Package 'MassWateR'

November 19, 2023

Title Quality Control and Analysis of Massachusetts Water Quality Data

Version 2.1.4

Description Methods for quality control and exploratory analysis of surface
water quality data collected in Massachusetts, USA. Functions are developed
to facilitate data formatting for the Water Quality Exchange Network
<https://www.epa.gov/waterdata/water-quality-data-upload-wqx> and reporting
of data quality objectives to state agencies. Quality control methods are
from Massachusetts Department of Environmental Protection (2020)
<https:
//www.mass.gov/orgs/massachusetts-department-of-environmental-protection>.

Depends R (>= 3.5)

Imports curl, dplyr, flextable, ggplot2, ggrepel, ggspatial, httr,
lubridate, prettymapr, raster, RColorBrewer, rmarkdown, readxl,
sf, tidyr, tibble, units, writexl

License CC0

Encoding UTF-8

RoxygenNote 7.2.3

URL <https://github.com/massbays-tech/MassWateR>,
<https://massbays-tech.github.io/MassWateR/>

BugReports https://github.com/massbays-tech/MassWateR/issues

LazyData true

LazyDataCompression xz

Suggests covr, knitr, patchwork, testthat (>= 3.0.0)

Config/testthat/edition 3

NeedsCompilation no

Author Marcus Beck [aut, cre] (<https://orcid.org/0000-0002-4996-0059>),
Jill Carr [aut],
Ben Wetherill [aut] (<https://orcid.org/0000-0002-8791-4256>)

Maintainer Marcus Beck <mbeck@tbep.org>

R **topics documented:**

---

anlzMWRdate   *Analyze trends by date in results file*

Description

Analyze trends by date in results file

Usage

```
anlzMWRdate(
  res = NULL,
  param,
  acc = NULL,
  sit = NULL,
  fset = NULL,
  thresh,
  group = c("site", "locgroup", "all"),
  threshlab = NULL,
  threshcol = "tan",
  site = NULL,
  resultatt = NULL,
  locgroup = NULL,
  dtrng = NULL,
  ptsize = 2,
  repel = FALSE,
  labsize = 3,
  expand = c(0.05, 0.1),
  confint = FALSE,
  palcol = "Set2",
  yscl = "auto",
  sumfun = yscl,
  colleg = FALSE,
  ttlsize = 1.2,
  runchk = TRUE,
  warn = TRUE
)
```

Arguments

res	character string of path to the results file or data.frame for results returned by [readMWRresults](#)
param	character string of the parameter to plot, must conform to entries in the "Simple Parameter" column of [paramsMWR](#)
acc	character string of path to the data quality objectives file for accuracy or data.frame returned by [readMWRacc](#)

sit	optional character string of path to the site metadata file or data.frame of site metadata returned by readMWRsites, required if locgroup is not NULL
fset	optional list of inputs with elements named res, acc, frecom, sit, or wqx overrides the other arguments
thresh	character indicating if relevant freshwater or marine threshold lines are included, one of "fresh", "marine", or "none", or a single numeric value to override the values included with the package
group	character indicating whether the results are grouped by site (default), combined across location groups, or combined across sites, see details
threshlab	optional character string indicating legend label for the threshold, required only if thresh is numeric
threshcol	character indicating color of threshold lines if available
site	character string of sites to include, default all
resultatt	character string of result attributes to plot, default all
locgroup	character string of location groups to plot from the "Location Group" column in the site metadata file, optional and only if sit is not NULL
dtrng	character string of length two for the date ranges as YYYY-MM-DD, default all
ptsize	numeric indicating size of the points
repel	logical indicating if overlapping site labels are offset, default FALSE
labsize	numeric indicating font size for the site labels, only if group = "site" or group = "locgroup"
expand	numeric of length two indicating expansion proportions on the x-axis to include labels outside of the plot range if repel = F and group = "site" or group = "locgroup"
confint	logical indicating if confidence intervals are shown, only applies if data are summarized using group as "locgroup" or "all"
palcol	character string indicating the color palette for points and lines from RColorBrewer, see details
yscl	character indicating one of "auto" (default), "log", or "linear", see details
sumfun	character indicating one of "auto", "mean", "geomean", "median", "min", or "max", see details
colleg	logical indicating if a color legend for sites or location groups is included if group = "site" or group = "locgroup"
ttlsize	numeric value indicating font size of the title relative to other text in the plot
runchk	logical to run data checks with checkMWRresults or checkMWRacc, applies only if res or acc are file paths
warn	logical to return warnings to the console (default)

Details

Results are shown for the selected parameter as continuous line plots over time. Specifying group = "site" plot a separate line for each site. Specifying group = "locgroup" will summarize results across sites in the locgroup argument based on the value passed to sumfun or yscl if no value is passed to sumfun. The site metadata file must be passed to the `sit` argument to use this option. Specifying group = "all" will summarize results across sites for each date based on the value passed to sumfun or yscl if no value is passed to sumfun. Summarized results will include confidence intervals if confint = TRUE and they can be calculated (i.e., more than one point is used in the summary and data are summarized using group as "locgroup" or "all").

Threshold lines applicable to marine or freshwater environments can be included in the plot by using the thresh argument. These thresholds are specific to each parameter and can be found in the [thresholdMWR](#) file. Threshold lines are plotted only for those parameters with entries in [thresholdMWR](#) and only if the value in `Result Unit` matches those in [thresholdMWR](#). The threshold lines can be suppressed by setting thresh = 'none'. A user-supplied numeric value can also be used for the thresh argument to override the default values. An appropriate label must also be supplied to threshlab if thresh is numeric.

Any acceptable color palette for from [RColorBrewer](#) for the points and lines can be used for palcol, which is passed to the palette argument in [scale_color_brewer](#). These could include any of the qualitative color palettes, e.g., "Set1", "Set2", etc. The continuous and diverging palettes will also work, but may return color scales for points and lines that are difficult to distinguish. The palcol argument does not apply if group = "all".

The y-axis scaling as arithmetic (linear) or logarithmic can be set with the yscl argument. If yscl = "auto" (default), the scaling is determined automatically from the data quality objective file for accuracy, i.e., parameters with "log" in any of the columns are plotted on log10-scale, otherwise arithmetic. Setting yscl = "linear" or yscl = "log" will set the axis as linear or log10-scale, respectively, regardless of the information in the data quality objective file for accuracy.

Similarly, the data will be summarized appropriately for group (only applies if group is not site) based on the value passed to sumfun. The default if no value is provided to sumfun is to use the appropriate summary based on the value provided to yscl. If yscl = "auto" (default), then sumfun = "auto", and the mean or geometric mean is used for the summary based on information in the data quality objective file for accuracy. Using yscl = "linear" or yscl = "log" will default to the mean or geometric mean summary if no value is provided to sumfun. Any other appropriate value passed to sumfun will override the value passed to yscl. Valid summary functions for sumfun include "auto", "mean", "geomean", "median", "min", or "max").

Any entries in resdat in the "Result Value" column as "BDL" or "AQL" are replaced with appropriate values in the "Quantitation Limit" column, if present, otherwise the "MDL" or "UQL" columns from the data quality objectives file for accuracy are used. Values as "BDL" use one half of the appropriate limit.

Value

A [ggplot](#) object that can be further modified.

Examples

```
# results data path
respth <- system.file('extdata/ExampleResults.xlsx', package = 'MassWateR')
```

```
# results data
resdat <- readMWRresults(respth)

# accuracy path
accpth <- system.file('extdata/ExampleDQOAccuracy.xlsx',
     package = 'MassWateR')

# accuracy data
accdat <- readMWRacc(accpth)

# site data path
sitpth <- system.file('extdata/ExampleSites.xlsx', package = 'MassWateR')

# site data
sitdat <- readMWRsites(sitpth)

# select sites
anlzMWRdate(res = resdat, param = 'DO', acc = accdat, group = 'site', thresh = 'fresh',
     site = c("ABT-026", "ABT-077"))
```

---

anlzMWRmap   *Analyze results with maps*

Description

Analyze results with maps

Usage

```
anlzMWRmap(
  res = NULL,
  param,
  acc = NULL,
  sit = NULL,
  fset = NULL,
  site = NULL,
  resultatt = NULL,
  locgroup = NULL,
  dtrng = NULL,
  ptsize = 4,
  repel = TRUE,
  labsize = 3,
  palcol = "Greens",
  palcolrev = FALSE,
  sumfun = "auto",
  crs = 4326,
  zoom = 11,
  addwater = "medium",
```

```
    watercol = "lightblue",
    maptype = NULL,
    buffdist = 2,
    scaledist = "km",
    northloc = "tl",
    scaleloc = "br",
    latlon = TRUE,
    ttlsize = 1.2,
    runchk = TRUE,
    warn = TRUE
)
```

Arguments

res	character string of path to the results file or `data.frame` for results returned by [readMWRresults](#)
param	character string of the parameter to plot, must conform to entries in the `"Simple Parameter"` column of [paramsMWR](#)
acc	character string of path to the data quality objectives file for accuracy or `data.frame` returned by [readMWRacc](#)
sit	character string of path to the site metadata file or `data.frame` of site metadata returned by [readMWRsites](#)
fset	optional list of inputs with elements named `res`, `acc`, `frecom`, `sit`, or `wqx` overrides the other arguments
site	character string of sites to include, default all
resultatt	character string of result attributes to plot, default all
locgroup	character string of location groups to plot from the `"Location Group"` column in the site metadata file, default all
dtrng	character string of length two for the date ranges as YYYY-MM-DD, default all
ptsize	numeric for size of the points, use a negative value to omit the points
repel	logical indicating if overlapping site labels are offset
labsize	numeric for size of the site labels
palcol	character string indicating the color palette to be used from [RColorBrewer](#), see details
palcolrev	logical indicating if color palette in `palcol` is reversed
sumfun	character indicating one of `"auto"` (default), `"mean"`, `"geomean"`, `"median"`, `"min"`, or `"max"`, see details
crs	numeric as a four-digit EPSG number for the coordinate reference system, see details
zoom	numeric indicating resolution of the base map, see details
addwater	character string as `"low"`, `"medium"` (default), `"high"`, or NULL (to suppress) to include water features with varying detail from the National Hydrography dataset, see details

watercol	character string of color for water objects if addwater is not NULL
maptype	character string as "cartolight", "cartodark", "osm", "hotstyle", or NULL (to suppress, default) indicating the basemap type, see details
buffdist	numeric for buffer around the bounding box for the selected sites in kilometers, see details
scaledist	character string indicating distance unit for the scale bar, "km" or "mi"
northloc	character string indicating location of the north arrow, see details
scaleloc	character string indicating location of the scale bar, see details
latlon	logical to include latitude and longitude labels on the plot, default TRUE
ttlsize	numeric value indicating font size of the title relative to other text in the plot
runchk	logical to run data checks with checkMWRresults, checkMWRacc, or checkMWRsites, applies only if res, acc, or sit are file paths
warn	logical to return warnings to the console (default)

Details

This function creates a map of summarized results for a selected parameter at each monitoring site. By default, all dates for the parameter are averaged. Options to filter by site, date range, and result attribute are provided. Only sites with spatial information in the site metadata file are plotted and a warning is returned for those that do not have this information. The site labels are also plotted next to each point. The labels can be suppressed by setting labsize = NULL.

Any acceptable color palette from RColorBrewer can be used for palcol, which is passed to the palette argument in scale_fill_distiller. These could include any of the sequential color palettes, e.g., "Greens", "Blues", etc. The diverging and qualitative palettes will also work, but may return uninterpretable color scales. The palette can be reversed by setting palcolrev = TRUE.

The default value for crs is EPSG 4326 for the WGS 84 projection in decimal degrees. The crs argument is passed to st_as_sf and any acceptable CRS appropriate for the data can be used.

The results shown on the map represent the parameter summary for each site within the date range provided by dtrng. If sumfun = "auto" (default), the mean is used where the distribution is determined automatically from the data quality objective file for accuracy, i.e., parameters with "log" in any of the columns are summarized with the geometric mean, otherwise arithmetic. Any other valid summary function will be applied if passed to sumfun ("mean", "geomean", "median", "min", "max"), regardless of the information in the data quality objective file for accuracy.

Using addwater = "medium" (default) will include lines and polygons of natural water bodies defined using the National Hydrography Dataset (NHD). The level of detail can be changed to low or high using addwater = "low" or addwater = "high", respectively. Use addwater = NULL to not show any water features.

A base map can be plotted using the maptype argument and is obtained from the annotation_map_tile function of ggspatial. The zoom value specifies the resolution of the map. Use higher values to download map tiles with greater resolution, although this increases the download time. The maptype argument describes the type of base map to download. Acceptable options include "cartolight", "cartodark", "osm", or "hotstyle". Use maptype = NULL to suppress the base map.

The area around the summarized points can be increased or decreased using the buffdist argument. This creates a buffered area around the bounding box for the points, where the units are kilometers.

A north arrow and scale bar are also placed on the map as defined by the northloc and scaleloc arguments. The placement for both can be chosen as "tl", "tr", "bl", or "br" for top-left, top-right, bottom-left, or bottom-right respectively. Setting either of the arguments to NULL will suppress the placement on the map.

Value

A [ggplot](#) object that can be further modified.

Examples

```
# results data path
respth <- system.file('extdata/ExampleResults.xlsx', package = 'MassWateR')

# results data
resdat <- readMWRresults(respth)

# accuracy path
accpth <- system.file('extdata/ExampleDQOAccuracy.xlsx', package = 'MassWateR')

# accuracy data
accdat <- readMWRacc(accpth)

# site data path
sitpth <- system.file('extdata/ExampleSites.xlsx', package = 'MassWateR')

# site data
sitdat <- readMWRsites(sitpth)


# map with NHD water bodies
anlzMWRmap(res = resdat, param = 'DO', acc = accdat, sit = sitdat, addwater = 'medium')
```

---

anlzMWRoutlier   *Analyze outliers in results file*

Description

Analyze outliers in results file

Usage

```
anlzMWRoutlier(
  res = NULL,
  param,
  acc = NULL,
  fset = NULL,
  type = c("box", "jitterbox", "jitter"),
```

```
  group,
  dtrng = NULL,
  repel = TRUE,
  outliers = FALSE,
  labsize = 3,
  fill = "lightgrey",
  alpha = 0.8,
  width = 0.8,
  yscl = "auto",
  ttlsize = 1.2,
  runchk = TRUE,
  warn = TRUE
)
```

Arguments

res	character string of path to the results file or data.frame for results returned by [readMWRresults](#)
param	character string of the parameter to plot, must conform to entries in the "Simple Parameter" column of [paramsMWR](#)
acc	character string of path to the data quality objectives file for accuracy or data.frame returned by [readMWRacc](#)
fset	optional list of inputs with elements named res, acc, frecom, sit, or wqx overrides the other arguments
type	character indicating "box", "jitterbox", or "jitter", see details
group	character indicating whether the summaries are grouped by month, site, or week of year
dtrng	character string of length two for the date ranges as YYYY-MM-DD, optional
repel	logical indicating if overlapping outlier labels are offset
outliers	logical indicating if outliers are returned to the console instead of plotting
labsize	numeric indicating font size for the outlier labels
fill	numeric indicating fill color for boxplots
alpha	numeric from 0 to 1 indicating transparency of fill color
width	numeric for width of boxplots
yscl	character indicating one of "auto" (default), "log", or "linear", see details
ttlsize	numeric value indicating font size of the title relative to other text in the plot
runchk	logical to run data checks with [checkMWRresults](#) or [checkMWRacc](#), applies only if res or acc are file paths
warn	logical to return warnings to the console (default)

Details

Outliers are defined following the standard ggplot definition as 1.5 times the inter-quartile range of each boxplot. The data frame returned if `outliers = TRUE` may vary based on the boxplot groupings defined by `group`.

Specifying `type = "box"` (default) will produce standard boxplots. Specifying `type = "jitterbox"` will produce boxplots with non-outlier observations jittered on top. Specifying `type = "jitter"` will suppress the boxplots and show only the jittered points and the outliers.

Specifying `group = "week"` will group the samples by week of year using an integer specifying the week. Note that there can be no common month/day indicating the start of the week between years and an integer is the only way to compare summaries if the results data span multiple years.

The y-axis scaling as arithmetic (linear) or logarithmic can be set with the `yscl` argument. If `yscl = "auto"` (default), the scaling is determined automatically from the data quality objective file for accuracy, i.e., parameters with "log" in any of the columns are plotted on log10-scale, otherwise arithmetic. Setting `yscl = "linear"` or `yscl = "log"` will set the axis as linear or log10-scale, respectively, regardless of the information in the data quality objective file for accuracy.

Any entries in `resdat` in the `"Result Value"` column as `"BDL"` or `"AQL"` are replaced with appropriate values in the `"Quantitation Limit"` column, if present, otherwise the `"MDL"` or `"UQL"` columns from the data quality objectives file for accuracy are used. Values as `"BDL"` use one half of the appropriate limit.

Value

A ggplot object that can be further modified if `outliers = FALSE`, otherwise a data frame of outliers is returned.

Examples

```
# results data path
respth <- system.file('extdata/ExampleResults.xlsx', package = 'MassWateR')

# results data
resdat <- readMWRresults(respth)

# accuracy path
accpth <- system.file('extdata/ExampleDQOAccuracy.xlsx',
    package = 'MassWateR')

# accuracy data
accdat <- readMWRacc(accpth)

# outliers by month
anlzMWRoutlier(res = resdat, param = 'DO', acc = accdat, group = 'month')

# outliers by site
anlzMWRoutlier(res = resdat, param = 'DO', acc = accdat, group = 'site')

# outliers by site, May through July 2021 only
anlzMWRoutlier(res = resdat, param = 'DO', acc = accdat, group = 'site',
    dtrng = c('2022-05-01', '2022-07-31'))
```

```
# outliers by month, type as jitterbox
anlzMWRoutlier(res = resdat, param = 'DO', acc = accdat, group = 'month', type = 'jitterbox')

# outliers by month, type as jitter
anlzMWRoutlier(res = resdat, param = 'DO', acc = accdat, group = 'month', type = 'jitter')

# data frame output
anlzMWRoutlier(res = resdat, param = 'DO', acc = accdat, group = 'month', outliers = TRUE)
```

---

anlzMWRoutlierall   *Analyze outliers in results file for all parameters*

Description

Analyze outliers in results file for all parameters

Usage

```
anlzMWRoutlierall(
  res = NULL,
  acc = NULL,
  fset = NULL,
  fig_height = 4,
  fig_width = 8,
  format = c("word", "png"),
  output_dir,
  output_file = NULL,
  type = c("box", "jitterbox", "jitter"),
  group,
  dtrng = NULL,
  repel = TRUE,
  outliers = FALSE,
  labsize = 3,
  fill = "lightgrey",
  alpha = 0.8,
  width = 0.8,
  yscl = "auto",
  ttlsize = 1.2,
  runchk = TRUE,
  warn = TRUE
)
```

Arguments

res   character string of path to the results file or data.frame for results returned by
               readMWRresults

acc	character string of path to the data quality objectives file for accuracy or data.frame returned by readMWRacc
fset	optional list of inputs with elements named res, acc, frecom, sit, or wqx overrides the other arguments
fig_height	numeric for plot heights in inches
fig_width	numeric for plot width in inches
format	character string indicating if results are placed in a word file or as separate png files in output_dir
output_dir	character string of the output directory for the results
output_file	optional character string for the file name if format = "word"
type	character indicating "box", "jitterbox", or "jitter", see details
group	character indicating whether the summaries are grouped by month, site, or week of year
dtrng	character string of length two for the date ranges as YYYY-MM-DD, optional
repel	logical indicating if overlapping outlier labels are offset
outliers	logical indicating if outliers are returned to the console instead of plotting
labsize	numeric indicating font size for the outlier labels
fill	numeric indicating fill color for boxplots
alpha	numeric from 0 to 1 indicating transparency of fill color
width	numeric for width of boxplots
yscl	character indicating one of "auto" (default), "log", or "linear", see details
ttlsize	numeric value indicating font size of the title relative to other text in the plot
runchk	logical to run data checks with checkMWRresults or checkMWRacc, applies only if res or acc are file paths
warn	logical to return warnings to the console (default)

Details

This function is a wrapper to anlzMWRoutlier to create plots for all parameters with appropriate data in the water quality monitoring results

Value

A word document named outlierall.docx (or name passed to output_file) if format = "word" or separate png files for each parameter if format = "png" will be saved in the directory specified by output_dir

Examples

```
# results data path
respth <- system.file('extdata/ExampleResults.xlsx', package = 'MassWateR')

# results data
resdat <- readMWRresults(respth)
```

```
# accuracy path
accpth <- system.file('extdata/ExampleDQOAccuracy.xlsx',
    package = 'MassWateR')

# accuracy data
accdat <- readMWRacc(accpth)


# create word output
anlzMWRoutlierall(resdat, accdat, group = 'month', format = 'word', output_dir = tempdir())

# create png output
anlzMWRoutlierall(resdat, accdat, group = 'month', format = 'png', output_dir = tempdir())
```

---

anlzMWRseason   *Analyze seasonal trends in results file*

Description

Analyze seasonal trends in results file

Usage

```
anlzMWRseason(
  res = NULL,
  param,
  acc = NULL,
  sit = NULL,
  fset = NULL,
  thresh,
  group = c("month", "week"),
  type = c("box", "jitterbox", "bar", "jitterbar", "jitter"),
  threshlab = NULL,
  threshcol = "tan",
  site = NULL,
  resultatt = NULL,
  locgroup = NULL,
  dtrng = NULL,
  confint = FALSE,
  fill = "lightblue",
  alpha = 0.8,
  width = 0.8,
  yscl = "auto",
  sumfun = yscl,
  ttlsize = 1.2,
  runchk = TRUE,
```

```
  warn = TRUE
)
```

Arguments

res	character string of path to the results file or data.frame for results returned by readMWRresults
param	character string of the parameter to plot, must conform to entries in the "Simple Parameter" column of paramsMWR
acc	character string of path to the data quality objectives file for accuracy or data.frame returned by readMWRacc
sit	optional character string of path to the site metadata file or data.frame of site metadata returned by readMWRsites, required if locgroup is not NULL
fset	optional list of inputs with elements named res, acc, frecom, sit, or wqx overrides the other arguments
thresh	character indicating if relevant freshwater or marine threshold lines are included, one of "fresh", "marine", or "none", or a single numeric value to override the values included with the package
group	character indicating whether the summaries are grouped by month (default) or week of year
type	character indicating "box", "jitterbox", "bar", "jitterbar" or "jitter", see details
threshlab	optional character string indicating legend label for the threshold, required only if thresh is numeric
threshcol	character indicating color of threshold lines if available
site	character string of sites to include, default all
resultatt	character string of result attributes to plot, default all
locgroup	character string of location groups to plot from the "Location Group" column in the site metadata file, optional and only if sit is not NULL
dtrng	character string of length two for the date ranges as YYYY-MM-DD, default all
confint	logical indicating if confidence intervals are shown, only applies if type = "bar"
fill	numeric indicating fill color for boxplots or barplots
alpha	numeric from 0 to 1 indicating transparency of fill color
width	numeric for width of boxplots or barplots
yscl	character indicating one of "auto" (default), "log", or "linear", see details
sumfun	character indicating one of "auto", "mean", "geomean", "median", "min", or "max", see details
ttlsize	numeric value indicating font size of the title relative to other text in the plot
runchk	logical to run data checks with checkMWRresults or checkMWRacc, applies only if res or acc are file paths
warn	logical to return warnings to the console (default)

Details

Summaries of a parameter are shown as boxplots if type = "box" or as barplots if type = "bar".
Points can be jittered over the boxplots by setting type = "jitterbox" or jittered over the barplots
by setting type = "jitterbar". Setting type = "jitter" will show only the jittered points. For
type = "bar" or type = "jitterbar", 95% confidence intervals can also be shown if confint =
TRUE and they can be estimated (i.e., more than one result value per bar and sumfun is "auto",
"mean", or "geomean").

Specifying group = "week" will group the samples by week of year using an integer specifying the
week. Note that there can be no common month/day indicating the start of the week between years
and an integer is the only way to compare summaries if the results data span multiple years.

Threshold lines applicable to marine or freshwater environments can be included in the plot by
using the thresh argument. These thresholds are specific to each parameter and can be found
in the thresholdMWR file. Threshold lines are plotted only for those parameters with entries in
thresholdMWR and only if the value in `Result Unit` matches those in thresholdMWR. The thresh-
old lines can be suppressed by setting thresh = 'none'. A user-supplied numeric value can also
be used for the thresh argument to override the default values. An appropriate label must also be
supplied to threshlab if thresh is numeric.

The y-axis scaling as arithmetic (linear) or logarithmic can be set with the yscl argument. If yscl
= "auto" (default), the scaling is determined automatically from the data quality objective file for
accuracy, i.e., parameters with "log" in any of the columns are plotted on log10-scale, otherwise
arithmetic. Setting yscl = "linear" or yscl = "log" will set the axis as linear or log10-scale,
respectively, regardless of the information in the data quality objective file for accuracy.

Similarly, the data will be summarized if type is "bar" or "jitterbar" based on the value passed
to sumfun. The default if no value is provided to sumfun is to use the appropriate summary based
on the value provided to yscl. If yscl = "auto" (default), then sumfun = "auto", and the mean
or geometric mean is used for the summary based on information in the data quality objective file
for accuracy. Using yscl = "linear" or yscl = "log" will default to the mean or geometric mean
summary if no value is provided to sumfun. Any other appropriate value passed to sumfun will
override the value passed to yscl. Valid summary functions for sumfun include "auto", "mean",
"geomean", "median", "min", or "max").

Any entries in resdat in the "Result Value" column as "BDL" or "AQL" are replaced with ap-
propriate values in the "Quantitation Limit" column, if present, otherwise the "MDL" or "UQL"
columns from the data quality objectives file for accuracy are used. Values as "BDL" use one half of
the appropriate limit.

Value

A ggplot object that can be further modified.

Examples

```
# results data path
respth <- system.file('extdata/ExampleResults.xlsx', package = 'MassWateR')

# results data
resdat <- readMWRresults(respth)
```

```
# accuracy path
accpth <- system.file('extdata/ExampleDQOAccuracy.xlsx',
      package = 'MassWateR')

# accuracy data
accdat <- readMWRacc(accpth)

# site data path
sitpth <- system.file('extdata/ExampleSites.xlsx', package = 'MassWateR')

# site data
sitdat <- readMWRsites(sitpth)

# seasonal trends by month, boxplot
anlzMWRseason(res = resdat, param = 'DO', acc = accdat, thresh = 'fresh', group = 'month',
      type = 'box')

# seasonal trends by week, boxplot
anlzMWRseason(res = resdat, param = 'DO', acc = accdat, thresh = 'fresh', group = 'week',
      type = 'box')

# seasonal trends by month, May to July only
anlzMWRseason(res = resdat, param = 'DO', acc = accdat, thresh = 'fresh', group = 'month',
      type = 'bar', dtrng = c('2022-05-01', '2022-07-31'))

# seasonal trends by month, barplot
anlzMWRseason(res = resdat, param = 'DO', acc = accdat, thresh = 'fresh', group = 'month',
      type = 'bar')

# seasonal trends by week, barplot
anlzMWRseason(res = resdat, param = 'DO', acc = accdat, thresh = 'fresh', group = 'week',
      type = 'bar')

# seasonal trends by location group, requires sitdat
anlzMWRseason(res = resdat, param = 'DO', acc = accdat, sit = sitdat, thresh = 'fresh',
      group = 'month', type = 'box', locgroup = 'Assabet')
```

---

anlzMWRsite	*Analyze data by sites in results file*

Description

Analyze data by sites in results file

Usage

```
anlzMWRsite(
  res = NULL,
  param,
  acc = NULL,
```

```
    sit = NULL,
    fset = NULL,
    type = c("box", "jitterbox", "bar", "jitterbar", "jitter"),
    thresh,
    threshlab = NULL,
    threshcol = "tan",
    site = NULL,
    resultatt = NULL,
    locgroup = NULL,
    dtrng = NULL,
    confint = FALSE,
    fill = "lightgreen",
    alpha = 0.8,
    width = 0.8,
    yscl = "auto",
    sumfun = yscl,
    byresultatt = FALSE,
    ttlsize = 1.2,
    runchk = TRUE,
    warn = TRUE
)
```

Arguments

res	character string of path to the results file or data.frame for results returned by [readMWRresults](#)
param	character string of the parameter to plot, must conform to entries in the "Simple Parameter" column of [paramsMWR](#)
acc	character string of path to the data quality objectives file for accuracy or data.frame returned by [readMWRacc](#)
sit	optional character string of path to the site metadata file or data.frame of site metadata returned by [readMWRsites](#), required if locgroup is not NULL
fset	optional list of inputs with elements named res, acc, frecom, sit, or wqx overrides the other arguments
type	character indicating "box", "jitterbox", "bar", "jitterbar" or "jitter", see details
thresh	character indicating if relevant freshwater or marine threshold lines are included, one of "fresh", "marine", or "none", or a single numeric value to override the values included with the package
threshlab	optional character string indicating legend label for the threshold, required only if thresh is numeric
threshcol	character indicating color of threshold lines if available
site	character string of sites to include, default all
resultatt	character string of result attributes to plot, default all
locgroup	character string of location groups to plot from the "Location Group" column in the site metadata file, optional and only if sit is not NULL

dtrng	character string of length two for the date ranges as YYYY-MM-DD, default all
confint	logical indicating if confidence intervals are shown, only applies if type is "bar" or "jitterbar"
fill	numeric indicating fill color for boxplots or barplots
alpha	numeric from 0 to 1 indicating transparency of fill color
width	numeric for width of boxplots or barplots
yscl	character indicating one of "auto" (default), "log", or "linear", see details
sumfun	character indicating one of "auto", "mean", "geomean", "median", "min", or "max", see details
byresultatt	logical indicating if the plot has sites grouped separately by result attributes, see details
ttlsize	numeric value indicating font size of the title relative to other text in the plot
runchk	logical to run data checks with checkMWRresults or checkMWRacc, applies only if res or acc are file paths
warn	logical to return warnings to the console (default)

Details

Summaries of a parameter for each site are shown as boxplots if type = "box" or as barplots if type = "bar". Points can be jittered over the boxplots by setting type = "jitterbox" or jittered over the barplots by setting type = "jitterbar". Setting type = "jitter" will show only the jittered points. For type = "bar" or type = "jitterbar", 95% confidence intervals can also be shown if confint = TRUE and they can be estimated (i.e., more than one result value per bar and sumfun is "auto", "mean", or "geomean").

Threshold lines applicable to marine or freshwater environments can be included in the plot by using the thresh argument. These thresholds are specific to each parameter and can be found in the thresholdMWR file. Threshold lines are plotted only for those parameters with entries in thresholdMWR and only if the value in `Result Unit` matches those in thresholdMWR. The threshold lines can be suppressed by setting thresh = 'none'. A user-supplied numeric value can also be used for the thresh argument to override the default values. An appropriate label must also be supplied to threshlab if thresh is numeric.

The y-axis scaling as arithmetic (linear) or logarithmic can be set with the yscl argument. If yscl = "auto" (default), the scaling is determined automatically from the data quality objective file for accuracy, i.e., parameters with "log" in any of the columns are plotted on log10-scale, otherwise arithmetic. Setting yscl = "linear" or yscl = "log" will set the axis as linear or log10-scale, respectively, regardless of the information in the data quality objective file for accuracy.

Similarly, the data will be summarized if type is "bar" or "jitterbar" based on the value passed to sumfun. The default if no value is provided to sumfun is to use the appropriate summary based on the value provided to yscl. If yscl = "auto" (default), then sumfun = "auto", and the mean or geometric mean is used for the summary based on information in the data quality objective file for accuracy. Using yscl = "linear" or yscl = "log" will default to the mean or geometric mean summary if no value is provided to sumfun. Any other appropriate value passed to sumfun will override the value passed to yscl. Valid summary functions for sumfun include "auto", "mean", "geomean", "median", "min", or "max").

Any entries in resdat in the ″Result Value″ column as ″BDL″ or ″AQL″ are replaced with appropriate values in the ″Quantitation Limit″ column, if present, otherwise the ″MDL″ or ″UQL″ columns from the data quality objectives file for accuracy are used. Values as ″BDL″ use one half of the appropriate limit.

The byresultatt argument can be used to group sites separately by result attributes. For example, sites with E. coli samples can be grouped by ″Dry″ or ″Wet″ conditions if present in the ″Result Attribute″ column. Filtering by sites first using the site argument is advised to reduce the amount of data that are plotted. The grouping can be filtered further by passing appropriate values in the ″Result Attribute″ column to the resultatt argument. Note that specifying result attributes with resultatt and setting byresultatt = FALSE will filter the plot data by the result attributes but will not plot the results separately.

Value

A [ggplot](#) object that can be further modified.

Examples

```
# results data path
respth <- system.file('extdata/ExampleResults.xlsx', package = 'MassWateR')

# results data
resdat <- readMWRresults(respth)

# accuracy path
accpth <- system.file('extdata/ExampleDQOAccuracy.xlsx',
     package = 'MassWateR')

# accuracy data
accdat <- readMWRacc(accpth)

# site data path
sitpth <- system.file('extdata/ExampleSites.xlsx', package = 'MassWateR')

# site data
sitdat <- readMWRsites(sitpth)

# site trends, boxplot
anlzMWRsite(res = resdat, param = 'DO', acc = accdat, type = 'box', thresh = 'fresh')

# site trends, barplot
anlzMWRsite(res = resdat, param = 'DO', acc = accdat, type = 'bar', thresh = 'fresh')

# site trends, May to July only
anlzMWRsite(res = resdat, param = 'DO', acc = accdat, type = 'box', thresh = 'fresh',
     dtrng = c('2022-05-01', '2022-07-31'))

# grouping by result attribute
anlzMWRsite(res = resdat, param = 'DO', acc = accdat, type = 'box', thresh = 'fresh',
     site = c('ABT-062', 'ABT-077'), byresultatt = TRUE)
```

```
# site trends by location group, requires sitdat
anlzMWRsite(res = resdat, param = 'DO', acc = accdat, sit = sitdat, type = 'box',
     thresh = 'fresh', locgroup = 'Assabet')
```

---

checkMWRacc   *Check data quality objective accuracy data*

Description

Check data quality objective accuracy data

Usage

```
checkMWRacc(accdat, warn = TRUE)
```

Arguments

accdat	input data frame
warn	logical to return warnings to the console (default)

Details

This function is used internally within readMWRacc to run several checks on the input data for completeness and conformance to WQX requirements

The following checks are made:

- Column name spelling: Should be the following: Parameter, uom, MDL, UQL, Value Range, Field Duplicate, Lab Duplicate, Field Blank, Lab Blank, Spike/Check Accuracy
- Columns present: All columns from the previous check should be present
- Column types: All columns should be characters/text, except for MDL and UQL
- Value Range column na check: The character string "na" should not be in the Value Range column, "all" should be used if the entire range applies
- Unrecognized characters: Fields describing accuracy checks should not include symbols or text other than $<=, \leq, <, >=, \geq, >, \pm$, "%", "BDL", "AQL", "log", or "all"
- Overlap in Value Range column: Entries in Value Range should not overlap for a parameter (excludes ascending ranges)
- Gap in Value Range column: Entries in Value Range should not include a gap for a parameter, warning only
- Parameter: Should match parameter names in the Simple Parameter or WQX Parameter columns of the paramsMWR data
- Units: No missing entries in units (uom), except pH which can be blank
- Single unit: Each unique Parameter should have only one type for the units (uom)
- Correct units: Each unique Parameter should have an entry in the units (uom) that matches one of the acceptable values in the Units of measure column of the paramsMWR data
- Empty columns: Columns with all missing or NA values will return a warning

Value

accdat is returned as is if no errors are found, otherwise an informative error message is returned prompting the user to make the required correction to the raw data before proceeding.

Examples

```
# accuracy path
accpth <- system.file('extdata/ExampleDQOAccuracy.xlsx',
     package = 'MassWateR')

# accuracy data with no checks
accdat <- readxl::read_excel(accpth, na = c('NA', ''), col_types = 'text')
accdat <- dplyr::mutate(accdat, dplyr::across(-c(`Value Range`), ~ dplyr::na_if(.x, 'na')))

checkMWRacc(accdat)
```

---

checkMWRfrecom   *Check data quality objective frequency and completeness data*

Description

Check data quality objective frequency and completeness data

Usage

```
checkMWRfrecom(frecomdat, warn = TRUE)
```

Arguments

frecomdat	input data frame
warn	logical to return warnings to the console (default)

Details

This function is used internally within [readMWRfrecom](#) to run several checks on the input data for frequency and completeness and conformance to WQX requirements

The following checks are made:

- Column name spelling: Should be the following: Parameter, Field Duplicate, Lab Duplicate, Field Blank, Lab Blank, Spike/Check Accuracy, % Completeness
- Columns present: All columns from the previous check should be present
- Non-numeric values: Values entered in columns other than the first should be numeric
- Values outside of 0 - 100: Values entered in columns other than the first should not be outside of 0 and 100
- Parameter: Should match parameter names in the Simple Parameter or WQX Parameter columns of the [paramsMWR](#) data
- Empty columns: Columns with all missing or NA values will return a warning

Value

frecomdat is returned as is if no errors are found, otherwise an informative error message is re-
turned prompting the user to make the required correction to the raw data before proceeding.

Examples

```
library(dplyr)

frecompth <- system.file('extdata/ExampleDQOFrequencyCompleteness.xlsx',
    package = 'MassWateR')

frecomdat <- suppressMessages(readxl::read_excel(frecompth,
    skip = 1, na = c('NA', 'na', ''),
  col_types = c('text', 'numeric', 'numeric', 'numeric', 'numeric', 'numeric', 'numeric')
  )) %>%
  rename(`% Completeness` = `...7`)

checkMWRfrecom(frecomdat)
```

---

checkMWRresults   *Check water quality monitoring results*

Description

Check water quality monitoring results

Usage

```
checkMWRresults(resdat, warn = TRUE)
```

Arguments

resdat	input data frame for results
warn	logical to return warnings to the console (default)

Details

This function is used internally within readMWRresults to run several checks on the input data for
completeness and conformance to WQX requirements.

The following checks are made:

- Column name spelling: Should be the following: Monitoring Location ID, Activity Type, Ac-
  tivity Start Date, Activity Start Time, Activity Depth/Height Measure, Activity Depth/Height
  Unit, Activity Relative Depth Name, Characteristic Name, Result Value, Result Unit, Quan-
  titation Limit, QC Reference Value, Result Measure Qualifier, Result Attribute, Sample Col-
  lection Method ID, Project ID, Local Record ID, Result Comment
- Columns present: All columns from the previous check should be present

- Activity Type: Should be one of Field Msr/Obs, Sample-Routine, Quality Control Sample-Field Blank, Quality Control Sample-Lab Blank, Quality Control Sample-Lab Duplicate, Quality Control Sample-Lab Spike, Quality Control-Calibration Check, Quality Control-Meter Lab Duplicate, Quality Control-Meter Lab Blank

- Date formats: Should be mm/dd/yyyy and parsed correctly on import

- Depth data present: Depth data should be included in Activity Depth/Height Measure or Activity Relative Depth Name for all rows where Activity Type is Field Msr/Obs or Sample-Routine

- Non-numeric Activity Depth/Height Measure: All depth values should be numbers, excluding missing values

- Activity Depth/Height Unit: All entries should be `ft`, `m`, or blank

- Activity Relative Depth Name: Should be either Surface, Bottom, Midwater, Near Bottom, or blank (warning only)

- Activity Depth/Height Measure out of range: All depth values should be less than or equal to 1 meter / 3.3 feet or entered as Surface in the Activity Relative Depth Name column (warning only)

- Characteristic Name: Should match parameter names in the `Simple Parameter` or `WQX Parameter` columns of the [paramsMWR](#) data (warning only)

- Result Value: Should be a numeric value or a text value as AQL or BDL

- QC Reference Value: Should be a numeric value or a text value as AQL or BDL

- Result Unit: No missing entries in `Result Unit`, except pH which can be blank

- Single Result Unit: Each unique parameter in `Characteristic Name` should have only one entry in `Result Unit` (excludes entries for lab spikes reported as `%` or `% recovery`)

- Correct Result Unit: Each unique parameter in `Characteristic Name` should have an entry in `Result Unit` that matches one of the acceptable values in the `Units of measure` column of the [paramsMWR](#) data (excludes entries for lab spikes reported as `%` or `% recovery`)

Value

`resdat` is returned as is if no errors are found, otherwise an informative error message is returned prompting the user to make the required correction to the raw data before proceeding. Checks with warnings can be fixed at the discretion of the user before proceeding.

Examples

```
library(dplyr)

respth <- system.file('extdata/ExampleResults.xlsx', package = 'MassWateR')

resdat <- suppressWarnings(readxl::read_excel(respth, na = c('NA', 'na', ''), guess_max = Inf)) %>%
  dplyr::mutate_if(function(x) !lubridate::is.POSIXct(x), as.character)

checkMWRresults(resdat)
```

checkMWRsites   *Check site metadata file*

Description

Check site metadata file

Usage

```
checkMWRsites(sitdat)
```

Arguments

sitdat   input data frame

Details

This function is used internally within [readMWRsites](#) to run several checks on the input data for completeness and conformance to WQX requirements

The following checks are made:

- Column name spelling: Should be the following: Monitoring Location ID, Monitoring Location Name, Monitoring Location Latitude, Monitoring Location Longitude, Location Group
- Columns present: All columns from the previous check should be present
- Missing longitude or latitude: No missing entries in Monitoring Location Latitude or Monitoring Location Longitude
- Non-numeric latitude values: Values entered in Monitoring Location Latitude must be numeric
- Non-numeric longitude values: Values entered in Monitoring Location Longitude must be numeric
- Positive longitude values: Values in Monitoring Location Longitude must be negative
- Missing Location ID: No missing entries for Monitoring Location ID

Value

sitdat is returned as is if no errors are found, otherwise an informative error message is returned prompting the user to make the required correction to the raw data before proceeding.

Examples

```
library(dplyr)

sitpth <- system.file('extdata/ExampleSites.xlsx', package = 'MassWateR')

sitdat <- readxl::read_excel(sitpth, na = c('NA', 'na', ''))

checkMWRsites(sitdat)
```

---

checkMWRwqx   *Check water quality exchange (wqx) metadata input*

Description

Check water quality exchange (wqx) metadata input

Usage

```
checkMWRwqx(wqxdat, warn = TRUE)
```

Arguments

wqxdat	input data frame
warn	logical to return warnings to the console (default)

Details

This function is used internally within readMWRwqx to run several checks on the input data for conformance with downstream functions

The following checks are made:

- Column name spelling: Should be the following: Parameter, Sampling Method Context, Method Speciation, Result Sample Fraction, Analytical Method, Analytical Method Context

- Columns present: All columns from the previous check should be present

- Unique parameters: Values in Parameter should be unique (no duplicates)

- Parameter: Should match parameter names in the Simple Parameter or WQX Parameter columns of the paramsMWR data (warning only)

Value

wqxdat is returned as is if no errors are found, otherwise an informative error message is returned prompting the user to make the required correction to the raw data before proceeding. Checks with warnings can be fixed at the discretion of the user before proceeding.

Examples

```
library(dplyr)

wqxpth <- system.file('extdata/ExampleWQX.xlsx', package = 'MassWateR')

wqxdat <- readxl::read_excel(wqxpth, na = c('NA', 'na', ''), col_types = 'text')

checkMWRwqx(wqxdat)
```

formMWRacc   *Format data quality objective accuracy data*

Description

Format data quality objective accuracy data

Usage

```
formMWRacc(accdat)
```

Arguments

accdat   input data fram

Details

This function is used internally within readMWRacc to format the input data for downstream analysis. The formatting includes:

- Minor formatting for units: For conformance to WQX, e.g., ppt is changed to ppth, s.u. is changed to NA in uom

- Convert Parameter: All parameters are converted to Simple Parameter in paramsMWR as needed

- Remove unicode: Remove or replace unicode characters with those that can be used in logical expressions in qcMWRacc, e.g., replace $\geq$ with $>=$

- Convert limits to numeric: Convert MDL and UQL columns to numeric

Value

A formatted data frame of the data quality objectives file for accuracy

Examples

```
accpth <- system.file('extdata/ExampleDQOAccuracy.xlsx', package = 'MassWateR')

accdat <- readxl::read_excel(accpth, na = c('NA', ''))
accdat <- dplyr::mutate(accdat, dplyr::across(-c(`Value Range`), ~ dplyr::na_if(.x, 'na')))

formMWRacc(accdat)
```

---

formMWRfrecom   *Format data quality objective frequency and completeness data*

Description

Format data quality objective frequency and completeness data

Usage

```
formMWRfrecom(frecomdat)
```

Arguments

frecomdat   input data frame

Details

This function is used internally within readMWRfrecom to format the input data for downstream analysis. The formatting includes:

- Convert Parameter: All parameters are converted to Simple Parameter in paramsMWR as needed

Value

A formatted data frame of the data quality objectives file for frequency and completeness

Examples

```
library(dplyr)

frecompth <- system.file('extdata/ExampleDQOFrequencyCompleteness.xlsx',
    package = 'MassWateR')

frecomdat <- suppressMessages(readxl::read_excel(frecompth,
     skip = 1, na = c('NA', 'na', ''),
   col_types = c('text', 'numeric', 'numeric', 'numeric', 'numeric', 'numeric', 'numeric')
   )) %>%
   rename(`% Completeness` = `...7`)

formMWRfrecom(frecomdat)
```

---

formMWRresults   *Format water quality monitoring results*

Description

Format water quality monitoring results

Usage

```
formMWRresults(resdat, tzone = "America/Jamaica")
```

Arguments

resdat   input data frame for results

tzone   character string for time zone

Details

This function is used internally within readMWRresults to format the input data for downstream analysis. The formatting includes:

- Fix date and time inputs: Activity Start Date is converted to YYYY-MM-DD as a date object, Actvity Start Time is convered to HH:MM as a character to fix artifacts from Excel import

- Minor formatting for Result Unit: For conformance to WQX, e.g., ppt is changed to ppth, s.u. is changed to NA

- Convert characteristic names: All parameters in Characteristic Name are converted to Simple Parameter in paramsMWR as needed

Value

A formatted data frame of the water quality monitoring results file

Examples

```
library(dplyr)

respth <- system.file('extdata/ExampleResults.xlsx', package = 'MassWateR')

resdat <- suppressWarnings(readxl::read_excel(respth, na = c('NA', 'na', ''), guess_max = Inf)) %>%
  dplyr::mutate_if(function(x) !lubridate::is.POSIXct(x), as.character)

formMWRresults(resdat)
```

formMWRwqx	*Format WQX metadata input*

Description

Format WQX metadata input

Usage

```
formMWRwqx(wqxdat)
```

Arguments

wqxdat	input data frame for wqx metadata

Details

This function is used internally within readMWRwqx to format the input data for downstream analysis. The formatting includes:

- Convert characteristic names: All parameters in Characteristic Name are converted to Simple Parameter in paramsMWR as needed

Value

A formatted data frame of the WQX metadata file

Examples

```
library(dplyr)

wqxpth <- system.file('extdata/ExampleWQX.xlsx', package = 'MassWateR')

wqxdat <- suppressWarnings(readxl::read_excel(wqxpth, na = c('NA', 'na', ''), col_types = 'text'))

formMWRwqx(wqxdat)
```

paramsMWR	*Master parameter list and units for Characteristic Name column in results data*

Description

Master parameter list and units for Characteristic Name column in results data

Usage

```
paramsMWR
```

Format

A data.frame

Details

This information is used to verify the correct format of input data and for formatting output data for upload to WQX. A column showing the corresponding WQX names is also included.

Examples

```
paramsMWR
```

---

qcMWRacc   *Run quality control accuracy checks for water quality monitoring results*

Description

Run quality control accuracy checks for water quality monitoring results

Usage

```
qcMWRacc(
  res = NULL,
  acc = NULL,
  frecom = NULL,
  fset = NULL,
  runchk = TRUE,
  warn = TRUE,
  accchk = c("Field Blanks", "Lab Blanks", "Field Duplicates", "Lab Duplicates",
    "Lab Spikes / Instrument Checks"),
  suffix = "%"
)
```

Arguments

res	character string of path to the results file or data.frame for results returned by [readMWRresults](#)
acc	character string of path to the data quality objectives file for accuracy or data.frame returned by [readMWRacc](#)
frecom	character string of path to the data quality objectives file for frequency and completeness or data.frame returned by [readMWRfrecom](#)
fset	optional list of inputs with elements named res, acc, frecom, sit, or wqx overrides the other arguments
runchk	logical to run data checks with [checkMWRresults](#) and [checkMWRacc](#), applies only if res or acc are file paths

warn	logical to return warnings to the console (default)
accchk	character string indicating which accuracy check to return, one to any of "Field Blanks", "Lab Blanks", "Field Duplicates", "Lab Duplicates", or "Lab Spikes / Instrument Checks"
suffix	character string indicating suffix to append to percentage values

Details

The function can be used with inputs as paths to the relevant files or as data frames returned by [readMWRresults](#) and [readMWRacc](#). For the former, the full suite of data checks can be evaluated with runkchk = T (default) or suppressed with runchk = F. In the latter case, downstream analyses may not work if data are formatted incorrectly. For convenience, a named list with the input arguments as paths or data frames can be passed to the fset argument instead. See the help file for [utilMWRinput](#).

Note that accuracy is only evaluated on parameters in the Parameter column in the data quality objectives accuracy file. A warning is returned if there are parameters in Parameter in the accuracy file that are not in Characteristic Name in the results file.

Similarly, parameters in the results file in the Characteristic Name column that are not found in the data quality objectives accuracy file are not evaluated. A warning is returned if there are parameters in Characteristic Name in the results file that are not in Parameter in the accuracy file.

The data quality objectives file for frequency and completeness is used to screen parameters in the results file for inclusion in the accuracy tables. Parameters with empty values in the frequency and completeness table are not returned.

Value

The output shows the accuracy checks from the input files returned as a list, with each element of the list corresponding to a specific accuracy check specified with accchk.

Examples

```
##
# using file paths

# results path
respth <- system.file('extdata/ExampleResults.xlsx', package = 'MassWateR')

# accuracy path
accpth <- system.file('extdata/ExampleDQOAccuracy.xlsx', package = 'MassWateR')

# frequency and completeness path
frecompth <- system.file('extdata/ExampleDQOFrequencyCompleteness.xlsx',
     package = 'MassWateR')

qcMWRacc(res = respth, acc = accpth, frecom = frecompth)
```

| qcMWRcom | *Run quality control completeness checks for water quality monitoring results* |

Description

Run quality control completeness checks for water quality monitoring results

Usage

```
qcMWRcom(res = NULL, frecom = NULL, fset = NULL, runchk = TRUE, warn = TRUE)
```

Arguments

res
: character string of path to the results file or data.frame for results returned by [readMWRresults](#)

frecom
: character string of path to the data quality objectives file for frequency and completeness or data.frame returned by [readMWRfrecom](#)

fset
: optional list of inputs with elements named res, acc, frecom, sit, or wqx overrides the other arguments

runchk
: logical to run data checks with [checkMWRresults](#) and [checkMWRfrecom](#), applies only if res or frecom are file paths

warn
: logical to return warnings to the console (default)

Details

The function can be used with inputs as paths to the relevant files or as data frames returned by [readMWRresults](#) and [readMWRfrecom](#). For the former, the full suite of data checks can be evaluated with runkchk = T (default) or suppressed with runchk = F. In the latter case, downstream analyses may not work if data are formatted incorrectly. For convenience, a named list with the input arguments as paths or data frames can be passed to the fset argument instead. See the help file for [utilMWRinput](#).

Note that frequency is only evaluated on parameters in the Parameter column in the data quality objectives frequency and completeness file. A warning is returned if there are parameters in Parameter in the frequency and completeness file that are not in Characteristic Name in the results file.

Similarly, parameters in the results file in the Characteristic Name column that are not found in the data quality objectives frequency and completeness file are not evaluated. A warning is returned if there are parameters in Characteristic Name in the results file that are not in Parameter in the frequency and completeness file.

Value

The output shows the completeness checks from the combined files. Each row applies to a completeness check for a parameter. The datarec and qualrec columns show the number of data records and qualified records, respectively. The datarec column specifically shows only records not for quality control by excluding those as duplicates, blanks, or spikes in the count. The standard column shows the relevant percentage required for the quality control check from the quality control objectives file, the complete column shows the calculated completeness taken from the input data, and the met column shows if the standard was met by comparing if complete is greater than or equal to standard.

Examples

```
##
# using file paths

# results path
respth <- system.file('extdata/ExampleResults.xlsx', package = 'MassWateR')

# frequency and completeness path
frecompth <- system.file('extdata/ExampleDQOFrequencyCompleteness.xlsx',
    package = 'MassWateR')

qcMWRcom(res = respth, frecom = frecompth)

##
# using data frames

# results data
resdat <- readMWRresults(respth)

# frequency and completeness data
frecomdat <- readMWRfrecom(frecompth)

qcMWRcom(res = resdat, frecom = frecomdat)
```

---

qcMWRfre   *Run quality control frequency checks for water quality monitoring re-*
                                 sults

Description

Run quality control frequency checks for water quality monitoring results

Usage

```
qcMWRfre(
  res = NULL,
  acc = NULL,
```

```
    frecom = NULL,
    fset = NULL,
    runchk = TRUE,
    warn = TRUE
)
```

Arguments

res	character string of path to the results file or data.frame for results returned by [readMWRresults](#)
acc	character string of path to the data quality objectives file for accuracy or data.frame returned by [readMWRacc](#)
frecom	character string of path to the data quality objectives file for frequency and completeness or data.frame returned by [readMWRfrecom](#)
fset	optional list of inputs with elements named res, acc, frecom, sit, or wqx overrides the other arguments
runchk	logical to run data checks with [checkMWRresults](#) and [checkMWRfrecom](#), applies only if res or frecom are file paths
warn	logical to return warnings to the console (default)

Details

The function can be used with inputs as paths to the relevant files or as data frames returned by [readMWRresults](#), [readMWRacc](#), and [readMWRfrecom](#). For the former, the full suite of data checks can be evaluated with runkchk = T (default) or suppressed with runchk = F. In the latter case, downstream analyses may not work if data are formatted incorrectly. For convenience, a named list with the input arguments as paths or data frames can be passed to the fset argument instead. See the help file for [utilMWRinput](#).

Note that frequency is only evaluated on parameters in the Parameter column in the data quality objectives frequency and completeness file. A warning is returned if there are parameters in Parameter in the frequency and completeness file that are not in Characteristic Name in the results file.

Similarly, parameters in the results file in the Characteristic Name column that are not found in the data quality objectives frequency and completeness file are not evaluated. A warning is returned if there are parameters in Characteristic Name in the results file that are not in Parameter in the frequency and completeness file.

Value

The output shows the frequency checks from the input files. Each row applies to a frequency check for a parameter. The Parameter column shows the parameter, the obs column shows the total records that apply to regular activity types, the check column shows the relevant activity type for each frequency check, the count column shows the number of records that apply to a check, the standard column shows the relevant percentage required for the quality control check from the quality control objectives file, and the met column shows if the standard was met by comparing if percent is greater than or equal to standard.

Examples

```
##
# using file paths

# results path
respth <- system.file('extdata/ExampleResults.xlsx', package = 'MassWateR')

# dqo accuracy data path
accpth <- system.file('extdata/ExampleDQOAccuracy.xlsx', package = 'MassWateR')

# frequency and completeness path
frecompth <- system.file('extdata/ExampleDQOFrequencyCompleteness.xlsx',
      package = 'MassWateR')

qcMWRfre(res = respth, acc = accpth, frecom = frecompth)

##
# using data frames

# results data
resdat <- readMWRresults(respth)

# accuracy data
accdat <- readMWRacc(accpth)

# frequency and completeness data
frecomdat <- readMWRfrecom(frecompth)

qcMWRfre(res = resdat, acc = accdat, frecom = frecomdat)
```

---

qcMWRreview   *Create the quality control review report*

Description

Create the quality control review report

Usage

```
qcMWRreview(
  res = NULL,
  acc = NULL,
  frecom = NULL,
  fset = NULL,
  output_dir,
  output_file = NULL,
  rawdata = TRUE,
  dqofontsize = 7.5,
  tabfontsize = 9,
```

```
  padding = 0,
  warn = TRUE,
  runchk = TRUE
)
```

Arguments

res	character string of path to the results file or data.frame for results returned by [readMWRresults](#)
acc	character string of path to the data quality objectives file for accuracy or data.frame returned by [readMWRacc](#)
frecom	character string of path to the data quality objectives file for frequency and completeness or data.frame returned by [readMWRfrecom](#)
fset	optional list of inputs with elements named res, acc, frecom, sit, or wqx overrides the other arguments
output_dir	character string of the output directory for the rendered file
output_file	optional character string for the file name
rawdata	logical to include quality control accuracy summaries for raw data, e.g., field blanks, etc.
dqofontsize	numeric for font size in the data quality objective tables in the first page of the review
tabfontsize	numeric for font size in the review tables
padding	numeric for row padding for table output
warn	logical indicating if warnings from the table functions are included in the file output
runchk	logical to run data checks with [checkMWRresults](#), [checkMWRacc](#), [checkMWRfrecom](#), applies only if res, acc, or frecom are file paths

Details

The function compiles a review report as a Word document for all quality control checks included in the MassWateR package. The report shows several tables, including the data quality objectives files for accuracy, frequency, and completeness, summary results for all accuracy checks, summary results for all frequency checks, summary results for all completeness checks, and individual results for all accuracy checks. The report uses the individual table functions (which can be used separately) to return the results, which include [tabMWRacc](#), [tabMWRfre](#), and [tabMWRcom](#). The help files for each of these functions can be consulted for a more detailed explanation of the quality control checks.

The workflow for using this function is to import the required data (results and data quality objective files) and to fix any errors noted on import prior to creating the review report. Additional warnings that may be of interest as returned by the individual table functions can be returned in the console by setting warn = TRUE.

Optional arguments that can be changed as needed include specifying the file name with output_file, suppressing the raw data summaries at the end of the report with rawdata = FALSE, and changing

the table font sizes (dqofontsize for the data quality objectives on the first page, tabfontsize for the remainder).

The function can be used with inputs as paths to the relevant files or as data frames returned by readMWRresults, readMWRacc, and readMWRfrecom. For the former, the full suite of data checks can be evaluated with runkchk = T (default) or suppressed with runchk = F, as explained in the relevant help files. In the latter case, downstream analyses may not work if data are formatted incorrectly. For convenience, a named list with the input arguments as paths or data frames can be passed to the fset argument instead. See the help file for utilMWRinput.

Value

A compiled review report named qcreview.docx (or name passed to output_file) will be saved in the directory specified by output_dir

Examples

```
# results data path
respth <- system.file('extdata/ExampleResults.xlsx', package = 'MassWateR')

# dqo accuracy data path
accpth <- system.file('extdata/ExampleDQOAccuracy.xlsx', package = 'MassWateR')

# dqo completeness data path
frecompth <- system.file('extdata/ExampleDQOFrequencyCompleteness.xlsx', package = 'MassWateR')

# results data
resdat <- readMWRresults(respth)

# accuracy data
accdat <- readMWRacc(accpth)

# frequency and completeness data
frecomdat <- readMWRfrecom(frecompth)


# create report
qcMWRreview(res = resdat, acc = accdat, frecom = frecomdat, output_dir = tempdir())
```

---

readMWRacc   *Read data quality objectives for accuracy from an external file*

Description

Read data quality objectives for accuracy from an external file

Usage

```
readMWRacc(accpth, runchk = TRUE, warn = TRUE)
```

Arguments

accpth	character string of path to the data quality objectives file for accuracy
runchk	logical to run data checks with checkMWRacc
warn	logical to return warnings to the console (default)

Details

Data are imported with read_excel and checked with checkMWRacc.

Value

A formatted data frame of data quality objectives for completeness that can be used for downstream analysis

Examples

```
accpth <- system.file('extdata/ExampleDQOAccuracy.xlsx', package = 'MassWateR')

accdat <- readMWRacc(accpth)
head(accdat)
```

---

readMWRfrecom	*Read data quality objectives for frequency and completeness from an external file*

Description

Read data quality objectives for frequency and completeness from an external file

Usage

```
readMWRfrecom(frecompth, runchk = TRUE, warn = TRUE)
```

Arguments

frecompth	character string of path to the data quality objectives file for frequency and completeness
runchk	logical to run data checks with checkMWRfrecom
warn	logical to return warnings to the console (default)

Details

Data are imported with read_excel and checked with checkMWRfrecom.

Value

A formatted data frame of data quality objectives for frequency and completeness that can be used for downstream analysis

Examples

```
frecompth <- system.file('extdata/ExampleDQOFrequencyCompleteness.xlsx',
    package = 'MassWateR')

frecomdat <- readMWRfrecom(frecompth)
head(frecomdat)
```

---

readMWRresults   *Read water quality monitoring results from an external file*

Description

Read water quality monitoring results from an external file

Usage

```
readMWRresults(respth, runchk = TRUE, warn = TRUE, tzone = "America/Jamaica")
```

Arguments

respth	character string of path to the results file
runchk	logical to run data checks with checkMWRresults
warn	logical to return warnings to the console (default)
tzone	character string for time zone, passed to formMWRresults

Details

Date are imported with read_excel, checked with checkMWRresults, and formatted with formMWRresults.

Value

A formatted water quality monitoring results data frame that can be used for downstream analysis

See Also

readMWRresultsview for troubleshooting import checks

Examples

```
respth <- system.file('extdata/ExampleResults.xlsx', package = 'MassWateR')

resdat <- readMWRresults(respth)
head(resdat)
```

---

readMWRresultsview   *Create summary spreadsheet of the water quality monitoring results*

Description

Create summary spreadsheet of unique values for each column in the water quality results file to check for data mistakes prior to running the readMWRresults function

Usage

```
readMWRresultsview(
  respth,
  columns = NULL,
  output_dir,
  output_file = NULL,
  maxlen = 8
)
```

Arguments

respth	character string of path to the results file
columns	character string indicating which columns to view, defaults to all
output_dir	character string of the output directory for the rendered file
output_file	optional character string for the name of the .csv file output, must include the file extension
maxlen	numeric to truncate numeric values to the specified length

Details

Acceptable options for the columns argument include any of the column names in the results file. The default setting (NULL) will show every column in the results file.

The output of this function can be useful to troubleshoot the checks when importing the water quality monitoring result file with readMWRresults (see https://massbays-tech.github.io/MassWateR/articles/MassWateR.html#data-import-and-checks).

Value

Creates a spreadsheet at the location specified by output_dir. Each column shows the unique values.

Examples

```
respth <- system.file('extdata/ExampleResults.xlsx', package = 'MassWateR')

# all columns
readMWRresultsview(respth, output_dir = tempdir())
```

```
# parameters and units
readMWRresultsview(respth, columns = c('Characteristic Name', 'Result Unit'),
    output_dir = tempdir())
```

---

readMWRsites   *Read site metadata from an external file*

Description

Read site metadata from an external file

Usage

```
readMWRsites(sitpth, runchk = TRUE)
```

Arguments

sitpth	character string of path to the site metadata file
runchk	logical to run data checks with checkMWRsites

Details

Data are imported with read_excel and checked with checkMWRsites.

Value

A formatted data frame of site metadata that can be used for downstream analysis

Examples

```
sitpth <- system.file('extdata/ExampleSites.xlsx', package = 'MassWateR')

sitdat <- readMWRsites(sitpth)
head(sitdat)
```

---

readMWRwqx   *Read water quality exchange (wqx) metadata input from an external*
                              file

Description

Read water quality exchange (wqx) metadata input from an external file

Usage

```
readMWRwqx(wqxpth, runchk = TRUE, warn = TRUE)
```

Arguments

wqxpth   character string of path to the wqx metadata file

runchk   logical to run data checks with checkMWRwqx

warn   logical to return warnings to the console (default)

Details

Date are imported with read_excel, checked with checkMWRwqx.

Value

A formatted data frame that can be used for downstream analysis

Examples

```
wqxpth <- system.file('extdata/ExampleWQX.xlsx', package = 'MassWateR')

wqxdat <- readMWRwqx(wqxpth)
head(wqxdat)
```

---

tabMWRacc   *Create a formatted table of quality control accuracy checks*

Description

Create a formatted table of quality control accuracy checks

Usage

```
tabMWRacc(
  res = NULL,
  acc = NULL,
  frecom = NULL,
  fset = NULL,
  runchk = TRUE,
  warn = TRUE,
  accchk = c("Field Blanks", "Lab Blanks", "Field Duplicates", "Lab Duplicates",
    "Lab Spikes / Instrument Checks"),
  type = c("individual", "summary", "percent"),
  pass_col = "#57C4AD",
  fail_col = "#DB4325",
  suffix = "%",
  caption = TRUE
)
```

Arguments

res	character string of path to the results file or data.frame for results returned by [readMWRresults](#)
acc	character string of path to the data quality objectives file for accuracy or data.frame returned by [readMWRacc](#)
frecom	character string of path to the data quality objectives file for frequency and completeness or data.frame returned by [readMWRfrecom](#), applies only if type = "summary" or type = "percent"
fset	optional list of inputs with elements named res, acc, frecom, sit, or wqx overrides the other arguments
runchk	logical to run data checks with [checkMWRresults](#) and [checkMWRacc](#), applies only if res or acc are file paths
warn	logical to return warnings to the console (default)
accchk	character string indicating which accuracy check to return, one to any of "Field Blanks", "Lab Blanks", "Field Duplicates", "Lab Duplicates", or "Lab Spikes / Instrument Checks"
type	character string indicating individual, summary or percent tabular output, see details
pass_col	character string (as hex code) for the cell color of checks that pass, applies only if type = 'percent'
fail_col	character string (as hex code) for the cell color of checks that fail, applies only if type = 'percent'
suffix	character string indicating suffix to append to percentage values
caption	logical to include a caption from accchk, only applies if type = "individual"

Details

The function can be used with inputs as paths to the relevant files or as data frames returned by [readMWRresults](#) and [readMWRacc](#). For the former, the full suite of data checks can be evaluated with runkchk = T (default) or suppressed with runchk = F, as explained in the relevant help files. In the latter case, downstream analyses may not work if data are formatted incorrectly. For convenience, a named list with the input arguments as paths or data frames can be passed to the fset argument instead. See the help file for [utilMWRinput](#).

Also note that accuracy is only evaluated on parameters that are shared between the results file and data quality objectives file for accuracy. A warning is returned for parameters that do not match between the files. This warning can be suppressed by setting warn = FALSE.

The function can return three types of tables as specified with the type argument: "individual", "summary", or "percent". The individual tables are specific to each type of accuracy check for each parameter (e.g., field blanks, lab blanks, etc.). The summary table summarizes all accuracy checks by the number of checks and how many hit/misses are returned for each across all parameters. The percent table is similar to the summary table, but showing only percentages with appropriate color-coding for hit/misses. The data quality objectives file for frequency and completeness is required if type = "summary" or type = "percent".

For type = "individual", the quality control tables for accuracy are retrieved by specifying the check with the accchk argument. The accchk argument can be used to specify one of the following values to retrieve the relevant tables: "Field Blanks", "Lab Blanks", "Field Duplicates", "Lab Duplicates", or "Lab Spikes / Instrument Checks".

For type = "summary", the function summarizes all accuracy checks by counting the number of quality control checks, number of misses, and percent acceptance for each parameter. All accuracy checks are used and the accchk argument does not apply.

For type = "percent", the function returns a similar table as for the summary option, except only the percentage of checks that pass for each parameter are shown in wide format. Cells are color-coded based on the percentage of checks that have passed using the percent thresholds from the % Completeness column of the data quality objectives file for frequency and completeness. Parameters without an entry for % Completeness are not color-coded and an appropriate warning is returned. All accuracy checks are used and the accchk argument does not apply.

Inputs for the results and data quality objectives for accuracy are processed internally with [qcMWRacc](#) and the same arguments are accepted for this function, in addition to others listed above.

Value

A [flextable](#) object with formatted results.

Examples

```
##
# using file paths

# results path
respth <- system.file('extdata/ExampleResults.xlsx', package = 'MassWateR')

# accuracy path
```

```
accpth <- system.file('extdata/ExampleDQOAccuracy.xlsx',
     package = 'MassWateR')

# frequency and completeness path
frecompth <- system.file('extdata/ExampleDQOFrequencyCompleteness.xlsx',
     package = 'MassWateR')

# table as individual
tabMWRacc(res = respth, acc = accpth, frecom = frecompth, type = 'individual',
     accchk = 'Field Blanks')
```

---

tabMWRcom   *Create a formatted table of quality control completeness checks*

Description

Create a formatted table of quality control completeness checks

Usage

```
tabMWRcom(
  res = NULL,
  frecom = NULL,
  fset = NULL,
  runchk = TRUE,
  warn = TRUE,
  pass_col = "#57C4AD",
  fail_col = "#DB4325",
  digits = 0,
  suffix = "%",
  parameterwd = 1.15,
  noteswd = 3
)
```

Arguments

res	character string of path to the results file or data.frame for results returned by readMWRresults
frecom	character string of path to the data quality objectives file for frequency and completeness or data.frame returned by readMWRfrecom
fset	optional list of inputs with elements named res, acc, frecom, sit, or wqx overrides the other arguments
runchk	logical to run data checks with checkMWRresults and checkMWRfrecom, applies only if res or frecom are file paths
warn	logical to return warnings to the console (default)
pass_col	character string (as hex code) for the cell color of checks that pass

fail_col	character string (as hex code) for the cell color of checks that fail
digits	numeric indicating number of significant digits to report for percentages
suffix	character string indicating suffix to append to percentage values
parameterwd	numeric indicating width of the parameter column
noteswd	numeric indicating width of notes column

Details

The function can be used with inputs as paths to the relevant files or as data frames returned by [readMWRresults](#) and [readMWRfrecom](#). For the former, the full suite of data checks can be evaluated with runkchk = T (default) or suppressed with runchk = F, as explained in the relevant help files. In the latter case, downstream analyses may not work if data are formatted incorrectly. For convenience, a named list with the input arguments as paths or data frames can be passed to the fset argument instead. See the help file for [utilMWRinput](#).

Also note that completeness is only evaluated on parameters that are shared between the results file and data quality objectives file for frequency and completeness. A warning is returned for parameters that do not match between the files. This warning can be suppressed by setting warn = FALSE.

A summary table showing the number of data records, number of qualified records, and percent completeness is created. The % Completeness column shows cells as green or red if the required percentage of observations for completeness are present as specified in the data quality objectives file. The Hit/ Miss column shows similar information but in text format, i.e., MISS is shown if the quality control standard for completeness is not met.

Inputs for the results and data quality objectives for frequency and completeness are processed internally with [qcMWRcom](#) and the same arguments are accepted for this function, in addition to others listed above.

Value

A [flextable](#) object with formatted results showing summary counts for all completeness checks for each parameter.

Examples

```
##
# using file paths

# results path
respth <- system.file('extdata/ExampleResults.xlsx', package = 'MassWateR')

# frequency and completeness path
frecompth <- system.file('extdata/ExampleDQOFrequencyCompleteness.xlsx',
     package = 'MassWateR')

tabMWRcom(res = respth, frecom = frecompth)

##
# using data frames
```

```
# results data
resdat <- readMWRresults(respth)

# frequency and completeness data
frecomdat <- readMWRfrecom(frecompth)

tabMWRcom(res = resdat, frecom = frecomdat)
```

---

tabMWRfre   *Create a formatted table of quality control frequency checks*

Description

Create a formatted table of quality control frequency checks

Usage

```
tabMWRfre(
  res = NULL,
  acc = NULL,
  frecom = NULL,
  fset = NULL,
  runchk = TRUE,
  warn = TRUE,
  type = c("summary", "percent"),
  pass_col = "#57C4AD",
  fail_col = "#DB4325",
  digits = 0,
  suffix = "%"
)
```

Arguments

res	character string of path to the results file or data.frame for results returned by [readMWRresults](#)
acc	character string of path to the data quality objectives file for accuracy or data.frame returned by [readMWRacc](#)
frecom	character string of path to the data quality objectives file for frequency and completeness or data.frame returned by [readMWRfrecom](#)
fset	optional list of inputs with elements named res, acc, frecom, sit, or wqx overrides the other arguments
runchk	logical to run data checks with [checkMWRresults](#) and [checkMWRfrecom](#), applies only if res or frecom are file paths
warn	logical to return warnings to the console (default)

type	character string indicating `summary` or `percent` tabular output, see details
pass_col	character string (as hex code) for the cell color of checks that pass, applies only if type = 'percent'
fail_col	character string (as hex code) for the cell color of checks that fail, applies only if type = 'percent'
digits	numeric indicating number of significant digits to report for percentages
suffix	character string indicating suffix to append to percentage values

Details

The function can be used with inputs as paths to the relevant files or as data frames returned by [readMWRresults](), [readMWRacc](), and [readMWRfrecom](). For the former, the full suite of data checks can be evaluated with runkchk = T (default) or suppressed with runchk = F, as explained in the relevant help files. In the latter case, downstream analyses may not work if data are formatted incorrectly. For convenience, a named list with the input arguments as paths or data frames can be passed to the fset argument instead. See the help file for [utilMWRinput]().

Also note that completeness is only evaluated on parameters that are shared between the results file and data quality objectives file for frequency and completeness. A warning is returned for parameters that do not match between the files. This warning can be suppressed by setting warn = FALSE.

The quality control tables for frequency show the number of records that apply to a given check (e.g., Lab Blank, Field Blank, etc.) relative to the number of "regular" data records (e.g., field samples or measures) for each parameter. A summary of all frequency checks for each parameter is provided if type = "summary" or a color-coded table showing similar information as percentages for each parameter is provided if type = "percent".

Inputs for the results and data quality objectives for accuracy and frequency and completeness are processed internally with [qcMWRcom]() and the same arguments are accepted for this function, in addition to others listed above.

Value

A [flextable]() object with formatted results.

Examples

```
##
# using file paths

# results path
respth <- system.file('extdata/ExampleResults.xlsx', package = 'MassWateR')

# dqo accuracy data path
accpth <- system.file('extdata/ExampleDQOAccuracy.xlsx', package = 'MassWateR')

# frequency and completeness path
frecompth <- system.file('extdata/ExampleDQOFrequencyCompleteness.xlsx',
     package = 'MassWateR')
```

```
# table as summary
tabMWRfre(res = respth, acc = accpth, frecom = frecompth, type = 'summary')

# table as percent
tabMWRfre(res = respth, acc = accpth, frecom = frecompth, type = 'percent')

##
# using data frames

# results data
resdat <- readMWRresults(respth)

# accuracy data
accdat <- readMWRacc(accpth)

# frequency and completeness data
frecomdat <- readMWRfrecom(frecompth)

# table as summary
tabMWRfre(res = resdat, acc = accdat, frecom = frecomdat, type = 'summary')

# table as percent
tabMWRfre(res = resdat, acc = accdat, frecom = frecomdat, type = 'percent')
```

---

tabMWRwqx   *Create and save tables in a single workbook for WQX upload*

Description

Create and save tables in a single workbook for WQX upload

Usage

```
tabMWRwqx(
  res = NULL,
  acc = NULL,
  sit = NULL,
  wqx = NULL,
  fset = NULL,
  output_dir,
  output_file = NULL,
  warn = TRUE,
  runchk = TRUE
)
```

Arguments

res   character string of path to the results file or data.frame for results returned by
               readMWRresults

acc	character string of path to the data quality objectives file for accuracy or data.frame returned by readMWRacc
sit	character string of path to the site metadata file or data.frame for site metadata returned by readMWRsites
wqx	character string of path to the wqx metadata file or data.frame for wqx metadata returned by readMWRwqx
fset	optional list of inputs with elements named res, acc, frecom, sit, or wqx overrides the other arguments
output_dir	character string of the output directory for the results
output_file	optional character string for the file name, must include .xlsx suffix
warn	logical to return warnings to the console (default)
runchk	logical to run data checks with checkMWRresults, checkMWRacc, checkMWRsites, checkMWRwqx, applies only if res, acc, sit, or wqx are file paths

Details

This function will export a single Excel workbook with three sheets, named "Project", "Locations", and "Results". The output is populated with as much content as possible based on information in the input files. The remainder of the information not included in the output will need to be manually entered before uploading the data to WQX. All required columns are present, but individual rows will need to be verified for completeness. It is the responsibility of the user to verify this information is complete and correct before uploading the data.

The workflow for using this function is to import the required data (results, data quality objectives file for accuracy, site metadata, and wqx metadata) and to fix any errors noted on import prior to creating the output. The function can be used with inputs as paths to the relevant files or as data frames returned by readMWRresults, readMWRacc, readMWRsites, and readMWRwqx. For the former, the full suite of data checks can be evaluated with runkchk = T (default) or suppressed with runchk = F, as explained in the relevant help files. In the latter case, downstream analyses may not work if data are formatted incorrectly. For convenience, a named list with the input arguments as paths or data frames can be passed to the fset argument instead. See the help file for utilMWRinput.

The name of the output file can also be changed using the output_file argument, the default being wqxtab.xlsx. Warnings can also be turned off or on (default) using the warn argument. This returns any warnings when data are imported and only applies if the file inputs are paths.

Value

An Excel workbook named wqxtab.xlsx (or name passed to output_file) will be saved in the directory specified by output_dir. The workbook will include three sheets names "Projects", "Locations", and "Results".

Examples

```
# results data path
respth <- system.file('extdata/ExampleResults.xlsx', package = 'MassWateR')

# dqo accuracy data path
accpth <- system.file('extdata/ExampleDQOAccuracy.xlsx', package = 'MassWateR')
```

```
# site data path
sitpth <- system.file('extdata/ExampleSites.xlsx', package = 'MassWateR')

# wqx data path
wqxpth <- system.file('extdata/ExampleWQX.xlsx', package = 'MassWateR')

# results data
resdat <- readMWRresults(respth)

# accuracy data
accdat <- readMWRacc(accpth)

# site data
sitdat <- readMWRsites(sitpth)

# wqx data
wqxdat <- readMWRwqx(wqxpth)

# create workbook
tabMWRwqx(res = resdat, acc = accdat, sit = sitdat, wqx = wqxdat, output_dir = tempdir())
```

---

thresholdMWR   *Master thresholds list for analysis of results data*

Description

Master thresholds list for analysis of results data

Usage

```
thresholdMWR
```

Format

A `data.frame` of 28 rows and 10 columns

Details

This file includes appropriate threshold values of water quality parameters for marine and freshwater
environments based on state standards or typical ranges in Massachusetts.

Examples

```
thresholdMWR
```

---

utilMWRfilter *Filter results data by parameter, date range, site, result attributes, and/or location group*

Description

Filter results data by parameter, date range, site, result attributes, and/or location group

Usage

```
utilMWRfilter(
  resdat,
  sitdat = NULL,
  param,
  dtrng = NULL,
  site = NULL,
  resultatt = NULL,
  locgroup = NULL,
  alllocgroup = FALSE,
  allresultatt = FALSE
)
```

Arguments

resdat	results data as returned by readMWRresults
sitdat	site metadata file as returned by readMWRresults
param	character string to filter results by a parameter in ″Characteristic Name″
dtrng	character string of length two for the date ranges as YYYY-MM-DD
site	character string of sites to include, default all
resultatt	character string of result attributes to include, default all
locgroup	character string of location groups to include from the ″Location Group″ column in the site metadata file
alllocgroup	logical indicating if results data are filtered by all location groups in ″Location Group″ in the site metadata file if locgroup = NULL, used only in anlzMWRdate
allresultatt	logical indicating if results data are filtered by all result attributes if resultatt = NULL, used only in anlzMWRsite

Value

resdat filtered by param, dtrng, site, resultatt, and/or locgroup, otherwise resdat filtered only by param if other arguments are NULL

Examples

```
# results file path
respth <- system.file('extdata/ExampleResults.xlsx', package = 'MassWateR')

# results data
resdat <- readMWRresults(respth)

# site data path
sitpth <- system.file('extdata/ExampleSites.xlsx', package = 'MassWateR')

# site data
sitdat <- readMWRsites(sitpth)

# filter by parameter, date range
utilMWRfilter(resdat, param = 'DO', dtrng = c('2022-06-01', '2022-06-30'))

# filter by parameter, site
utilMWRfilter(resdat, param = 'DO', site = c('ABT-026', 'ABT-062', 'ABT-077'))

# filter by parameter, result attribute
utilMWRfilter(resdat, param = 'DO', resultatt = 'DRY')

# filter by parameter, location group, date range
utilMWRfilter(resdat, param = 'DO', sitdat = sitdat,
     locgroup = 'Assabet', dtrng = c('2022-06-01', '2022-06-30'))
```

---

utilMWRfiltersurface   *Filter results data to surface measurements*

Description

Filter results data to surface measurements

Usage

```
utilMWRfiltersurface(resdat)
```

Arguments

resdat   results data as returned by [readMWRresults](#)

Details

This function is used internally for all analysis functions

Value

resdat filtered by `Activity Depth/Height Measure` less than or equal to 1 meter or 3.3 feet or
`Activity Relative Depth Name` as "Surface"

Examples

```
# results file path
respth <- system.file('extdata/ExampleResults.xlsx', package = 'MassWateR')

# results data
resdat <- readMWRresults(respth)

# filter surface data
utilMWRfiltersurface(resdat)
```

---

utilMWRfre   *Prep results data for frequency checks*

Description

Prep results data for frequency checks

Usage

```
utilMWRfre(resdat, param, accdat, warn = TRUE)
```

Arguments

resdat	results data as returned by readMWRresults
param	character string to filter results and check if a parameter in the "Characteristic Name" column in the results file is also found in the data quality objectives file for accuracy, see details
accdat	data.frame for data quality objectives file for accuracy as returned by readMWRacc
warn	logical to return warnings to the console (default)

Details

This function is similar to utilMWRlimits with some additional processing appropriate for creating the frequency table in tabMWRfree. The param argument is used to identify the appropriate "MDL" or "UQL" values in the data quality objectives file for accuracy. A warning is returned to the console if the accuracy file does not contain the appropriate information for the parameter. Results will be filtered by param regardless of any warning.

Value

resdat filtered by param with any entries in "Result Value" as "BDL" or "AQL" replaced with appropriate values in the "Quantitation Limit" column, if present, otherwise the "MDL" or "UQL" columns from the data quality objectives file for accuracy are used. Values as "BDL" use one half of the appropriate limit. Values not in the "Value Range" column of the accuracy file are removed from the output.

Examples

```
# results file path
respth <- system.file('extdata/ExampleResults.xlsx', package = 'MassWateR')

# results data
resdat <- readMWRresults(respth)

# accuracy path
accpth <- system.file('extdata/ExampleDQOAccuracy.xlsx',
     package = 'MassWateR')

# accuracy data
accdat <- readMWRacc(accpth)

# apply to total phosphorus
utilMWRfre(resdat, accdat, param = 'TP')

# apply to E.coli
utilMWRfre(resdat, accdat, param = 'E.coli')
```

---

utilMWRhttpgrace   *Load external file from remote source, fail gracefully*

Description

Load external file from remote source, fail gracefully

Usage

```
utilMWRhttpgrace(remote_file)
```

Arguments

remote_file   URL of the external file

Value

The external file as an RData object

Examples

```
# fails gracefully
utilMWRhttpgrace('http://httpbin.org/status/404')

# imports data or fails gracefully
fl <- 'https://github.com/massbays-tech/MassWateRdata/raw/main/data/streamsMWR.RData'
utilMWRhttpgrace(fl)
```

---

utilMWRinput   *Utility function to import data as paths or data frames*

Description

Utility function to import data as paths or data frames

Usage

```
utilMWRinput(
  res = NULL,
  acc = NULL,
  frecom = NULL,
  sit = NULL,
  wqx = NULL,
  fset = NULL,
  runchk = TRUE,
  warn = TRUE
)
```

Arguments

res	character string of path to the results file or data.frame for results returned by [readMWRresults](#)
acc	character string of path to the data quality objectives file for accuracy or data.frame returned by [readMWRacc](#)
frecom	character string of path to the data quality objectives file for frequency and completeness or data.frame returned by [readMWRfrecom](#)
sit	character string of path to the site metadata file or data.frame for site metadata returned by [readMWRsites](#)
wqx	character string of path to the wqx metadata file or data.frame for wqx metadata returned by [readMWRwqx](#)
fset	optional list of inputs with elements named res, acc, frecom, sit, or wqx, overrides the other arguments, see details
runchk	logical to run data checks with [checkMWRresults](#), [checkMWRacc](#), [checkMWRfrecom](#), [checkMWRsites](#), or [checkMWRwqx](#), applies only if res, acc, frecom, sit, or wqx are file paths
warn	logical to return warnings to the console (default)

Details

The function is used internally by others to import data from paths to the relevant files or as data frames returned by [readMWRresults](#), [readMWRacc](#), [readMWRfrecom](#), [readMWRsites](#), or [readMWRwqx](#). For the former, the full suite of data checks can be evaluated with runkchk = T (default) or suppressed with runchk = F.

The fset argument can used in place of the preceding arguments. The argument accepts a list with named elements as res, acc, frecom, sit, or wqx, where the elements are either character strings of the path or data frames to the corresponding inputs. Missing elements will be interpreted as NULL values. This argument is provided as convenience to apply a single list as input versus separate inputs for each argument.

Any of the arguments for the data files can be NULL, used as a convenience for downstream functions that do not require all.

Value

A five element list with the imported results, data quality objective files, site metadata, and wqx metadata, named "resdat", "accdat", "frecomdat", "sitdat", and "wqxdat", respectively.

Examples

```
##
# using file paths

# results path
respth <- system.file('extdata/ExampleResults.xlsx', package = 'MassWateR')

# accuracy path
accpth <- system.file('extdata/ExampleDQOAccuracy.xlsx', package = 'MassWateR')

# frequency and completeness path
frecompth <- system.file('extdata/ExampleDQOFrequencyCompleteness.xlsx',
    package = 'MassWateR')

# site path
sitpth <- system.file('extdata/ExampleSites.xlsx', package = 'MassWateR')

# wqx path
wqxpth <- system.file('extdata/ExampleWQX.xlsx', package = 'MassWateR')

inp <- utilMWRinput(res = respth, acc = accpth, frecom = frecompth, sit = sitpth, wqx = wqxpth)
inp$resdat
inp$accdat
inp$frecomdat
inp$sitdat
inp$wqxdat

##
# using data frames

# results data
resdat <- readMWRresults(respth)

# accuracy data
accdat <- readMWRacc(accpth)

# frequency and completeness data
frecomdat <- readMWRfrecom(frecompth)
```

```
# site data
sitdat <- readMWRsites(sitpth)

# wqx data
wqxdat <- readMWRwqx(wqxpth)

inp <- utilMWRinput(res = resdat, acc = accdat, frecom = frecomdat, sit = sitdat, wqx = wqxpth)
inp$resdat
inp$accdat
inp$frecomdat
inp$sitdat
inp$wqxdat

##
# using fset as list input

# input with paths to files
fset <- list(
  res = respth,
  acc = accpth,
  frecom = frecompth,
  sit = sitpth,
  wqx = wqxpth
)
utilMWRinput(fset = fset)
```

---

utilMWRinputcheck *Check if required inputs are present for a function*

Description

Check if required inputs are present for a function

Usage

```
utilMWRinputcheck(inputs)
```

Arguments

inputs	list of arguments passed from the parent function

Value

NULL if all inputs are present, otherwise an error message indicating which inputs are missing

Examples

```
inputchk <- formals(tabMWRcom)
inputchk$res <- system.file('extdata/ExampleResults.xlsx', package = 'MassWateR')
inputchk$frecom <- system.file('extdata/ExampleDQOFrequencyCompleteness.xlsx',
  package = 'MassWateR')

utilMWRinputcheck(inputchk)
```

---

utilMWRlimits   *Fill results data as BDL or AQL with appropriate values*

Description

Fill results data as BDL or AQL with appropriate values

Usage

```
utilMWRlimits(resdat, param, accdat, warn = TRUE)
```

Arguments

resdat   results data as returned by [readMWRresults](#)

param   character string to filter results and check if a parameter in the "Characteristic
                Name" column in the results file is also found in the data quality objectives file
                for accuracy, see details

accdat   data.frame for data quality objectives file for accuracy as returned by [readMWRacc](#)

warn   logical to return warnings to the console (default)

Details

The param argument is used to identify the appropriate "MDL" or "UQL" values in the data quality
objectives file for accuracy. A warning is returned to the console if the accuracy file does not contain
the appropriate information for the parameter. Results will be filtered by param regardless of any
warning.

Value

resdat filtered by param with any entries in "Result Value" as "BDL" or "AQL" replaced with
appropriate values in the "Quantitation Limit" column, if present, otherwise the "MDL" or "UQL"
columns from the data quality objectives file for accuracy are used. Values as "BDL" use one half
of the appropriate limit. Output only includes rows with the activity type as "Field Msr/Obs" or
"Sample-Routine".

Examples

```
# results file path
respth <- system.file('extdata/ExampleResults.xlsx', package = 'MassWateR')

# results data
resdat <- readMWRresults(respth)

# accuracy path
accpth <- system.file('extdata/ExampleDQOAccuracy.xlsx',
     package = 'MassWateR')

# accuracy data
accdat <- readMWRacc(accpth)

# apply to total phosphorus
utilMWRlimits(resdat, accdat, param = 'TP')

# apply to E.coli
utilMWRlimits(resdat, accdat, param = 'E.coli')
```

---

utilMWRoutlier   *Identify outliers in a numeric vector*

Description

Identify outliers in a numeric vector

Usage

```
utilMWRoutlier(x, logscl)
```

Arguments

x	numeric vector of any length
logscl	logical to indicate if vector should be log10-transformed first

Details

Outliers are identified as 1.5 times the interquartile range

Value

A logical vector equal in length to x indicating TRUE for outliers or FALSE for within normal range

Examples

```
x <- rnorm(20)
utilMWRoutlier(x, logscl = FALSE)
```

---

utilMWRsumfun   *Verify summary function*

Description

Verify summary function

Usage

```
utilMWRsumfun(accdat, param, sumfun = "auto")
```

Arguments

accdat   data.frame for data quality objectives file for accuracy as returned by readMWRacc

param   character string for the parameter to evaluate as provided in the "Parameter"
                column of "accdat"

sumfun   character indicating one of "auto" (default), "mean", "geomean", "median",
                "min", or "max", see details

Details

This function verifies appropriate summary functions are passed from sumfun. The mean or geo-
metric mean output is used for sumfun = "auto" based on information in the data quality objective
file for accuracy, i.e., parameters with "log" in any of the columns are summarized with the geomet-
ric mean, otherwise arithmetic. Using "mean" or "geomean" for sumfun will apply the appropriate
function regardless of information in the data quality objective file for accuracy.

Value

Character indicating the appropriate summary function based on the value passed to sumfun.

Examples

```
# accuracy path
accpth <- system.file('extdata/ExampleDQOAccuracy.xlsx',
     package = 'MassWateR')

# accuracy data
accdat <- readMWRacc(accpth)

# geomean auto
utilMWRsumfun(accdat, param = 'E.coli')

# mean force
utilMWRsumfun(accdat, param = 'E.coli', sumfun = 'mean')

# mean auto
utilMWRsumfun(accdat, param = 'DO')
```

```
# geomean force
utilMWRsumfun(accdat, param = 'DO', sumfun = 'geomean')
```

---

utilMWRsummary   *Summarize a results data frame by a grouping variable*

Description

Summarize a results data frame by a grouping variable

Usage

```
utilMWRsummary(dat, accdat, param, sumfun = "auto", confint)
```

Arguments

dat	input data frame
accdat	data.frame for data quality objectives file for accuracy as returned by [readMWRacc](#)
param	character string for the parameter to evaluate as provided in the "Parameter" column of "accdat"
sumfun	character indicating one of "auto" (default), "mean", "geomean", "median", "min", or "max", see details
confint	logical if user expects a confidence interval to be returned with the summary

Details

This function summarizes a results data frame by an existing grouping variable using the function supplied to sumfun. The mean or geometric mean is used for sumfun = "auto" based on information in the data quality objective file for accuracy, i.e., parameters with "log" in any of the columns are summarized with the geometric mean, otherwise arithmetic. Using "mean" or "geomean" for sumfun will apply the appropriate function regardless of information in the data quality objective file for accuracy.

Value

A summarized data frame, a warning will be returned if the confidence interval cannot be estimated and confint = TRUE

Examples

```
library(dplyr)

# results data path
respth <- system.file('extdata/ExampleResults.xlsx', package = 'MassWateR')

# results data
```

```
resdat <- readMWRresults(respth)

# accuracy path
accpth <- system.file('extdata/ExampleDQOAccuracy.xlsx',
     package = 'MassWateR')

# accuracy data
accdat <- readMWRacc(accpth)

# fill BDL, AQL
resdat <- utilMWRlimits(resdat = resdat, accdat = accdat, param = "DO")

dat <- resdat %>%
  group_by(`Monitoring Location ID`)

# summarize sites by mean
utilMWRsummary(dat, accdat, param = 'DO', sumfun = 'auto', confint = TRUE)

# summarize sites by minimum
utilMWRsummary(dat, accdat, param = 'DO', sumfun = 'min', confint = FALSE)
```

---

utilMWRthresh   *Get threshold lines from thresholdMWR*

Description

Get threshold lines from thresholdMWR

Usage

```
utilMWRthresh(resdat, param, thresh, threshlab = NULL)
```

Arguments

resdat	results data as returned by [readMWRresults](#)
param	character string to first filter results by a parameter in "Characteristic Name"
thresh	character indicating if relevant freshwater or marine threshold lines are included, one of "fresh", "marine", or "none", or a single numeric value to override the values included with the package
threshlab	optional character string indicating legend label for the threshold, required only if thresh is numeric

Value

If thresh is not numeric and thresholds are available for param, a data.frame of relevant marine or freshwater thresholds, otherwise NULL. If thresh is numeric, a data.frame of the threshold with the appropriate label from threshlabel.

Examples

```
# results file path
respth <- system.file('extdata/ExampleResults.xlsx', package = 'MassWateR')

# results data
resdat <- readMWRresults(respth)

# get threshold lines
utilMWRthresh(resdat = resdat, param = 'E.coli', thresh = 'fresh')

# user-defined numeric threshold line
utilMWRthresh(resdat = resdat, param = 'TP', thresh = 5, threshlab = 'My threshold')
```

---

utilMWRtitle   *Format the title for analyze functions*

Description

Format the title for analyze functions

Usage

```
utilMWRtitle(
  param,
  accdat = NULL,
  sumfun = NULL,
  site = NULL,
  dtrng = NULL,
  resultatt = NULL,
  locgroup = NULL
)
```

Arguments

param	character string of the parameter to plot
accdat	optional data.frame for data quality objectives file for accuracy as returned by [readMWRacc](#)
sumfun	optional character indicating one of "auto", "mean", "geomean", "median", "min", or "max"
site	character string of sites to include
dtrng	character string of length two for the date ranges as YYYY-MM-DD
resultatt	character string of result attributes to plot
locgroup	character string of location groups to plot from the "Location Group" column in the site metadata file

Details

All arguments are optional except `param`, appropriate text strings are appended to the `param` argument for all other optional arguments indicating the level of filtering used in the plot and data summary if appropriate

Value

A formatted character string used for the title in analysis plots

Examples

```
# no filters
utilMWRtitle(param = 'DO')

# filter by date only
utilMWRtitle(param = 'DO', dtrng = c('2021-05-01', '2021-07-31'))

# filter by all
utilMWRtitle(param = 'DO', site = 'test', dtrng = c('2021-05-01', '2021-07-31'),
    resultatt = 'test', locgroup = 'test')

# title using summary
accpth <- system.file('extdata/ExampleDQOAccuracy.xlsx', package = 'MassWateR')
accdat <- readMWRacc(accpth, runchk = FALSE)
utilMWRtitle(param = 'DO', accdat = accdat, sumfun = 'auto', site = 'test',
    dtrng = c('2021-05-01', '2021-07-31'), resultatt = 'test', locgroup = 'test')
```

---

utilMWRvaluerange   *Check if incomplete range in* Value Range *column*

Description

Check if incomplete range in Value Range column

Usage

```
utilMWRvaluerange(accdat)
```

Arguments

accdat   data.frame for data quality objectives file for accuracy as returned by readMWRacc

Details

The function evaluates if an incomplete or overlapping range is present in the Value Range column of the data quality objectives file for accuracy

Value

A named vector of ″gap″, ″nogap″, or ″overlap″ indicating if a gap is present, no gap is present, or an overlap is present in the ranges provided by the value range for each parameter. The names correspond to the parameters.

Examples

```
# accuracy path
accpth <- system.file('extdata/ExampleDQOAccuracy.xlsx',
     package = 'MassWateR')

# accuracy data with no checks
accdat <- readxl::read_excel(accpth, na = c('NA', ''), col_types = 'text')
accdat <- dplyr::mutate(accdat, dplyr::across(-c(`Value Range`), ~ dplyr::na_if(.x, 'na')))

utilMWRvaluerange(accdat)
```

---

utilMWRyscale   *Get logical value for y axis scaling*

Description

Get logical value for y axis scaling

Usage

```
utilMWRyscale(accdat, param, yscl = ″auto″)
```

Arguments

accdat	data.frame for data quality objectives file for accuracy as returned by [readMWRacc](#)
param	character string for the parameter to evaluate as provided in the ″Parameter″ column of ″accdat″
yscl	character indicating one of ″auto″ (default), ″log″, or ″linear″

Value

A logical value indicating TRUE for log10-scale, FALSE for arithmetic (linear)

Examples

```
# accuracy path
accpth <- system.file('extdata/ExampleDQOAccuracy.xlsx',
     package = 'MassWateR')

# accuracy data
accdat <- readMWRacc(accpth)
```

```
# log auto
utilMWRyscale(accdat, param = 'E.coli')

# linear force
utilMWRyscale(accdat, param = 'E.coli', yscl = 'linear')

# linear auto
utilMWRyscale(accdat, param = 'DO')

# log force
utilMWRyscale(accdat, param = 'DO', yscl = 'log')
```

Index