object of class NpdeObject
----------------------------------------
----        Component data           ----
----------------------------------------
Object of class NpdeData
    Structured data: dv ~ time | id
    Covariates: amt wt sex
This object has the following components:
     data: data
     with 32 subjects
      247 observations
The data has the following components
     X: time (hr)
     Y: dv (ug/L)
     missing data: mdv  (1=missing)
----------------------------------------
----        Component results        ----
----------------------------------------
Object of class NpdeRes
  containing the following elements:
    predictions (ypred)
    prediction discrepancies (pd)
    normalised prediction distribution errors (npde)
    completed responses (ycomp) for censored data
    decorrelated responses (ydobs)
  the dataframe has  247 non-missing observations .
First 10 lines of results, removing missing observations:
        ypred ycomp    pd       ydobs       npde
1   0.5814627   0.0 0.359 -0.43862113 -0.3611330
2   3.1220995   1.9 0.454 -0.05116685  0.1636585
3   8.4386880   3.3 0.126 -1.35607837 -1.4985131
4  11.2936700   6.6 0.140  0.04661369  0.1560419
5  12.6249280   9.1 0.127 -0.56691520 -0.5417366
6  11.7504645  10.8 0.355  0.89677342  0.9230138
7  10.8386881   8.6 0.141 -1.08057832 -1.1358962
8   8.5881834   5.6 0.026 -1.89143682 -1.8521799
9   7.0369975   4.0 0.019 -1.06275059 -1.0278933
10  5.7174611   2.7 0.017 -0.60300349 -0.5947658
         tnpde
1   4.79538371
2   6.98209215
3   0.05613313
4   6.95035522
5   4.04284249
6  10.14618430
7   1.56708994
```

```
8  -1.41753046
9   2.01711779
10  3.82187933
> wcov
Object of class NpdeObject
----------------------------------------
----        Component data           ----
----------------------------------------
Object of class NpdeData
    Structured data: dv ~ time | id
    Covariates: amt wt sex
This object has the following components:
     data: data
     with 32 subjects
      247 observations
The data has the following components
     X: time (hr)
     Y: dv (ug/L)
     missing data: mdv  (1=missing)
----------------------------------------
----        Component results        ----
----------------------------------------
Object of class NpdeRes
  containing the following elements:
    predictions (ypred)
    prediction discrepancies (pd)
    normalised prediction distribution errors (npde)
    completed responses (ycomp) for censored data
    decorrelated responses (ydobs)
  the dataframe has  247 non-missing observations .
First 10 lines of results, removing missing observations:
        ypred ycomp    pd       ydobs       npde      tnpde
1   0.3746216   0.0 0.398 -0.34893746 -0.2585273  5.2792639
2   2.6728119   1.9 0.500 -0.03334944  0.1535051  6.8242576
3   8.0288884   3.3 0.115 -1.43486957 -1.5300676  0.5113814
4  10.9755216   6.6 0.106 -0.32327872 -0.2845355  5.1817410
5  12.2572079   9.1 0.063 -0.96158402 -0.9944579  2.5197520
6  11.3470900  10.8 0.381  0.71955362  0.7421442  9.0314719
7  10.4732161   8.6 0.107 -1.05944210 -1.0581216  2.2810327
8   8.0557699   5.6 0.026 -1.82921396 -1.9268366 -0.9763794
9   6.3939722   4.0 0.030 -0.99871382 -0.9862713  2.5504491
10  5.0984787   2.7 0.026 -0.69323563 -0.6588377  3.7782238
```

2 Types of graphs

Each of the default diagnostic plots, as well as a number of additional plots not shown by default, can also be produced on its own, using the argument `plot.type="type"`. Table 1 lists the plots that can be created in this way.

Plot type	Description
data	Plots the observed data in the dataset
hist	Histogram of the npde
qqplot	QQ-plot of the npdeversus its theoretical distribution
ecdf	Empirical distribution function of the npde
x.scatter	Scatterplot of the npde versus the predictor X
pred.scatter	Scatterplot of the npde versus the population predicted values
cov.scatter	Scatterplot of the npde versus covariates
vpc	Plots a Visual Predictive Check (VPC)
loq	Plots the probability for an observation to be BQL, versus the predictor X

Table 1: Plot types available in the npde library. QQ-plots, histograms, cumulative cdf, and scatter plots can be produced for npde, pd or npd.

This function for plotting is:

```
> plot(x,plot.type="data")
```

The first argument, x, is a NpdeObject and plot.type argument is used to specify the type of graph.

3 Changing graphical parameters

Default layout for graphs in the npde library can be modified through the use of many options. An additional document, demo npde2.0.pdf, is included in the inst directory of the package, presenting additional examples of graphs and how to change the options. Tables Graphical parameters that can be passed on the plot function: titles and axes.,Graphical parameters that can be passed on the plot function: colours and symbols. Graphical parameters that can be passed on the plot function: colours and symbols.,Graphical options for VPC and residual plots.,Boolean parameters that can be passed on the plot function., following table shows the options that can be set, either by specifying them on the fly in a call to plot applied to a NpdeObject object, or by storing them in the prefs component of the object. Note that not all of the graphical parameters in par() can be used, but it is possible for instance to use the xaxt="n" option below to suppress plotting of the X-axis, and to then add back the axis with the R function axis() to tailor the tickmarks or change colours as wanted. It is also possible of course to extract npde, fitted values or original data to produce any of these plots by hand if the flexibility provided in the library isn't sufficient.

Argument	Description	Default value
verbose	Output is produced for some plots (most notably when binning is used, this prints out the boundaries of the binning intervals) if TRUE	FALSE
main	Title	depends on plot
sub	Subtitle	empty
size.main	Size of the main title	14
size.sub	Size of the title for covariate	12
xlab	Label for the X-axis	depends on plot
ylab	Label for the Y-axis	depends on plot
size.xlab	Size of the label for the X-axis	12
size.ylab	Size of the label for the Y-axis	12
breaks.x	Number of tick marks on the X-axis	10
breaks.y	Number of tick marks on the Y-axis	10
size.x.text	Size of tick marks and tick labels on the X-axis	10
size.y.text	Size of tick marks and tick labels on the Y-axis	10
xlim	Range of values on the X-axis	empty, adjusts to the data
ylim	Range of values on the Y-axis	empty, adjusts to the data
xaxt	A character whether to plot the X axis. Specifying "n" suppresses plotting of the axis	"y"
yaxt	A character whether to plot the Y axis. Specifying "n" suppresses plotting of the axis	"y"
xlog	Scale for the X-axis (TRUE: logarithmic scale)	FALSE
ylog	Scale for the Y-axis (TRUE: logarithmic scale)	FALSE

Table 2: Graphical parameters that can be passed on the plot function: titles and axes.

Argument	Description	Default value
col	Main colour for observed data (applied to lines and symbols pertaining to observations if no other option is given to supersede this value)	"slategray4"
lty	Line type for observed data	1
lwd	Line width for observed data	0.5
pch	Symbol used to plot observed data	20
alpha	Transparencyfor observed data	1
size	Symbol size to plot observed data	1
fill	Colour used to fill area elements related to observed data (such as histogram bars)	"white
fill.outliers.med	Color for the outliers of the median confidence interval	"red"
fill.outliers.bands	Color for the outliers of the bounds of the confidence interval	"red"
alpha.outliers.med	Transparency of the color for the outliers of the median confidence interval	1
alpha.outliers.bands	Transparency of the color for the outliers the bounds of the confidence interval	1
col.bands	Colour for the lines of the bounds of the confidence interval	"steelblue4"
lty.bands	Type for the lines of bounds of the confidence interval	2
lwd.bands	Width of the lines of bounds of the confidence interva	0.25
alpha.bands	Transparency of the bounds of the confidence interval	"0.3"
fill.bands	Colour of the confidence interval	"steelblue2"
col.med	Colour for the lines of the median of the confidence interval	"salmon4"
lty.med	Type for the lines of the median of the confidence interval	2
lwd.med	Width of the lines of the median of the confidence interva	0.5
alpha.med	Transparency of of the median confidence interval	"0.5"
fill.med	Colour of the median confidence interval	"pink"
type	Type for the line for qqplot and scatter. Display line and points.	"b"

Table 3: Graphical parameters that can be passed on the plot function: colours and symbols.

Argument	Description	Default value
col.pobs	Colour for observed data	"slategray4"
pch.pobs	Symbol used to plot observed data	20
size.pobs	Symbol size to plot observed data	1.5
alpha.pobs	Transparencyfor observed data	1.5
col.lobs	Colour for the line of observed data	"slategray4"
lty.lobs	Line type for the line of observed data	20
lwd.lobs	Line width for the line of observed data	1.5
col.pcens	Colour for the censored data	"red"
pch.pcens	Symbol for the censored data	8
size.pcens	Symbol size for the censored data	1.2
alpha.pcens	Transparency for the censored data	1
col.line.loq	Colour for the LOQ line	"black"
lty.line.loq	Symbol type for the LOQ line	5
lwd.line.loq	Symbol size for the LOQ line	0.5

Table 4: Graphical parameters that can be passed on the plot function: colours and symbols.

Graphical options for VPC and residual plots		
Parameter	**Description**	**Default value**
`bands`	Whether prediction intervals should be plotted	TRUE
`approx.pi`	If TRUE, samples from $\mathcal{N}(0,1)$ are used to plot prediction intervals, while if FALSE, prediction bands are obtained using pd/npde computed for the simulated data	TRUE
bin.method	Method used to bin points (one of "equal", "width", "user" or "optimal"); at least the first two letters of the method need to be specified	"equal"
`bin.number`	Number of binning intervals	10
`vpc.interval`	Size of interval	0.95
`bin.breaks`	Vector of breaks used with user-defined breaks (vpc.method="user")	NULL
`bin.extreme`	Can be set to a vector of 2 values to fine-tune the behaviour of the binning algorithm at the boundaries; specifying c(0.01,0.99) with the "equal" binning method and vpc.bin=10 will create 2 extreme bands containing 1% of the data on the X-interval, then divide the region within the two bands into the remaining 8 intervals each containing the same number of data; in this case the intervals will all be equal except for the two extreme intervals, the size of which is fixed by the user; complete fine-tuning can be obtained by setting the breaks with the vpc.method="user"	NULL
`pi.size`	Width of the prediction interval on the quantiles	0.95
`bin.lambda`	Value of lambda used to select the optimal number of bins through a penalised criterion	0.3
`bin.beta`	Value of beta used to compute the variance-based criterion (Jopt,beta(I)) in the clustering algorithm	0.2
`bands.rep`	Number of simulated datasets used to compute prediction bands	200

Table 5: Graphical options for VPC and residual plots.

Argument	Description	Default value
plot.box	If TRUE, boxplots are produced instead of scatterplots	FALSE
covsplit	If TRUE, plot are split by covariates	FALSE
plot.loq	If TRUE, data under the LOQ are plotted	FALSE
line.loq	If TRUE, horizontal line should is plotted at Y=LOQ in data and VPC plots	FALSE
impute.loq	If TRUE, the imputed values are plotted for data under the LOQ	FALSE
plot.obs	If TRUE, observations, pd/ndpe should are plotted on top of the prediction bands	TRUE
grid	If TRUE, display a grid on the background of the plot	FALSE
box	If TRUE, boxplots are be performed instead of scatterplots (see documentation)	FALSE

Table 6: Boolean parameters that can be passed on the plot function.

4 Default plots

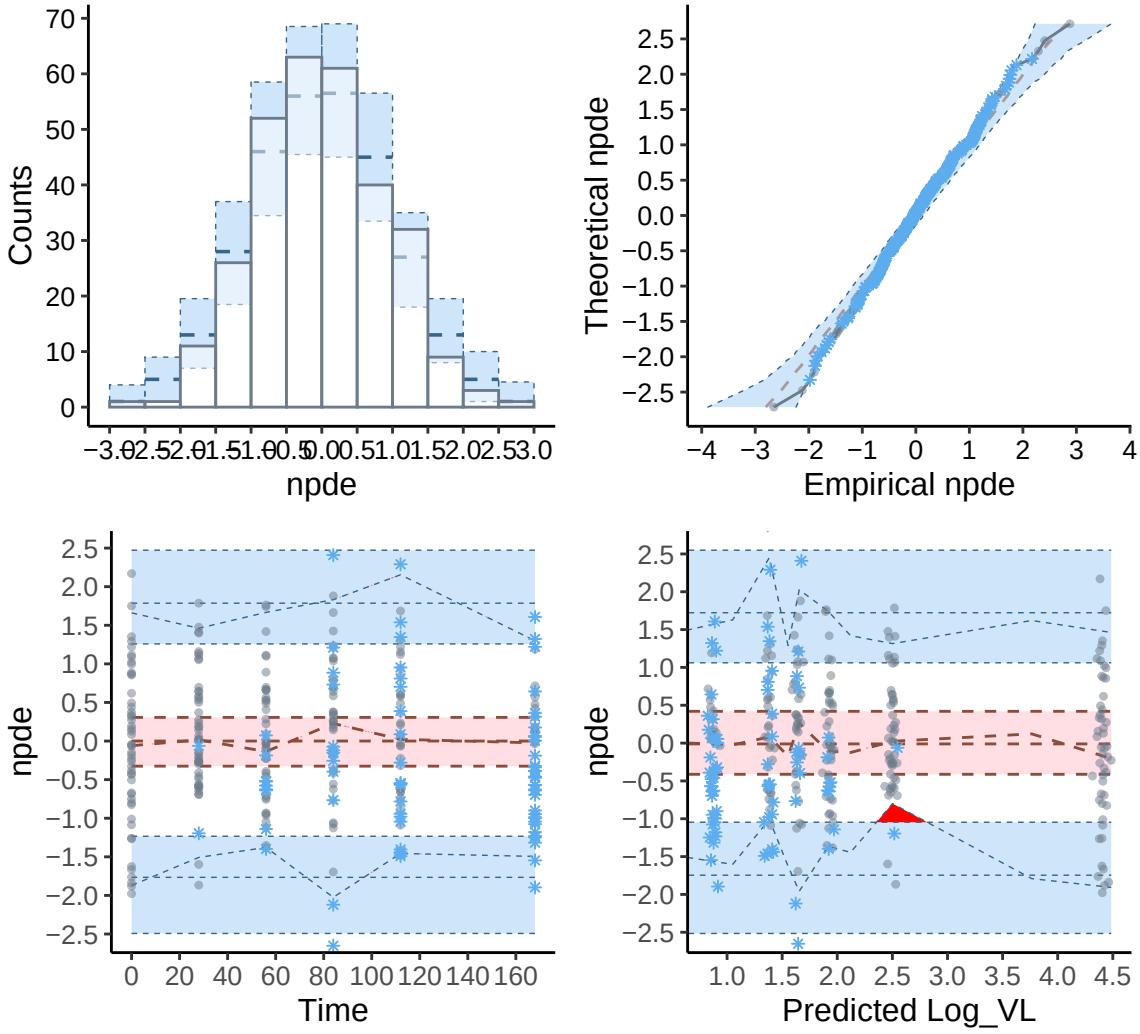
4.1 Waffle plot

By default, the package produces and saves the four graphs:

1. a quantile-quantile plot: plot of the npde versus the corresponding quantiles of a normal distribution
   - the line y = x is also drawn
2. a histogram of the npde
   - the shape of the normal distribution $\mathcal{N}(0,1)$ is also shown
3. a plot of the npde versus the independent variable X
4. a plot of the npde versus ypred
   - for these last two graphs, we plot the lines corresponding to y = 0 and to the critical values 5% and 95% (delimiting the 90% confidence interval in which we expect to nd the bulk of the npde).

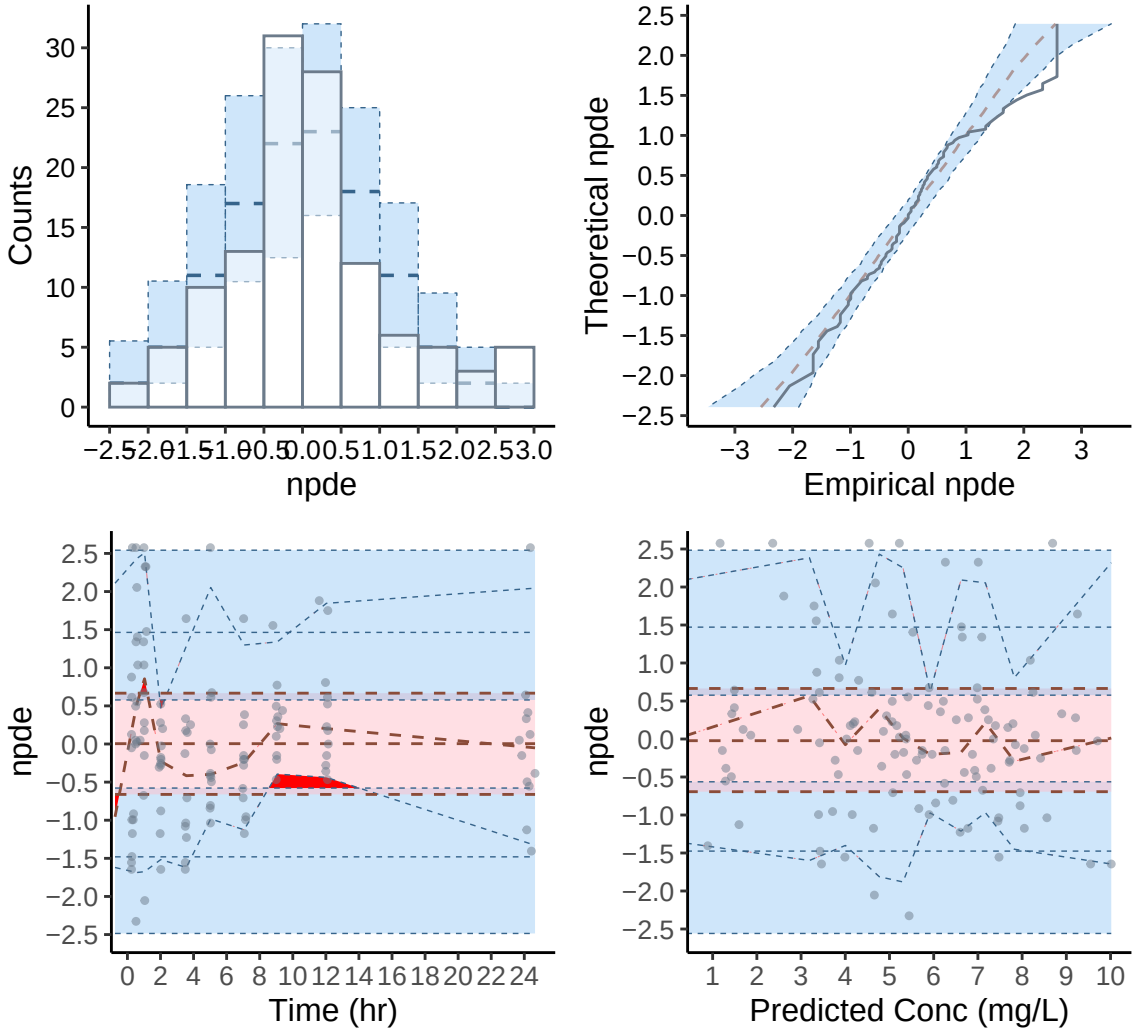
```
> plot(yvir20)
```

Figure 2: Default plot for yvir20.



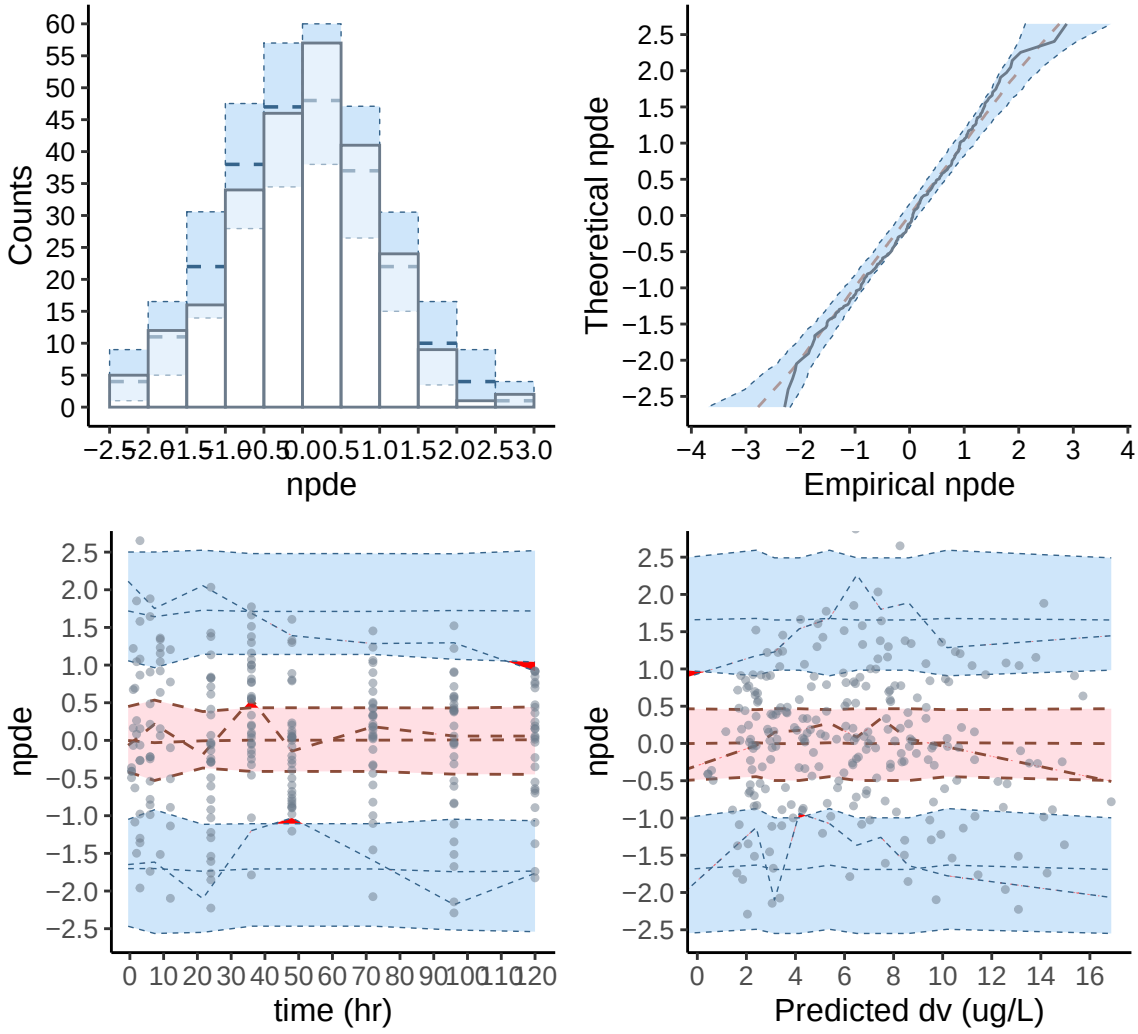

```
> plot(theofit1)
```

Figure 3: Default plot for theofit1.



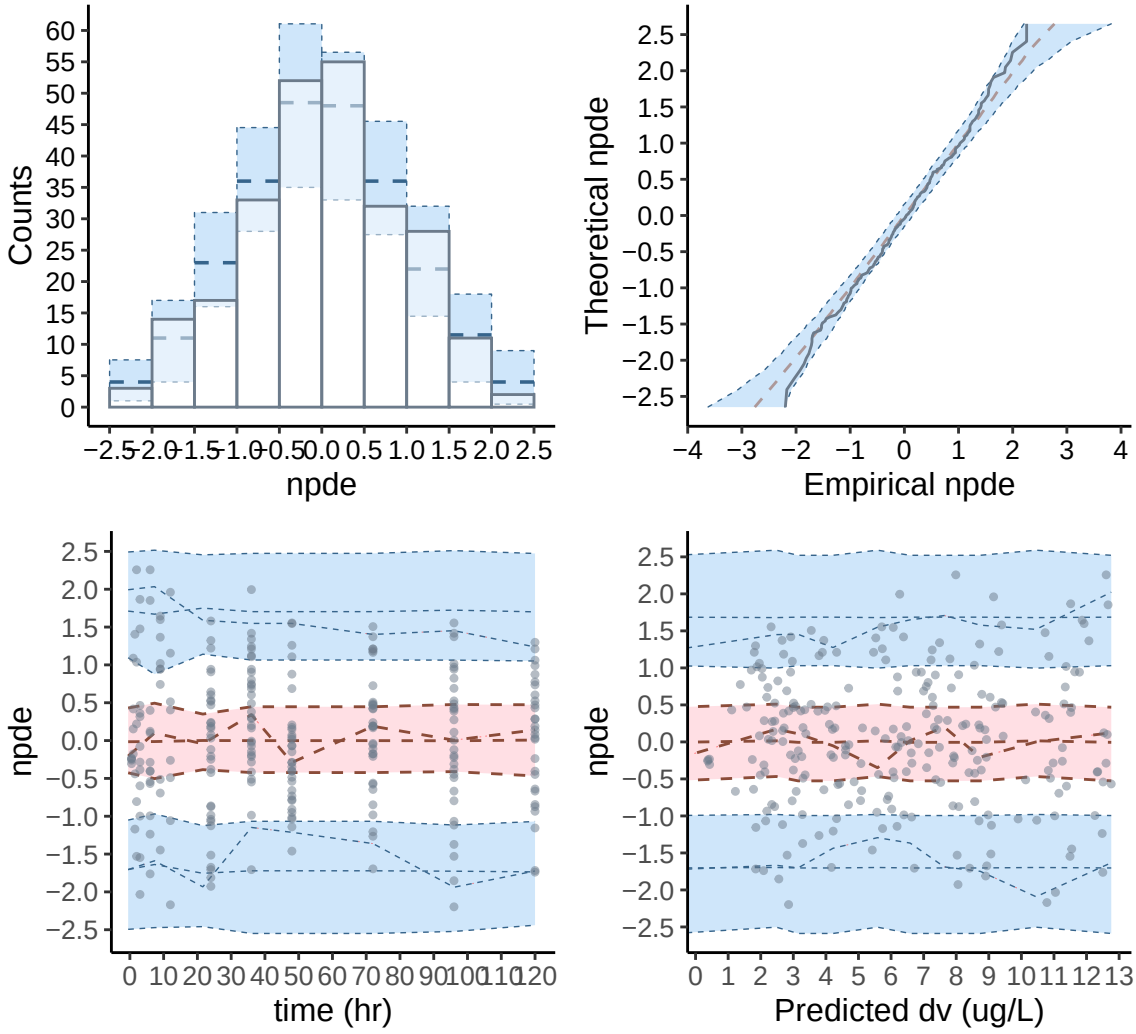

```
> plot(wbase)
```

Figure 4: Default plot for wbase.



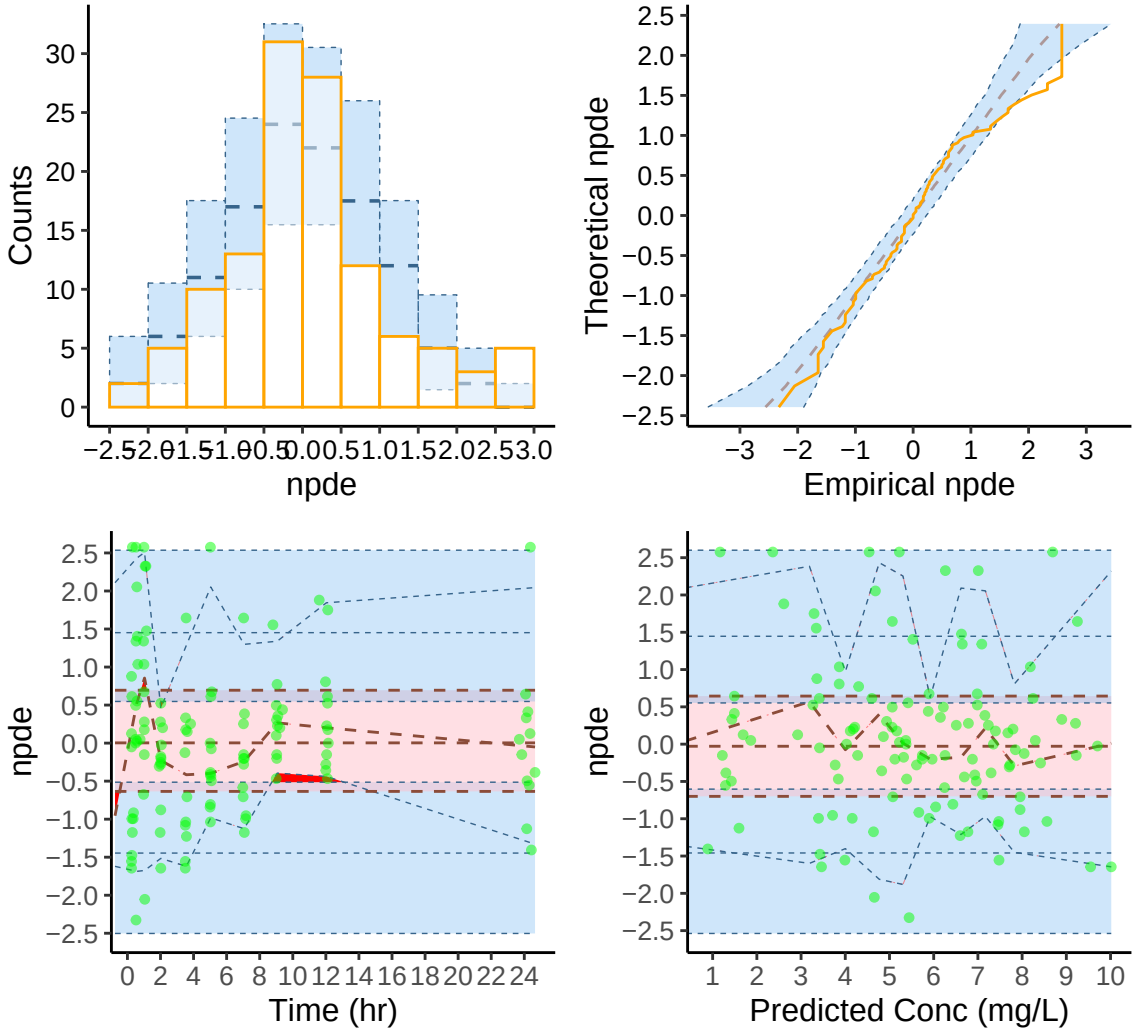

```
> plot(wcov)
```

Figure 5: Default plot for wcov.




```
> plot(theofit1,col.pobs="green",pch.pobs=3,size.pobs=2)
```

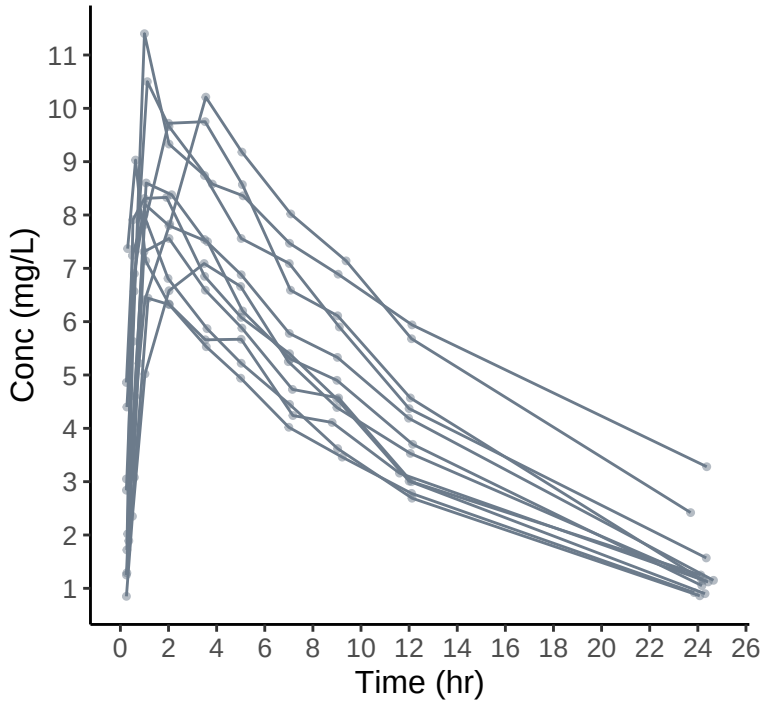
Figure 6: Default plot for theofit1 with user defined options for the observations.



4.2 Data

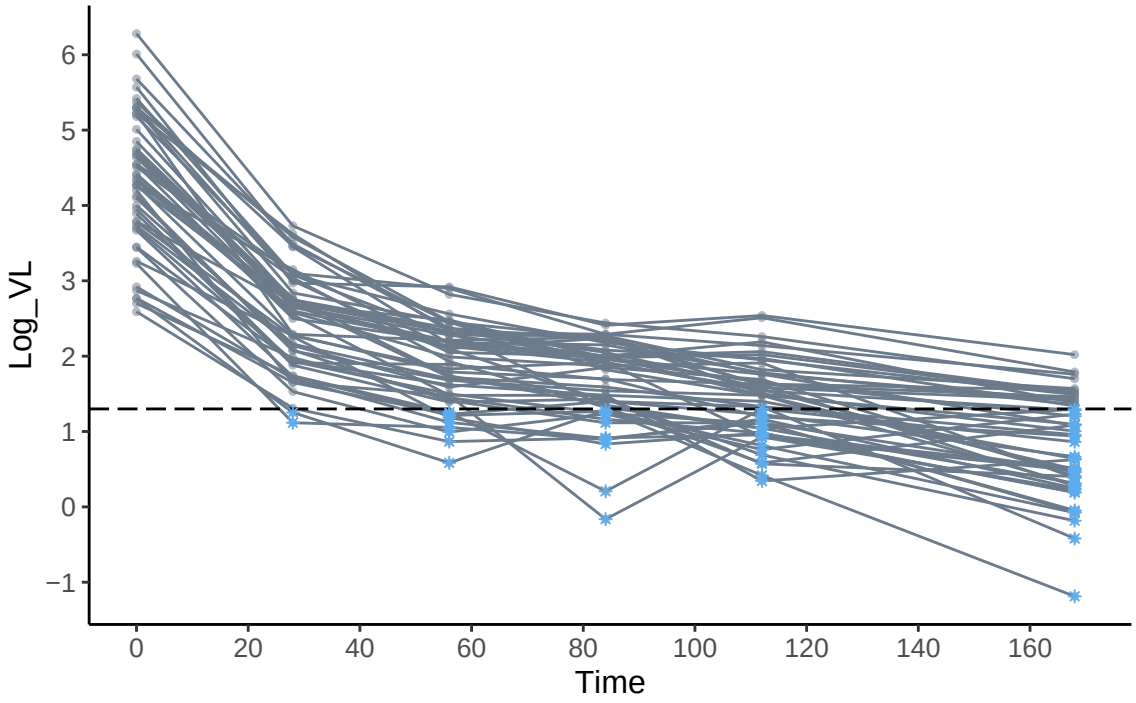
```
> plot(theofit1,plot.type="data")
```

Figure 7: Default plot for theofit1 - data.




```
> plot(yvir20,plot.type="data",plot.loq=TRUE,line.loq=TRUE)
```

Figure 8: Default plot for yvir20 - data.



4.3 VPC (Visual Predictive Check)

```
> plot(theofit1,plot.type="vpc")
> plot(theofit1, plot.type="vpc", plot.box=TRUE)
```

Figure 9: Default plot for theofit1 - VPC.

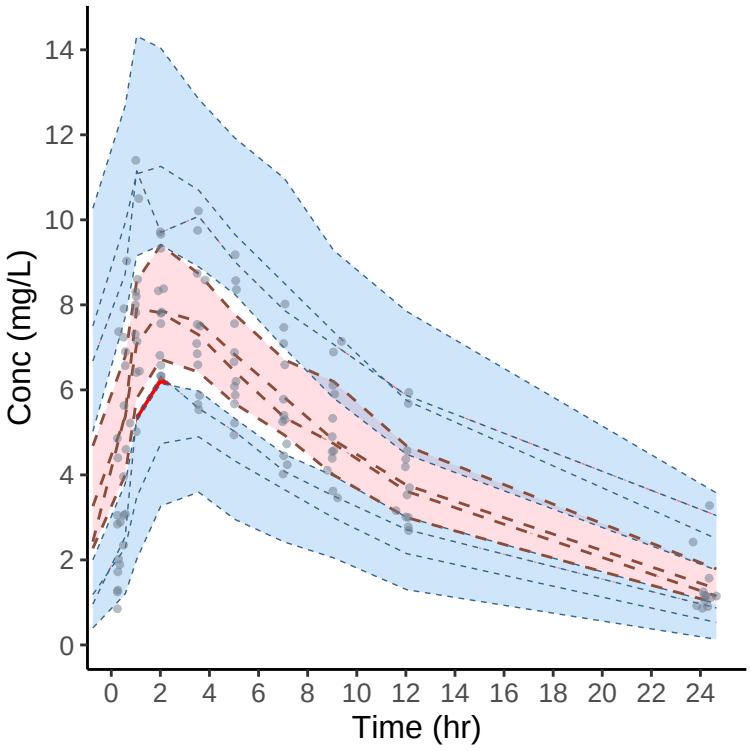
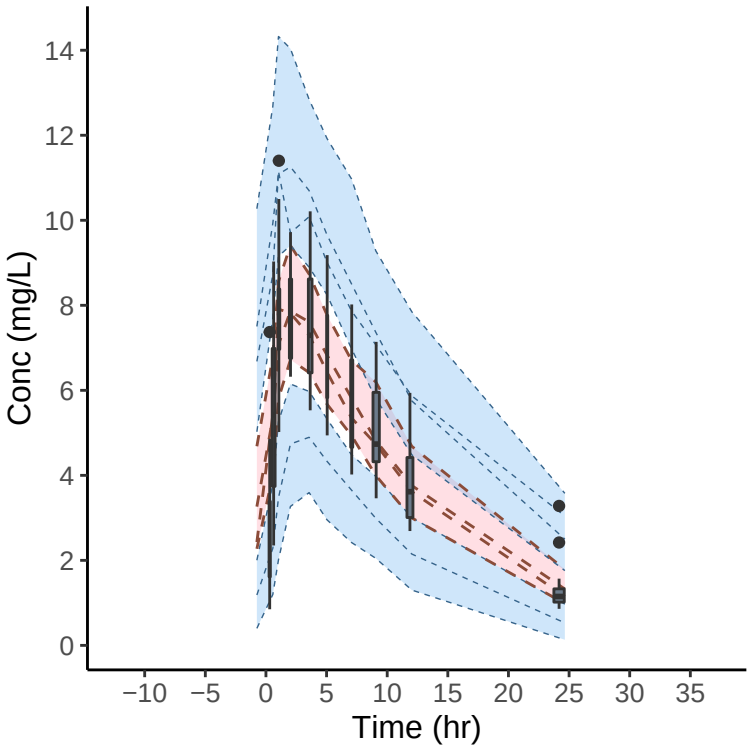


Figure 10: Default plot for theofit1 with boxplots.



4.4 Scatterplots: scatterplots of npde versus X or predictions

```
> plot(theofit1,plot.type="x.scatter")
> plot(theofit1,plot.type="pred.scatter")
```

Figure 11: Default plot for theofit1 - scatterplot of npde vs X.

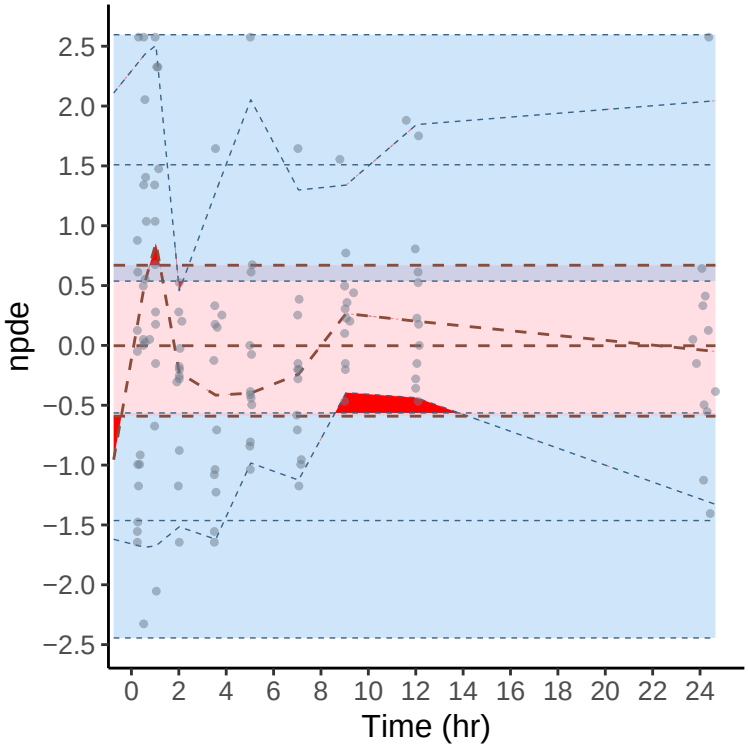
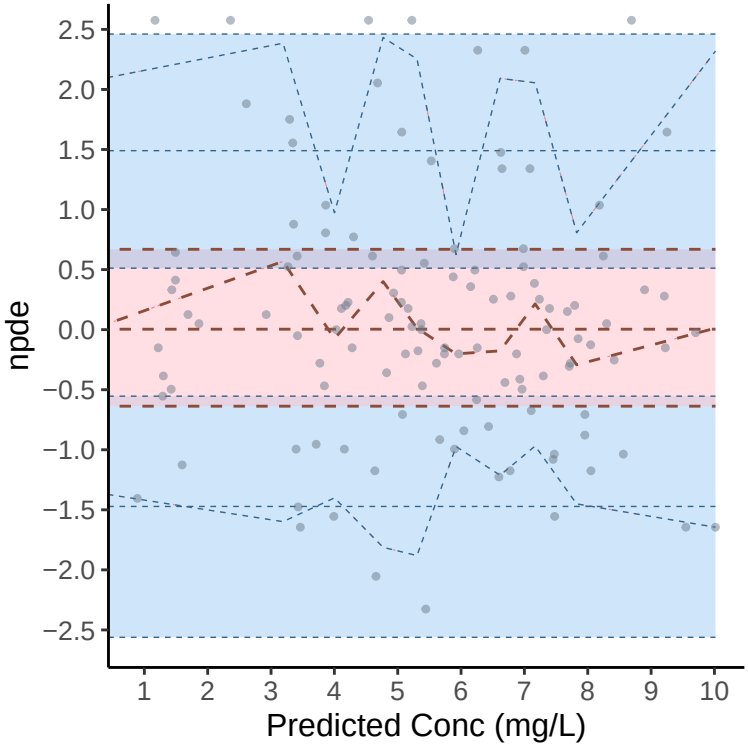
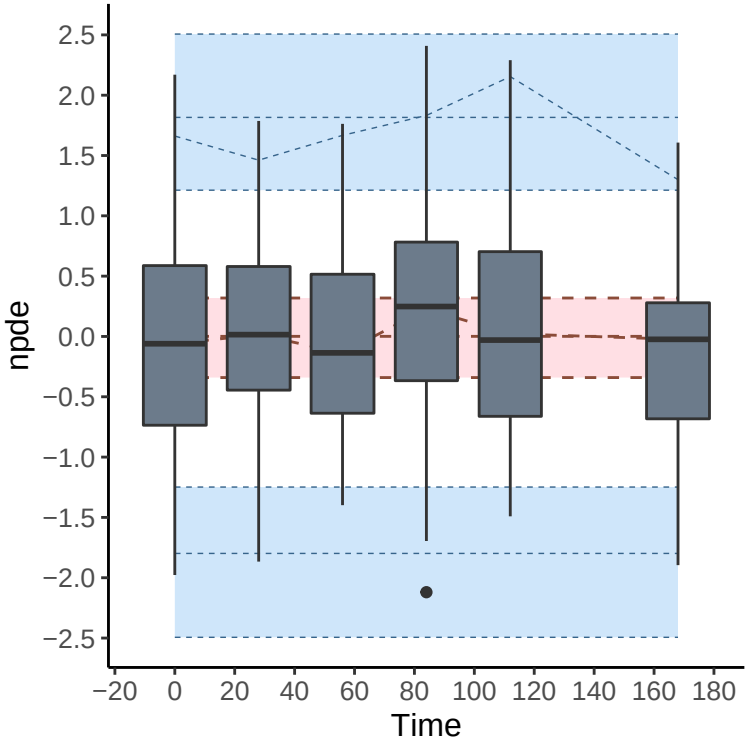


Figure 12: Default plot for theofit1 - scatterplot of npde vs predictions.




```
> plot(yvir20 , plot.type="x.scatter", plot.box=TRUE)
```

Figure 13: Default plot for theofit1 - scatterplot of npde vs X with boxplot.

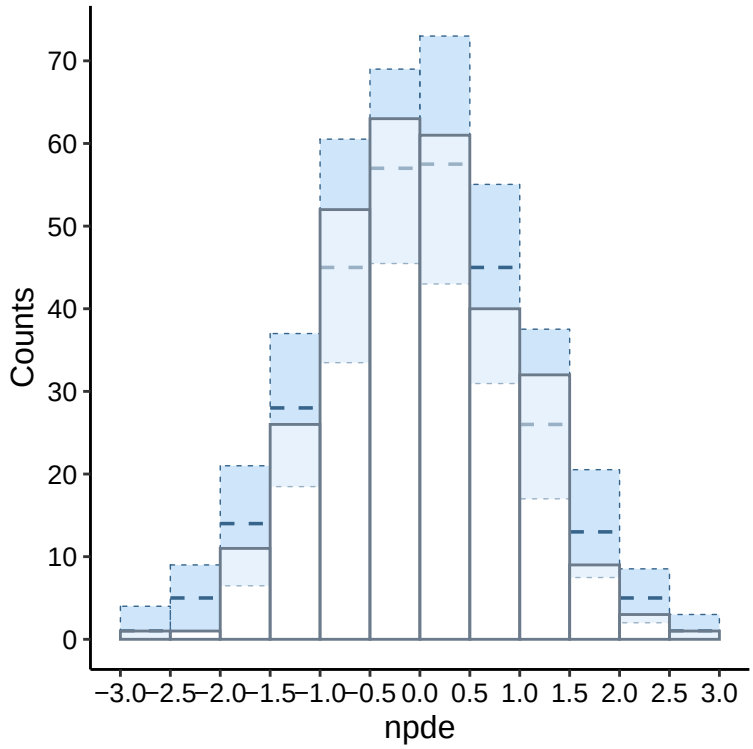


5 Histogram, QQ-plot and Ecdf plots.

5.1 Histogram

```
> plot(yvir20,plot.type="hist")
```

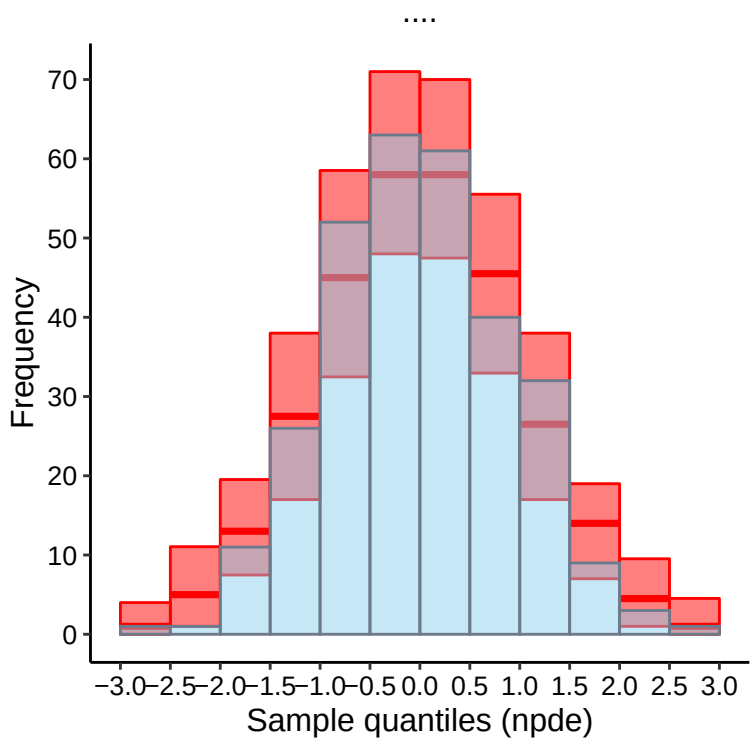
Figure 14: Histogram plot for yvir20.




```
> plot(yvir20,plot.type="hist",which="npde",
      main = "Title",
      size.main = 14,
      sub = "",
      size.sub = "",
      xlab= "Sample quantiles (npde)",
      ylab= "Frequency",
      size.xlab = 12, ## cex.axis = 12
      size.ylab = 12,
      #xlim=c(0,25),
      #ylim=c(-3,3.5),
      xlim=c(),
      ylim=c(),
      bands=TRUE, # TRUE, FALSE
      lty.grid="dotted",
      col = "green",
      fill = "skyblue",
      alpha = 0.95,
      lty = 2,
      lwd = 1,
      col.bands = "red",
      fill.bands = "red" ,
      alpha.bands = 0.5,
      lty.bands = 1,
      lwd.bands = 0.5,
      col.pobs = "darkblue",
```

```
      size.pobs = 2,
      size.text.x = 10,
      size.text.y = 10,
      breaks.x = 10,
      breaks.y = 10,
      xlog = FALSE,
      ylog = FALSE)
```

Figure 15: Histogram plot for yvir20 with user defined options.



5.2 QQ-plot

5.3 Ecdf plots

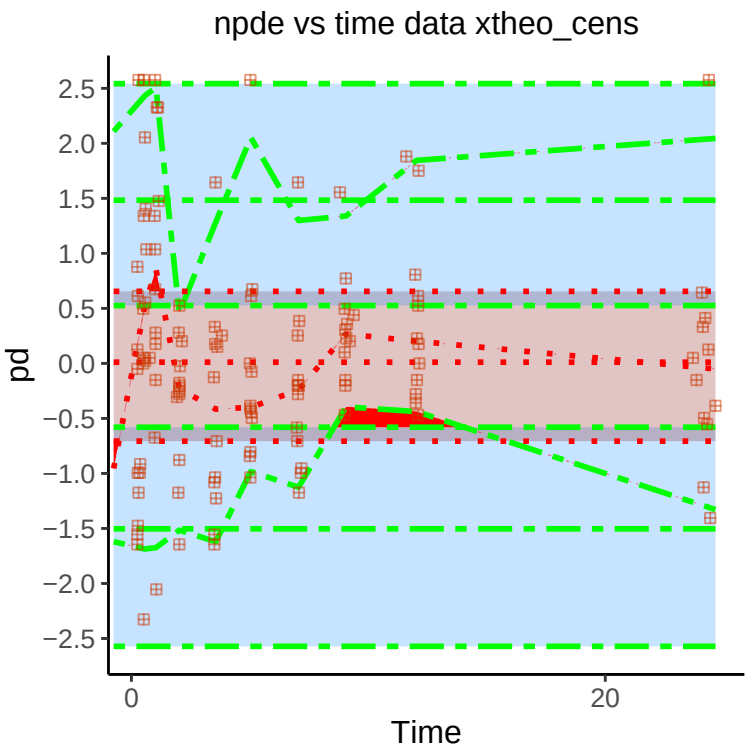
6 x.scatter and pred.scatter plots

6.1 x.scatter plots

```
> plot(theofit1, plot.type="x.scatter",

      main = "npde vs time data xtheo_cens",
      size.main = 14,
      sub = "",
      size.sub = "",

      xlab= "Time",
```

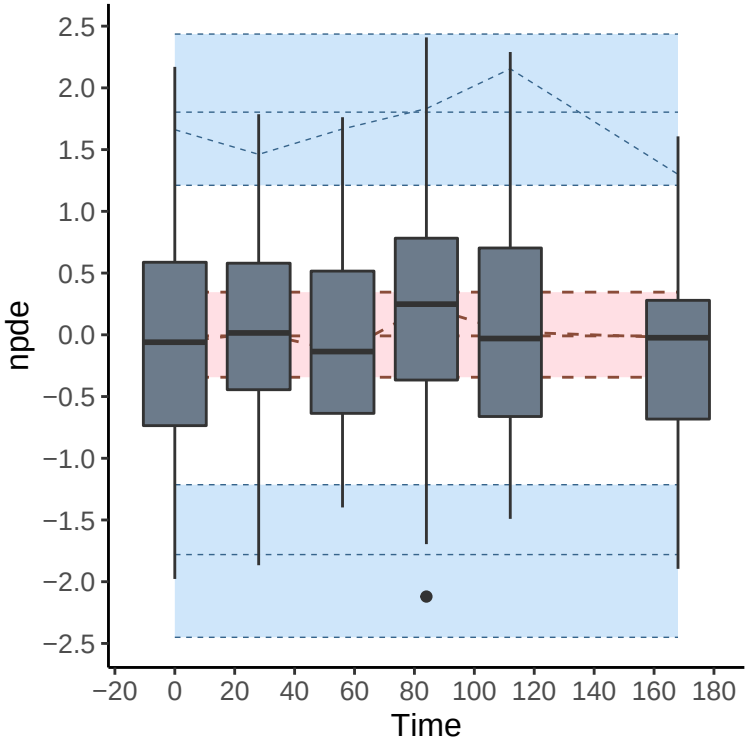
```
      ylab= "pd",
      size.xlab = 12, ## cex.axis = 12
      size.ylab = 12,
      #xlim=c(0,25),
      #ylim=c(-3,3.5),
      xlim=c(),
      ylim=c(),

      approx.pi=TRUE,
      bands=TRUE,
      plot.obs=TRUE,

      alpha.med = 0.25,
      fill.med = "firebrick4",
      col.med="red",
      lty.med=3,
      lwd.med=1,

      alpha.bands = 0.25,
      fill.bands = "dodgerblue",
      col.bands="green",
      lty.bands=6,
      lwd.bands=1,

      col.med = "red",
      lty.med = 1,
      lwd.med = 1,

      col.bands = "blue",
      lty.bands = 1,
      lwd.bands = 1,

      col.pobs = "orangered3",
      pch.pobs = 12,
      size.pobs = 1.5,

      col.pcens = "yellow",
      pch.pcens = 15,
      size.pcens = 1.75,

      size.text.x = 10,
      size.text.y = 10,

      breaks.x = 1,
      breaks.y = 10,

      xlog = FALSE,
      ylog = FALSE)
```

Figure 16: x.scatter plot for the theofit1 with user defined options



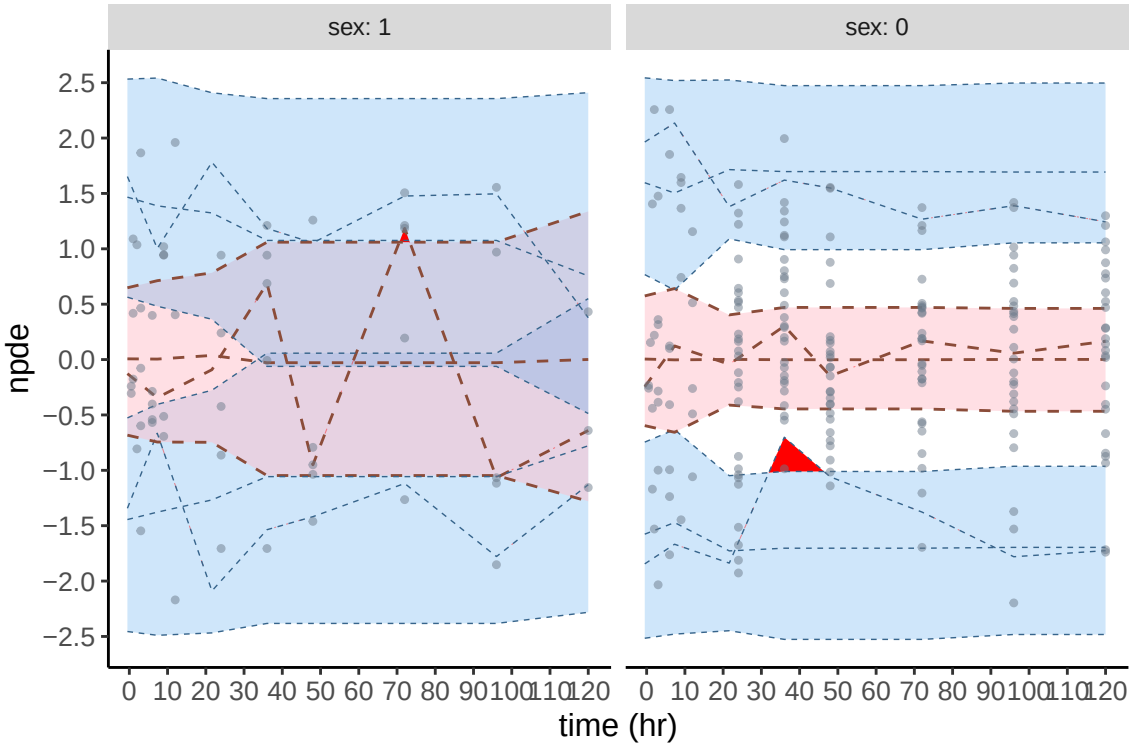

```
> plot(yvir20, plot.type="x.scatter", plot.box=TRUE)
```

Figure 17: x.scatter plot for the theofit1 with boxplots




```
> plot(wcov, plot.type="x.scatter", covsplit=TRUE, which.cov=c("sex","wt"))
```

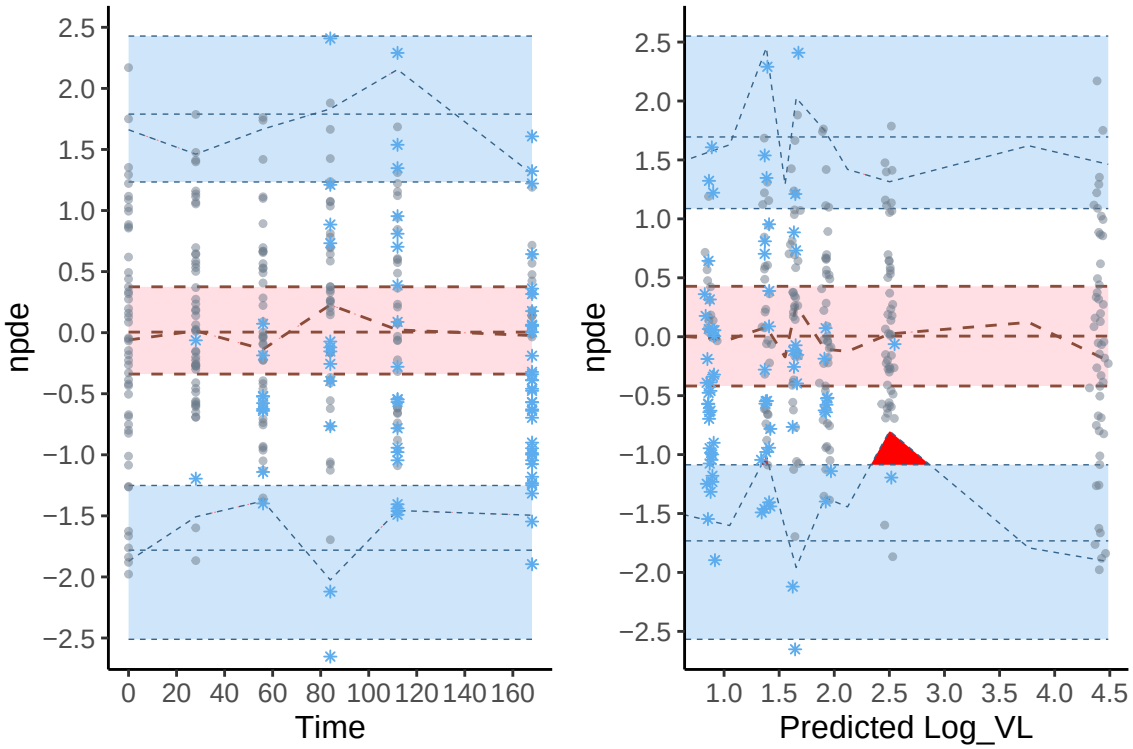
Figure 18: Wcov data - npde versus X, split according to sex and wt



7 Waffle plot

```
> plot.x.scatter.yvir20 = plot(yvir20,plot.type="x.scatter")
> plot.pred.scatter.yvir20 = plot(yvir20,plot.type="pred.scatter")
> do.call(grid.arrange, c(plot.x.scatter.yvir20, plot.pred.scatter.yvir20, list
    ( nrow=1, ncol=2)))
```

Figure 19: yvir20 data - Waffle plot



8 Plot LOQ

```
> plot(yvir20, plot.type="loq")
> plot(yvir20, plot.type="loq", lwd.bands=1.75)
```

Figure 20: Default plots - fraction of data below LOQ.

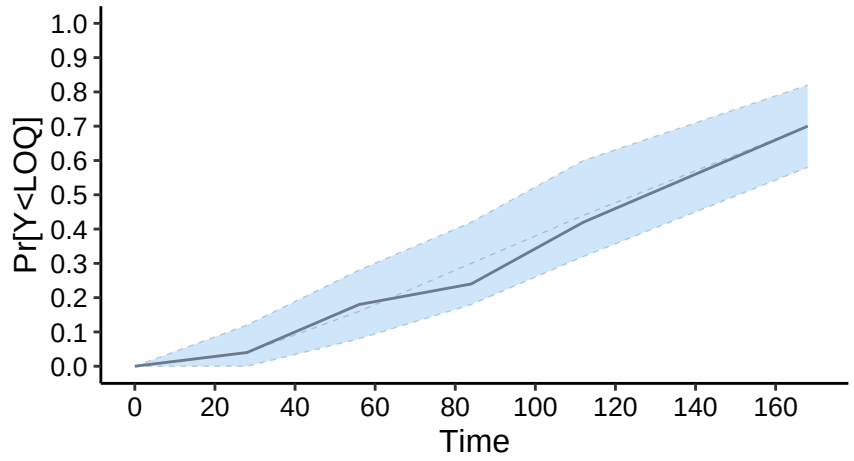
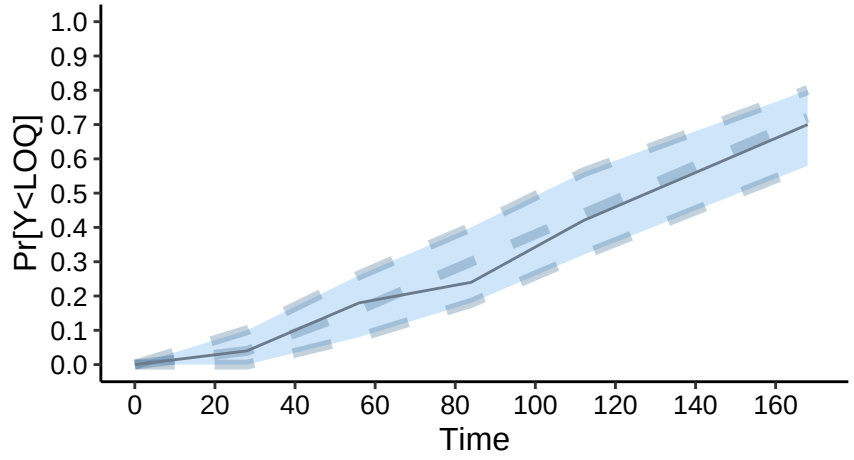


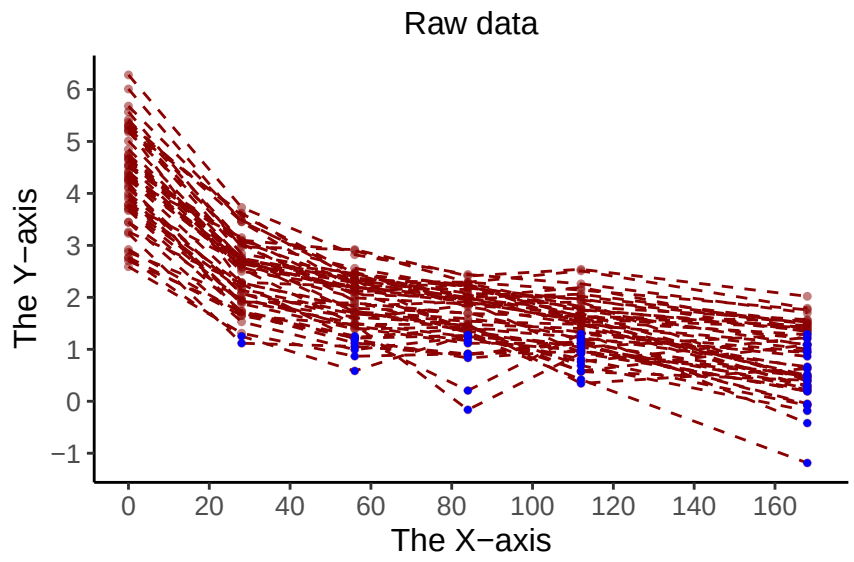
Figure 21: Default plots - fraction of data below LOQ, with user defined options.



9 Options for plots

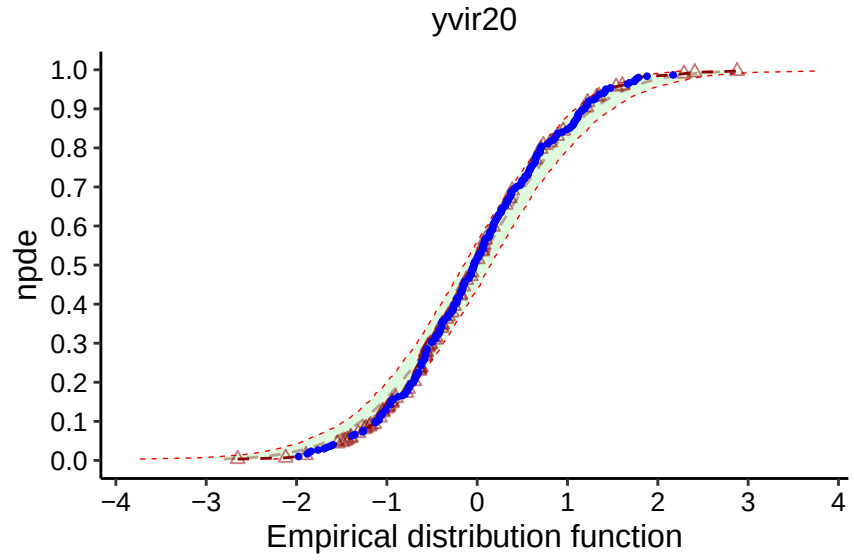
```
> plot(yvir20,plot.type="data",main="Raw data",
 col.pobs="DarkRed",col.lobs="DarkRed", col.pcens="blue", pch.pobs=2,pch.pcens
    =20,lty.lobs=2,xlab="The X-axis", ylab="The Y-axis",sub="Some changes",cex
    =0.8, cex.lab=1.5, plot.loq=TRUE)
```

Figure 22: Plots - options for data.




```
> plot(yvir20,plot.type="ecdf",main="yvir20",col.pobs="DarkRed",col.lobs="
    DarkRed",
 col.pcens="blue",pch.pobs=2,pch.pcens=20,lty.lobs=2,xlab="The X-axis",
 ylab="The Y-axis",sub="Some changes",cex=0.8, cex.lab=1.5,bands=TRUE,
 col.fillpi="lightgreen",col.lpi="green")
```

Figure 23: Plots - options for ecdf.




```
> plot(yvir20,plot.type="vpc",
      main = "yvir20 vpc",
```

```
      size.main = 14,
      sub = "",
      size.sub = "",

      xlab="The X-axis",
      ylab="The Y-axis",

      size.xlab = 12, ## cex.axis = 12
      size.ylab = 12,

      bands=TRUE,
      plot.obs=TRUE,

      col.line.loq = "orange",
      lty.line.loq  = 4,
      lwd.line.loq  = 1,

      alpha.med = 0.25,
      fill.med = "firebrick4",
      col.med="red",
      lty.med=3,
      lwd.med=1,

      alpha.bands = 0.25,
      fill.bands = "dodgerblue",
      col.bands="green",
      lty.bands=6,
      lwd.bands=1,

      col.med = "red",
      lty.med = 1,
      lwd.med = 1,

      col.bands = "blue",
      lty.bands = 1,
      lwd.bands = 1,

      col.pobs = "orangered3",
      pch.pobs = 12,
      size.pobs = 1.5,

      col.pcens = "yellow",
      pch.pcens = 15,
      size.pcens = 1.75,

      size.text.x = 10,
      size.text.y = 10,

      breaks.x = 10,
      breaks.y = 10)
```

Figure 24: Plots - options for VPC.