

# Package ‘MassWaterR’

May 29, 2025

**Title** Quality Control and Analysis of Massachusetts Water Quality Data

**Version** 2.2.0

**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, maptiles, RColorBrewer, rmarkdown, readxl, sf, tidyr, tidyterra, tibble, units, writexl

**License** CC0

**Encoding** UTF-8

**RoxygenNote** 7.3.2

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

**BugReports** <https://github.com/massbays-tech/MassWaterR/issues>

**LazyData** true

**LazyDataCompression** xz

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

**Config/testthat/edition** 3

**NeedsCompilation** no

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

**Maintainer** Marcus Beck <mbeck@tbep.org>

**Repository** CRAN

**Date/Publication** 2025-05-29 12:30:02 UTC

## Contents

anlzMWRdate . . . . .	3
anlzMWRmap . . . . .	6
anlzMWRoutlier . . . . .	9
anlzMWRoutlierall . . . . .	12
anlzMWRseason . . . . .	14
anlzMWRsite . . . . .	17
checkMWRacc . . . . .	21
checkMWRcens . . . . .	22
checkMWRfrecom . . . . .	23
checkMWRresults . . . . .	24
checkMWRsites . . . . .	26
checkMWRwqx . . . . .	27
formMWRacc . . . . .	28
formMWRcens . . . . .	29
formMWRfrecom . . . . .	30
formMWRresults . . . . .	31
formMWRwqx . . . . .	32
paramsMWR . . . . .	32
qcMWRacc . . . . .	33
qcMWRcom . . . . .	35
qcMWRfre . . . . .	37
qcMWRreview . . . . .	39
readMWRacc . . . . .	41
readMWRcens . . . . .	42
readMWRfrecom . . . . .	43
readMWRresults . . . . .	43
readMWRresultsview . . . . .	44
readMWRsites . . . . .	45
readMWRwqx . . . . .	46
tabMWRacc . . . . .	47
tabMWRcom . . . . .	49
tabMWRfre . . . . .	51
tabMWRwqx . . . . .	53
thresholdMWR . . . . .	55
utilMWRfilter . . . . .	56
utilMWRfiltersurface . . . . .	57
utilMWRfre . . . . .	58
utilMWRhttpgrace . . . . .	59
utilMWRinput . . . . .	60
utilMWRinputcheck . . . . .	63
utilMWRlimits . . . . .	64
utilMWRoutlier . . . . .	65
utilMWRsheet . . . . .	66
utilMWRsumfun . . . . .	68
utilMWRsummary . . . . .	69
utilMWRthresh . . . . .	71

*anlzMWRdate* 3

utilMWRtitle . . . . . 72  
utilMWRvaluerange . . . . . 73  
utilMWRyscale . . . . . 74

**Index** 76

---

*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,  
  bssize = 11,  
  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 <a href="#">paramsMWR</a>
acc	character string of path to the data quality objectives file for accuracy or data.frame returned by <a href="#">readMWRacc</a>
sit	optional character string of path to the site metadata file or data.frame of site metadata returned by <a href="#">readMWRsites</a> , required if locgroup is not NULL
fset	optional list of inputs with elements named res, acc, frecom, sit, or wxq 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
repe1	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 repe1 = 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 <a href="#">RColorBrewer</a> , 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
bssize	numeric for overall plot text scaling, passed to <a href="#">theme_minimal</a>
runchk	logical to run data checks with <a href="#">checkMWRresults</a> or <a href="#">checkMWRacc</a> , 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 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 = 'MassWaterR')
```

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

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

# accuracy data
accdat <- readMWRacc(accpth)

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

# 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,
  psize = 4,
  repel = TRUE,
  labsize = 3,
  palcol = "Greens",
  palcolrev = FALSE,
  sumfun = "auto",
  crs = 4326,
  zoom = 11,
  addwater = "medium",
```

```

    watercol = "lightblue",
    matype = NULL,
    buffdist = 2,
    scaledist = "km",
    northloc = "tl",
    scaleloc = "br",
    latlon = TRUE,
    ttlsize = 1.2,
    bssize = 11,
    runchk = TRUE,
    warn = TRUE
)

```

### Arguments

res	character string of path to the results file or data.frame for results returned by <a href="#">readMWRresults</a>
param	character string of the parameter to plot, must conform to entries in the "Simple Parameter" column of <a href="#">paramsMWR</a>
acc	character string of path to the data quality objectives file for accuracy or data.frame returned by <a href="#">readMWRacc</a>
sit	character string of path to the site metadata file or data.frame of site metadata returned by <a href="#">readMWRsites</a>
fset	optional list of inputs with elements named res, acc, frecom, sit, or wxq 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 <a href="#">RColorBrewer</a> , 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 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
bssize	numeric for overall plot text scaling, passed to <a href="#">theme_gray</a>
runchk	logical to run data checks with <a href="#">checkMWRresults</a> , <a href="#">checkMWRacc</a> , or <a href="#">checkMWRsites</a> , 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. 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 "OpenStreetMap", "OpenStreetMap.DE", "OpenStreetMap.France", "OpenStreetMap.HOT", "OpenTopoMap", "Esri.WorldStreetMap", "Esri.DeLorme", "Esri.WorldTopoMap", "Esri.WorldImagery", "Esri.WorldTerrain", "Esri.WorldShadedRelief", "Esri.OceanBasemap", "Esri.NatGeoWorldMap", "Esri.WorldGrayCanvas", "CartoDB.Positron", "CartoDB.PositronNoLabels",



"CartoDB.PositronOnlyLabels", "CartoDB.DarkMatter", "CartoDB.DarkMatterNoLabels", "CartoDB.DarkMatterOnlyLabels", "CartoDB.Voyager", "CartoDB.VoyagerNoLabels", or "CartoDB.VoyagerOnlyLabels". Use `mapttype = 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 = 'MassWaterR')

# results data
resdat <- readMWRresults(respth)

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

# accuracy data
accdat <- readMWRacc(accpth)

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

# 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**

```
anzMWRoutlier(
  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,
  bssize = 11,
  runchk = TRUE,
  warn = TRUE
)
```

**Arguments**

res	character string of path to the results file or data.frame for results returned by <a href="#">readMWRresults</a>
param	character string of the parameter to plot, must conform to entries in the "Simple Parameter" column of <a href="#">paramsMWR</a>
acc	character string of path to the data quality objectives file for accuracy or data.frame returned by <a href="#">readMWRacc</a>
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

bssize	numeric for overall plot text scaling, passed to <code>theme_minimal</code>
runchk	logical to run data checks with <code>checkMWRresults</code> or <code>checkMWRacc</code> , 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 = 'MassWaterR')

# results data
resdat <- readMWRresults(respth)

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

# accuracy data
accdat <- readMWRacc(accpth)

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

```
# outliers by site
anlzMWRoutlier(res = resdat, param = 'D0', acc = accdat, group = 'site')

# outliers by site, May through July 2021 only
anlzMWRoutlier(res = resdat, param = 'D0', acc = accdat, group = 'site',
  dtrng = c('2022-05-01', '2022-07-31'))

# outliers by month, type as jitterbox
anlzMWRoutlier(res = resdat, param = 'D0', acc = accdat, group = 'month', type = 'jitterbox')

# outliers by month, type as jitter
anlzMWRoutlier(res = resdat, param = 'D0', acc = accdat, group = 'month', type = 'jitter')

# data frame output
anlzMWRoutlier(res = resdat, param = 'D0', acc = accdat, group = 'month', outliers = TRUE)
```

---

anlzMWRoutlierall	<i>Analyze outliers in results file for all parameters</i>
-------------------	--

---

## 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", "zip"),
  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",
  ttlsz = 1.2,
  bssz = 11,
  runchk = TRUE,
```

```
warn = TRUE
)
```

### Arguments

res	character string of path to the results file or data.frame for results returned by <a href="#">readMWRresults</a>
acc	character string of path to the data quality objectives file for accuracy or data.frame returned by <a href="#">readMWRacc</a>
fset	optional list of inputs with elements named res, acc, frecom, sit, or wxq 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, as separate png files, or as a zipped file of 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
repe1	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
bssize	numeric for overall plot text scaling, passed to <a href="#">theme_minimal</a>
runchk	logical to run data checks with <a href="#">checkMWRresults</a> or <a href="#">checkMWRacc</a> , 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 = 'MassWaterR')

# results data
resdat <- readMWRresults(respth)

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

# 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())

# create zipped png output
anlzMWRoutlierall(resdat, accdat, group = 'month', format = 'zip', 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,
bssize = 11,
runchk = TRUE,
warn = TRUE
)

```

### Arguments

res	character string of path to the results file or data.frame for results returned by <a href="#">readMWRresults</a>
param	character string of the parameter to plot, must conform to entries in the "Simple Parameter" column of <a href="#">paramsMWR</a>
acc	character string of path to the data quality objectives file for accuracy or data.frame returned by <a href="#">readMWRacc</a>
sit	optional character string of path to the site metadata file or data.frame of site metadata returned by <a href="#">readMWRsites</a> , required if locgroup is not NULL
fset	optional list of inputs with elements named res, acc, frecom, sit, or wxq 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
bssize	numeric for overall plot text scaling, passed to <a href="#">theme_minimal</a>
runchk	logical to run data checks with <a href="#">checkMWRresults</a> or <a href="#">checkMWRacc</a> , 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 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.



**Value**

A `ggplot` object that can be further modified.

**Examples**

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

# results data
resdat <- readMWRresults(respth)

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

# accuracy data
accdat <- readMWRacc(accpth)

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

# 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')
```

**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,
  bssize = 11,
  runchk = TRUE,
  warn = TRUE
)
```

**Arguments**

res	character string of path to the results file or data.frame for results returned by <a href="#">readMWRresults</a>
param	character string of the parameter to plot, must conform to entries in the "Simple Parameter" column of <a href="#">paramsMWR</a>
acc	character string of path to the data quality objectives file for accuracy or data.frame returned by <a href="#">readMWRacc</a>
sit	optional character string of path to the site metadata file or data.frame of site metadata returned by <a href="#">readMWRsites</a> , 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
bssize	numeric for overall plot text scaling, passed to <code>theme_minimal</code>
runchk	logical to run data checks with <code>checkMWRresults</code> or <code>checkMWRacc</code> , 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 = 'MassWaterR')

# results data
resdat <- readMWRresults(respth)

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

# accuracy data
accdat <- readMWRacc(accpth)

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

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

```

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

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

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

# grouping by result attribute
anzMWRsite(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
anzMWRsite(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 = 'MassWaterR')

# 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)
```

---

checkMWRcens

*Check censored data*

---

### Description

Check censored data

### Usage

```
checkMWRcens(censdat, warn = TRUE)
```

### Arguments

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

## Details

This function is used internally within `readMWRcens` to run several checks on the input data for completeness and conformance.

The following checks are made:

- Column name spelling: Should be the following: Parameter, Missed and Censored Records
- Columns present: All columns from the previous check should be present
- Non-numeric or empty entries in Missed and Censored Records: All values should be numbers
- Negative Missed and Censored Records: All values should be greater than or equal to zero
- Parameter: Should match parameter names in the Simple Parameter or WQX Parameter columns of the `paramsMWR` data (warning only)

## Value

`censdat` 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

```
censpth <- system.file('extdata/ExampleCensored.xlsx', package = 'MassWater')  
censdat <- suppressWarnings(readxl::read_excel(censpth, na = c('NA', 'na', ''), guess_max = Inf))  
checkMWRcens(censdat)
```

---

checkMWRfrecom

*Check data quality objective frequency and completeness data*

---

## Description

Check data quality objective frequency and completeness data

## Usage

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

## Arguments

<code>frecomdat</code>	input data frame
<code>warn</code>	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 returned prompting the user to make the required correction to the raw data before proceeding.

## Examples

```
library(dplyr)

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

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, Activity Start Date, Activity Start Time, Activity Depth/Height Measure, Activity Depth/Height Unit, Activity Relative Depth Name, Characteristic Name, Result Value, Result Unit, Quantitation Limit, QC Reference Value, Result Measure Qualifier, Result Attribute, Sample Collection 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
- Non-numeric Quantitation Limit: All values should be numbers, excluding missing values
- 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	<i>Check site metadata file</i>
---------------	---------------------------------

---

**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 = 'MassWaterR')

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 = 'MassWaterR')

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

formMWRacc(accdat)
```

---

formMWRcens	<i>Format censored data</i>
-------------	-----------------------------

---

## Description

Format censored data

## Usage

```
formMWRcens(censdat)
```

## Arguments

censdat            input data frame

## Details

This function is used internally within [readMWRcens](#) to format the input data for downstream analysis. The formatting includes:

- Convert Parameter: All parameters are converted to Simple Parameter in [paramsMWR](#) as needed,
- Convert Missed and Censored Records: All values are converted to numeric

## Value

A formatted data frame of the censored data

## Examples

```
library(dplyr)

censpth <- system.file('extdata/ExampleCensored.xlsx',
  package = 'MassWaterR')

censdat <- suppressMessages(readxl::read_excel(censpth,
  na = c('NA', 'na', ''))
))

formMWRcens(censdat)
```

---

`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	<i>Format water quality monitoring results</i>
----------------	--

---

## Description

Format water quality monitoring results

## Usage

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

## Arguments

resdat	input data frame for results
tzzone	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, Activity Start Time is converted 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	<i>Format WQX metadata input</i>
------------	----------------------------------

---

**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	<i>Master parameter list and units for Characteristic Name column in results data</i>
-----------	---

---

**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	<i>Run quality control accuracy checks for water quality monitoring results</i>
----------	---

---

**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 <code>data.frame</code> for results returned by <a href="#">readMWRresults</a>
acc	character string of path to the data quality objectives file for accuracy or <code>data.frame</code> returned by <a href="#">readMWRacc</a>
frecom	character string of path to the data quality objectives file for frequency and completeness or <code>data.frame</code> returned by <a href="#">readMWRfrecom</a>
fset	optional list of inputs with elements named <code>res</code> , <code>acc</code> , <code>frecom</code> , <code>sit</code> , or <code>wqx</code> overrides the other arguments
runchk	logical to run data checks with <a href="#">checkMWRresults</a> and <a href="#">checkMWRacc</a> , applies only if <code>res</code> or <code>acc</code> 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 `runkchk = 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	<i>Run quality control completeness checks for water quality monitoring results</i>
----------	---

---

## Description

Run quality control completeness checks for water quality monitoring results

## Usage

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

## Arguments

res	character string of path to the results file or data.frame for results returned by <a href="#">readMWRresults</a>
frecom	character string of path to the data quality objectives file for frequency and completeness or data.frame returned by <a href="#">readMWRfrecom</a>
cens	character string of path to the censored data file or data.frame returned by <a href="#">readMWRcens</a> , optional
fset	optional list of inputs with elements named res, acc, frecom, sit, or wxq overrides the other arguments
runchk	logical to run data checks with <a href="#">checkMWRresults</a> and <a href="#">checkMWRfrecom</a> , 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](#), [readMWRfrecom](#), and [readMWRcens](#) (optional). For the former, the full suite of data checks can be evaluated with `runchk = 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 CharacteristicName 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.

A similar warning is returned if there are parameters in the censored data, if provided, that are not in the results file. However, an error is returned if there are parameters in the data quality objectives frequency and completeness file that are not in the censored data file.

All warnings can be suppressed by setting `warn = FALSE`.

## 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 = 'MassWaterR')

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

# censored path
censpth <- system.file('extdata/ExampleCensored.xlsx',
  package = 'MassWaterR')

qcMWRcom(res = respth, frecom = frecompth, cens = censpth)

##
# using data frames

# results data
resdat <- readMWRresults(respth)

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

# censored data
censdat <- readMWRcens(censpth)

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

---

qcMWRfre	<i>Run quality control frequency checks for water quality monitoring results</i>
----------	--

---

### 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 <a href="#">readMWRresults</a>
acc	character string of path to the data quality objectives file for accuracy or data . frame returned by <a href="#">readMWRacc</a>
frecom	character string of path to the data quality objectives file for frequency and completeness or data . frame returned by <a href="#">readMWRfrecom</a>
fset	optional list of inputs with elements named res, acc, frecom, sit, or wxq overrides the other arguments
runchk	logical to run data checks with <a href="#">checkMWRresults</a> and <a href="#">checkMWRfrecom</a> , 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 `runchk = 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 = 'MassWaterR')

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

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

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,
  cens = NULL,
  fset = NULL,
  output_dir,
  output_file = NULL,
  savesheet = FALSE,
  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 <a href="#">readMWRresults</a>
acc	character string of path to the data quality objectives file for accuracy or data.frame returned by <a href="#">readMWRacc</a>
frecom	character string of path to the data quality objectives file for frequency and completeness or data.frame returned by <a href="#">readMWRfrecom</a>
cens	character string of path to the censored data file or data.frame returned by <a href="#">readMWRcens</a>
fset	optional list of inputs with elements named res, acc, frecom, and cens, 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
savesheet	logical indicating if a spreadsheet of the tables in the report is also saved (default FALSE)
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 <a href="#">checkMWRresults</a> , <a href="#">checkMWRacc</a> , <a href="#">checkMWRfrecom</a> , 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 MassWaterR 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, data quality objective, and censored 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). Set `savesheet = TRUE` to also save a spreadsheet of the tables in the report.

The function can be used with inputs as paths to the relevant files or as data frames returned by [readMWRresults](#), [readMWRacc](#), [readMWRfrecom](#), and [readMWRcens](#) (optional). For the former, the full suite of data checks can be evaluated with `runchk = 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 = 'MassWaterR')

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

# dqo completeness data path
```



```
frecompth <- system.file('extdata/ExampleDQ0FrequencyCompleteness.xlsx', package = 'MassWater')

# censored data path
censpth <- system.file('extdata/ExampleCensored.xlsx', package = 'MassWater')

# results data
resdat <- readMWRresults(respth)

# accuracy data
accdat <- readMWRacc(accpth)

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

# censored data
censdat <- readMWRcens(censpth)

# create report
qcMWRreview(res = resdat, acc = accdat, frecom = frecomdat, cens = censdat, 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 <a href="#">checkMWRacc</a>
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 = 'MassWaterR')  
  
accdat <- readMWRacc(accpth)  
head(accdat)
```

---

readMWRcens	<i>Read censored data from an external file</i>
-------------	---

---

## Description

Read censored data from an external file

## Usage

```
readMWRcens(censpth, runchk = TRUE, warn = TRUE)
```

## Arguments

censpth	character string of path to the censored file
runchk	logical to run data checks with <a href="#">checkMWRcens</a>
warn	logical to return warnings to the console (default)

## Details

Data are imported with [read\\_excel](#), checked with [checkMWRcens](#), and formatted with [formMWRcens](#). The input file includes rows for each parameter and two columns indicating the parameter name and number of missed or censored records for that parameter. The data are used to complete the number of missed and censored records column for the completeness table created with [tabMWRcom](#) and is an optional input. The parameters in this file must match those in the data quality objectives file for frequency and completeness.

## Value

A formatted censored data frame that can be used for downstream analysis

## Examples

```
censpth <- system.file('extdata/ExampleCensored.xlsx', package = 'MassWaterR')  
  
censdat <- readMWRcens(censpth)  
head(censdat)
```

---

readMWRfrecom	<i>Read data quality objectives for frequency and completeness from an external file</i>
---------------	--

---

### 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 <a href="#">checkMWRfrecom</a>
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 = 'MassWaterR')  
  
frecomdat <- readMWRfrecom(frecompth)  
head(frecomdat)
```

---

readMWRresults	<i>Read water quality monitoring results from an external file</i>
----------------	--

---

### Description

Read water quality monitoring results from an external file

### Usage

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

## Arguments

respth	character string of path to the results file
runchk	logical to run data checks with <a href="#">checkMWRresults</a>
warn	logical to return warnings to the console (default)
tzzone	character string for time zone, passed to <a href="#">formMWRresults</a>

## Details

Data 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	<i>Create summary spreadsheet of the water quality monitoring results</i>
--------------------	---

---

## 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#import-and-checks>).

Unique entries for the Result Value column will include NA entries if present, all other columns will not.

## 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	<i>Read site metadata from an external file</i>
--------------	---

---

## 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	<i>Read water quality exchange (wqx) metadata input from an external file</i>
------------	---

---

**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 = 'MassWaterR')

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 <a href="#">readMWRresults</a>
acc	character string of path to the data quality objectives file for accuracy or data.frame returned by <a href="#">readMWRacc</a>
frecom	character string of path to the data quality objectives file for frequency and completeness or data.frame returned by <a href="#">readMWRfrecom</a> , applies only if type = "summary" or type = "percent"
fset	optional list of inputs with elements named res, acc, frecom, sit, or wxq overrides the other arguments
runchk	logical to run data checks with <a href="#">checkMWRresults</a> and <a href="#">checkMWRacc</a> , 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 `runkchk = 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 = 'MassWaterR')

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

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

# 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,
  cens = 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 <a href="#">readMWRresults</a>
frecom	character string of path to the data quality objectives file for frequency and completeness or data.frame returned by <a href="#">readMWRfrecom</a>
cens	character string of path to the censored data file or data.frame returned by <a href="#">readMWRcens</a> , optional
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 <a href="#">checkMWRresults</a> and <a href="#">checkMWRfrecom</a> , 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](#), [readMWRfrecom](#), and [readMWRcens](#) (optional). For the former, the full suite of data checks can be evaluated with `runchk = 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. A similar warning is returned if there are parameters in the censored data, if provided, that are not in the results file and vice versa. These warnings 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 = 'MassWaterR')

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

# censored path
censpth <- system.file('extdata/ExampleCensored.xlsx', package = 'MassWaterR')

tabMWRcom(res = respth, frecom = frecompth, cens = censpth)

##
# using data frames

# results data
resdat <- readMWRresults(respth)

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

# censored data
censdat <- readMWRcens(censpth)

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

---

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 <a href="#">readMWRresults</a>
acc	character string of path to the data quality objectives file for accuracy or data . frame returned by <a href="#">readMWRacc</a>
frecom	character string of path to the data quality objectives file for frequency and completeness or data . frame returned by <a href="#">readMWRfrecom</a>
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 <a href="#">checkMWRresults</a> and <a href="#">checkMWRfrecom</a> , 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 `runchk = 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 = 'MassWaterR')

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

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

# 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,
  listout = FALSE,
  warn = TRUE,
  runchk = TRUE
)

```

**Arguments**

res	character string of path to the results file or data.frame for results returned by <a href="#">readMWRresults</a>
acc	character string of path to the data quality objectives file for accuracy or data.frame returned by <a href="#">readMWRacc</a>
sit	character string of path to the site metadata file or data.frame for site metadata returned by <a href="#">readMWRsites</a>
wqx	character string of path to the wqx metadata file or data.frame for wqx metadata returned by <a href="#">readMWRwqx</a>
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
listout	logical to return a list of the output for each sheet of the workbook (default is FALSE)
warn	logical to return warnings to the console (default)
runchk	logical to run data checks with <a href="#">checkMWRresults</a> , <a href="#">checkMWRacc</a> , <a href="#">checkMWRsites</a> , <a href="#">checkMWRwqx</a> , 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 `runkchk = 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 = 'MassWaterR')

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

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

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

# results data
resdat <- readMWRresults(respth)

# accuracy data
accdat <- readMWRacc(accpth)

# site data
sitdat <- readMWRsites(sitpth)

# wxq data
wqxdat <- readMWRwqx(wqxpth)

# create workbook
tabMWRwqx(res = resdat, acc = accdat, sit = sitdat, wxq = wxqdat, 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	<i>Filter results data by parameter, date range, site, result attributes, and/or location group</i>
---------------	---

---

**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 <a href="#">readMWRresults</a>
sitdat	site metadata file as returned by <a href="#">readMWRresults</a>
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
allocgroup	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 <a href="#">anzMWRdate</a>
allresultatt	logical indicating if results data are filtered by all result attributes if resultatt = NULL, used only in <a href="#">anzMWRsite</a>

### 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 = 'MassWaterR')

# results data
resdat <- readMWRresults(respth)

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

# 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 = 'MassWaterR')

# 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 = 'MassWaterR')

# results data
resdat <- readMWRresults(respth)

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

# 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	<i>Utility function to import data as paths or data frames</i>
--------------	--

---

**Description**

Utility function to import data as paths or data frames

**Usage**

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

**Arguments**

res	character string of path to the results file or data.frame for results returned by <a href="#">readMWRresults</a>
acc	character string of path to the data quality objectives file for accuracy or data.frame returned by <a href="#">readMWRacc</a>
frecom	character string of path to the data quality objectives file for frequency and completeness or data.frame returned by <a href="#">readMWRfrecom</a>

sit	character string of path to the site metadata file or data.frame for site metadata returned by <a href="#">readMWRsites</a>
wqx	character string of path to the wqx metadata file or data.frame for wqx metadata returned by <a href="#">readMWRwqx</a>
cens	character string of path to the censored data file or data.frame for censored data returned by <a href="#">readMWRcens</a>
fset	optional list of inputs with elements named res, acc, frecom, sit, wqx, or cens, overrides the other arguments, see details
runchk	logical to run data checks with <a href="#">checkMWRresults</a> , <a href="#">checkMWRacc</a> , <a href="#">checkMWRfrecom</a> , <a href="#">checkMWRsites</a> , <a href="#">checkMWRwqx</a> , or <a href="#">checkMWRcens</a> , applies only if res, acc, frecom, sit, wqx, or cens 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](#), [readMWRwqx](#), or [readMWRcens](#). For the former, the full suite of data checks can be evaluated with `runchk = T` (default) or suppressed with `runchk = F`.

The `fset` argument can be used in place of the preceding arguments. The argument accepts a list with named elements as `res`, `acc`, `frecom`, `sit`, `wqx`, or `cens`, 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 a 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 six element list with the imported results, data quality objective files, site metadata, wqx metadata, and censored data named `"resdat"`, `"accdat"`, `"frecomdat"`, `"sitdat"`, `"wxdat"`, and `"censdat"` respectively.

### Examples

```
##
# using file paths

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

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

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

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

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

# censored path
censpth <- system.file('extdata/ExampleCensored.xlsx', package = 'MassWaterR')

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

##
# 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)

# censored data
censdat <- readMWRcens(censpth)

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

##
# using fset as list input

# input with paths to files
fset <- list(
```

```
    res = respth,  
    acc = accpth,  
    frecom = frecompth,  
    sit = sitpth,  
    wqx = wqxpth,  
    cens = censpth  
  )  
  utilMWRinput(fset = fset)
```

---

utilMWRinputcheck	<i>Check if required inputs are present for a function</i>
-------------------	--

---

### Description

Check if required inputs are present for a function

### Usage

```
utilMWRinputcheck(inputs, nocheck = NULL)
```

### Arguments

inputs	list of arguments passed from the parent function
nocheck	optional character vector of inputs not to check, allows for optional inputs

### 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')  
inputchk$cens <- system.file('extdata/ExampleCensored.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 <a href="#">readMWRresults</a>
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 <a href="#">readMWRacc</a>
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 = 'MassWaterR')

# results data
resdat <- readMWRresults(respth)

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



```
# 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	<i>Identify outliers in a numeric vector</i>
----------------	--

---

### Description

Identify outliers in a numeric vector

### Usage

```
utilMWRoutlier(x, logsc1)
```

### Arguments

x	numeric vector of any length
logsc1	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, logsc1 = FALSE)
```

---

`utilMWRsheet`*Format a list of QC tables for spreadsheet export*

---

### Description

Format a list of QC tables for spreadsheet export

### Usage

```
utilMWRsheet(datin, rawdata = TRUE)
```

### Arguments

<code>datin</code>	list of input QC tables
<code>rawdata</code>	logical to include quality control accuracy summaries for raw data, e.g., field blanks, etc.

### Details

The function is used internally with [qcMWRreview](#) to format data quality objective and quality control tables for export into an Excel spreadsheet. These changes are specific to the spreadsheet format and may not reflect the formatting in the Word document produced by [qcMWRreview](#).

The `datin` list is expected to contain the following elements:

- `fecomdat` Data Quality Objectives for frequency and completeness data frame as returned by [readMWRfecom](#)
- `accdat` Data Quality Objectives for accuracy data frame as returned by [readMWRacc](#)
- `tabfreper` Frequency checks percent table, created with [tabMWRfre](#)
- `tabfresum` Frequency checks summary table, created with [tabMWRfre](#)
- `tabaccper` Accuracy checks percent table, created with [tabMWRacc](#)
- `tabaccsum` Accuracy checks summary table, created with [tabMWRacc](#)
- `tabcom` Completeness table, created with [tabMWRcom](#)
- `indflddup` Individual accuracy checks table for field duplicates, created with [tabMWRacc](#), can be NULL
- `indlabdup` Individual accuracy checks table for lab duplicates, created with [tabMWRacc](#), can be NULL
- `indfldblk` Individual accuracy checks table for field blanks, created with [tabMWRacc](#), can be NULL
- `indlabblk` Individual accuracy checks table for lab blanks, created with [tabMWRacc](#), can be NULL
- `indlabins` Individual accuracy checks table for lab spikes and instrument checks, created with [tabMWRacc](#), can be NULL

**Value**

A list similar to the input with formatting applied

**Examples**

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

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

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

# censored data path
censpth <- system.file('extdata/ExampleCensored.xlsx', package = 'MassWaterR')

# results data
resdat <- readMWRresults(respth)

# accuracy data
accdat <- readMWRacc(accpth)

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

# censored data
censdat <- readMWRcens(censpth)

# frequency table percent
tabfreper <- tabMWRfre(res = resdat, acc = accdat, frecom = frecomdat, type = 'percent',
  warn = FALSE)

# frequency summary table
tabfresum <- tabMWRfre(res = resdat, acc = accdat, frecom = frecomdat, type = 'summary',
  warn = FALSE)

# accuracy table percent
tabaccper <- tabMWRacc(res = resdat, acc = accdat, frecom = frecomdat, type = 'percent',
  warn = FALSE)

# accuracy table summary
tabaccsum <- tabMWRacc(res = resdat, acc = accdat, frecom = frecomdat, type = 'summary',
  warn = FALSE)

# completeness table
tabcom <- tabMWRcom(res = resdat, frecom = frecomdat, cens = censdat, warn = FALSE,
  parameterwd = 1.15, noteswd = 2)

# individual accuracy checks for raw data
indflddup <- tabMWRacc(res = resdat, acc = accdat, frecom = frecomdat, type = 'individual',
  accchk = 'Field Duplicates', warn = FALSE, caption = FALSE)
```

```

indlabdup <- tabMWRacc(res = resdat, acc = accdat, frecom = frecomdat, type = 'individual',
  accchk = 'Lab Duplicates', warn = FALSE, caption = FALSE)
indfldblk <- tabMWRacc(res = resdat, acc = accdat, frecom = frecomdat, type = 'individual',
  accchk = 'Field Blanks', warn = FALSE, caption = FALSE)
indlabblk <- tabMWRacc(res = resdat, acc = accdat, frecom = frecomdat, type = 'individual',
  accchk = 'Lab Blanks', warn = FALSE, caption = FALSE)
indlabins <- tabMWRacc(res = resdat, acc = accdat, frecom = frecomdat, type = 'individual',
  accchk = 'Lab Spikes / Instrument Checks', warn = FALSE, caption = FALSE)

# input
datin <- list(
  frecomdat = frecomdat,
  accdat = accdat,
  tabfreper = tabfreper,
  tabfresum = tabfresum,
  tabaccper = tabaccper,
  tabaccsum = tabaccsum,
  tabcom = tabcom,
  indflldup = indflldup,
  indlabdup = indlabdup,
  indfldblk = indfldblk,
  indlabblk = indlabblk,
  indlabins = indlabins
)

utilMWRsheet(datin)

```

---

 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 <a href="#">readMWRacc</a>
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 geometric 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 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

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

## Examples

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

# 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	<i>Summarize a results data frame by a grouping variable</i>
----------------	--

---

## 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 <a href="#">readMWRacc</a>

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 = 'MassWaterR')

# results data
resdat <- readMWRresults(respth)

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

# 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	<i>Get threshold lines from thresholdMWR</i>
---------------	--

---

## Description

Get threshold lines from thresholdMWR

## Usage

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

## Arguments

resdat	results data as returned by <a href="#">readMWRresults</a>
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 threshlab.

## Examples

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

# 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,  
  acmdat = NULL,  
  sumfun = NULL,  
  site = NULL,  
  dtrng = NULL,  
  resultatt = NULL,  
  locgroup = NULL  
)
```

### Arguments

param	character string of the parameter to plot
acmdat	optional data . frame for data quality objectives file for accuracy as returned by <a href="#">readMWRacc</a>
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 = 'MassWaterR')
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	<i>Check if incomplete range in Value Range column</i>
-------------------	--

---

## 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 = 'MassWaterR')

# 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	<i>Get logical value for y axis scaling</i>
---------------	---

---

**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 <a href="#">readMWRacc</a>
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 = 'MassWaterR')

# 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

- \* **datasets**
  - paramsMWR, 32
  - thresholdMWR, 55
- anlzMWRdate, 3, 57
- anlzMWRmap, 6
- anlzMWRoutlier, 9, 13
- anlzMWRoutlierall, 12
- anlzMWRseason, 14
- anlzMWRsite, 17, 57
  
- checkMWRacc, 4, 8, 11, 13, 16, 19, 21, 33, 40, 41, 47, 54, 61
- checkMWRcens, 22, 42, 61
- checkMWRfrecom, 23, 35, 37, 40, 43, 50, 52, 61
- checkMWRresults, 4, 8, 11, 13, 16, 19, 24, 33, 35, 37, 40, 44, 47, 50, 52, 54, 61
- checkMWRsites, 8, 26, 46, 54, 61
- checkMWRwqx, 27, 46, 54, 61
  
- flextable, 49, 50, 53
- formMWRacc, 28
- formMWRcens, 29, 42
- formMWRfrecom, 30
- formMWRresults, 31, 44
- formMWRwqx, 32
  
- ggplot, 5, 9, 11, 17, 20
  
- paramsMWR, 4, 7, 10, 15, 18, 22–25, 27–32, 32
  
- qcMWRacc, 28, 33, 48
- qcMWRcom, 35, 50, 52
- qcMWRfre, 37
- qcMWRreview, 39, 66
  
- read\_excel, 41–44, 46
- readMWRacc, 4, 7, 10, 13, 15, 18, 21, 28, 33, 34, 37, 39, 40, 41, 47, 48, 52, 54, 55, 58, 60, 61, 64, 66, 68, 69, 72–74
- readMWRcens, 23, 29, 35, 39, 40, 42, 50, 61
- readMWRfrecom, 24, 30, 33, 35, 37, 39, 40, 43, 47, 50, 52, 60, 61, 66
- readMWRresults, 3, 7, 10, 13, 15, 18, 25, 31, 33–35, 37, 39, 40, 43, 44, 47, 48, 50, 52, 54–56, 58, 60, 61, 64, 71
- readMWRresultsview, 44, 44
- readMWRsites, 4, 7, 15, 18, 26, 45, 54, 55, 61
- readMWRwqx, 27, 32, 46, 54, 55, 61
  
- scale\_color\_brewer, 5
- scale\_fill\_distiller, 8
- st\_as\_sf, 8
  
- tabMWRacc, 40, 47, 66
- tabMWRcom, 40, 42, 49, 66
- tabMWRfre, 40, 51, 66
- tabMWRwqx, 53
- theme\_gray, 8
- theme\_minimal, 4, 11, 13, 16, 19
- thresholdMWR, 5, 16, 19, 55
  
- utilMWRfilter, 56
- utilMWRfiltersurface, 57
- utilMWRfre, 58
- utilMWRhttpgrace, 59
- utilMWRinput, 34, 35, 37, 40, 48, 50, 52, 55, 60
- utilMWRinputcheck, 63
- utilMWRlimits, 64
- utilMWRoutlier, 65
- utilMWRsheet, 66
- utilMWRsumfun, 68
- utilMWRsummary, 69
- utilMWRthresh, 71
- utilMWRtitle, 72
- utilMWRvaluerange, 73
- utilMWRyscale, 74