

# Package ‘EBASE’

January 20, 2025

**Title** Estuarine Bayesian Single-Station Estimation Method for Ecosystem Metabolism

**Version** 1.1.0

**Date** 2024-09-25

**Description** Estimate ecosystem metabolism in a Bayesian framework for individual water quality monitoring stations with continuous dissolved oxygen time series. A mass balance equation is used that provides estimates of parameters for gross primary production, respiration, and gas exchange. Methods adapted from Grace et al. (2015) <[doi:10.1002/lom3.10011](https://doi.org/10.1002/lom3.10011)> and Wanninkhof (2014) <[doi:10.4319/lom.2014.12.351](https://doi.org/10.4319/lom.2014.12.351)>. Details in Beck et al. (2024) <[doi:10.1002/lom3.10620](https://doi.org/10.1002/lom3.10620)>.

**Depends** R (>= 3.5)

**Imports** doParallel, dplyr, foreach, ggplot2 (>= 3.4.0), lubridate, R2jags (>= 0.6.1), rjags (>= 4.10), tidyr, truncnorm, zoo

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

**License** CC0

**SystemRequirements** JAGS 4.x.y (<https://mcmc-jags.sourceforge.net>)

**Encoding** UTF-8

**RoxygenNote** 7.2.3

**URL** <https://fawda123.github.io/EBASE/>,  
<https://github.com/fawda123/EBASE/>

**BugReports** <https://github.com/fawda123/EBASE/issues>

**LazyData** true

**LazyDataCompression** xz

**Config/testthat/edition** 3

**VignetteBuilder** knitr

**NeedsCompilation** no

**Author** Marcus Beck [aut, cre] (<<https://orcid.org/0000-0002-4996-0059>>),  
Maria Herrmann [aut],  
Jill Arriola [aut] (<<https://orcid.org/0000-0003-2173-8349>>),  
Raymond Najjar [aut] (<<https://orcid.org/0000-0002-2960-5965>>)

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

**Repository** CRAN

**Date/Publication** 2024-09-25 17:30:02 UTC

## Contents

|                         |    |
|-------------------------|----|
| credible_plot . . . . . | 2  |
| credible_prep . . . . . | 3  |
| ebase . . . . .         | 4  |
| ebase_eqboxy . . . . .  | 8  |
| ebase_form . . . . .    | 9  |
| ebase_plot . . . . .    | 10 |
| ebase_prep . . . . .    | 11 |
| ebase_rho . . . . .     | 12 |
| ebase_schmidt . . . . . | 13 |
| ebase_years . . . . .   | 13 |
| exdat . . . . .         | 16 |
| exres . . . . .         | 16 |
| fit_plot . . . . .      | 18 |
| interp_plot . . . . .   | 19 |
| metab_update . . . . .  | 20 |
| prior_plot . . . . .    | 21 |

**Index** **23**

---

|               |  |
|---------------|--|
| credible_plot | <i>Plot credible intervals for a, R, and b</i> |
|---------------|--|

---

### Description

Plot credible intervals for a, R, and b

### Usage

```
credible_plot(res, params = c("a", "R", "b"))
```

### Arguments

|        |   |
|--------|---|
| res    | output data frame from <a href="#">ebase</a>  |
| params | character vector indicating which parameters to plot, one to any of a, R, or b, (default all) |

### Details

This function plots 95% credible intervals (2.5th to 97.5th percentiles, approximate posterior distributions) for a, R and/or b using the output from [ebase](#). Results in the plot are grouped by the ndays argument that was used in [ebase](#).

**Value**

A `ggplot` object

**Examples**

```
# plot credible intervals
credible_plot(exres)
```

---

|               |   |
|---------------|---|
| credible_prep | <i>Get credible intervals for a, R, b</i> |
|---------------|---|

---

**Description**

Get credible intervals for a, R, b

**Usage**

```
credible_prep(res, params = c("a", "R", "b"), labels = FALSE)
```

**Arguments**

|        |  |
|--------|--|
| res    | output data frame from <a href="#">ebase</a>   |
| params | character vector indicating which parameters to plot, one to any of a, R, or b (default all)                 |
| labels | logical indicating if parameter labels are output as an expression for parsing in plot facets, default FALSE |

**Details**

This function gets 95% credible intervals (2.5th to 97.5th percentiles, approximate posterior distributions) for a, R, and/or b using the output from [ebase](#). The function is used in [credible\\_plot](#), but is provided as a separate function for convenience.

**Value**

A data frame

**Examples**

```
# get credible intervals
credible_prep(exres)
```

---

|       |   |
|-------|---|
| ebase | <i>Estuarine Bayesian Single-station Estimation method for ecosystem metabolism</i> |
|-------|---|

---

## Description

Estuarine Bayesian Single-station Estimation method for ecosystem metabolism

## Usage

```
ebase(
  dat,
  Z,
  interval,
  ndays = 1,
  aprior = c(4, 2),
  rprior = c(300, 150),
  bprior = c(0.251, 0.125),
  bmax = 0.502,
  nogas = FALSE,
  doave = TRUE,
  maxinterp = 43200/interval,
  n.iter = 10000,
  update.chains = TRUE,
  n.burnin = n.iter * 0.5,
  n.chains = 3,
  n.thin = 10,
  model_file = NULL
)
```

## Arguments

|          |  |
|----------|--|
| dat      | input data frame   |
| Z        | numeric as single value for water column depth (m) or vector equal in length to number of rows in dat                                |
| interval | timestep interval in seconds   |
| ndays    | numeric for number of days in dat for optimizing the metabolic equation, see details   |
| aprior   | numeric vector of length two indicating the mean and standard deviation for the prior distribution of the $a$ parameter, see details |
| rprior   | numeric vector of length two indicating the mean and standard deviation for the prior distribution of the $R$ parameter, see details |
| bprior   | numeric vector of length two indicating the mean and standard deviation for the prior distribution of the $b$ parameter, see details |

|               |   |
|---------------|---|
| bmax          | numeric value for the upper limit on the prior distribution for bprior, set as twice the default value of the mean  |
| nogas         | logical indicating if gas exchange is not included in the metabolic model, see details  |
| doave         | logical indicating if the average dissolved oxygen concentration is used as the starting value for the estimation (default), otherwise the first observation will be used if FALSE, see details |
| maxinterp     | numeric value for minimum number of continuous observations that must not be interpolated within a group defined by ndays to assign as NA in output, see details                                |
| n.iter        | number of MCMC iterations, passed to <a href="#">jags</a>   |
| update.chains | logical to run <a href="#">metab_update</a> if chains do not converge   |
| n.burnin      | number of MCMC chains to delete, passed to <a href="#">jags</a>   |
| n.chains      | number of MCMC chains to run, passed to <a href="#">jags</a>  |
| n.thin        | number of nth iterations to save for each chain, passed to <a href="#">jags</a>   |
| model_file    | NULL to use the model file included with the package or a path to a model text file can be used   |

### Details

Required input data are time series for dissolved oxygen (mg L<sup>-1</sup>), water temperature (C), salinity (psu), total PAR (W m<sup>-2</sup>), and wind speed (m s<sup>-1</sup>). See the [exdat](#) example data file for a representation of the required data. Data are typically from continuously monitored water quality and weather parameters are hourly or sub-hourly time steps. Oxygen concentrations are converted to mmol/m<sup>3</sup> prior to metabolic estimation. Water column depth is also required. This can be supplied as a single value or a vector of length equal to the number of rows in dat.

The metabolic estimates are based on a mass balance equation in Grace et al. 2015 with the gas exchange estimate from Wanninkhof 2004. It is similar to that provided by the BASEmetab R package at <https://github.com/dgiling/BASEmetab>, with modifications to estimate different parameters. The equation optimized in the JAGS model is:

$$Z \frac{dC_d}{dt} = aPAR - R + bU_{10}^2 \left( \frac{Sc}{600} \right)^{-0.5} (C_{Sat} - C_d)$$

More simply:

$$Z \frac{dC_d}{dt} = P - R + D$$

Net ecosystem metabolism (NEM) is then estimated as:

$$NEM = P - R$$

Gross production is provided by  $aPAR$ , respiration is provided by  $R$ , and gas exchange is provided by the remainder. The likelihood of the parameters  $a$ ,  $R$ , and  $b$  given the observed data are estimated from the JAGS model using prior distributions shown in the model file. At each time step, the

change in oxygen concentration between time steps is calculated from the equation using model inputs and parameter guesses, and then a finite difference approximation is used to estimate modeled oxygen concentration. The first modeled value starts at the mean oxygen concentration for all measurements in the optimization period. The estimated concentration at each time step is also returned for comparison to observed as one measure of model performance.

The prior distributions for the  $a$ ,  $R$ , and  $b$  parameters are defined in the model file included with the package as normal distributions with mean and standard deviations provided by the `aprior`, `rprior`, and `bprior` arguments. The default values were chosen based on approximate values from national syntheses of metabolic estimates. An additional constraint is that all the prior distributions are truncated to be positive values as required by the core metabolism equation above. The upper limit for  $b$  is set as two times 0.251, as given in eqn. 4 in Wanninkhof 2014. Units for each parameter are  $(\text{mmol m}^{-2} \text{ d}^{-1})/(\text{W m}^{-2})$  for  $a$ ,  $\text{mmol m}^{-2} \text{ d}^{-1}$  for  $R$ , and  $(\text{cm hr}^{-1})/(\text{m}^2 \text{ s}^{-2})$  for  $b$ .

The `ndays` argument defines the model optimization period as the number of days that are used for optimizing the above mass balance equation. By default, this is done each day, i.e., `ndays=1` such that a loop is used that applies the model equation to observations within each day, evaluated iteratively from the first observation in a day to the last. Individual parameter estimates for  $a$ ,  $R$ , and  $b$  are then returned for each day. However, more days can be used to estimate the unknown parameters, such that the loop can be evaluated for every `ndays` specified by the argument. The `ndays` argument will separate the input data into groups of consecutive days, where each group has a total number of days equal to `ndays`. The final block may not include the complete number of days specified by `ndays` if the number of unique dates in the input data includes a remainder when divided by `ndays`, e.g., if seven days are in the input data and `ndays = 5`, there will be two groups where the first has five days and the second has two days. The output data from `ebase` includes a column that specifies the grouping that was used based on `ndays`.

Missing values in the input data are also interpolated prior to estimating metabolism. It is the responsibility of the user to verify that these interpolated values are not wildly inaccurate. Missing values are linearly interpolated between non-missing values at the time step specified by the value in `interval`. This works well for small gaps, but can easily create inaccurate values at gaps larger than a few hours. The `interp_plot` function can be used to visually assess the interpolated values. Records at the start or end of the input time series that do not include a full day are also removed. A warning is returned to the console if gaps are found or dangling records are found.

The `maxinterp` argument specifies a minimum number of observations that must not be interpolated within groups defined by `ndays` that are assigned NA in the output (except `Date` and `DateTimeStamp`). Groups with continuous rows of interpolated values with length longer than this argument are assigned NA. The default value is half a day, i.e., 43200 seconds divided by the value in `interval`.

The `doave` argument can be used to define which dissolved oxygen value is used as the starting point in the Bayesian estimation for the optimization period. The default setting (`doave = TRUE`) will use the average of all the dissolved oxygen values in the optimization period defined by `ndays`. For example, the average of all dissolved oxygen values in each 24 hour period will be used if `doave = TRUE` and `ndays = 1`. The first dissolved oxygen observation of the time series in the optimization period will be used as the starting point if `doave = F`. In this case, the simulated dissolved oxygen time series will always start at the first observed dissolved oxygen value for each optimization period.

## Value

A data frame with metabolic estimates for areal gross production (P, O<sub>2</sub> mmol m<sup>-2</sup> d<sup>-1</sup>), respiration (R, O<sub>2</sub> mmol m<sup>-2</sup> d<sup>-1</sup>), and gas exchange (D, O<sub>2</sub> mmol m<sup>-2</sup> d<sup>-1</sup>, positive values as ingassing, negative values as outgassing). NEM (net ecosystem metabolism, O<sub>2</sub> mmol m<sup>-2</sup> d<sup>-1</sup>) is also returned as P - R. Additional parameters estimated by the model that are returned include a and b. The a parameter is a constant that represents the primary production per quantum of light with units of (mmol m<sup>-2</sup> d<sup>-1</sup>)/(W m<sup>-2</sup>) and is used to estimate gross production (Grace et al., 2015). The b parameter is a constant used to estimate gas exchange in (cm hr<sup>-1</sup>)/(m<sup>2</sup> s<sup>-2</sup>) (provided as 0.251 in eqn. 4 in Wanninkhof 2014). Observed dissolved oxygen (DO\_obs, mmol m<sup>-3</sup>), modeled dissolved oxygen (DO\_mod, mmol m<sup>-3</sup>), and delta dissolved oxygen of the modeled results (dDO, mmol m<sup>-3</sup> d<sup>-1</sup>) are also returned. Note that delta dissolved oxygen is a daily rate.

95% credible intervals for a, b, and R are also returned in the corresponding columns a\_lo, a\_hi, b\_lo, b\_hi, R\_lo, and R\_hi, for the 2.5th and 97.5th percentile estimates for each parameter, respectively. These values indicate the interval within which there is a 95% probability that the true parameter is in this range. Note that all values for these parameters are repeated across rows, although only one estimate for each is returned based on the number of days defined by ndays.

Model fit can also be assessed using the converge and rsq columns. The values in these columns apply to each group in the grp column as specified with the ndays argument. The converge column indicates "Check convergence" or "Fine" if the JAGS estimate converged at that iteration (repeated across rows for the group). The n.chains argument can be increased if convergence is not achieved. Similarly, the rsq column shows the r-squared values of the linear fit between the modeled and observed dissolved oxygen (repeated across rows for the group). These values can also be viewed with `fit_plot`.

The nogas argument can be set to TRUE to exclude gas exchange from the metabolic estimates. This will force the prior distribution for b as mean 0 and standard deviation approximately 0.

## References

- Grace, M.R., Giling, D.P., Hladyz, S., Caron, V., Thompson, R.M., Nally, R.M., 2015. Fast processing of diel oxygen curves: Estimating stream metabolism with BASE (BAYesian Single-station Estimation). *Limnology and Oceanography: Methods* 13, e10011. <https://doi.org/10.1002/lom3.10011>
- Wanninkhof, R., 2014. Relationship between wind speed and gas exchange over the ocean revisited. *Limnology and Oceanography: Methods* 12, 351–362. <https://doi.org/10.4319/lom.2014.12.351>

## Examples

```
# get one day of data
dat <- exdat[as.Date(exdat$DateTimeStamp, tz = 'America/Jamaica') == as.Date('2012-06-01'), ]

# run ebase, use more chains and iterations for a better fit, update.chains as T
ebase(dat, interval = 900, Z = 1.85, n.chains = 2, n.iter = 50,
      update.chains = FALSE)
```

---

|              |                          |
|--------------|--------------------------|
| ebase_eqboxy | <i>Oxygen saturation</i> |
|--------------|--------------------------|

---

**Description**

Oxygen saturation

**Usage**

```
ebase_eqboxy(temp, salt)
```

**Arguments**

|      |                             |
|------|-----------------------------|
| temp | numeric for temperature (C) |
| salt | numeric for salinity (PSU)  |

**Details**

Function to calculate equilibrium OXYGEN concentration in seawater, from water temperature (C) and salinity (PSU)

**Value**

oxysat (mmol/m<sup>3</sup>)

**References**

Garcia, H., Gordon, L.I., 1992. Oxygen solubility in seawater: Better fitting equations. *Limnology and Oceanography* 37, 1307-1312. <https://doi.org/10.4319/lo.1992.37.6.1307>

**Examples**

```
temp <- c(10, 20, 30)
salt <- c(30, 35, 40)
ebase_eqboxy(temp = temp, salt = salt)
```



---

|            |                            |
|------------|----------------------------|
| ebase_form | <i>Format ebase output</i> |
|------------|----------------------------|

---

## Description

Format ebase output

## Usage

```
ebase_form(out, dat, interval, maxinterp = 43200/interval)
```

## Arguments

|           |   |
|-----------|---|
| out       | data.frame for model output   |
| dat       | data.frame as returned by <a href="#">ebase_prep</a>  |
| interval  | timestep interval in seconds  |
| maxinterp | numeric value for minimum number of continuous observations that must not be interpolated within a group defined by ndays to assign as NA in output |

## Details

This function is used internally with [ebase](#) and should not be called by itself.

## Value

Formatted output for [ebase](#) with interpolated rows as NA (except Date and DateTimeStamp as defined by maxinterp)

## Examples

```
library(dplyr)

# get four days of data
dat <- exdat %>%
  filter(lubridate::month(DateTimeStamp) == 6) %>%
  filter(lubridate::day(DateTimeStamp) %in% 1:4)
dat <- ebase_prep(dat, Z = 1.85, interval = 900, ndays = 1)

ebase_form(exres, dat, interval = 900, maxinterp = 48)
```

---

`ebase_plot`*Plot results from EBASE*

---

**Description**

Plot results from EBASE

**Usage**

```
ebase_plot(res, asnem = FALSE, instantaneous = TRUE)
```

**Arguments**

|                            |  |
|----------------------------|--|
| <code>res</code>           | output data frame from <a href="#">ebase</a>                                   |
| <code>asnem</code>         | logical indicating if NEM is plotted with P and negative R, see details        |
| <code>instantaneous</code> | logical indicating if results are instantaneous (default) or averaged to daily |

**Details**

The plot shows P, R, and D over time as positive values if `asnem = F` (default). Positive values for D are ingassing, negative outgassing.

If `asnem = T`, NEM is plotted as a separate line as the difference between P and R. R is shown as negative values to indicate the negative influence on NEM. D is also excluded.

If `instantaneous = F`, the plot shows daily-averaged values. The y-axis units are the same for daily or instantaneous values.

**Value**

A [ggplot](#)

**Examples**

```
# plot instantaneous
ebase_plot(exres)

# plot NEM, negative R, exclude D
ebase_plot(exres, asnem = TRUE)

# plot daily-averaged
ebase_plot(exres, instantaneous = FALSE)
```

*ebase\_prep*                      *Prepare data for ebase*

**Description**

Prepare data for ebase

**Usage**

`ebase_prep(dat, Z, interval, ndays = 1)`

**Arguments**

|                       |  |
|-----------------------|--|
| <code>dat</code>      | input data frame   |
| <code>Z</code>        | numeric as single value for water column depth (m) or vector equal in length to number of rows in <code>dat</code> |
| <code>interval</code> | timestep interval in seconds   |
| <code>ndays</code>    | numeric for number of days in <code>dat</code> for optimizing the metabolic equation, see details                  |

**Details**

Checks if all columns are present by matching those in `exdat`, checks if `DateTimeStamp` is in ascending order, converts dissolved oxygen from mg/L to mmol/m<sup>3</sup>, calculates the Schmidt number (unitless) from water temp (C) and salinity (psu), and calculates dissolved oxygen equilibrium concentration (mmol/m<sup>3</sup>) from salinity and temperature

The `ndays` argument defines the number of days that are used for optimizing the above mass balance equation. By default, this is done each day, i.e., `ndays= 1` such that a loop is used that applies the model equation to observations within each day, evaluated iteratively from the first observation in a day to the last. Individual parameter estimates for *a*, *R*, and *b* are then returned for each day. However, more days can be used to estimate the unknown parameters, such that the loop can be evaluated for every `ndays` specified by the argument. The `ndays` argument will separate the input data into groups of consecutive days, where each group has a total number of days equal to `ndays`. The final block may not include the complete number of days specified by `ndays` if the number of unique dates in the input data includes a remainder when divided by `ndays`, e.g., if seven days are in the input data and `ndays = 5`, there will be two groups where the first has five days and the second has two days. The output data from `ebase` includes a column that specifies the grouping that was used based on `ndays`.

Missing values are interpolated at the interval specified by the `interval` argument for conformance with the core model equation. Records at the start or end of the input time series that do not include a full day are also removed. A warning is returned to the console if gaps are found or dangling records are found.

**Value**

A data frame with additional columns required for [ebase](#). Dissolved oxygen as a volumetric concentration in dat as mg/L is returned in areal units as mmol/m<sup>2</sup>. If multiple time steps are identified, the number of rows in data frame is expanded based on the time step define by `interval`. Numeric values in the expanded rows will be interpolated if `interp = TRUE`, otherwise they will remain as NA values.

**Examples**

```
dat <- ebase_prep(exdat, Z = 1.85, interval = 900)
head(dat)
```

---

|           |                                     |
|-----------|-------------------------------------|
| ebase_rho | <i>Seawater density calculation</i> |
|-----------|-------------------------------------|

---

**Description**

Seawater density calculation

**Usage**

```
ebase_rho(temp, salt, P)
```

**Arguments**

|      |   |
|------|---|
| temp | numeric for temperature (C)                   |
| salt | numeric for salinity (PSU)                    |
| P    | numeric for pressure above atmospheric (dbar) |

**Details**

Density of seawater is calculated according to the internationally accepted (UNESCO) equations. The standard error of the equation is  $3.6 \times 10^{-3}$  kg/m<sup>3</sup>.

**Value**

Rho (kg/m<sup>3</sup>)

**References**

Millero, F.J., Poisson, A., 1981. International one-atmosphere equation of state of seawater. Deep Sea Research 28, 625-629. [https://doi.org/10.1016/0198-0149\(81\)90122-9](https://doi.org/10.1016/0198-0149(81)90122-9)

**Examples**

```
temp <- c(10, 20, 30)
salt <- c(30, 35, 40)
ebase_rho(temp = temp, salt = salt, P = 0)
```

---

|               |                                   |
|---------------|-----------------------------------|
| ebase_schmidt | <i>Schmidt number calculation</i> |
|---------------|-----------------------------------|

---

**Description**

Schmidt number calculation

**Usage**

```
ebase_schmidt(temp, salt)
```

**Arguments**

|      |                             |
|------|-----------------------------|
| temp | numeric for temperature (C) |
| salt | numeric for salinity (PSU)  |

**Details**

The Schmidt number is calculated for the air-sea gas transfer velocity.

**Value**

sc (unitless)

**Examples**

```
temp <- c(10, 20, 30)
salt <- c(30, 35, 40)
ebase_schmidt(temp = temp, salt = salt)
```

---

|             |  |
|-------------|--|
| ebase_years | <i>Estuarine Bayesian Single-station Estimation method for ecosystem metabolism for long time series</i> |
|-------------|--|

---

**Description**

Estuarine Bayesian Single-station Estimation method for ecosystem metabolism for long time series

**Usage**

```

ebase_years(
  dat,
  Z,
  interval,
  ndays = 1,
  aprior = c(4, 2),
  rprior = c(300, 150),
  bprior = c(0.251, 0.125),
  bmax = 0.502,
  nogas = FALSE,
  doave = TRUE,
  maxinterp = 43200/interval,
  n.iter = 10000,
  update.chains = TRUE,
  n.burnin = n.iter * 0.5,
  n.chains = 3,
  n.thin = 10,
  model_file = NULL,
  ncores = NULL,
  quiet = TRUE,
  maxtry = 5
)

```

**Arguments**

|                       |   |
|-----------------------|---|
| <code>dat</code>      | input data frame  |
| <code>Z</code>        | numeric as single value for water column depth (m) or vector equal in length to number of rows in <code>dat</code>  |
| <code>interval</code> | timestep interval in seconds  |
| <code>ndays</code>    | numeric for number of days in <code>dat</code> for optimizing the metabolic equation, see details   |
| <code>aprior</code>   | numeric vector of length two indicating the mean and standard deviation for the prior distribution of the $a$ parameter, see details  |
| <code>rprior</code>   | numeric vector of length two indicating the mean and standard deviation for the prior distribution of the $R$ parameter, see details  |
| <code>bprior</code>   | numeric vector of length two indicating the mean and standard deviation for the prior distribution of the $b$ parameter, see details  |
| <code>bmax</code>     | numeric value for the upper limit on the prior distribution for <code>bprior</code> , set as twice the default value of the mean  |
| <code>nogas</code>    | logical indicating if gas exchange is not included in the metabolic model, see details  |
| <code>doave</code>    | logical indicating if the average dissolved oxygen concentration is used as the starting value for the estimation (default), otherwise the first observation will be used if FALSE, see details |

|                            |   |
|----------------------------|---|
| <code>maxinterp</code>     | numeric value for minimum number of continuous observations that must not be interpolated within a group defined by <code>ndays</code> to assign as NA in output, see details |
| <code>n.iter</code>        | number of MCMC iterations, passed to <code>jags</code>  |
| <code>update.chains</code> | logical to run <code>metab_update</code> if chains do not converge  |
| <code>n.burnin</code>      | number of MCMC chains to delete, passed to <code>jags</code>  |
| <code>n.chains</code>      | number of MCMC chains to run, passed to <code>jags</code>   |
| <code>n.thin</code>        | number of nth iterations to save for each chain, passed to <code>jags</code>  |
| <code>model_file</code>    | NULL to use the model file included with the package or a path to a model text file can be used   |
| <code>ncores</code>        | numeric for number of cores to use for parallel processing, use NULL to suppress  |
| <code>quiet</code>         | logical to suppress progress messages to the console  |
| <code>maxtry</code>        | numeric for maximum number of times to retry the model if it fails  |

### Details

`ebase` is run for each year in the supplied data. This facilitates running `ebase` on long time series by running the model sequentially on each year of data, with progress messages printed to the console if `quiet = FALSE`. The model run for each year will restart if it fails, up to `maxtry` times, after which processing continues with the next year. The model is run in parallel using the number of cores used set by `ncores`. If `ncores = NULL`, sequential processing is used. All other arguments are passed to `ebase`.

Similar results can be obtained by running `ebase` on the entire data set, but this function is useful for long time series where the model may fail for some years, e.g., when weather data may be missing for some years.

### Value

Output identical to that returned by `ebase`, where the results for each year are appended to the data frame as the function progresses through the years. Note that the `grp` column that specifies the optimization period defined by `ndays` is unique to each year, e.g., values will be repeated across years.

### Examples

```
# get one day of data
dat <- exdat[as.Date(exdat$DateTimeStamp, tz = 'America/Jamaica') == as.Date('2012-06-01'), ]

# run ebase, use more chains and iterations for a better fit, update.chains as T
ebase_years(dat, Z = 1.85, interval = 900, n.chains = 2, n.iter = 50,
  update.chains = FALSE)
```

exdat

*Sample data from Apalachicola NERRS*

---

**Description**

Sample data from Apalachicola NERRS

**Usage**

exdat

**Format**

A data.frame object with 27648 rows and 6 columns

**DateTimeStamp** date and time, America/Jamaica time zone, 15 minute time step

**DO\_obs** dissolved oxygen, mg/L

**Temp** water temperature, C

**Sal** salinity, psu

**PAR** total PAR, W/m2

**WSpd** num, m/s

**See Also**

Other utilities: [exres](#)

**Examples**

```
head(exdat)
```

---

exres*Example results for four days from Apalachicola NERRS*

---

**Description**

Example results for four days from Apalachicola NERRS

**Usage**

exres



**Format**

A data frame with 384 observations and 29 variables:

**DateTimeStamp** POSIXct, format: "2012-06-01 00:00:00" "2012-06-01 00:15:00" ...

**Date** Date, format: "2012-06-01" "2012-06-01" ...

**grp** Numeric, grouping variable defined by ndays in [ebase](#)

**Z** Numeric, depth in meters

**DO\_obs** Numeric, observed dissolved oxygen, mmol m-3

**DO\_mod** Numeric, modelled dissolved oxygen, mmol m-3

**DO\_modlo** Numeric, lower credible interval of modelled dissolved oxygen

**DO\_modhi** Numeric, upper credible interval of modelled dissolved oxygen

**dDO** Numeric, change in dissolved oxygen, mmol m-3 d-1

**converge** Character, convergence status

**rsq** Numeric, R-squared value

**a** Numeric, parameter a (mmol m-2 d-1)/(W m-2)

**alo** Numeric, lower credible interval of parameter a

**ahi** Numeric, upper credible interval of parameter a

**b** Numeric, parameter b, (cm hr-1)/(m2 s-2)

**blo** Numeric, lower credible interval of parameter b

**bhi** Numeric, upper credible interval of parameter b

**P** Numeric, production, O2 mmol m-2 d-1

**Plo** Numeric, lower credible interval of parameter P

**Phi** Numeric, upper credible interval of parameter P

**R** Numeric, respiration, O2 mmol m-2 d-1

**Rlo** Numeric, lower credible interval of parameter R

**Rhi** Numeric, upper credible interval of parameter R

**NEM** Numeric, net ecosystem metabolism, O2 mmol m-2 d-1

**NEMlo** Numeric, lower credible interval of parameter NEM

**NEMhi** Numeric, upper credible interval of parameter NEM

**D** Numeric, gas exchange, O2 mmol m-2 d-1)

**Dlo** Numeric, lower credible interval of parameter D

**Dhi** Numeric, upper credible interval of parameter D

**See Also**

Other utilities: [exdat](#)

**Examples**

```
head(exres)
```

---

|          |   |
|----------|---|
| fit_plot | <i>Plot observed and modeled dissolved oxygen</i> |
|----------|---|

---

### Description

Plot observed and modeled dissolved oxygen

### Usage

```
fit_plot(res, bygroup = FALSE, scatter = FALSE, showfit = TRUE)
```

### Arguments

|         |   |
|---------|---|
| res     | output data frame from <a href="#">ebase</a>  |
| bygroup | logical indicating if the plot is faceted by group  |
| scatter | logical indicating if a scatter plot of modeled versus estimated dissolved oxygen is returned |
| showfit | logical indicating if a linear fit is shown in the plot, applies only if scatter = TRUE       |

### Details

Dissolved oxygen (mmol/m<sup>3</sup>) is plotted as observed from the input data (points) and modeled (lines) based on inputs to [ebase](#) if scatter = FALSE. A scatter plot of modeled versus estimated dissolved oxygen is returned if scatter = TRUE, including a linear fit if showfit = TRUE. The plot is faceted by group based on the ndays argument to [ebase](#) if bygroup = TRUE. The r-squared value of the fit between modeled and observed dissolved oxygen is also shown in the facet label for the group if bygroup = TRUE.

### Value

A [ggplot](#) object

### Examples

```
# plot observed and modeled DO
fit_plot(exres)

# plot observed and modeled DO by group
fit_plot(exres, bygroup = TRUE)

# as scatter plot
fit_plot(exres, scatter = TRUE)

# as scatter plot by group
fit_plot(exres, scatter = TRUE, bygroup = TRUE)
```

---

|             |   |
|-------------|---|
| interp_plot | <i>Create a diagnostic plot showing interpolated values prior to metabolism estimates</i> |
|-------------|---|

---

## Description

Create a diagnostic plot showing interpolated values prior to metabolism estimates

## Usage

```
interp_plot(  
  dat,  
  param = c("DO_obs", "DO_sat", "Z", "Temp", "Sal", "PAR", "WSpd", "sc"),  
  Z,  
  interval,  
  ndays = 1  
)
```

## Arguments

|          |   |
|----------|---|
| dat      | input data frame  |
| param    | character string of the parameter to plot, one of DO_obs, DO_sat, Z, Temp, Sal, PAR, WSpd, or sc      |
| Z        | numeric as single value for water column depth (m) or vector equal in length to number of rows in dat |
| interval | timestep interval in seconds  |
| ndays    | numeric for number of days in dat for optimizing the metabolic equation, see details                  |

## Details

Missing values in the input data can also be interpolated prior to estimating metabolism. This is the default behavior and it is the responsibility of the user to verify that these interpolated values are not wildly inaccurate. Missing values are linearly interpolated between non-missing values at the time step specified by the value in `interval`. This works well for small gaps, but can easily create inaccurate values at gaps larger than a few hours. The plot from this function can be used to visually assess the interpolated gaps.

## Value

A `ggplot` object

## Examples

```
library(dplyr)

# get four days of data
dat <- exdat %>%
  filter(lubridate::month(DateTimeStamp) == 6) %>%
  filter(lubridate::day(DateTimeStamp) %in% 1:4)

# create missing values
set.seed(222)
dat <- dat %>%
  slice_sample(prop = 0.9) %>%
  arrange(DateTimeStamp)

interp_plot(dat, Z = 1.85, interval = 900, param = 'D0_sat')
```

---

|              |                                   |
|--------------|-----------------------------------|
| metab_update | <i>Update metabolism jags fit</i> |
|--------------|-----------------------------------|

---

## Description

Update metabolism jags fit

## Usage

```
metab_update(metabfit, update.chains, n.iter)
```

## Arguments

|               |                                 |
|---------------|---------------------------------|
| metabfit      | initial jags metabolism output  |
| update.chains | logical to update, only if TRUE |
| n.iter        | number of iterations            |

## Details

This function is used by [ebase](#) and is not to be called directly by the user. It provides additional model iterations if convergence is not achieved.

## Value

Updated jags metabolism output

---

prior\_plot

*Plot prior distributions for a, R, and b*

---

### Description

Plot prior distributions for a, R, and b

### Usage

```
prior_plot(  
  aprior = c(4, 2),  
  rprior = c(300, 150),  
  bprior = c(0.251, 0.125),  
  bmax = 0.502,  
  n = 1000  
)
```

### Arguments

|        |   |
|--------|---|
| aprior | numeric vector of length two indicating the mean and standard deviation for the prior distribution of the <i>a</i> parameter, see details |
| rprior | numeric vector of length two indicating the mean and standard deviation for the prior distribution of the <i>R</i> parameter, see details |
| bprior | numeric vector of length two indicating the mean and standard deviation for the prior distribution of the <i>b</i> parameter, see details |
| bmax   | numeric value for the upper limit on the prior distribution for bprior, set as twice the default value of the mean                        |
| n      | numeric indicating number of random samples to draw from prior distributions  |

### Details

This function produces a plot of the prior distributions that are used in [ebase](#) for the *a*, *R*, and *b* parameters for the optimization equation for estimating metabolism. The [ebase](#) function uses the same default values for the arguments for aprior, rprior, and bprior as required for this function. If the default values are changed for [ebase](#), this function can be used to assess how changing characteristics of the prior distributions could influence the resulting parameter estimates and their posterior distributions (e.g., as shown with [credible\\_plot](#)).

All parameters follow a normal Gaussian distribution for the priors with the means and standard deviations defined by the arguments. All distributions are truncated to include only values greater than zero as required by the core metabolism equation. The upper limit for *b* is also set as twice the default value of the mean in the bprior argument. Truncated normal distributions are obtained using the [rtruncnorm](#) function with the number of random samples defined by the n argument.

The density curves for each parameter are normalized such that the peak values are always equal to 1.

**Value**

A `ggplot` object

**Examples**

```
# default plot  
prior_plot()
```

```
# changing the mean and standard deviation for the b parameter  
prior_plot(bprior = c(0.2, 0.05))
```

# Index

## \* datasets

exdat, 16

exres, 16

## \* utilities

exdat, 16

exres, 16

credible\_plot, 2, 3, 21

credible\_prep, 3

ebase, 2, 3, 4, 9–12, 15, 17, 18, 20, 21

ebase\_eqboxy, 8

ebase\_form, 9

ebase\_plot, 10

ebase\_prep, 9, 11

ebase\_rho, 12

ebase\_schmidt, 13

ebase\_years, 13

exdat, 5, 11, 16, 17

exres, 16, 16

fit\_plot, 7, 18

ggplot, 3, 10, 18, 19, 22

interp\_plot, 6, 19

jags, 5, 15

metab\_update, 5, 15, 20

prior\_plot, 21

rtruncnorm, 21