Package 'seawaveQ'

August 4, 2020

Version 2.0.1

Date 2020-07-31

Title SEAWAVE-Q Model

Author Karen R. Ryberg [aut, cre],
Aldo V. Vecchia [aut],
Benjamin York [ctb]

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

Depends R (>= 3.0.0), survival, rms, lubridate, reshape2, plyr

Suggests waterData

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.

License Unlimited | file LICENSE

URL https://dx.doi.org/10.3133/ofr20131255

LazyData true

Copyright This software is in the public domain because it contains
materials that originally came from the United States
Geological Survey, an agency of the United States Department of
Interior. For more information, see the official USGS copyright
policy at
https://www.usgs.gov/visual-id/credit_usgs.html#copyright

NeedsCompilation no

RoxygenNote 7.1.1

Repository CRAN

Date/Publication 2020-08-04 12:30:06 UTC

R topics documented:

---

seawaveQ–package | *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*

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), Vecchia and others (2009), Ryberg and others (2010), Ryberg and others (2014), Ryberg and Gilliom (2015), and Oelsner and others (2017) and for applications of the model.

Details

Package:	seawaveQ
Type:	Package
Version:	2.0.1
Date:	2020–07–31
License:	Unlimited | file LICENSE

Author(s)

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

References

Oelsner, G.P., Sprague, L.A., Murphy, J.C., Zuellig, R.E., Johnson, H.M., Ryberg, K.R., Falcone, J.A., Stets, E.G., Vecchia, A.V., Riskin, M.L., De Cicco, L.A., Mills, T.J., and Farmer, W.H., 2017, Water-quality trends in the nation's rivers and streams, 1972–2012—Data preparation, statistical methods, and trend results: U. S. Geological Survey Scientific Investigations Report 2017–5006, 136 p., https://doi.org/10.3133/sir20175006.

Ryberg, K.R. and Gilliom, R.J., 2015, Trends in pesticide concentrations and use for major rivers of the United States: Science of The Total Environment, v. 538, p. 431–444, https://doi.org/10.1016/j.scitotenv.2015.06.095.

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, https://dx.doi.org/10.3133/ofr20131255.

Ryberg, K.R. and York, B.C., 2020, 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, Version 2.0.0: U.S. Geological Survey Open-File Report 2020–1082, 25 p., with 4 appendixes.

Ryberg, K.R., Vecchia, A.V., Gilliom, R.J., and Martin, J.D., 2014, Pesticide trends in major rivers of the United States, 1992–2010: U.S. Geological Survey Scientific Investigations Report 2014–5135, 74 p., https://pubs.er.usgs.gov/publication/sir20145135.

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., https://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., https://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, no. 24, p. 9096–9102.

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, <https://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 column heading (parameter name) for the particular water-quality constituent to be plotted (omit 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 value has been estimated; and a less than symbol, "<", indicates the value has been censored as less than a minimum reporting level. See Oblinger Childress and others (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., https://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.

Usage

```
combineData(qwdat, cqwdat, qwcols = c("staid", "dates", "R", "P"))
```

Arguments

qwdat
: is the dataset containing water-quality sample data with columns that begin with a P (or other user-defined indicator) followed by alphanumeric characters. These columns are concentration data. In addition there need to be columns that begin with an R (or other user- defined indicator) followed by alphanumeric characters that match those of the associated concentration data. The R columns contain data qualification codes. See example datasets for more information about the data format, IllRivValleyCty and qwMoRivOmaha. This package assumes that the data are in micrograms per liter.

cqwdat
: is the dataset containing variables that can be used as explanatory variables for the seawaveQ model. See example dataset for more information about the data format cqwMoRivOmaha. These are daily values with no missing values allowed between the first and the last date in the dataset.

qwcols
: is a character vector with column headings for a station (location) identifier, a dates column identifier, beginning of column headers for remarks code (default is R), and beginning of column headers for concentration data (default is P for parameter).

Format

A data frame with the number of rows equal to the number of rows in the data frame indicated by qwdat. The number of columns depend on the two input data frames. Minimally there will be a station identification column, a dates column, a column of qualification codes, and a column of water-quality data.

Value

A data frame

Note

The columns indicated by qwcols[1:2] are used to combine the datasets. The first column is the station identifier and the second column is the date column. These two column headings must be the same in the two datasets being combined and the dates in the datasets being combined must be of class Date and use the same format.

Author(s)

Karen R. Ryberg and Aldo V. Vecchia

Examples

```
data(swData)
MoRivOmaha <- combineData(qwdat = qwMoRivOmaha, cqwdat = cqwMoRivOmaha,
qwcols = c("staid", "dates", "R", "P"))
```

---

compwaveconv   *Seasonal Wave Computation*

Description

Function to compute seasonal wave.

Usage

```
compwaveconv(cmaxt, jmod, hlife)
```

Arguments

cmaxt	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

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, https://doi.org/10.1111/j.1752-1688.2008.00225.x.

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(cmaxt = 0.483, jmod = 2, hlife = 4)
swave[floor(360 * dseas)]
plot(seq(0, 1, 1/360), swave, typ = "l")
```

---

cqwMoRivOmaha	*Continuously monitored (daily) data for 06610000 Missouri River at Omaha, Neb.*

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 *Example continuous ancillary variable data.*

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 fitswavecav

| 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,
                      iwcav=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   *Example continuous ancillary variable matrix.*

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,
                 iwcav=c("flowa30","flowa1"), dcol="dates",
                 qwcols=c("R","P"))
```

Examples

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

---

examplecdatsub   *Example water-quality data.*

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.

yrc	numeric	Year
moc	numeric	Month
dac	numeric	Day
jdayc	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,
                 iwcav=c("flowa30","flowa1"), dcol="dates",
                 qwcols=c("R","P"))
```

See Also

qwMoRivOmaha cqwMoRivOmaha

Examples

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

---

examplecentmp   *Example logical vector.*

Description

This is an example of data that is passed internally to subfunctions of 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,
                    iwcav=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,
                    iwcav=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,
                    iwcav=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 15 columns.
column description
1 mclass, model class, 1 for linear models, 2 for those with restricted cubic splines
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

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,
                         iwcav=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,
                      iwcav=c("flowa30","flowa1"), dcol="dates",
                      qwcols=c("R","P"))
```

Examples

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

---

exampletndlinpr   *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 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,
                      iwcav=c("flowa30","flowa1"), dcol="dates",
                      qwcols=c("R","P"))
```

Examples

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

---

exampletseas   *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 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,
                    iwcav=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,
                    iwcav=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,
                    iwcav=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,
                    iwcav=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(
  cdatsub,
  cavdat,
  yrstart,
  yrend,
  tndbeg,
  tndend,
  tanm,
  pnames,
  qwcols,
  mclass = 1,
  numk = 4,
  plotfile = FALSE,
  textfile = FALSE
)
```

Arguments

cdatsub	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 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	indicates the class of model to use. A class 1 model is the the traditional SEAWAVE-Q model that has a linear time trend. A class 2 model is a newer option for longer trend periods that uses a set of restricted cubic splines on the time variable to provide a more flexible model. The default is 1. (Harrell, 2010, 2018).
numk	is the number of knots in the restricted cubic spline model (mclass = 2). The default is 4, and the recommended number is 3–7.
plotfile	is by default FALSE. True will write pdf files of plots to the user's file system.
textfile	is by default FALSE. True will write text output files to the user's file system. These files are useful for detailed model comparisons, documenting session information, and for model archives.

Value

A PDF file (if plotfile is TRUE) 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 or seawaveQPlots2 to generate the plots.

Author(s)

Aldo V. Vecchia and Karen R. Ryberg

References

Allison, P.D., 1995, Survival analysis using the SAS system—A practical guide: Cary, N.C., SAS Institute, Inc., 304 p.

Harrell, F.E., Jr., 2010, Regression modeling strategies—With applications to linear models, logistic regression, and survival analysis: New York, Springer-Verlag, 568 p.

Harrell, F.E., Jr., 2018, rms—Regression modeling strategies: R package version 5.1-2, https://CRAN.R-project.org/package=rms.

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"), plotfile = FALSE,
textfile = FALSE)
```

---

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,
```

```
    iwcav = c("none"),
    dcol = "dates",
    qwcols = c("R", "P"),
    mclass = 1,
    numk = 4,
    alpha = 0.1,
    bootRCS = FALSE,
    nboot = 1000,
    plotfile = FALSE,
    textfile = FALSE
)
```

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 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 iwcav=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	indicates the class of model to use. A class 1 model is the traditional SEAWAVE-Q model that has a linear time trend. A class 2 model is a newer option for longer trend periods that uses a set of restricted cubic splines on the time variable to provide a more flexible model.
numk	is the number of knots in the restricted cubic spline model. The default is 4, and the recommended number is 3–7.
alpha	is the significance level or alpha values for statistical significance and confidence intervals.

bootRCS		is a logical value indicating whether or not to perform block bootstrapping for an attained significance level for the trend with restricted cubic splines. No bootstrapping is performed for the linear trend model.
nboot		is the number of bootstrap replicates. A large number, 10,000, is recommended, but this takes a long time.
plotfile		is by default FALSE. True will write pdf files of plots to the user's file system.
textfile		is by default FALSE. True will write text output files to the user's file system. These files are useful for detailed model comparisons, documenting session information, and for model archives.

Format

The data frame returned as the first element of the output list has one row for each chemical analyzed and the number of columns depends on the number of continuous ancillary variables used. The general format is as follows:

pname	character	parameter analyzed
mclass	numeric	a value of 1 or 2
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
cmaxt	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[alpahnumeric]	numeric	coefficient(s) for the trend component(s) 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[alphanumeric]	numeric	standard error for the trend component(s)
se[alphanumeric]	numeric	0 or more standard errors for the continuous ancillary variables
pvaltnd[alphanumeric]	numeric	the *p*-value for the trend component(s)

The data frame returned as the sixth element of the output list has one row for each chemical analyzed. The general format for linear models is described in pesticideTrendCalcs. The format for restricted cubic spline models is as described as follows:

pname	character	parameter analyzed
mclass	numeric	A value of 1 or 2
baseConc	numeric	the concentration at the beginning of the trend period
endCon	numeric	the concentration at the end of the trend period
rcsctndPpor	numeric	the concentration trend in percent over the trend period
rcsctndOrigPORPercentBase	numeric	the conc. trend in original units over period of record (calc. based on percent per y
pvalrcstnd	numeric	*p*-value, attained significance level, based on bootstrapping
ctndlklhd	numeric	trend likelihood

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. The version in the 2.0.0 update to the R package has an option to use restricted cubic splines as a more flexible definition of the temporal trend.

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. The sixth element is a data frame that provides a summary of the trends with the columns pname (parameter name), mclass (a value of 1, indicating a linear trend model, a value of 2 for models using restricted cubic splines), and the columns describing the trends. For linear models see `pesticideTrendCalcs`. For models with restricted cubic splines, see the format section. See Ryberg and York (2020) for additional details.

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, 2018b). When the data are retrieved through the National Water Information System: Web Interface (`https://waterdata.usgs.gov/nwis`; U.S. Geological Survey, 2018a), 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 formats that can be modified to use with this function. For example, a user may have concentration values and qualification codes in one column, such as a column labeled simazine 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).

None of the variations of SEAWAVE is a simple model. The model complexity increases with flow anomalies and with the addition of restricted cubic splines. As the number of parameters in the model increases or the degree of censoring increases, the sample size must also increase. See Ryberg and others (2020) for more details on sample size.

Author(s)

Aldo V. Vecchia and Karen R. Ryberg

References

Ryberg, K.R. and York, B.C., 2020, 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 2020–1082, 25 p., with 4 appendixes.

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., <https://pubs.usgs.gov/sir/2010/5139/>.

U.S. Geological Survey, 2018a, National Water Information System: Web Interface, accessed July 7, 2018, at <https://waterdata.usgs.gov/nwis>.

U.S. Geological Survey, 2018b, Parameter code definition: National Water Information System: Web Interface, accessed July 18, 2018, at <https://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, <https://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)
myfitLinearTrend <- fitswavecav(cdat = modMoRivOmaha, cavdat = cqwMoRivOmaha,
tanm = "myfitLinearTrend", pnames = c("04035", "04037", "04041"), yrstart = 1995,
yrend = 2003, tndbeg = 1995, tndend = 2003, iwcav = c("flowa30", "flowa1"),
dcol = "dates", qwcols = c("R", "P"))
# trend model results
myfitLinearTrend[[1]]
# example regression call
myfitLinearTrend[[2]][[1]]
# first few lines of observed concentrations
head(myfitLinearTrend[[3]])
# first few lines of predicted concentrations
head(myfitLinearTrend[[4]])
# summary statistics for predicted concentrations
myfitLinearTrend[[5]]
```

```
# summary of trends
myfitLinearTrend[[6]]
myfitRCSTrend <- fitswavecav(cdat = modMoRivOmaha, cavdat = cqwMoRivOmaha,
tanm = "myfitRCSTrend", pnames = c("04035", "04037", "04041"), yrstart = 1995,
yrend = 2003, tndbeg = 1995, tndend = 2003, iwcav = c("flowa30", "flowa1"),
dcol = "dates", qwcols = c("R", "P"), mclass = 2, numk = 4, bootRCS = FALSE,
plotfile = FALSE, textfile = FALSE)
```
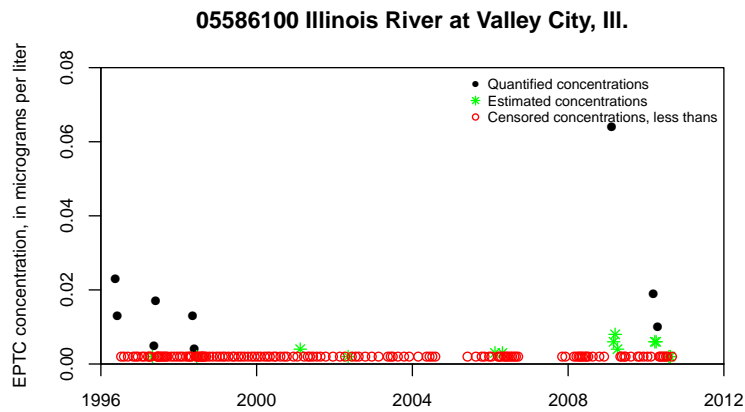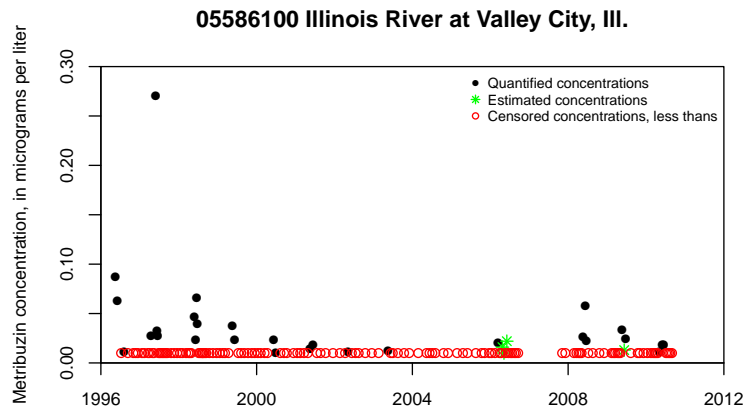

IllRivValleyCty   *Water-quality data for 05586100 Illinois River at Valley City, Ill.*

Description

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

05586100 Illinois River at Valley City, Ill.

05586100 Illinois River at Valley City, Ill.
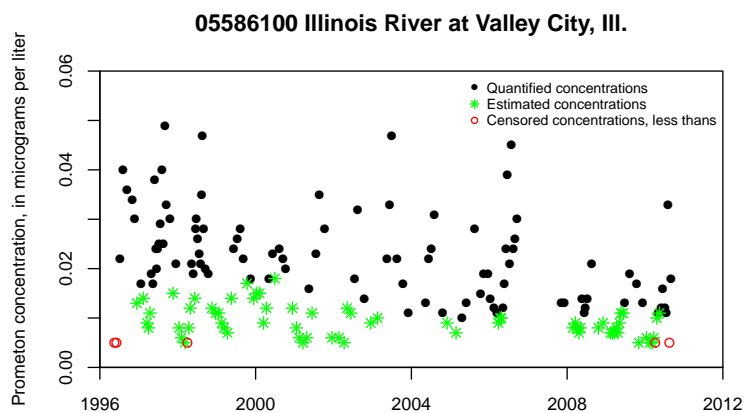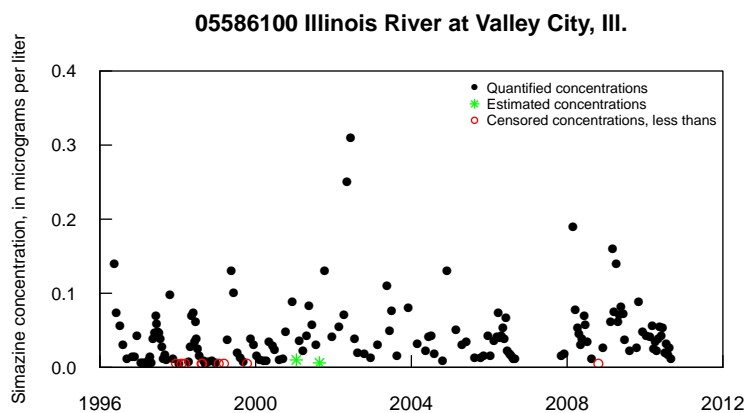
Usage

```
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
yrc	numeric	Year
moc	numeric	Month
dac	numeric	Day
jdayc	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 https://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)

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

```
# Simazine 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",
            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"))

# 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",
            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"))

# 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",
            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"))
```

---

loadCalculations   *Calculate annual loads and summarize trends*

Description

Function to calculate pesticide loads in kilograms per year and summarize trends.

Usage

```
loadCalculations(
  dailyDat,
  pestPredict,
  modRes,
  concTrends,
  yrtype = 1,
  alpha = 0.1
)
```

Arguments

dailyDat	is the daily streamflow data in the form of a data frame with three columns representing a station ID, date, and streamflow.
pestPredict	is the continuous (daily) estimation of pesticide concentrations for one or more pesticides at a single site. This should be in the form of the fourth element of the list returned by fitswavecav.
modRes	is the first element of the list returned by fitswavecav and includes the scale parameter for one or more pesticide trend models at a single site. The scale parameter is used in the bias correction.
concTrends	the SEAWAVE-Q trend in flow-normalized annual load. Cannot be different (computationally) from the trend in flow-normalized annual concentration when there is no trend in flow (Oelsner and others, 2017).
yrtype	allows one to calculate annual loads based on a calendar year or a water year, where a water year is the 12-month period October 1 through September 30 designated by the calendar year in which it ends. A yrtype of 1 represents a calendar year and is the default because that is the way the original model was developed. A yrtype of 2 represents a water year.
alpha	is the significance level or alpha value for statistical significance and confidence intervals.

Format

The first data frame returned has one row for each pesticide-year at a particular site and four columns. The general format is as follows:

pstaid	character	the station identification number
pcode	character	the parameter code for which a load was calculated
year or wyear	numeric	the year or water year for which a load was calculated
load	numeric	the load in kilograms per year

The second data frame returned has one row for each pesticide at a particular site and 11 columns. The general format is as follows:

pcode	character	the parameter code for which a load trend was calculated
mclass	numeric	a value of 1 or 2
mclass	numeric	a value of 1 or 2
alpha	numeric	a significance level
ltndPpor	numeric	the load trend in percent over the period of record
luciPpor	numeric	the load upper confidence interval for the trend in percent over the period of record
llciPpor	numeric	the load lower confidence interval for the trend in percent over the period of record
baseLoad	numeric	the base load, the load for the first year of trend period
ltndOrigPORPercentBase	numeric	the load trend in original units over the period of record (calculation based on percent per year and base load)
luciOrigPORPercentBase	numeric	the load trend upper confidence interval for the trend in original units over the period of record (calculation based on percent per year and base load)
llciOrigPORPercentBase	numeric	the load trend lower confidence interval for the trend in original units over the period of record (calculation based on percent per year and base load)
ltndlklhd	numeric	is the load trend likelihood

Details

Parameter load (mass) is the product of water-quality concentration (a mass per volume) and an associated streamflow rate (volume per time). This function generates an annual time series of pesticide loads on either a calendar year basis or a water year basis and summarizes load trends.

Value

Two data frames, the first contains the annual loads, the second contains the trend summary.

Note

In this load calculation function, daily pesticide concentration estimates provided by the fitswavecav function are corrected for retransformation bias (the concentration model is built on the base-10 logarithm of concentration; therefore, a bias correction is required when transforming back to the original units) and then used to calculate daily loads. The bias correction is based on the quasi-maximum likelihood estimator (Cohn and others, 1989) that was developed for natural logarithms, with an adjustment for the base-10 logarithm of the concentration. To calculate loads, the bias-corrected concentration estimates (assumed to be in micrograms per liter) are multiplied by daily streamflow and a constant, 0.892998605, which converts the load units (micrograms per liter * cubic feet per second) to kilograms per year. Daily loads are summed to annual values. See page 70 and equation 26 of Oelsner and others (2017) for further details regarding the load calculation and bias correction. Users may modify this function to convert to units other than kilograms per year.

Author(s)

Karen R. Ryberg

References

Cohn, T.A., DeLong, L.L., Gilroy, E.J., Hirsch, R.M., and Wells, D.K., 1989, Estimating constituent loads: Water Resources Research, v. 25, no. 5, p. 937–942.

Oelsner, G.P., Sprague, L.A., Murphy, J.C., Zuellig, R.E., Johnson, H.M., Ryberg, K.R., Falcone, J.A., Stets, E.G., Vecchia, A.V., Riskin, M.L., De Cicco, L.A., Mills, T.J., and Farmer, W.H., 2017, Water-quality trends in the Nation's rivers and streams, 1972–2012—Data preparation, statistical methods, and trend results (ver. 2.0, October 2017): U.S. Geological Survey Scientific Investigations Report 2017–5006, 136 p., https://doi.org/10.3133/sir20175006.

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, iwcav = c("flowa30", "flowa1"),
dcol = "dates", qwcols = c("R", "P"))
MoRivOmahaLoadsYr <- loadCalculations(cqwMoRivOmaha[, 1:3], myfit1[[4]],
myfit1[[1]], myfit1[[6]])
MoRivOmahaLoadsYr
```

---

pesticideTrendCalcs　　　　　*Summarize linear trends*

Description

Internal function to summarize the trend results.

Usage

```
pesticideTrendCalcs(
  tndbeg,
  tndend,
  ctnd,
  pval,
  alpha,
  setnd,
  scl,
  baseConc,
  mclass
)
```

Arguments

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.
ctnd	is the concentration trend, the coefficient on the time variable.
pval	is the p-value for the linear trend component.
alpha	is the significance level or alpha value for statistical significance and confidence intervals.
setnd	is the standard error for the linear trend component.
scl	is the scale factor from the survreg.object.
baseConc	is the base concentration, the median concentration (midpoint of the trend line) for the first year of the trend analysis .
mclass	indicates the class of model to use. A class 1 model is the the traditional SEAWAVE-Q model that has a linear time trend. A class 2 model is a newer option for longer trend periods that uses a set of restricted cubic splines on the time variable to provide a more flexible model.

Format

The data frame returned has one row for each chemical analyzed and the number of columns are defined as follows:

pname	character	parameter analyzed
mclass	numeric	a value of 1 or 2
alpha	numeric	a significance level
ctndPpor	numeric	the concentration trend in percent over the period of record
cuciPpor	numeric	the concentration upper confidence interval for the trend in percent over the period of record
clciPpor	numeric	the concentration lower confidence interval for the trend in percent over the period of record
baseConc	numeric	the base concentration, median concentration or midpoint of trend line, for first year of trend period
ctndOrigPORPercentBase	numeric	the concentration trend in original units over the period of record (calculation based on percent per year and base concentration)
cuciOrigPORPercentBase	numeric	the concentration trend upper confidence interval for the trend in original units over the period of record (calculation based on percent per year and base concentration)
clciOrigPORPercentBase	numeric	the concentration trend lower confidence interval for the trend in original units over the period of record (calculation based on percent per year and base concentration)
ctndlklhd	numeric	is the concentration trend likelihood

Details

pesticideTrendCalcs is called from within [fitswavecav](#)

Value

The data frame returned has one row for each chemical analyzed and summaries of the trend.

Note

Based on trend calculations used to display and summarize pesticide trends here https://nawqatrends.wim.usgs.gov/swtrends/. A likelihood value that is the functional equivalent of the two-sided p-value associated with the significance level of the trend was determined as follows: Likelihood = (1 - (p-value / 2)), where p-value is the p-value for the trend coefficient (Oelsner and others, 2017).

Author(s)

Karen R. Ryberg

References

Oelsner, G.P., Sprague, L.A., Murphy, J.C., Zuellig, R.E., Johnson, H.M., Ryberg, K.R., Falcone, J.A., Stets, E.G., Vecchia, A.V., Riskin, M.L., De Cicco, L.A., Mills, T.J., and Farmer, W.H., 2017, Water-quality trends in the Nation's rivers and streams, 1972–2012—Data preparation, statistical methods, and trend results (ver. 2.0, October 2017): U.S. Geological Survey Scientific Investigations Report 2017–5006, 136 p., https://doi.org/10.3133/sir20175006.

| 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 concentration and continuous ancillary data to be used by [fitswavecav](#) and its other internal functions.

Usage

```
prepData(cdat, cavdat, yrstart, yrend, dcol, pnames, iwcav, 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 starting 'P', such as "00945" for sulfate).
iwcav	is a character variable indicating which continuous ancillary variables to include, if none use iwcav=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)

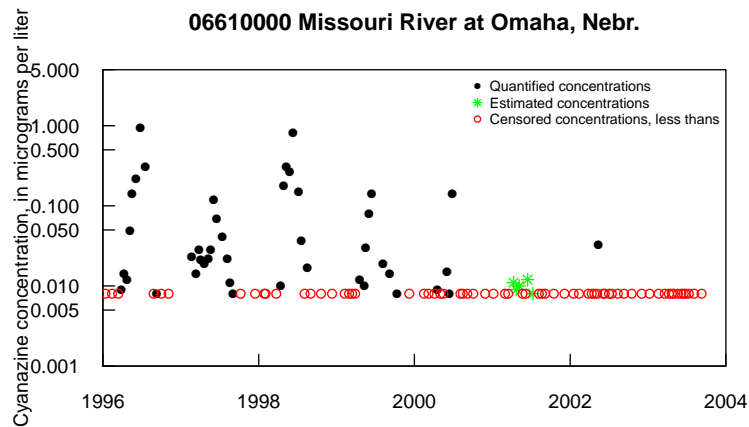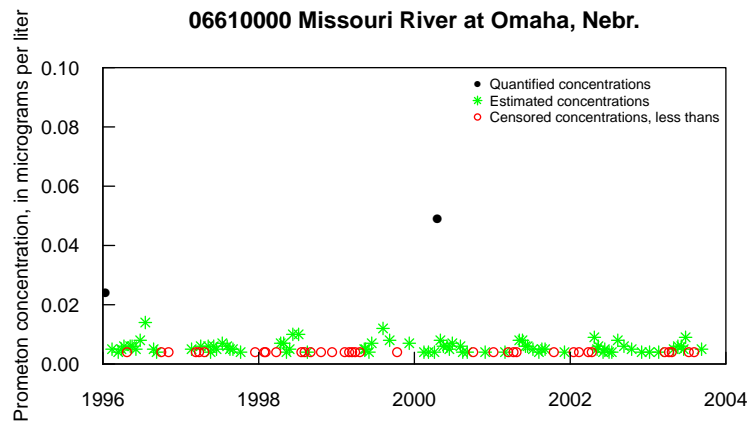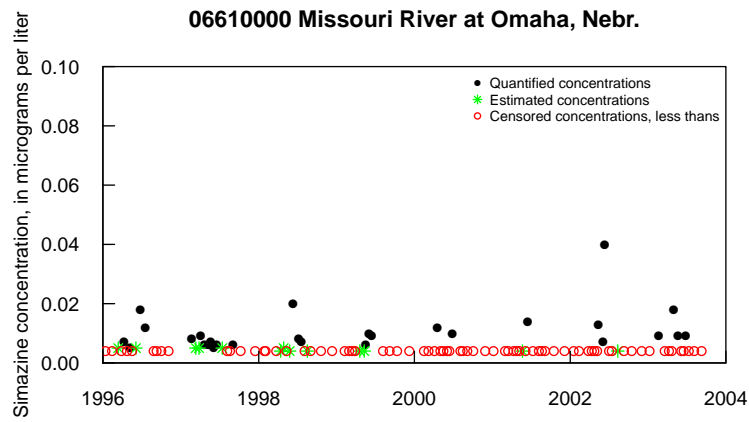Aldo V. Vecchia and Karen R. Ryberg

Examples

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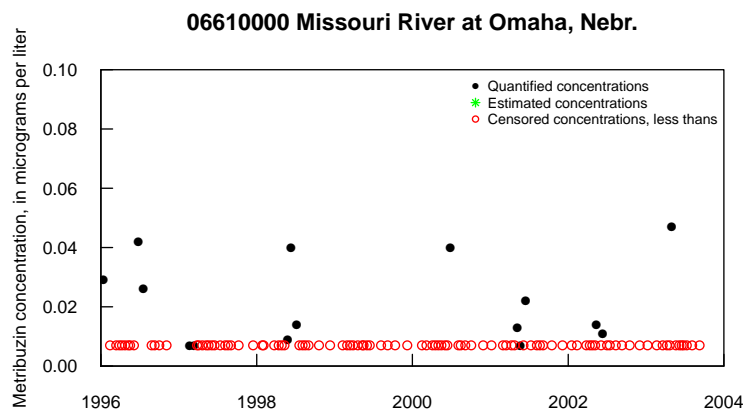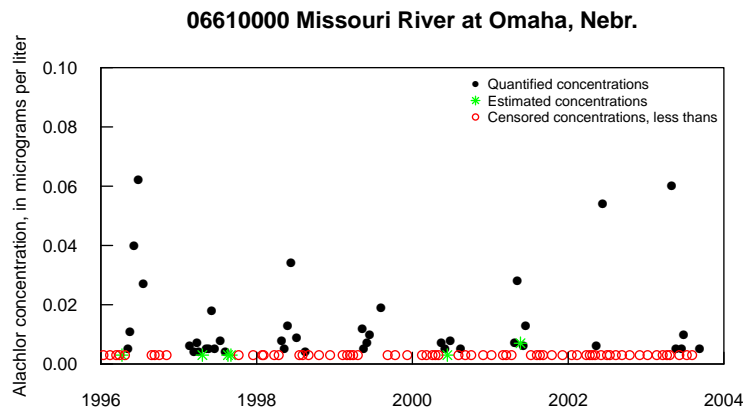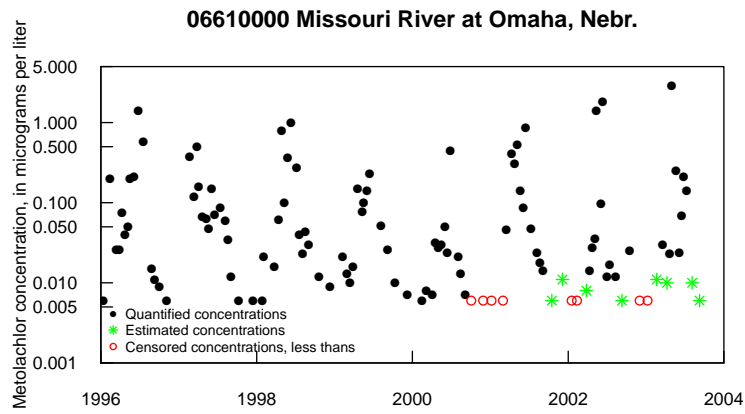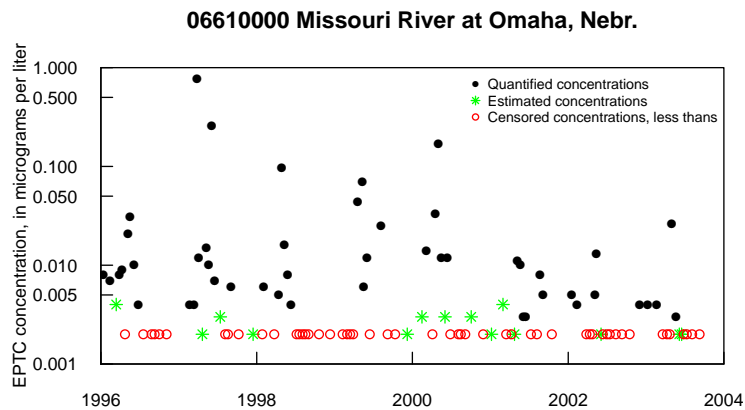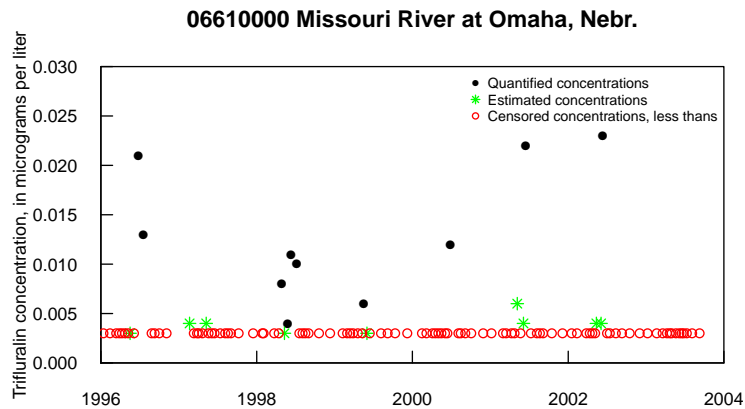
---

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

Description

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

06610000 Missouri River at Omaha, Nebr.

- Quantified concentrations
- Estimated concentrations
- Censored concentrations, less thans

06610000 Missouri River at Omaha, Nebr.

- Quantified concentrations
- Estimated concentrations
- Censored concentrations, less thans

06610000 Missouri River at Omaha, Nebr.

- Quantified concentrations
- Estimated concentrations
- Censored concentrations, less thans

06610000 Missouri River at Omaha, Nebr.

- Quantified concentrations
- Estimated concentrations
- Censored concentrations, less thans

06610000 Missouri River at Omaha, Nebr.

- Quantified concentrations
- Estimated concentrations
- Censored concentrations, less thans

06610000 Missouri River at Omaha, Nebr.

- Quantified concentrations
- Estimated concentrations
- Censored concentrations, less thans

06610000 Missouri River at Omaha, Nebr.

06610000 Missouri River at Omaha, Nebr.

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)

# 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",
            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"))

# 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",
            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"))

# 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",
            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"))

# 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",
            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"))

# 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",
                  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"))

     # 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",
                  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"))

     # 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",
                  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"))

     # 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",
                  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"))
```

---

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,
  cmaxt,
  tseas,
  tseaspr,
  tndlin,
  tndlinpr,
  cdatsub,
  cavdat,
  cavmat,
  clog,
  centmp,
  yrstart,
  yrend,
  tyr,
  tyrpr,
  pnames,
  tanm,
  mclass = 1,
  plotfile = FALSE
)
```

Arguments

stpars	is a matrix of information about the best seawaveQ model for the concentration data, see [examplestpars]().
cmaxt	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, cdatasub.
tndlinpr	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 starting 'P', such as "00945" for sulfate).
tanm	is 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.
plotfile	is by default FALSE. True will write pdf files of plots to the user's file system.

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, logC < L, then the residual for that concentration is censored as well, logC -fitted(logC) < L -fitted(logC) = 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 fitted(logC) + 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 (3rd ed.): New York, McGraw-Hill, Inc., 564 p.

Examples

```
data(swData)
myPlots <- seawaveQPlots(stpars=examplestpars, cmaxt=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", plotfile = FALSE)
```

---

seawaveQPlots2   *Internal function that generates plots of data and model results.*

Description

seawaveQPlots2 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. This is the
version for models that use restricted cubic splines.

Usage

```
seawaveQPlots2(
  stpars,
  cmaxt,
  tseas,
  tseaspr,
  tndrcs,
  tndrcspr,
  cdatsub,
  cavdat,
  cavmat,
  clog,
  centmp,
  yrstart,
  yrend,
  tyr,
  tyrpr,
  pnames,
  tanm,
  mclass = 2,
  numk,
  plotfile = FALSE
)
```

Arguments

stpars	is a matrix of information about the best seawaveQ model for the concentration data, see [examplestpars]().
cmaxt	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.
tndrcs	is the decimal time centered on the midpoint of the trend for the sample data, cdatasub, then converted to a linear tail-restricted cubic spline with a particular number of knots (Harrell, 2010, 2018).
tndrcspr	is the decimal time centered on the midpoint of the trend for the continuous data, cavdat, then converted to a linear tail-restricted cubic spline using the knots from tndrcs.
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 starting 'P', such as "00945" for sulfate).
tanm	is a character identifier that names the trend analysis run. It is used to label output files.
mclass	indicates the class of model to use. A class 1 model is the traditional SEAWAVE-Q model that has a linear time trend. A class 2 model is a newer option for longer trend periods that uses a set of restricted cubic splines on the time variable to provide a more flexible model.
numk	is the number of knots in the restricted cubic spline model. The default is 4, and the recommended number is 3–7.
plotfile	is by default FALSE. True will write pdf files of plots to the user's file system.

Value

A PDF file (if plotfile is TRUE) 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 seawaveQPlots2 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, logC < L, then the residual for that concentration is censored as well, logC -fitted(logC) < L -fitted(logC) = 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 for a censored concentration is fitted(logC) + 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

Harrell, F.E., Jr., 2010, Regression modeling strategies—With applications to linear models, logistic regression, and survival analysis: New York, Springer-Verlag, 568 p.

Harrell, F.E., Jr., 2018, rms—Regression modeling strategies: R package version 5.1-2, https://CRAN.R-project.org/package=rms.

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.

Index