

Package ‘seawaveQ’

December 13, 2013

Type Package

Title U.S. Geological Survey seawaveQ model

Version 1.0.0

Date 2013-12-13

Author Karen R. Ryberg and Aldo V. Vecchia

Maintainer Karen R. Ryberg <kryberg@usgs.gov>

License Unlimited | file LICENSE

LazyLoad yes

Description

A model and utilities for analyzing trends in chemical concentrations in streams with a seasonal wave (seawave) and adjustment for streamflow (Q) and other ancillary variables

Depends R (>= 2.15.3), survival, NADA, lubridate

Suggests waterData

URL <http://dx.doi.org/10.3133/ofr20131255>

R topics documented:

seawaveQ-package	2
cenScatPlot	3
combineData	4
compwaveconv	6
cqwMoRivOmaha	7
examplecavdat	8
examplecavmat	9
examplecdatsub	9
examplecentmp	11
exampleclog	12
exampleqwcols	12
examplestpars	13
exampletndlin	14
exampletndlinpr	15
exampletseas	15

exampletseaspr	16
exampletyr	17
exampletyrpr	17
fitMod	18
fitswavecav	19
IllRivValleyCty	22
prepData	27
qwMoRivOmaha	28
rosBoxPlot	34
seawaveQPlots	36

Index 39

seawaveQ-package	<i>A model and utilities for analyzing trends in chemical concentrations in streams with a seasonal wave (seawave) and adjustment for stream-flow (Q) and other ancillary variables (Q) and other ancillary variables</i>
------------------	---

Description

An R package for the U.S. Geological Survey seawaveQ model, a parametric regression model specifically designed for analyzing seasonal- and flow-related variability and trends in pesticide concentrations. See Vecchia and others (2008) for the original description of the model and see Sullivan and others (2009), Ryberg and others (2010), and Vecchia and others (2009) for applications of the model.

Details

Package:	seawaveQ
Type:	Package
Version:	1.0.0
Date:	2013-12-13
License:	Unlimited file LICENSE
LazyLoad:	yes

Author(s)

Karen R. Ryberg <kryberg@usgs.gov> and Aldo V. Vecchia <avecchia@usgs.gov>

References

Ryberg, K.R. and Vecchia, A.V., 2013, seawaveQ—An R package providing a model and utilities for analyzing trends in chemical concentrations in streams with a seasonal wave (seawave) and adjustment for streamflow (Q) and other ancillary variables: U.S. Geological Survey Open-File Report 2013-1255, 13 p., with 3 appendixes, <http://dx.doi.org/10.3133/ofr20131255>.

Ryberg, K.R., Vecchia, A.V., Martin, J.D., Gilliom, R.J., 2010, Trends in pesticide concentrations in urban streams in the United States, 1992-2008: U.S. Geological Survey Scientific Investigations

Report 2010-5139, 101 p. <http://pubs.usgs.gov/sir/2010/5139/>.

Sullivan, D.J., Vecchia, A.V., Lorenz, D.L., Gilliom, R.J., Martin, J.D., 2009, Trends in pesticide concentrations in corn-belt streams, 1996-2006: U.S. Geological Survey Scientific Investigations Report 2009-5132, 75 p. <http://pubs.usgs.gov/sir/2009/5132/>.

Vecchia, A.V., Gilliom, R.J., Sullivan, D.J., Lorenz, D.L., and Martin, J.D., 2009, Trends in concentrations and use of agricultural herbicides for Corn Belt rivers, 1996-2006: Environmental Science and Technology, v. 43, p. 9,096-9,102.

Vecchia, A.V., Martin, J.D., and Gilliom, R.J., 2008, Modeling variability and trends in pesticide concentrations in streams: Journal of the American Water Resources Association, v. 44, no. 5, p. 1308-1324, <http://dx.doi.org/10.1111/j.1752-1688.2008.00225.x>.

cenScatPlot

Scatter plot of water-quality data

Description

Function to generate a scatter plot that indicates censored and estimated water-quality concentrations.

Usage

```
cenScatPlot(data, datescol = "dates", pname,
            qwcols = c("R", "P"), site = "", xlabel = "",
            ylabel = "Concentration", legpos = "topright",
            legcex = 1, ...)
```

Arguments

data	is the dataset with columns that begin with P followed by alphanumeric characters indicating concentration data and columns that begin with R followed by alphanumeric characters that match those of the concentration data indicating qualification codes. See example datasets for more information about the data format, see IllRivValleyCty and qwMoRivOmaha .
datescol	is the column label for the dates column.
pname	is the the column heading (parameter name) for the particular water-quality constituent to be plotted (omit the the starting character, for example for sulfate data indicated by P00945, enter "00945").
qwcols	is a character vector with the beginning of the column headers for remarks code (default is R), and beginning of column headers for concentration data (default is P for parameter).
site	is a label for the plot title indicating the site where the water-quality samples were collected.
xlabel	is the label for the x-axis, defaults to no label.
ylabel	is the label for the y-axis.
legpos	is the position of the legend, see legend .
legcex	is a numerical value giving the amount by which the legend text and symbols should be magnified relative to the default, 1.
...	arguments to be passed to plot method.

Details

This function uses the qualification, or remark, column associated with water-quality concentration values to indicate which samples are unqualified, which are estimated, and which are censored. A blank remark field or an "_" indicates that the concentration value is not qualified; an "E" indicates the values has been estimated; and a less than symbol, "<", indicates the value has been censored as less than a minimum reporting level. See Oblinger Childress (1999) for information on the minimum reporting level and the definition of "E" for U.S. Geological Survey data. Other users may have a different definition of the minimum reporting level, but censored values need to be qualified with a "<". Using the "E" code is optional.

Value

a scatter plot

Author(s)

Karen R. Ryberg

References

Oblinger Childress, C.J., Foreman, W.T., Connor, B.F., and Maloney, T.J., 1999, New reporting procedures based on long-term method detection levels and some considerations for interpretations of water-quality data provided by the U.S. Geological Survey: U.S. Geological Survey Open-File Report 99-193, 19 p. (Also available at http://water.usgs.gov/owq/OFR_99-193/index.html.)

Examples

```
data(swData)
# scatter plot of Simazine concentrations
cenScatPlot(IllRivValleyCty, pname="04035")
# scatter plot with many additional plotting arguments
par(las=1, tcl=0.5)
cenScatPlot(IllRivValleyCty, pname="04035",
            site="05586100 Illinois River at Valley City, IL",
            ylabel="Simazine concentration, in micrograms per liter",
            legcex=0.7, ylim=c(0,0.4), yaxs="i", cex.lab=0.9,
            cex.axis=0.9, xlim=c(as.Date("1996-01-01", "%Y-%m-%d"),
            as.Date("2012-01-01", "%Y-%m-%d")), xaxs="i", xaxt="n")
axdates<-c("1996-01-01", "2000-01-01", "2004-01-01", "2008-01-01",
            "2012-01-01")
axis(1, as.Date(axdates, "%Y-%m-%d"),
      labels=c("1996", "2000", "2004", "2008", "2012"), cex.axis=0.9)
```

combineData

Combine water-quality sample data and continuous ancillary variables

Description

Function to combine water-quality sample data and continuous (daily) ancillary variables and drop unnecessary columns.

compwaveconv *Seasonal Wave Computation*

Description

Function to compute seasonal wave.

Usage

```
compwaveconv(cmxt, jmod, hlife, mclass = 1)
```

Arguments

cmxt	is the time of the maximum chemical concentration, decimal time in years.
jmod	is the choice of model or pulse input function, an integer 1 through 14.
hlife	is the model half-life in months, 1 to 4 months
mclass	has not been implemented yet, but will provide additional model options.

Value

a numeric vector of size 361 with discrete values of the seasonal wave for decimal season seq(0, 1, 1/360).

Note

The seasonal wave is a dimensionless, periodic function of time with an annual cycle, similar to a mixture of sine and cosine functions often used to model seasonality in concentration data. However, the seasonal wave is better suited for modeling seasonal behavior of pesticide data than a mixture of sines and cosines. The pulse input function, represented by jmod, has either one or two distinct application seasons (when pesticides may be transported to the stream) of lengths from 1 to 6 months. Therefore, 56 (14x4) choices for the wave function are available. The numeric vector is a discrete approximation of the continuous wave function defined on the interval 0 to 1.

Author(s)

Aldo V. Vecchia

References

Vecchia, A.V., Martin, J.D. and Gilliom, R.J., 2008, Modeling variability and trends in pesticide concentrations in streams: JAWRA Journal of the American Water Resources Association, v. 44, p. 1308–1324, <http://onlinelibrary.wiley.com/doi/10.1111/j.1752-1688.2008.00225.x/abstract>.

Examples

```
# evaluate seasonal wave for specified decimal seasons
# these example decimal dates represent days at points 0.25, 0.5, and
# 0.75 percent of the way through the year and the end of the year
dseas <- c(0.25, 0.5, 0.75, 1)
swave <- compwaveconv(cmxt=0.483, jmod=2, hlife=4, mclass=1)
swave[floor(360 * dseas)]
plot(seq(0,1,1/360),swave, typ="l")
```

cqwMoRivOmaha	<i>Continuously monitored (daily) data for 06610000 Missouri River at Omaha, Neb.</i>
---------------	---

Description

Continuously monitored (daily) streamflow and sediment data for U.S. Geological Survey stream-gage 06610000 Missouri River at Omaha, Neb., and streamflow and sediment anomalies.

Usage

```
cqwMoRivOmaha
```

Format

A data frame containing 2,922 daily observations of two hydrologic variables, streamflow and sediment, and streamflow and sediment anomalies. There are eight variables.

staid	character	USGS Station identification number
dates	date	date water-quality sample collected
dflow	numeric	daily mean streamflow, cubic feet per second
flowa30	numeric	30-day streamflow anomaly
flowa1	numeric	1-day streamflow anomaly
dsed	numeric	daily mean sediment concentration, milligrams per liter
seda30	numeric	30-day sediment anomaly
seda1	numeric	1-day sediment anomaly

Details

The streamflow and sediment anomalies were generated using the R package `waterData` (Ryberg and Vecchia, 2012).

Note

See Ryberg and Vecchia (2012) for more information on calculating the anomalies and for additional references documenting the use of streamflow anomalies in water-quality trend analysis studies.

Source

Data provided by Patrick Phillips, U.S. Geological Survey, New York Water Science Center.

References

Ryberg, K.R. and Vecchia, A.V., 2012, `waterData`—An R package for retrieval, analysis, and anomaly calculation of daily hydrologic time series data, version 1.0.0: U.S. Geological Survey Open-File Report 2012–1168, 8 p. (Also available at <http://pubs.usgs.gov/of/2012/1168/>.)

Examples

```
data(swData)
```

```
# summary of water-quality concentrations
apply(cqwMoRivOmaha[,3:8], 2, summary)
```

examplecavdat	<i>Example continuous ancillary variable data.</i>
---------------	--

Description

This is an example of the continuous ancillary data that is passed internally to subfunctions of `fitswavecav`. It is provided here for use with examples of internal functions.

Usage

```
examplecavdat
```

Format

A data frame containing 2,893 data variables and 30-day and 1-day streamflow anomalies (Ryberg and Vecchia, 2012).

yrx	numeric	Year
mox	numeric	Month
dax	numeric	Day
jdayx	numeric	Julian day from first day water year for start year in <code>fitswavecav</code>
flowa30	numeric	30-day streamflow anomaly
flowa1	numeric	1-day streamflow anomaly

Source

Internal data captured from the following function call:

```
fitswavecav(cdat=modMoRivOmaha, cavdat=cqwMoRivOmaha,
            tanm="myexample", pnames=c("04041"), yrstart=1995,
            yrend=2003, tndbeg=1995, tndend=2003,
            iwca=c("flowa30", "flowa1"), dcol="dates",
            qwcols=c("R", "P"))
```

References

Ryberg, K.R., and Vecchia, A.V., 2012, waterData—An R package for retrieval, analysis, and anomaly calculation of daily hydrologic time series data, version 1.0: U.S. Geological Survey Open-File Report 2012–1168, 8 p. (Also available at <http://pubs.usgs.gov/of/2012/1168/>.)

Examples

```
data(swData)
head(examplecavdat)
```

examplecavmat	<i>Example continuous ancillary variable matrix.</i>
---------------	--

Description

This is an example of the continuous ancillary matrix that is passed internally to subfunctions of [fitswavecav](#). It is provided here for use with examples of internal functions.

Usage

```
examplecavmat
```

Format

A matrix containing 115 30-day and 1-day streamflow anomalies (Ryberg and Vecchia, 2012).

flowa30	numeric	30-day streamflow anomaly
flowa1	numeric	1-day streamflow anomaly

Source

Internal data captured from the following function call:

```
fitswavecav(cdat=modMoRivOmaha, cavdat=cqwMoRivOmaha,
            tanm="myexample", pnames=c("04041"), yrstart=1995,
            yrend=2003, tndbeg=1995, tndend=2003,
            iwca=c("flowa30","flowa1"), dcol="dates",
            qwcols=c("R","P"))
```

Examples

```
data(swData)
head(examplecavmat)
```

examplecdatsub	<i>Example water-quality data.</i>
----------------	------------------------------------

Description

This is an example of the water-quality data that is passed internally to subfunctions of [fitswavecav](#). It is provided here for use with examples of internal functions.

Usage

```
examplecdatsub
```

Format

A data frame containing 115 observations of 10 variables. The date variables were internally calculated. The columns R04041 and P04041 are a subset of [qwMoRivOmaha](#) and the 30-day and 1-day streamflow and sediment anomalies are a subset of [cqWMoRivOmaha](#).

yr	numeric	Year
mo	numeric	Month
day	numeric	Day
jday	numeric	Julian day from first day of start year in fitswavecav
flowa30	numeric	30-day streamflow anomaly
flowa1	numeric	1-day streamflow anomaly
seda30	numeric	30-day sediment anomaly
seda1	numeric	1-day sediment anomaly

Source

Internal data captured from the following function call:

```
fitswavecav(cdat=modMoRivOmaha, cavdat=cqwMoRivOmaha,
            tanm="myexample", pnames=c("04041"), yrstart=1995,
            yrend=2003, tndbeg=1995, tndend=2003,
            iwca=c("flowa30","flowa1"), dcol="dates",
            qwcols=c("R","P"))
```

See Also

[qwMoRivOmaha](#) [cqwMoRivOmaha](#)

Examples

```
data(swData)
head(examplecdatsub)
```

examplecentmp	<i>Example logical vector.</i>
---------------	--------------------------------

Description

This is an example of data that is passed internally to subfunctions of `link{fitswavecav}`. This logical vector indicates which water-quality values are censored. It is provided here for use with examples of the internal functions.

Usage

```
examplecentmp
```

Format

A logical vector of 115 observations.

Source

Internal data captured from the following function call:

```
fitswavecav(cdat=modMoRivOmaha, cavdat=cqwMoRivOmaha,
            tanm="myexample", pnames=c("04041"), yrstart=1995,
            yrend=2003, tndbeg=1995, tndend=2003,
            iwca=c("flowa30","flowa1"), dcol="dates",
            qwcols=c("R","P"))
```

Examples

```
data(swData)
examplecentmp
```

```
exampleclog
```

Example of logarithmically transformed concentration data.

Description

This is an example of data that is used internally by [fitMod](#) and passed to its subfunction [seawaveQPlots](#). This numeric vector represents the base-10 logarithm of the water-quality concentrations. It is provided here for use with examples of the internal functions.

Usage

```
exampleclog
```

Format

A numeric vector of 115 observations.

Source

Internal data captured from the following function call:

```
fitswavecav(cdat=modMoRivOmaha, cavdat=cqwMoRivOmaha,
            tanm="myexample", pnames=c("04041"), yrstart=1995,
            yrend=2003, tndbeg=1995, tndend=2003,
            iwca=c("flowa30","flowa1"), dcol="dates",
            qwcols=c("R","P"))
```

Examples

```
data(swData)
exampleclog
```

```
exampleqwcols
```

Example data indicators.

Description

This is an example of the character vector used to indicate which columns represent qualification codes and which represent water-quality concentration data. It is provided here for use with examples of the internal functions.

Usage

```
exampleqwcols
```

Format

A numeric vector of 115 observations.

Source

Internal data captured from the following function call:

```
fitswavecav(cdat=modMoRivOmaha, cavdat=cqwMoRivOmaha,
            tanm="myexample", pnames=c("04041"), yrstart=1995,
            yrend=2003, tndbeg=1995, tndend=2003,
            iwca=c("flowa30","flowa1"), dcol="dates",
            qwcols=c("R","P"))
```

See Also

[prepData fitMod](#)

Examples

```
data(swData)
exampleqwcols
```

examplestpars

Example matrix for internal use.

Description

This is an example of data that is passed internally to subfunctions of [fitswavecav](#). It is provided here for use with examples of the internal functions.

Usage

```
examplestpars
```

Format

A numeric matrix of two rows and 14 columns.

column description

1 mclass, model class has not been implemented yet and is equal to 1

2 model chosen (a number 1-56), this number represents both the pulse input function and the half-life

3 is the scale factor from the survreg object 4 is the likelihood for the model chosen

5 is the coefficient for the model intercept

6 is the coefficient for the seasonal wave component of the model

7 is the coefficient for the trend component of the model

8 is the coefficient for the 30-day flow anomaly

9 is the coefficient for the 1-day flow anomaly

10 is the standard error for the intercept term

11 is the standard error for the seasonal wave term

12 is the standard error for the trend term

13 is the standard error for the 30-day flow anomaly term

14 is the standard error for the 1-day flow anomaly term
 15 is cmaxt, the decimal season of maximum concentration
 16 is the p-value for the trend line

Source

Internal data captured from the following function call:

```
fitswavecav(cdat=modMoRivOmaha, cavdat=cqwMoRivOmaha,
            tanm="myexample", pnames=c("04041"), yrstart=1995,
            yrend=2003, tndbeg=1995, tndend=2003,
            iwca=c("flowa30", "flowa1"), dcol="dates",
            qwcols=c("R", "P"))
```

See Also

[fitswavecav](#)

Examples

```
data(swData)
examplestpars
```

exampletndlin

Example numeric vector used internally.

Description

This is an example of data that is passed internally to [seawaveQPlots](#). This numeric vector contains trend coefficients for the water-quality samples. It is provided here for use with examples of the internal functions.

Usage

```
exampletndlin
```

Format

A numeric vector of 115 observations.

Source

Internal data captured from the following function call:

```
fitswavecav(cdat=modMoRivOmaha, cavdat=cqwMoRivOmaha,
            tanm="myexample", pnames=c("04041"), yrstart=1995,
            yrend=2003, tndbeg=1995, tndend=2003,
            iwca=c("flowa30", "flowa1"), dcol="dates",
            qwcols=c("R", "P"))
```

Examples

```
data(swData)
head(exampletndlin)
```

exampletndlinpr	<i>Example numeric vector used internally.</i>
-----------------	--

Description

This is an example of data that is passed internally to [seawaveQPlots](#). This numeric vector contains trend coefficients for a continuous water-quality prediction based on the continuous ancillary variables. It is provided here for use with examples of the internal functions.

Usage

```
exampletndlinpr
```

Format

A numeric vector of 2,893 observations.

Source

Internal data captured from the following function call:

```
fitswavecav(cdat=modMoRivOmaha, cavdat=cqwMoRivOmaha,  
            tanm="myexample", pnames=c("04041"), yrstart=1995,  
            yrend=2003, tndbeg=1995, tndend=2003,  
            iwca=c("flowa30", "flowa1"), dcol="dates",  
            qwcols=c("R", "P"))
```

Examples

```
data(swData)  
head(exampletndlinpr)
```

exampletseas	<i>Example numeric vector used internally.</i>
--------------	--

Description

This is an example of data that is passed internally to [seawaveQPlots](#). This numeric vector contains decimal seasonal (0-1) values for the water-quality samples. It is provided here for use with examples of the internal functions.

Usage

```
exampletseas
```

Format

A numeric vector of 115 observations.

Source

Internal data captured from the following function call:

```
fitswavecav(cdat=modMoRivOmaha, cavdat=cqwMoRivOmaha,  
            tanm="myexample", pnames=c("04041"), yrstart=1995,  
            yrend=2003, tndbeg=1995, tndend=2003,  
            iwca=c("flowa30","flowa1"), dcol="dates",  
            qwcols=c("R","P"))
```

Examples

```
data(swData)  
head(exampletseas)
```

exampletseaspr

Example numeric vector used internally.

Description

This is an example of data that is passed internally to [seawaveQPlots](#). This numeric vector contains decimal seasonal (0-1) values for the continuous ancillary data. It is provided here for use with examples of the internal functions.

Usage

```
exampletseaspr
```

Format

A numeric vector of 2,893 observations.

Source

Internal data captured from the following function call:

```
fitswavecav(cdat=modMoRivOmaha, cavdat=cqwMoRivOmaha,  
            tanm="myexample", pnames=c("04041"), yrstart=1995,  
            yrend=2003, tndbeg=1995, tndend=2003,  
            iwca=c("flowa30","flowa1"), dcol="dates",  
            qwcols=c("R","P"))
```

Examples

```
data(swData)  
head(exampletseaspr)
```

`exampletyr`*Example numeric vector used internally.*

Description

This is an example of data that is passed internally to [seawaveQPlots](#). This numeric vector contains decimal dates for the water-quality samples. It is provided here for use with examples of the internal functions.

Usage`exampletyr`**Format**

A numeric vector of 115 observations.

Source

Internal data captured from the following function call:

```
fitswavecav(cdat=modMoRivOmaha, cavdat=cqwMoRivOmaha,
            tanm="myexample", pnames=c("04041"), yrstart=1995,
            yrend=2003, tndbeg=1995, tndend=2003,
            iwca=c("flowa30", "flowa1"), dcol="dates",
            qwcols=c("R", "P"))
```

Examples

```
data(swData)
head(exampletyr)
```

`exampletyrpr`*Example numeric vector used internally.*

Description

This is an example of data that is passed internally to [seawaveQPlots](#). This numeric vector contains decimal dates for continuous ancillary variables. It is provided here for use with examples of the internal functions.

Usage`exampletyrpr`**Format**

A numeric vector of 2,893 observations.

Source

Internal data captured from the following function call:

```
fitswavecav(cdat=modMoRivOmaha, cavdat=cqwMoRivOmaha,
            tanm="myexample", pnames=c("04041"), yrstart=1995,
            yrend=2003, tndbeg=1995, tndend=2003,
            iwca=c("flowa30", "flowa1"), dcol="dates",
            qwcols=c("R", "P"))
```

Examples

```
data(swData)
head(exampletyrpr)
```

 fitMod

Internal function that fits the seawaveQ model.

Description

fitMod is called from within [fitswavecav](#) but can be invoked directly. It fits the seawaveQ model and returns the results.

Usage

```
fitMod(cdat, cavdat, yrstart, yrend, tndbeg, tndend,
       tanm, pnames, qwcols, mclass = 1)
```

Arguments

cdat	is the concentration data
cavdat	is the continuous (daily) ancillary data
yrstart	is the starting year of the analysis (treated as January 1 of that year).
yrend	is the ending year of the analysis (treated as December 31 of that year).
tndbeg	is the beginning (in whole or decimal years) of the trend period.
tndend	is the end (in whole or decimal years) of the trend period.
tanm	is a character identifier that names the trend analysis run. It is used to label output files.
pnames	is the parameter (water-quality constituents) to analyze (if using USGS parameters, omit the the starting 'P', such as "00945" for sulfate).
qwcols	is a character vector with the beginning of the column headers for remarks code (default is R), and beginning of column headers for concentration data (default is P for parameter).
mclass	has not been implemented yet, but will provide additional model options.

Value

a pdf file containing plots (see [seawaveQPlots](#)), a text file showing the best model survival regression call and results, and a list. The first element of the list contains information about the data and the model(s) selected (see [examplestpars](#)). The second element of the list contains the summary of the survival regression call. The third element of the list is itself a list containing the observed concentrations (censored and uncensored) and the predicted concentrations used by [seawaveQPlots](#) to generate the plots.

Author(s)

Aldo V. Vecchia and Karen R. Ryberg

Examples

```
data(swData)
myRes <- fitMod(cdatsub=examplecdatsub, cavdat=examplecavdat,
  yrstart=1995, yrend=2003, tndbeg=1995, tndend=2003, tanm="myfit3",
  pnames=c("04041"), qwcols=c("R", "P"))
```

 fitswavecav

Fit seasonal wave and continuous ancillary data for trend analysis

Description

Function to prepare data and fit the seawaveQ model.

Usage

```
fitswavecav(cdat, cavdat, tanm = "trend1", pnames,
  yrstart = 0, yrend = 0, tndbeg = 0, tndend = 0,
  iwca = c("none"), dcol = "dates",
  qwcols = c("R", "P"), mclass = 1)
```

Arguments

cdat	is the concentration data
cavdat	is the continuous (daily) ancillary data
tanm	is a character identifier that names the trend analysis run. It is used to label output files.
pnames	are the parameters (water-quality constituents) to analyze (omit the the starting character, for example for sulfate data indicated by P00945, enter "00945").
yrstart	is the starting year of the analysis (treated as January 1 of that year). Zero means the start date will be determined by the start date of cavdat, the continuous ancillary data.
yrend	is the ending year of the analysis (treated as December 31 of that year). Zero means the end date will be determined by the end date of cavdat, the continuous ancillary data.
tndbeg	is the beginning (in whole or decimal years) of the trend period. Zero means the begin date will be the beginning of the concentration data, cdat.

tndend	is the end of the trend (treated as December 31 of that year). Zero means the end date will be the end of the concentration data, cdat.
iwcav	is a character vector indicating which continuous ancillary variables to include, if none are used for analysis, use iwca=c("none").
dcol	is the column name for the dates, should be the same for both cdat and cavdat
qwcols	is a character vector with the beginning of the column headers for remarks code (default is R), and beginning of column headers for concentration data (default is P for parameter).
mclass	has not been implemented yet but will provide additional model options.

Format

The data frame returned has one row for each parameter analyzed and the number of columns depend on the number of continuous ancillary variables used. The general format is as follows:

pname	character	Parameter analyzed
mclass	numeric	Currently a value of 1
jmod	numeric	The choice of pulse input function, an integer 1–14.
hlife	numeric	the model half-life in months, an integer, 1 to 4 months
cmxt	numeric	the decimal season of maximum concentration
scl	numeric	the scale factor from the survreg.object
loglik	numeric	the log-likelihood for the model
cint	numeric	coefficient for model intercept
cwave	numeric	coefficient for the seasonal wave
ctnd	numeric	coefficient for the trend component of model
c[alphanumeric]	numeric	0 or more coefficients for the continuous ancillary variables
seint	numeric	standard error for the intercept
sewave	numeric	standard error for the seasonal wave
setnd	numeric	standard error for the trend
se[alphanumeric]	numeric	0 or more standard errors for the continuous ancillary variables
pvaltnd	numeric	the p-value for the trend line

Details

Fits the seawaveQ model (Vecchia and others, 2008) using a seasonal wave and continuous ancillary variables (streamflow anomalies and other continuous variables such as conductivity or sediment) to model water quality.

Value

a pdf file containing plots of the data and modeled concentration, a text file containing a summary of the survival regression call for each model selected, and a list. The first element of the list is a data frame described under format. The second element of the list is the summary of the survival regression call. The third element is the observed concentration data (censored and uncensored). The fourth element is the concentration data predicted by the model. The fifth element provides summary statistics for the predicted concentrations.

Note

The assumed data format is one with columns for water-quality concentration values and a related column for qualification of those values, such as in the case of left-censored values less than a particular value. For example, a water-quality sample was collected and the laboratory analysis indicated that the concentration was less than 0.01 micrograms per liter. The USGS parameter code for simazine is 04035 (U.S. Geological Survey, 2013b). When the data are retrieved through the National Water Information System: Web Interface (<http://waterdata.usgs.gov/nwis>; U.S. Geological Survey, 2013a), the concentration values are in a column labeled P04035 and the qualification information, or remark codes, are in a column labeled R04035. To use this function, the argument `pnames` would be the unique identifier for simazine values and qualifications, 04035, and the `qwcols` argument would be `c("R", "P")` to indicate that the qualification column starts with an R and the values column starts with a P.

Other users may have data in different format that can be changed to use with this function. For example, a user may have concentration values and qualification codes in one column, such as a column labeled `simzaine` with the values 0.05, 0.10, <0.01, <0.01, and 0.90. In this case, the less thans and any other qualification codes should be placed in a separate column. The column names for the qualification codes and the concentration values should be the same with the exception of different beginning letters to indicate which column is which. The columns could be named `Rsimazine` and `Psimazine`. Then the argument `pnames = "simazine"` and the argument `qwcols = c("R", "P")`.

Users should exercise caution when their water-quality data have multiple censoring limits and may want to recensor the data to a single censoring level. Censoring and recensoring issues are discussed in the text and Appendix 1 of Ryberg and others (2010).

Author(s)

Aldo V. Vecchia and Karen R. Ryberg

References

Ryberg, K.R., Vecchia, A.V., Martin, J.D., and Gilliom, R.J., 2010, Trends in pesticide concentrations in urban streams in the United States, 1992–2008: U.S. Geological Survey Scientific Investigations Report 2010-5139, 101 p. (Also available at <http://pubs.usgs.gov/sir/2010/5139/>.)

U.S. Geological Survey, 2013a, National Water Information System: Web Interface, accessed February 26, 2013, at <http://waterdata.usgs.gov>.

U.S. Geological Survey, 2013b, Parameter code definition: National Water Information System: Web Interface, accessed February 26, 2013, at <http://nwis.waterdata.usgs.gov/usa/nwis/pmcodes>.

Vecchia, A.V., Martin, J.D., and Gilliom, R.J., 2008, Modeling variability and trends in pesticide concentrations in streams: *Journal of the American Water Resources Association*, v. 44, no. 5, p. 1308-1324, <http://dx.doi.org/10.1111/j.1752-1688.2008.00225.x>.

See Also

The functions that `fitswavecav` calls internally:
[prepData](#) and [fitMod](#).

Examples

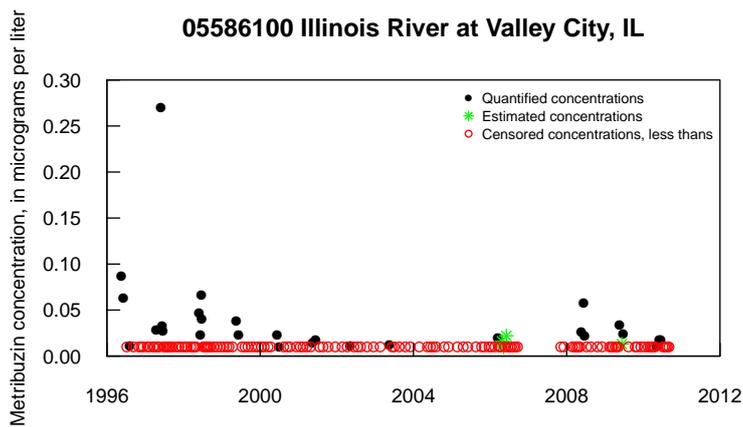
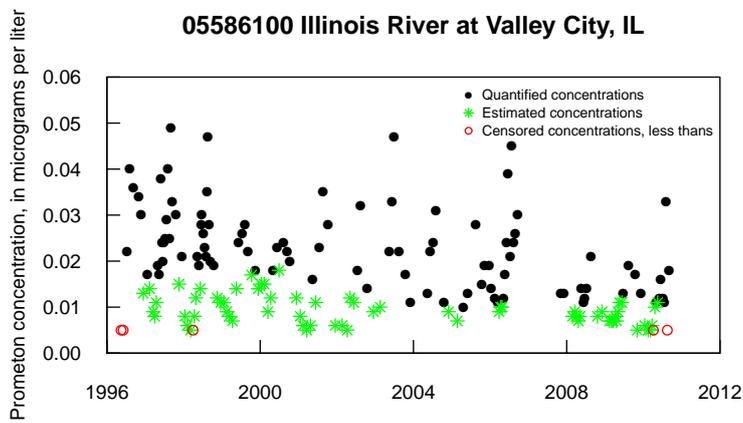
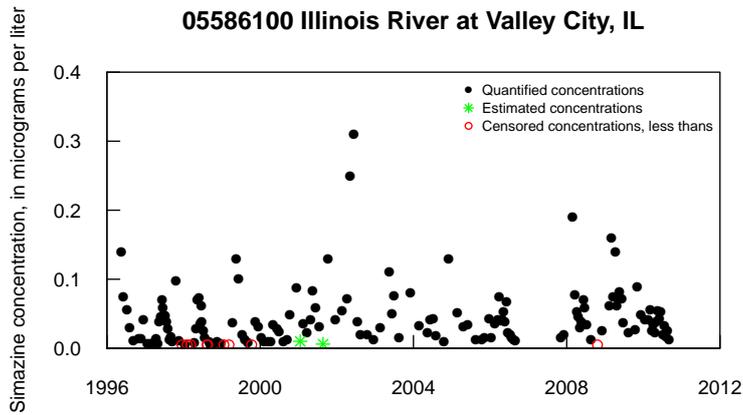
```
data(swData)
modMoRivOmaha<-combineData(qwdat=qwMoRivOmaha, cqwdat=cqwMoRivOmaha)
myfit1 <- fitswavecav(cdat=modMoRivOmaha, cavdat=cqwMoRivOmaha,
```

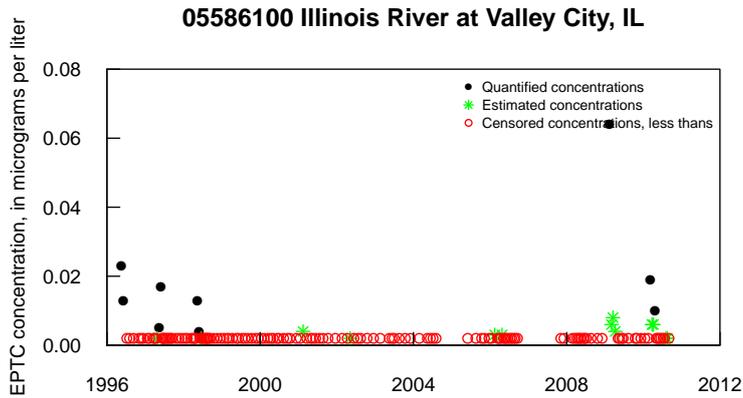
```
tanm="myfit1", pnames=c("04035", "04037", "04041"), yrstart=1995,
yrend=2003, tndbeg=1995, tndend=2003, iwca=c("flowa30","flowa1"),
dcol="dates", qwcols=c("R","P"))
myfit2 <- fitswavecav(cdat=modMoRivOmaha, cavdat=cqwMoRivOmaha,
tanm="myfit2", pnames=c("04035", "04037", "04041"), yrstart=1995,
yrend=2003, tndbeg=1995, tndend=2003, iwca=c("seda30","seda1"),
dcol="dates", qwcols=c("R","P"))
## Not run:
myfit3 <- fitswavecav(cdat=modMoRivOmaha, cavdat=cqwMoRivOmaha,
tanm="myfit3", pnames=c("04035", "04037", "04041"), yrstart=1995,
yrend=2003, tndbeg=1995, tndend=2003, iwca=c("flowa30","flowa1",
"seda30", "seda1"), dcol="dates", qwcols=c("R","P"))

## End(Not run)
# trend model results
myfit1[[1]]
# example regression call
myfit1[[2]][[1]]
# first few lines of observed concentrations
head(myfit1[[3]])
# first few lines of predicted concentrations
head(myfit1[[4]])
# summary statistics for predicted concentrations
head(myfit1[[5]])
```

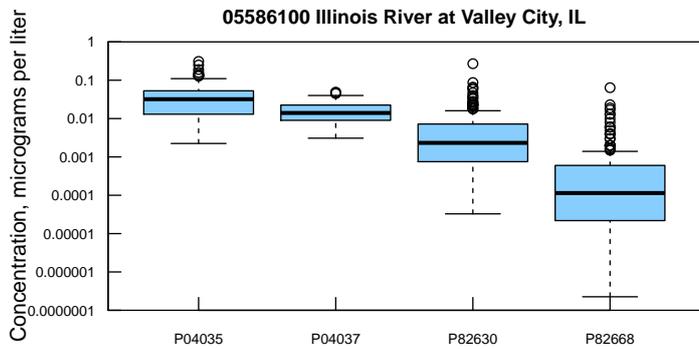
Description

Scatterplots of water-quality data for 05586100 Illinois River at Valley City, Ill.





Boxplot of data using regression on order statistics (Helsel, 2005; Lee, 2010) to estimate censored values.



Censored values estimated using regression on order statistics

Usage

IllRivValleyCty

Format

A data frame containing 168 water-quality samples for 4 constituents. There are 20 variables.

site	character	Site abbreviation for study
staid	character	USGS Station identification number
dates	date	Date water-quality sample collected
ycr	numeric	Year
mo	numeric	Month
day	numeric	Day
jday	numeric	Julian day from first day of associated streamflow data used
R04035	character	Remark code (blank, _, <, or E)
P04035	numeric	Simazine, water, filtered, recoverable, micrograms per liter
R04037	character	Remark code (blank, _, <, or E)

P04037	numeric	Prometon, water, filtered, recoverable, micrograms per liter
R82630	character	Remark code (blank, _, <, or E)
P82630	numeric	Metribuzin, water, filtered, recoverable, micrograms per liter
R82668	character	Remark code (blank, _, <, or E)
P82668	numeric	EPTC, water, filtered (0.7 micron glass fiber filter), recoverable, micrograms per liter
dflow	numeric	Streamflow, cubic feet per second
flowa30	numeric	30-day streamflow anomaly
flowa1	numeric	1-day streamflow anomaly
dsed	numeric	Sediment concentration, milligrams per liter
seda30	numeric	30-day sediment anomaly
seda1	numeric	1-day sediment anomaly

Details

Chemical concentration data are in the columns that start with a P and are followed by a number. Qualification codes for the concentration data are in the columns that start with an R followed by the same numbers as the associated concentration data. For example, column P04035 indicates simazine data, 04035, being the U.S. Geological Survey parameter code for simazine. The qualification codes for the simazine concentrations are found in the column R04035, indicating a U.S. Geological Survey remark code. Remark codes include _ or nothing, indicating no qualification of the value in the associated concentration field; <, indicating a censored value that is less than the number reported in the associated concentration field; and E, indicating that the value has been estimated. See Oblinger Childress and others (1999) for information on the remark codes used by the U.S. Geological Survey. The streamflow and sediment anomalies were generated using the R package waterData (Ryberg and Vecchia, 2012).

Source

Data provided by Patrick Phillips, U.S. Geological Survey, New York Water Science Center.

References

Oblinger Childress, C.J., Foreman, W.T., Connor, B.F., and Maloney, T.J., 1999, New reporting procedures based on long-term method detection levels and some considerations for interpretations of water-quality data provided by the U.S. Geological Survey Open-File Report 99-193, 19 p. (Also available at http://water.usgs.gov/owq/OFR_99-193/index.html.)

Examples

```
data(swData)

# summary of water-quality concentrations
apply(IllRivValleyCty[,grep("P[[:digit:]]"),
dimnames(IllRivValleyCty)[[2]]], 2, summary)

# simple boxplot of water-quality concentrations
rosBoxPlot(IllRivValleyCty)

# same boxplot function with many additional plotting arguments
rosBoxPlot(IllRivValleyCty,
  site="05586100 Illinois River at Valley City, Ill.", log="y",
  yaxt="n", ylim=c(0.0000001, 1), qwcols=c("R", "P"),
  ylab=c("Concentration, micrograms per liter"), col="skyblue1",
  cex.axis=0.7, cex.sub=0.8, par(tcl=0.5, las=1,
                                yaxs="i",
```

```

                                mgp=c(3,0.5,0),
                                mar=c(5,5,2,2),
                                cex.main=0.9)
axis(2, at=c(0.000001, 0.00001, 0.0001, 0.001, 0.01, 0.1, 1),
      labels=c("0.000001", "0.00001", "0.0001", "0.001", "0.01", "0.1",
              "0.1", "1"), cex.axis=0.7)

# scatter plot of simazine concentrations
cenScatPlot(IllRivValleyCty, pname="04035")

# scatter plot with many additional plotting arguments
par(las=1, tcl=0.5)
cenScatPlot(IllRivValleyCty, pname="04035",
            site="05586100 Illinois River at Valley City, Ill.",
            ylabel="Simazine concentration, in micrograms per liter",
            legcex=0.7,
            ylim=c(0,0.4), yaxs="i", cex.lab=0.9, cex.axis=0.9,
            xlim=c(as.Date("1996-01-01"), as.Date("2012-01-01")),
            xaxs="i", xaxt="n")
axdates<-c("1996-01-01", "2000-01-01", "2004-01-01", "2008-01-01",
           "2012-01-01")
axis(1, as.Date(axdates), labels=c("1996", "2000", "2004", "2008",
                                   "2012"), cex.axis=0.9)

# Prometon scatter plot
cenScatPlot(IllRivValleyCty, pname="04037",
            site="05586100 Illinois River at Valley City, Ill.",
            ylabel="Prometon concentration, in micrograms per liter",
            legcex=0.7,
            ylim=c(0,0.06), yaxs="i", cex.lab=0.9, cex.axis=0.9,
            xlim=c(as.Date("1996-01-01"),
                  as.Date("2012-01-01")), xaxs="i",
            xaxt="n")
axdates<-c("1996-01-01", "2000-01-01", "2004-01-01", "2008-01-01",
           "2012-01-01")
axis(1, as.Date(axdates), labels=c("1996", "2000", "2004", "2008",
                                   "2012"), cex.axis=0.9)

# Metribuzin scatter plot
cenScatPlot(IllRivValleyCty, pname="82630",
            site="05586100 Illinois River at Valley City, Ill.",
            ylabel="Metribuzin concentration, in micrograms per liter",
            legcex=0.7,
            ylim=c(0,0.3), yaxs="i", cex.lab=0.9, cex.axis=0.9,
            xlim=c(as.Date("1996-01-01"),
                  as.Date("2012-01-01")), xaxs="i",
            xaxt="n")
axdates<-c("1996-01-01", "2000-01-01", "2004-01-01", "2008-01-01",
           "2012-01-01")
axis(1, as.Date(axdates), labels=c("1996", "2000", "2004", "2008",
                                   "2012"), cex.axis=0.9)

# EPTC scatter plot
cenScatPlot(IllRivValleyCty, pname="82668",
            site="05586100 Illinois River at Valley City, Ill.",
            ylabel="EPTC concentration, in micrograms per liter",
            legcex=0.7, ylim=c(0,0.08), yaxs="i", cex.lab=0.9,
            cex.axis=0.9, xlim=c(as.Date("1996-01-01"),

```

```

      as.Date("2012-01-01")), xaxs="i", xaxt="n")
axdates<-c("1996-01-01", "2000-01-01", "2004-01-01", "2008-01-01",
           "2012-01-01")
axis(1, as.Date(axdates), labels=c("1996", "2000", "2004", "2008", "2012"),
     cex.axis=0.9)

```

prepData

Prepares concentration data and continuous ancillary data

Description

prepData is usually called from within [fitswavecav](#) but can be invoked directly. It performs some date calculations, removes rows with missing values for concentration or continuous variables, and returns the the concentration and continuous ancillary data to be used by [fitswavecav](#) and its other internal functions.

Usage

```

prepData(cdat, cavdat, yrstart, yrend, dcol, pnames,
         iwca, qwcols)

```

Arguments

cdat	is the concentration data.
cavdat	is the continuous (daily) ancillary data.
yrstart	is the starting year of the analysis (treated as January 1 of that year). Zero means the start date will be determined by the start date of cavdat, the continuous ancillary data.
yrend	is the ending year of the analysis (treated as December 31 of that year). Zero means the end date will be determined by the end date of cavdat, the continuous ancillary data.
dcol	is the column name for the dates, should be the same for both cdat and cavdat.
pnames	are the parameters (water-quality constituents) to analyze (if using USGS parameters, omit the the starting 'P', such as "00945" for sulfate).
iwca	is a character variable indicating which continuous ancillary variables to include, if none use iwca=c("none").
qwcols	is a character vector with the beginning of the column headers for remarks code (default is R), and beginning of column headers for concentration data (default is P for parameter).

Value

a list. The first element is the concentration data with additional date information, missing values removed, and extra columns removed. The second element is the continuous ancillary data with additional date information, missing values removed, and extra columns removed.

Author(s)

Aldo V. Vecchia and Karen R. Ryberg

Examples

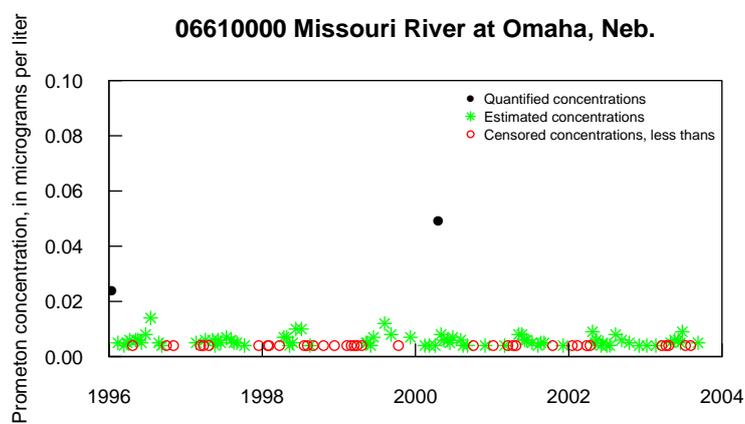
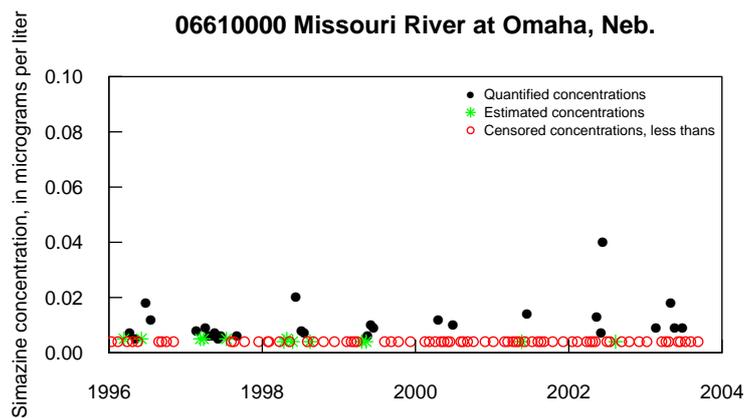
```
data(swData)
modMoRivOmaha<-combineData(qwdat=qwMoRivOmaha, cqwdat=cqwMoRivOmaha)
preppedDat <- prepData(modMoRivOmaha, cqwMoRivOmaha, yrstart=1995,
  yrend=2003, dcol="dates", pnames=c("04035", "04037", "04041"),
  iwca=c("flowa30","flowa1"), qwcols=c("R","P"))
```

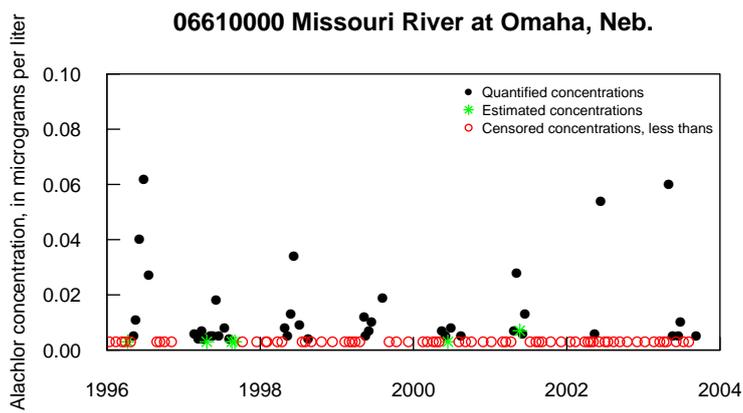
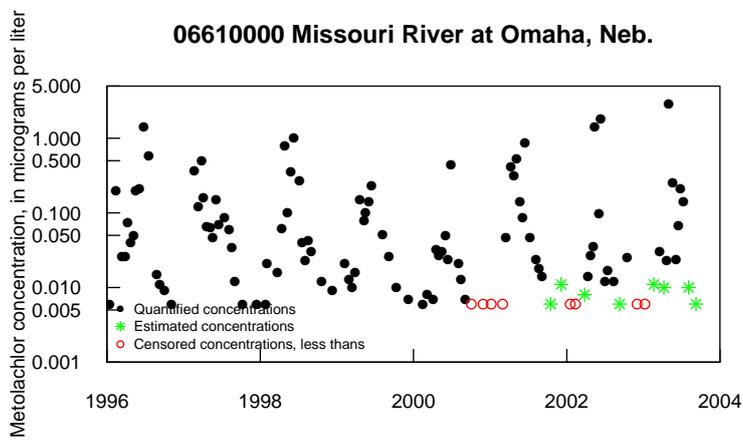
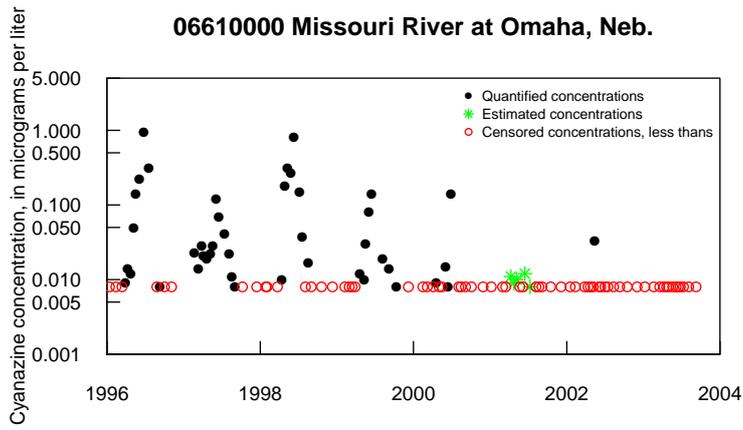
qwMoRivOmaha

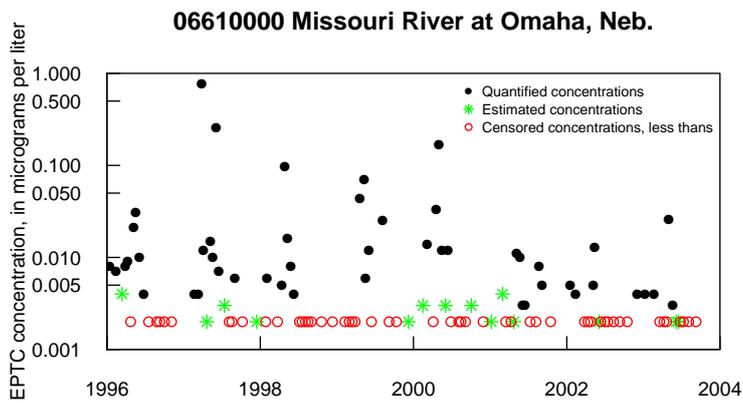
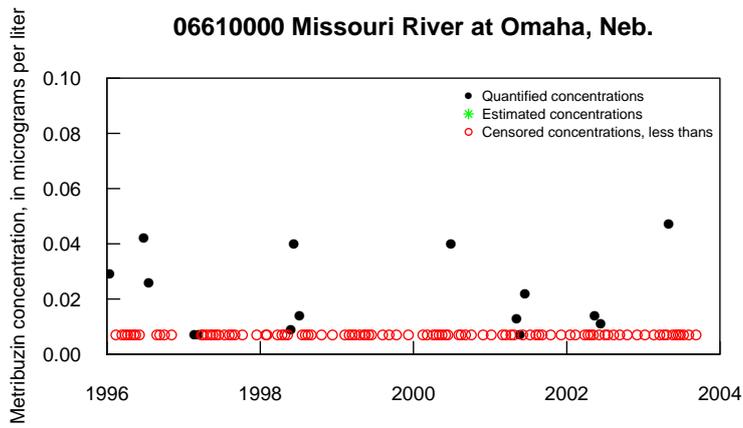
Water-quality data for 06610000 Missouri River at Omaha, Nebr.

Description

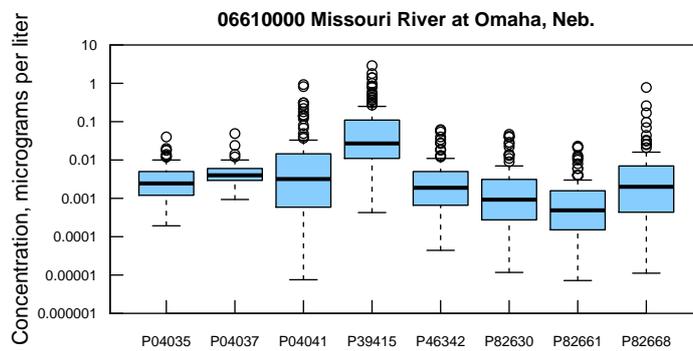
Scatterplots of water-quality data for 06610000 Missouri River at Omaha, Nebr.







Boxplot of data using regression on order statistics (Helsel, 2005; Lee, 2010) to estimate censored values.



Censored values estimated using regression on order statistics

Usage

qwMoRivOmaha

Format

A data frame containing 115 water-quality samples for eight chemical constituents. There are 20 variables.

staid	character	USGS Station identification number
dates	date	Date water-quality sample collected
times	numeric	Time sample was collected
R04035	character	Remark code (blank, _, <, or E)
P04035	numeric	Simazine, water, filtered, recoverable, micrograms per liter
R04037	character	Remark code (blank, _, <, or E)
P04037	numeric	Prometon, water, filtered, recoverable, micrograms per liter
R04041	character	Remark code (blank, _, <, or E)
P04041	numeric	Cyanazine, water, filtered, recoverable, micrograms per liter
R39415	character	Remark code (blank, _, <, or E)
P39415	numeric	Metolachlor, water, filtered, recoverable, micrograms per liter
R46342	character	Remark code (blank, _, <, or E)
P46342	numeric	Alachlor, water, filtered, recoverable, micrograms per liter
R82630	character	Remark code (blank, _, <, or E)
P82630	numeric	Metribuzin, water, filtered, recoverable, micrograms per liter
R82661	character	Remark code (blank, _, <, or E)
P82661	numeric	Trifluralin, water, filtered (0.7 micron glass fiber filter), recoverable, micrograms per liter
R82668	character	Remark code (blank, _, <, or E)
P82668	numeric	EPTC, water, filtered (0.7 micron glass fiber filter), recoverable, micrograms per liter

Details

Chemical concentration data are in the columns that start with a P and are followed by a number. Qualification codes for the concentration data are in the columns that start with an R followed by the same numbers as the associated concentration data. For example, column P04035 indicates simazine data, 04035, being the U.S. Geological Survey parameter code for simazine. The qualification codes for the simazine concentrations are found in the column R04035, indicating a U.S. Geological Survey remark code. Remark codes include _ or nothing, indicating no qualification of the value in the associated concentration field; <, indicating a censored value that is less than the number reported in the associated concentration field; and E, indicating that the value has been estimated. See Oblinger Childress and others (1999) for information on the remark codes used by the U.S. Geological Survey.

Source

Data provided by Patrick Phillips, U.S. Geological Survey, New York Water Science Center.

References

Oblinger Childress, C.J., Foreman, W.T., Connor, B.F., and Maloney, T.J., 1999, New reporting procedures based on long-term method detection levels and some considerations for interpretations of water-quality data provided by the U.S. Geological Survey Open-File Report 99-193, 19 p. (Also available at http://water.usgs.gov/owq/OFR_99-193/index.html.)

Examples

```

data(swData)

# summary of water-quality concentrations
apply(qwMoRivOmaha[,grep("P[[:digit:]]",dimnames(qwMoRivOmaha)[[2]])],
2, summary)

# simple boxplot of water-quality concentrations
rosBoxPlot(qwMoRivOmaha, qwcols=c("R", "P"))

# same boxplot function with many additional plotting arguments
rosBoxPlot(qwMoRivOmaha, site="06610000 Missouri River at Omaha, Nebr.",
log="y", yaxt="n", ylim=c(0.000001, 10), qwcols=c("R", "P"),
ylab=c("Concentration, micrograms per liter"), col="skyblue1",
cex.axis=0.7, cex.sub=0.8, par(tcl=0.5, las=1,
yaxs="i",
mgp=c(3,0.5,0),
mar=c(5,5,2,2),
cex.main=0.9))
axis(2, at=c(0.000001, 0.00001, 0.0001, 0.001, 0.01, 0.1, 1, 10),
labels=c("0.000001", "0.00001", "0.0001", "0.001", "0.01",
"0.1", "1", "10"), cex.axis=0.7)

# scatter plot of Simazine concentrations
cenScatPlot(qwMoRivOmaha, pname="04035")

# scatter plot with many additional plotting arguments
par(las=1, tcl=0.5)
cenScatPlot(qwMoRivOmaha, pname="04035",
site="06610000 Missouri River at Omaha, Nebr.",
ylabel="Simazine concentration, in micrograms per liter",
legcex=0.7, qwcols=c("R", "P"),
ylim=c(0,0.1), yaxs="i", cex.lab=0.9, cex.axis=0.9,
xlim=c(as.Date("1996-01-01"), as.Date("2004-01-01")),
xaxs="i", xaxt="n")
axdates<-c("1996-01-01", "1998-01-01", "2000-01-01", "2002-01-01",
"2004-01-01")
axis(1, as.Date(axdates), labels=c("1996", "1998", "2000", "2002",
"2004"), cex.axis=0.9)

# Prometon scatter plot
cenScatPlot(qwMoRivOmaha, pname="04037",
site="06610000 Missouri River at Omaha, Nebr.",
ylabel="Prometon concentration, in micrograms per liter",
legcex=0.7, qwcols=c("R", "P"),
ylim=c(0,0.1), yaxs="i", cex.lab=0.9, cex.axis=0.9,
xlim=c(as.Date("1996-01-01"),
as.Date("2004-01-01")), yaxs="i",
xaxt="n")
axdates<-c("1996-01-01", "1998-01-01", "2000-01-01", "2002-01-01",
"2004-01-01")
axis(1, as.Date(axdates), labels=c("1996", "1998", "2000", "2002",
"2004"), cex.axis=0.9)

# Cyanazine scatter plot
cenScatPlot(qwMoRivOmaha, pname="04041",

```

```

site="06610000 Missouri River at Omaha, Nebr.",
ylabel="Cyanazine concentration, in micrograms per liter",
legcex=0.7, qwcols=c("R", "P"),
ylim=c(0.001,5), yaxs="i", cex.lab=0.9, cex.axis=0.9,
xlim=c(as.Date("1996-01-01"),
        as.Date("2004-01-01")), xaxs="i",
xaxt="n", log="y")
axdates<-c("1996-01-01", "1998-01-01", "2000-01-01", "2002-01-01",
           "2004-01-01")
axis(1, as.Date(axdates), labels=c("1996", "1998", "2000", "2002",
                                   "2004"), cex.axis=0.9)

# Metolachlor scatter plot
cenScatPlot(qwMoRivOmaha, pname="39415",
            site="06610000 Missouri River at Omaha, Nebr.",
            ylabel="Metolachlor concentration, in micrograms per liter",
            legcex=0.7, qwcols=c("R", "P"),
            ylim=c(0.001,5), yaxs="i", cex.lab=0.9, cex.axis=0.9,
            xlim=c(as.Date("1996-01-01"),
                    as.Date("2004-01-01")), xaxs="i",
            xaxt="n", log="y", legpos="bottomleft")
axdates<-c("1996-01-01", "1998-01-01", "2000-01-01", "2002-01-01",
           "2004-01-01")
axis(1, as.Date(axdates), labels=c("1996", "1998", "2000", "2002",
                                   "2004"), cex.axis=0.9)

# Alachlor scatter plot
cenScatPlot(qwMoRivOmaha, pname="46342",
            site="06610000 Missouri River at Omaha, Nebr.",
            ylabel="Alachlor concentration, in micrograms per liter",
            legcex=0.7, qwcols=c("R", "P"),
            ylim=c(0, 0.1), yaxs="i", cex.lab=0.9, cex.axis=0.9,
            xlim=c(as.Date("1996-01-01"),
                    as.Date("2004-01-01")), xaxs="i",
            xaxt="n")
axdates<-c("1996-01-01", "1998-01-01", "2000-01-01", "2002-01-01",
           "2004-01-01")
axis(1, as.Date(axdates), labels=c("1996", "1998", "2000", "2002",
                                   "2004"), cex.axis=0.9)

# Metribuzin scatter plot
cenScatPlot(qwMoRivOmaha, pname="82630",
            site="06610000 Missouri River at Omaha, Nebr.",
            ylabel="Metribuzin concentration, in micrograms per liter",
            legcex=0.7, qwcols=c("R", "P"),
            ylim=c(0, 0.1), yaxs="i", cex.lab=0.9, cex.axis=0.9,
            xlim=c(as.Date("1996-01-01"),
                    as.Date("2004-01-01")), xaxs="i",
            xaxt="n")
axdates<-c("1996-01-01", "1998-01-01", "2000-01-01", "2002-01-01",
           "2004-01-01")
axis(1, as.Date(axdates), labels=c("1996", "1998", "2000", "2002",
                                   "2004"), cex.axis=0.9)

# Trifluralin scatter plot
cenScatPlot(qwMoRivOmaha, pname="82661",
            site="06610000 Missouri River at Omaha, Nebr.",

```

```

      ylabel="Trifluralin concentration, in micrograms per liter",
      legcex=0.7, qwcols=c("R", "P"),
      ylim=c(0,0.03), yaxs="i", cex.lab=0.9, cex.axis=0.9,
      xlim=c(as.Date("1996-01-01"),
             as.Date("2004-01-01")), xaxs="i",
             xaxt="n")
axdates<-c("1996-01-01", "1998-01-01", "2000-01-01", "2002-01-01",
           "2004-01-01")
axis(1, as.Date(axdates), labels=c("1996", "1998", "2000", "2002",
                                   "2004"), cex.axis=0.9)

# EPTC scatter plot
cenScatPlot(qwMoRivOmaha, pname="82668",
            site="06610000 Missouri River at Omaha, Nebr.",
            ylabel="EPTC concentration, in micrograms per liter",
            legcex=0.7, qwcols=c("R", "P"),
            ylim=c(0.001,1), yaxs="i", cex.lab=0.9, cex.axis=0.9,
            xlim=c(as.Date("1996-01-01"),
                   as.Date("2004-01-01")), xaxs="i",
                   xaxt="n", log="y")
axdates<-c("1996-01-01", "1998-01-01", "2000-01-01", "2002-01-01",
           "2004-01-01")
axis(1, as.Date(axdates), labels=c("1996", "1998", "2000", "2002",
                                   "2004"), cex.axis=0.9)

```

rosBoxPlot

Boxplot of water-quality data

Description

Function to create boxplots of water-quality data that include censored values.

Usage

```
rosBoxPlot(data, site = "", qwcols = c("R", "P"), ...)
```

Arguments

data	is the dataset with columns that begin with P followed by a number indicating concentration data and columns that begin with R followed by numbers that match those of the concentration data indicating qualification codes. See example data sets for more information about the data format, IIIRivValleyCty and qwMoRivOmaha .
qwcols	is a character vector with the beginning of the column headers for remarks code (default is R), and beginning of column headers for concentration data (default is P for parameter).
site	is a label for the plot title indicating the site where the water-quality samples were collected.
...	arguments to be passed to plot method.

Details

This function determines the columns within the data set that have concentration data, based on them having column headings that start with P (or a user-specified indicator in the second element of the `qwcols` argument) followed by a number. The function determines the associated remark, or qualification columns, based on them having column headings that start with R (or a user-specified indicator in the first element of the `qwcols` argument) followed by numbers that match the associated concentration data. Then it determines which values are censored, indicated by a less than symbol in the R columns, performs regression on order statistics, `ros`, using the NADA package, and estimates values for the censored concentrations for constituents with less than 90-percent censoring. The water-quality concentrations are then depicted by boxplots.

Value

a boxplot

Note

The regression on order statistics function in R package NADA (Lee, 2012), `ros`, is an implementation of a regression on order statistics designed for multiply-censored analytical-chemistry data (Helsel, 2005). The method assumes data contains zero to many left-censored (less-than) values. For highly censored data, `ros` may produce a warning message. Such as,

```
Warning messages: 1: In ros(my.list$obs,
  my.list$cen) : Input > 80% censored -- Results are
  tenuous.
```

The boxplot will still be generated, but the user should consider the warning message when interpreting the plots. See Oblinger Childress and others (1999) for information on the remark codes used by the U.S. Geological Survey.

Author(s)

Karen R. Ryberg

References

Helsel, D.R., 2005, Nondetects and data analysis: New York, John Wiley and Sons.

Lee, Lopaka, 2012, Nondetects and data analysis for environmental data: R package version 1.5-4, <http://CRAN.R-project.org/package=NADA>.

Oblinger Childress, C.J., Foreman, W.T., Connor, B.F., and Maloney, T.J., 1999, New reporting procedures based on long-term method detection levels and some considerations for interpretations of water-quality data provided by the U.S. Geological Survey: U.S. Geological Survey Open-File Report 99-193, 19 p. (Also available at http://water.usgs.gov/owq/OFR_99-193/index.html.)

Examples

```
data(swData)
# summary of water-quality concentrations
apply(I11RivValleyCty[, grep("P[[:digit:]]",
  dimnames(I11RivValleyCty)[[2]])], 2, summary)
# simple boxplot of water-quality concentrations
rosBoxPlot(I11RivValleyCty)
# same boxplot function with many additional plotting arguments
```

```

rosBoxPlot(IllRivValleyCty,
           site="05586100 Illinois River at Valley City, IL",
           log="y", yaxt="n", ylim=c(0.0000001, 1), qwcols=c("R", "P"),
           ylab=c("Concentration, micrograms per liter"),
           col="skyblue1", cex.axis=0.7, cex.sub=0.8,
           par(tcl=0.5, las=1, yaxs="i", mgp=c(3, 0.5, 0),
              mar=c(5, 5, 2, 2),cex.main=0.9))
axis(2,
     at=c(0.0000001, 0.000001, 0.00001, 0.0001, 0.001, 0.01, 0.1, 1),
     labels=c("0.0000001", "0.000001", "0.00001", "0.0001", "0.001", "0.001",
              "0.01", "0.1", "1"), cex.axis=0.7)

```

seawaveQPlots

Internal function that generates plots of data and model results.

Description

seawaveQPlots is usually called from within `fitMod` but can be invoked directly. It generates plots of data and model results, as well as diagnostic plots, and returns the observed and predicted concentrations so that users may plot the concentrations using their own functions.

Usage

```

seawaveQPlots(stpars, cmxt, tseas, tseaspr, tndlin,
              tndlinpr, cdatsub, cavdat, cavmat, clog, centmp,
              yrstart, yrend, tyr, tyrpr, pnames, tanm, mclass = 1)

```

Arguments

stpars	is a matrix of information about the best seawaveQ model for the concentration data, see examplestpars .
cmxt	is the decimal season of maximum chemical concentration.
tseas	is the decimal season of each concentration value in cdatsub.
tseaspr	is the decimal season date used to model concentration using the continuous data set cavdat.
tndlin	is the decimal time centered on the midpoint of the trend for the sample data, cdatsub.
tndlinpr	is is the decimal time centered on the midpoint of the trend for the continuous data, cavdat.
cdatsub	is the concentration data
cavdat	is the continuous (daily) ancillary data
cavmat	is a matrix containing the continuous ancillary variables.
clog	is a vector of the base-10 logarithms of the concentration data.
centmp	is a logical vector indicating which concentration values are censored.
yrstart	is the starting year of the analysis (treated as January 1 of that year).
yrend	is the ending year of the analysis (treated as December 31 of that year).
tyr	is a vector of decimal dates for the concentration data
tyrpr	is a vector of decimal dates for the continuous ancillary variables.

pnames	is the parameter (water-quality constituents) to analyze (if using USGS parameters, omit the the starting 'P', such as "00945" for sulfate).
tanm	is an a character identifier that names the trend analysis run. It is used to label output files.
mclass	has not been implemented yet, but will provide additional model options.

Value

a pdf file containing plots of the data and modeled concentrations and regression diagnostic plots and a list containing the observed concentrations (censored and uncensored) and the predicted concentrations used for the plot.

Note

The plotting position used for representing censored values in the plots produced by [seawaveQPlots](#) is an important consideration for interpreting model fit. Plotting values obtained by using the censoring limit, or something smaller such as one-half of the censoring limit, produce plots that are difficult to interpret if there are a large number of censored values. Therefore, to make the plots more representative of diagnostic plots used for standard (non-censored) regression, a method for substituting randomized residuals in place of censored residuals was used. If a log-transformed concentration is censored at a particular limit, $\log C < L$, then the residual for that concentration is censored as well, $\log C - \text{fitted}(\log C) < L - \text{fitted}(\log C) = \text{rescen}$. In that case, a randomized residual was generated from a conditional normal distribution

```
resran <- scl * qnorm(runif(1) * pnorm(rescen / scl)),
```

where *scl* is the scale parameter from the survival regression model, *pnorm* is the R function for computing cumulative normal probabilities, *runif* is the R function for generating a random variable from the uniform distribution, and *qnorm* is the R function for computing quantiles of the normal distribution. Under the assumption that the model residuals are uncorrelated, normally distributed random variables with mean zero and standard deviation *scl*, the randomized residuals generated in this manner are an unbiased sample of the true (but unknown) residuals for the censored data. This is an application of the probability integral transform (Mood and others, 1974) to generate random variables from continuous distributions. The plotting position used a censored concentration is $\text{fitted}(\log C) + \text{resran}$. Note that each time a new model fit is performed, a new set of randomized residuals is generated and thus the plotting positions for censored values can change.

Author(s)

Aldo V. Vecchia and Karen R. Ryberg

References

Mood, A.M., Graybill, F.A., and Boes, D.C., 1974, Introduction to the theory of statistics (3d ed.): New York, McGraw-Hill, Inc., 564 p.

Examples

```
data(swData)
myPlots <- seawaveQPlots(stpars=examplestpars, cmxt=0.4808743,
tseas=exampletseas, tseaspr=exampletseaspr, tndlin=exampletndlin,
tndlinpr=exampletndlinpr, cdatsub=examplecdatsub, cavdat=examplecavdat,
```

```
cavmat=examplecavmat, clog=exampleclog, centmp=examplecentmp,  
yrstart=1995, yrend=2003, tyr=exampletyr, tyrpr=exampletyrpr,  
pnames=c("04041"), tanm="examplePlots04041")
```

Index

- *Topic **datagen**
 - compwaveconv, 6
 - *Topic **datasets**
 - cqwMoRivOmaha, 7
 - examplecavdat, 8
 - examplecavmat, 9
 - examplecdatsub, 9
 - IllRivValleyCty, 22
 - qwMoRivOmaha, 28
 - *Topic **dplot**
 - seawaveQPlots, 36
 - *Topic **hplot**
 - cenScatPlot, 3
 - rosBoxPlot, 34
 - seawaveQPlots, 36
 - *Topic **manip**
 - combineData, 4
 - prepData, 27
 - *Topic **models**
 - fitMod, 18
 - fitswavecav, 19
 - *Topic **package**
 - seawaveQ-package, 2
 - *Topic **regression**
 - fitMod, 18
 - fitswavecav, 19
 - *Topic **survival**
 - fitMod, 18
 - fitswavecav, 19
 - *Topic **ts**
 - fitMod, 18
 - fitswavecav, 19
- cenScatPlot, 3
combineData, 4
compwaveconv, 6
cqwMoRivOmaha, 5, 7, 10, 11
- examplecavdat, 8
examplecavmat, 9
examplecdatsub, 9
examplecentmp, 11
exampleclog, 12
exampleqwcols, 12
- examplestpars, 13, 19, 36
exampletndlin, 14
exampletndlinpr, 15
exampletseas, 15
exampletseaspr, 16
exampletyr, 17
exampletyrpr, 17
- fitMod, 12, 13, 18, 21, 36
fitswavecav, 8, 9, 11, 13, 14, 18, 19, 27
- IllRivValleyCty, 3, 5, 22, 34
- legend, 3
- plot, 3, 34
prepData, 13, 21, 27
- qwMoRivOmaha, 3, 5, 10, 11, 28, 34
- ros, 35
rosBoxPlot, 34
- seawaveQ (seawaveQ-package), 2
seawaveQ-package, 2
seawaveQPlots, 12, 14–17, 19, 36, 37