

Package: ldc (via r-universe)

July 12, 2024

Type Package

Title Calculate and Plot Pollutant Load Duration Curves

Version 0.1.0

Maintainer Michael Schramm <mpschrmm@gmail.com>

Description Load duration curves are a method for visualizing pollutant loads in freshwater streams based on the assumed relationship between streamflow and load. Functions are provided for calculating exceedance probabilities, pollutant loading, and plotting load duration curves.

License MIT + file LICENSE

Encoding UTF-8

LazyData true

Depends R (>= 2.10),

RoxygenNote 7.1.2

Imports DescTools, dplyr (>= 0.8.0), ggplot2, Hmisc, magrittr, purrr, rlang, units

URL <https://github.com/TxWRI/ldc>, <https://txwri.github.io/ldc/>

BugReports <https://github.com/TxWRI/ldc/issues>

Suggests rmarkdown, knitr, testthat (>= 3.0.0), vdiff, covr

VignetteBuilder knitr

Config/testthat/edition 3

Config/Needs/website TxWRI/twriTemplates

Repository <https://txwri.r-universe.dev>

RemoteUrl <https://github.com/TxWRI/ldc>

RemoteRef HEAD

RemoteSha d44d15def0a2af678de622e164d706eb19568c99

Contents

calc_annual_ldc	2
calc_ldc	4
draw_ldc	6
summ_ldc	8
tres_palacios	9

Index	11
--------------	-----------

calc_annual_ldc	<i>Calculate annualized load duration curve</i>
-----------------	---

Description

Calculates the median annual ldc with confidence intervals.

Usage

```
calc_annual_ldc(
  .tbl,
  Q = NULL,
  C = NULL,
  Date = NULL,
  allowable_concentration = NULL,
  breaks = c(1, 0.8, 0.4, 0),
  labels = c("High Flows", "Medium Flows", "Low Flows"),
  conf_level = 0.9,
  estimator = 6,
  n = 500
)
```

Arguments

.tbl	data frame with at least three columns Q (discharge or flow), C (associated pollutant concentration), and Date.
Q	variable name in .tbl for discharge or flow. This must be of class 'units', typically with a units value of "ft^3/s".
C	variable name in .tbl for associated pollutant concentration at a given flow value. This must be of class 'units', typically with a units value of "mg/L" or "cfu/100mL".
Date	variable name in .tbl for the event Date. This variable must be of class 'Date'.
allowable_concentration	an object of class units specifying the allowable pollutant concentration.
breaks	a numeric vector of break points for flow categories. Must be of length of labels + 1. defaults to c(1, 0.8, 0.4, 0).
labels	labels for the categories specified by breaks.

conf_level	numeric, confidence level (default is 0.9) of the median interval at given exceedance probability.
estimator	one of c(5, 6, 7, 8, 9, "hd"). 6 is the default method corresponding to the Weibull plotting position. Further details are provided in quantile . "hd" uses the Harrell-Davis Distribution-Free Quantile Estimator (see: hdquantile).
n	numeric, the length of generated probability points. Larger n may result in a slightly smoother curve at a cost of increased processing time. The probability points are used to generate the continuous sample quantiles types 5 to 9 (see quantile).

Details

The median annual ldc is calculated by computing the flow duration curve for each individual year in the dataset. Exceedance probabilities are calculated from the descending order of Daily Flows. By default, the Weibull plotting position is used:

$$p = P(Q > q_i) = \frac{i}{n + 1}$$

where $q_i, i = 1, 2, \dots, n$, is the i -th sorted streamflow value.

The median streamflow +/- chosen confidence interval is calculated at each exceedance probability. The load duration curve is calculated by multiplying the median streamflow by the allowable concentration and appropriate conversions.

Value

list of two tibbles (Q and C). Includes variables in .tbl and Daily_Flow_Volume (discharge volume), Daily_Load (pollutant sample volume), P_Exceedance (exceedance probability), Flow_Category (as defined by breaks and labels).

References

Vogel, Richard M., and Neil M. Fennessey. "Flow-duration curves. I: New interpretation and confidence intervals." *Journal of Water Resources Planning and Management* 120, no. 4 (1994): 485-504. doi:10.1061/(ASCE)07339496(1994)120:4(485)

Examples

```
# Basic example using built in Tres Palacios data
library(dplyr)
library(units)
# Format data
install_unit("cfu")
df <- as_tibble(tres_palacios) %>%
  ## filter data so this run quicker
  filter(!is.na(Indicator_Bacteria)) %>%
  ## flow must have units, here is is in cfs
  mutate(Flow = set_units(Flow, "ft^3/s")) %>%
  ## pollutant concentration must have units
  mutate(Indicator_Bacteria = set_units(Indicator_Bacteria, "cfu/100mL"))
```

```

# Calculate LDC

## specify the allowable concentration
allowable_concentration <- 126
## set the units
units(allowable_concentration) <- "cfu/100mL"
df_ldc <- calc_annual_ldc(df,
                        Q = Flow,
                        C = Indicator_Bacteria,
                        Date = Date,
                        allowable_concentration = allowable_concentration,
                        estimator = 5,
                        n = 1000)

df_ldc$Q

## cleanup
remove_unit("cfu")

```

calc_ldc

Calculate load duration curve

Description

Calculates the period of record load duration curve from a data frame that includes mean daily flow and associated point measurements of pollutant concentration.

Usage

```

calc_ldc(
  .tbl,
  Q = NULL,
  C = NULL,
  allowable_concentration = NULL,
  breaks = c(1, 0.8, 0.4, 0),
  labels = c("High Flows", "Medium Flows", "Low Flows"),
  estimator = 6
)

```

Arguments

.tbl	data frame with at least two columns Q (discharge or flow) and C (associated pollutant concentration).
Q	variable name in .tbl for discharge or flow. This must have unit set, typically "ft ³ /s".
C	variable name in .tbl for associated pollutant concentration at a given flow value. This must have a unit set, typically "mg/L" or "cfu/100mL".
allowable_concentration	an object of class units specifying the allowable pollutant concentration.

breaks	a numeric vector of break points for flow categories. Must be of length of labels + 1. defaults to c(1, 0.8, 0.4, 0).
labels	labels for the categories specified by breaks.
estimator	numeric, one of c(5,6,7,8,9). 6 is the default method corresponding to the Weibull plotting position. Further details are provided in stats::quantile().

Details

The exceedance probability is calculated from the descending order of Daily Flows. By default, the Weibull plotting position is used:

$$p = P(Q > q_i) = \frac{i}{n+1}$$

where $q_i, i = 1, 2, \dots, n$, is the i -th sorted streamflow value.

Value

object of class tibble. Includes variables in .tbl and Daily_Flow_Volume (discharge volume), Daily_Load (pollutant sample volume), P_Exceedance (exceedance probability), Flow_Category (as defined by breaks and labels).

Examples

```
# Basic example using built in Tres Palacios data
library(dplyr)
library(units)
# Format data
install_unit("cfu")
df <- as_tibble(tres_palacios) %>%
  ## filter data so this run quicker
  filter(!is.na(Indicator_Bacteria)) %>%
  ## flow must have units, here is is in cfs
  mutate(Flow = set_units(Flow, "ft^3/s")) %>%
  ## pollutant concentration must have units
  mutate(Indicator_Bacteria = set_units(Indicator_Bacteria, "cfu/100mL"))
# Calculate LDC

## specify the allowable concentration
allowable_concentration <- 126
## set the units
units(allowable_concentration) <- "cfu/100mL"
df_ldc <- calc_ldc(df,
  Q = Flow,
  C = Indicator_Bacteria,
  allowable_concentration = allowable_concentration)

df_ldc

## cleanup
remove_unit("cfu")
```

draw_ldc	<i>Draw a load duration curve</i>
----------	-----------------------------------

Description

Creates a load duration curve visualization from the outputs of `calc_ldc` and `summ_ldc` as a ggplot object.

Usage

```
draw_ldc(
  .tbl_calc,
  .tbl_summ,
  y_lab = NULL,
  ldc_legend_name = "Allowable Load at State Water Quality Standard",
  measurement_name = "Measurement Value",
  measurement_shape = 21,
  measurement_color = "dodgerblue",
  measurement_alpha = 1,
  summary_name = "Summarized Measured Load",
  summary_stat_shape = 12,
  summary_stat_color = "red",
  label_nudge_y = 0,
  label_font_family = "Arial",
  label_font_size = 3,
  label_break = TRUE
)
```

Arguments

<code>.tbl_calc</code>	data frame object created by <code>calc_ldc</code>
<code>.tbl_summ</code>	data frame object created by <code>summ_ldc</code>
<code>y_lab</code>	optional string for y-axis label name, will be appended with units automatically. default is NULL.
<code>ldc_legend_name</code>	string, provides the name used for the allowable pollutant load line in the legend. required.
<code>measurement_name</code>	string, provides the name used for measured load values in the legend. required.
<code>measurement_shape</code>	aesthetic value passed to the layer plotting measured load values. defaults to 21.
<code>measurement_color</code>	aesthetic value passed to the layer plotting measured load values. defaults to "dodgerblue".
<code>measurement_alpha</code>	aesthetic value passed to the layer plotting measured load values. defaults to 1.

summary_name string, provides the name used for summary statistic values in the legend. required.

summary_stat_shape aesthetic value passed to the layer plotting summary statistic values. defaults to 12.

summary_stat_color aesthetic value passed to the layer plotting summary statistic values. defaults to "red".

label_nudge_y numeric value to vertically nudge flow category labels. If a log10 transformed scale is being used, a log value is probably appropriate for example $\log_{10}(1000)$.

label_font_family string specifying font family to use in flow category labels.

label_font_size numeric value specifying font size to use in flow category labels.

label_break logical, add line breaks to flow category labels. Labels will break at spaces.

Value

ggplot object

Examples

```
# Basic example using built in Tres Palacios data
library(dplyr)
library(units)
library(ggplot2)
# Format data
install_unit("cfu")
df <- as_tibble(tres_palacios) %>%
  ## filter data so this run quicker
  filter(!is.na(Indicator_Bacteria)) %>%
  ## flow must have units, here is is in cfs
  mutate(Flow = set_units(Flow, "ft^3/s")) %>%
  ## pollutant concentration must have units
  mutate(Indicator_Bacteria = set_units(Indicator_Bacteria, "cfu/100mL"))
# Calculate LDC

## specify the allowable concentration
allowable_concentration <- 126
## set the units
units(allowable_concentration) <- "cfu/100mL"
df_ldc <- calc_ldc(df,
  Q = Flow,
  C = Indicator_Bacteria,
  allowable_concentration = allowable_concentration)

# Summarize LDC
df_sum <- summ_ldc(df_ldc,
  Q = Flow,
  C = Indicator_Bacteria,
```

```

      Exceedance = P_Exceedance,
      groups = Flow_Category,
      method = "geomean")

# Create ggplot object
draw_ldc(df_ldc,
        df_sum,
        y_lab = expression(paste(italic("E. coli"))),
        label_nudge_y = log10(1000) +
        scale_y_log10() +
        theme(legend.title = element_blank(),
              legend.direction = "vertical",
              legend.position = "bottom")

## cleanup
remove_unit("cfu")

```

summ_ldc

Summarize load duration curve

Description

Calculates summary statistics for flow and pollutant concentrations for desired flow categories. Estimates "average" pollutant load per category based on average concentration times the median flow.

Usage

```
summ_ldc(.tbl, Q, C, Exceedance, groups, method = "geomean")
```

Arguments

.tbl	data frame, preferably the output from <code>calc_ldc()</code> .
Q	variable name in .tbl for discharge or flow. This must have unit set, typically "ft ³ /s".
C	variable name in .tbl for associated pollutant concentration at a given flow value. This must have a unit set, typically "mg/L" or "cfu/100mL".
Exceedance	variable name in .tbl with flow/load exceedance probabilities.
groups	variable name in .tbl with categorized flow names.
method	string that describes the summary statistic used for the pollutant concentration. Must be one of <code>c('geomean', 'mean', 'median')</code> .

Value

object of class tibble. Includes Flow Category grouping variable, median flow and exceedance values, geometric mean/mean/median pollutant concentration, and estimated average load based on median flow times the average pollutant concentration per flow category.

Examples

```

# Basic example using built in Tres Palacios data
library(dplyr)
library(units)
# Format data
install_unit("cfu")
df <- as_tibble(tres_palacios) %>%
  ## filter data so this run quicker
  filter(!is.na(Indicator_Bacteria)) %>%
  ## flow must have units, here is is in cfs
  mutate(Flow = set_units(Flow, "ft^3/s")) %>%
  ## pollutant concentration must have units
  mutate(Indicator_Bacteria = set_units(Indicator_Bacteria, "cfu/100mL"))
# Calculate LDC

## specify the allowable concentration
allowable_concentration <- 126
## set the units
units(allowable_concentration) <- "cfu/100mL"
df_ldc <- calc_ldc(df,
  Q = Flow,
  C = Indicator_Bacteria,
  allowable_concentration = allowable_concentration)

# Summarize LDC
df_sum <- summ_ldc(df_ldc,
  Q = Flow,
  C = Indicator_Bacteria,
  Exceedance = P_Exceedance,
  groups = Flow_Category,
  method = "geomean")

df_sum

## cleanup
remove_unit("cfu")

```

tres_palacios

Mean daily flow and point E. coli bacteria measurements.

Description

A dataset containing the mean daily flow and E. coli bacteria concentrations on the Tres Palacios River from 2000 through 2020.

Usage

```
tres_palacios
```

Format

A data frame with 7671 rows and 4 variables:

site_no USGS gage number

Date Observation Date

Flow Mean Daily Flow in cfs

Indicator_Bacteria Bacteria concentration measured on the given day in MPN/100mL

Source

USGS NWIS <https://waterdata.usgs.gov/nwis> and TCEQ SWQM <https://www.tceq.texas.gov/waterquality/monitoring>

Index

* datasets

tres_palacios, 9

calc_annual_ldc, 2

calc_ldc, 4

draw_ldc, 6

hdquantile, 3

quantile, 3

summ_ldc, 8

tres_palacios, 9