Package 'dycdtools'

May 8, 2022

Title Tools for DYRESM-CAEDYM Model Development: Calibration Assistant and Post-Processing

Version 0.4.1

Description Tools for DYRESM-CAEDYM model development, including assisting with calibrating selected model parameters and visualising model output through time series plot, profile plot, contour plot, and scatter plot.

License MIT + file LICENSE

Encoding UTF-8

LazyData true

Imports dplyr, ncdf4, tidyr, hydroGOF, ggplot2, RColorBrewer, lubridate, R.utils, parallel

RoxygenNote 7.1.2

Depends R (>= 2.10)

Suggests testthat (>= 3.0.0)

Config/testthat/edition 3

NeedsCompilation no

Author Songyan Yu [aut, cre] (<https://orcid.org/0000-0001-5765-7060>), Christopher McBride [ctb], Marieke Frassl [ctb]

Maintainer Songyan Yu <sunny.yu@griffith.edu.au>

Repository CRAN

Date/Publication 2022-05-08 10:00:02 UTC

R topics documented:

`calib_assist` *Assist calibration of DYRESM-CAEDYM model. Before using this function, make sure that you have set up "Bin" and "Files" sub-folders under the DYCD model folder.*

Description

This function tries different combinations of selected parameter values and outputs corresponding values of fit-of-goodness by calculating some objective functions. Then users can choose the optimal set of parameter values to calibrate the model.

Usage

```
calib_assist(
  cal.para,
  combination = "random",
  n,
  model.var,
  phyto.group = NA,
  obs.data,
  objective.function = c("NSE", "RMSE"),
  start.date,
  end.date,
  dycd.wd,
  dycd.output,
  file.name,
  verbose = TRUE,
  parallel = FALSE,
  n.cores = NULL,
  write.out = TRUE
)
```

Arguments

cal.para	a data frame or a character string naming an external .csv file where below columns are included: "Parameter" where parameter names (abbreviation is allowed), "Min", "Max", and "Increment" describing the minimum and maximum parameter values and expected increment in the value range, "Input_file" and "Line_NO" listing in which configuration file at which line can the parameter can be found.
combination	a vector of string character of how to pick up combinations of parameter values."random" - the function randomly picks up given number of combinations; "all" - the function tries all possible combinations of parameter values.
n	the number of randomly selections. Must be provided if combination = "random".
model.var	a vector of string character of modelled variables for calibration. If the chlorophyll of multiple phytoplankton groups is used collectively for calibration, use "CHLA" and further specify which phytoplankton groups are to be combined in the argument of "phyto.group". If phytoplankton groups are separately calibrated, list their abbreviation in this argument. Five abbreviations are supported: CHLOR, FDIAT, NODUL, CYANO and CRYPT.
phyto.group	a vector of simulated phytoplankton groups, including CHLOR, FDIAT, NODUL, CYANO and CRYPT.
obs.data	a character string naming a file of observed lake data. This file needs to be prepared in a given format (see example data).
objective.function	a vector of string character claiming what objective function(s) to be used for calibration. Selected from the following five functions: "NSE": Nash-Sutcliffe efficiency coefficient, "RMSE": Root Mean Square Error, "MAE": Mean Absolute Error, "RAE": Relative Absolute Error, "Pearson": Pearson's r.
start.date, end.date	the beginning and ending simulation dates for the intended DYRESM-CAEDYM calibration. The date format must be "%Y-%m-%d".
dycd.wd	the directory where input files (including the bat file) to DYRESM-CAEDYM are stored. either relative or absolute path is allowed.
dycd.output	a character string naming the output file of the model calibration.
file.name	a character string naming a .csv file for writing out the auto-calibration results.
verbose	if TRUE, the auto-calibration information is printed.
parallel	if TRUE, the calibration process can be run on multiple cores.
n.cores	When parallel is TRUE, n.cores is the number of cores the calibration function will be run on. If not provided, the default value is the number of available cores on the computer -1.
write.out	if TRUE, the auto-calibration results are saved a file with a file name set by the "file.name" argument.

Value

a dataframe of trialed values of parameters and corresponding values of objective function(s).

Note

No executable examples are provided to illustrate the use of this function, as this function relies on the DYRESM-CAEDYM executables to work.

`change_input_file` *change parameter value of input files to DYRESM_CAEDYM model.*

Description

change parameter value of input files to DYRESM_CAEDYM model.

Usage

```
change_input_file(input_file, row_no, new_value)
```

Arguments

`input_file`	vector of input format, such as "par","cfg".
`row_no`	the number of row where the variable of interest is in the input file
`new_value`	the new value that will be assigned to the variable of interest.

`delete_space` *Delete all whitespace until a non-whitespace character.*

Description

Delete all whitespace until a non-whitespace character.

Usage

```
delete_space(extract_val)
```

Arguments

`extract_val`	a vector.

ext_output *Extract simulations from DYRESM-CAEDYM output.*

Description

Extract simulations from DYRESM-CAEDYM output. It is recommended to use the following example code to assign values of each simulated variable to a corresponding matrix.

Usage

```
ext_output(dycd.output, var.extract, verbose = FALSE)
```

Arguments

dycd.output	a string of characters of the output netcdf file names of DYRESM-CAEDYM model
var.extract	a vector of variables to be extracted from the output file. Options include TEMP, DO, TP, NO3, NH4, TN, CHLA, FDIAT, SSOL1, SSOL2, and PO4. Apart from nominated variables, simulation period and layer height data are also extracted. For a full list of variables that can be extracted, use "data("output_name")".
verbose	if TRUE, the information on extraction of simulation outputs is printed.

Value

a list of values of those variables of interest, as well as two compulsory variables (i.e. simulation period, water level)

Examples

```
# extract simulated temperature values from DYRESM-CAEDYM simulation file
var.values<-ext_output(dycd.output=system.file("extdata", "dysim.nc", package = "dycdtools"),
                       var.extract=c("TEMP"))

for(i in 1:length(var.values)){
 expres<-paste0(names(var.values)[i],"<-data.frame(var.values[[",i,"]])")
 eval(parse(text=expres))
}
```

hgt_to_dpt *convert from height to depth*

Description

convert from height to depth

Usage

```
hgt_to_dpt(height)
```

Arguments

height	a vector of height profile

interpol *Interpolation of simulation for a series of user-defined depths.*

Description

Convert simulated variable at irregular layer heights to a dataframe of the same variable at sequenced layer heights.

Usage

```
interpol(layerHeights, var, min.depth, max.depth, by.value)
```

Arguments

layerHeights	layer heights, extracted from DYCD outputs
var	simulated variable values, extracted from DYCD outputs
min.depth, max.depth, by.value	minimum and maximum depth for interpolation at the depth increment of by.value.

Value

a matrix of interpolated values of such variable.

Examples

```
# extract simulated temperature values from DYRESM-CAEDYM simulation file
 var.values<-ext_output(dycd.output=system.file("extdata", "dysim.nc", package = "dycdtools"),
                        var.extract=c("TEMP"))

 for(i in 1:length(var.values)){
  expres<-paste0(names(var.values)[i],"<-data.frame(var.values[[",i,"]])")
  eval(parse(text=expres))
 }
 # interpolate temperature for depths from 0 to 13 m at increment of 0.5 m
 temp.interpolated<-interpol(layerHeights = dyresmLAYER_HTS_Var,
                             var = dyresmTEMPTURE_Var,
                             min.dept = 0,max.dept = 13,by.value = 0.5)
```

`objective_fun` *Calculate measures of goodness of fit for DYCD model simulations.*

Description

calculate any of the five objective functions that are commonly used to measure goodness of fit: 1) Nash-Sutcliffe efficiency coefficient (NSE), 2) Root Mean Square Error (RMSE), 3) Mean Absolute Error (MAE), 4) Relative Absolute Error (RAE), and 5) Pearson's r (Pearson).

Usage

```
objective_fun(
  sim,
  obs,
  fun = c("NSE", "RMSE"),
  start.date,
  end.date,
  min.depth,
  max.depth,
  by.value
)
```

Arguments

`sim`	a matrix of bio-geochemical variable values with column of time and row of depth.
`obs`	a data frame of observed value, with three columns: Date, depth, value.
`fun`	objective function to be calculated. select any from 'NSE', 'RMSE', 'MAE', 'RAE', and 'Pearson'. Multiple is allowed.
`start.date, end.date`	the beginning and ending simulation dates for the intended DYRESM-CAEDYM model run. The date format must be "%Y-%m-%d".

`min.depth, max.depth`
: the minimum and maximum depths of the simulation matrix .

`by.value`
: the value of increment for depth of the simulation matrix.

Value

a vector of objective function values. The first is NSE and the second is RMSE.

`obs_temp` *Example observed profiling temperature data across different depths over the period of 6-11 June 2017.*

Description

A table has three columns. The first column name is Date in the form of dd-mm-YY. The second column is Depth where the temperature data was monitored. The third column is monitored temperature value.

Usage

```
data(obs_temp)
```

Format

A data frame with 77 rows and 3 variables:

Date date when the monitoring happened

Depth depth of monitoring

TEMP temperature value

Source

self-made.

`output_name` *Default DYCD simulation variable names with their variable name*

Description

A table has two columns. The first column name is var.name, meaning variable names that are used in the extract.output function. The second column is the default DYCD simulation variable names, such as "dyresmLAYER_HTS_Var".

Usage

```
data(output_name)
```

Format

A data frame with 65 rows and 2 variables:

var.name variable name

output.name default DYCD simulation variable name

Source

self-made.

`plot_cont` *Contour plot of simulation results of a bio-geochemical variable.*

Description

Contour plot a matrix of values of a bio-geochemical variable, which can be generated through "interpol" function.

Usage

```
plot_cont(
  sim,
  sim.start,
  sim.end,
  legend.title,
  min.depth,
  max.depth,
  by.value,
  nlevels
)
```

Arguments

`sim`	a matrix of simulated variables that have been interpolated
`sim.start, sim.end`	the start and end of the simulation period for the DYRESM-CAEDYM model run of interest. The date format must be "%Y-%m-%d".
`legend.title`	the legend title of the contour figure.
`min.depth, max.depth, by.value`	minimum and maximum depth used to be the start of y axis of the contour plot, at the increment of by.value.
`nlevels`	a set of levels which are used to partition the range of simulation variable.

Value

This function returns a filled.contour object.

Examples

```
 # extract simulated temperature values from DYRESM-CAEDYM simulation file
 var.values<-ext_output(dycd.output=system.file("extdata", "dysim.nc", package = "dycdtools"),
                      var.extract=c("TEMP"))

 for(i in 1:length(var.values)){
  expres<-paste0(names(var.values)[i],"<-data.frame(var.values[[",i,"]])")
  eval(parse(text=expres))
 }

# interpolate temperature for depths from 0 to 13 m at increment of 0.5 m
  temp.interpolated<-interpol(layerHeights = dyresmLAYER_HTS_Var,
                              var = dyresmTEMPTURE_Var,
                              min.dept = 0,max.dept = 13,by.value = 0.5)

# contour plot of temperature simulations
  p <- plot_cont(sim=temp.interpolated,
            sim.start="2017-06-06",
            sim.end="2017-06-15",
            legend.title="T \u00B0C",
            min.depth=0,max.depth=13,by.value=0.5,
            nlevels=20)

  p
```

`plot_cont_comp` *Contour plot of a variable simulation, with observed data shown as dots in the generated contour plot.*

Description

Contour plot a matrix of a bio-geochemical variable values, which can be generated through "interpol" function.

Usage

```
plot_cont_comp(
  sim,
  obs,
  file.name,
  sim.start,
  sim.end,
  plot.start,
  plot.end,
  legend.title,
  min.depth,
  max.depth,
```

```
  by.value,
  nlevels
)
```

Arguments

sim	a matrix of simulated variables that have been interpolated
obs	observed values of variable.
file.name	the file path to save the generated contour figure.
sim.start, sim.end	the start and end of the simulation period for the DYRESM-CAEDYM model run of interest. The date format must be "%Y-%m-%d".
plot.start, plot.end	the start and end of the plot period, in the format of "%Y-%m-%d"
legend.title	the legend title of the contour figure.
min.depth, max.depth, by.value	minimum and maximum depth used to be the start of y axis of the contour plot, at the increment of by.value.
nlevels	a set of levels which are used to partition the range of simulation variable.

Value

This function returns a filled.contour object.

Examples

```
# extract simulated temperature values from DYRESM-CAEDYM simulation file
 var.values<-ext_output(dycd.output=system.file("extdata", "dysim.nc", package = "dycdtools"),
                        var.extract=c("TEMP"))

  for(i in 1:length(var.values)){
     expres<-paste0(names(var.values)[i],"<-data.frame(var.values[[",i,"]])")
     eval(parse(text=expres))
  }

# interpolate temperature for depths from 0 to 13 m at increment of 0.5 m
  temp.interpolated<-interpol(layerHeights = dyresmLAYER_HTS_Var,
                              var = dyresmTEMPTURE_Var,
                              min.dept = 0,max.dept = 13,by.value = 0.5)

  data(obs_temp)
# contour plot of temperature simulations with observed data shown as colour-coded dots
  p <- plot_cont_comp(sim=temp.interpolated,
                 obs=obs_temp,
                 sim.start = "2017-06-06",
                 sim.end = "2017-06-15",
                 plot.start="2017-06-06",
                 plot.end="2017-06-15",
                 legend.title="T \u00B0C",
                 min.depth=0,max.depth=13,by.value=0.5,
```

```
                nlevels=20)

  p
```

---

plot_prof *Profile plot of simulated variable values vs. depth*

Description

A post-processing function used to visualise model output in a profile graph.

Usage

```
plot_prof(
  sim,
  obs,
  sim.start,
  sim.end,
  plot.start,
  plot.end,
  xlabel,
  min.depth,
  max.depth,
  by.value
)
```

Arguments

sim	interpolated values of variable.
obs	observed values of variable.
sim.start, sim.end	the beginning and ending simulation dates for the intended DYRESM-CAEDYM model run. The date format must be "%Y-%m-%d".
plot.start, plot.end	the beginning and ending dates for the plotting purpose. The date format must be "%Y-%m-%d".
xlabel	the x axis label of the profile figure
min.depth, max.depth, by.value	minimum and maximum depth for the profile plot at the depth increment of by.value.

Value

This function returns a ggplot object that can be modified with ggplot package functions.

Examples

```
 var.values<-ext_output(dycd.output=system.file("extdata", "dysim.nc", package = "dycdtools"),
                      var.extract=c("TEMP"))

 for(i in 1:length(var.values)){
   expres<-paste0(names(var.values)[i],"<-data.frame(var.values[[",i,"]])")
   eval(parse(text=expres))
  }

# interpolate temperature for depths from 0 to 13 m at increment of 0.5 m
  temp.interpolated<-interpol(layerHeights = dyresmLAYER_HTS_Var,
                              var = dyresmTEMPTURE_Var,
                              min.dept = 0,max.dept = 13,by.value = 0.5)

 data(obs_temp)
# profile plot of temperature sim and obs
  p <- plot_prof(sim=temp.interpolated,
            obs = obs_temp,
            sim.start="2017-06-06",
            sim.end="2017-06-15",
            plot.start="2017-06-06",
            plot.end="2017-06-15",
            xlabel = "Temperature \u00B0C",
            min.depth = 0,max.depth = 13,by.value = 0.5)
 p
```

`plot_scatter` *Scatter plot of sim and obs var values*

Description

Scatter plot of sim and obs var values

Usage

```
plot_scatter(
  sim,
  obs,
  sim.start,
  sim.end,
  plot.start,
  plot.end,
  min.depth,
  max.depth,
  by.value
)
```

Arguments

sim	interpolated values of variable.
obs	observed values of variable. This data need to have fixed types of colnames and orders.
sim.start, sim.end	the beginning and ending simulation dates for the intended DYRESM-CAEDYM model run. The date format must be "%Y-%m-%d".
plot.start, plot.end	the beginning and ending dates for the plotting purpose. The date format must be "%Y-%m-%d".
min.depth, max.depth, by.value	minimum and maximum depth for the profile plot at the depth increment of by.value.

Value

This function returns a ggplot object that can be modified with ggplot package functions.

Examples

```
 var.values<-ext_output(dycd.output=system.file("extdata", "dysim.nc", package = "dycdtools"),
                        var.extract=c("TEMP"))

 for(i in 1:length(var.values)){
   expres<-paste0(names(var.values)[i],"<-data.frame(var.values[[",i,"]])")
   eval(parse(text=expres))
  }

# interpolate temperature for depths from 0 to 13 m at increment of 0.5 m
  temp.interpolated<-interpol(layerHeights = dyresmLAYER_HTS_Var,
                              var = dyresmTEMPTURE_Var,
                              min.dept = 0,max.dept = 13,by.value = 0.5)

 data(obs_temp)

# scatter plot of sim and obs temperature
 p <- plot_scatter(sim=temp.interpolated,
              obs=obs_temp,
              sim.start="2017-06-06",
              sim.end="2017-06-15",
              plot.start="2017-06-06",
              plot.end="2017-06-15",
              min.depth = 0,max.depth = 13,by.value = 0.5)

 p
```

plot_ts *Time series plot of simulated and observed values*

Description

Time series plot of simulated and observed values

Usage

```
plot_ts(
  sim,
  obs,
  target.depth,
  sim.start,
  sim.end,
  plot.start,
  plot.end,
  min.depth,
  max.depth,
  by.value,
  ylabel
)
```

Arguments

`sim`	interpolated values of variable
`obs`	observed values of variable
`target.depth`	a vector of depth (unit:m) to be used to extract and plot variable values.
`sim.start, sim.end`	the beginning and ending simulation dates for the intended DYRESM-CAEDYM model run. The date format must be "%Y-%m-%d".
`plot.start, plot.end`	the beginning and ending dates for the plotting purpose. The date format must be "%Y-%m-%d".
`min.depth, max.depth, by.value`	minimum and maximum depth for the profile plot at the depth increment of by.value.
`ylabel`	the y axis title.

Value

This function returns a ggplot object that can be modified with ggplot package functions.

Examples

```
 var.values<-ext_output(dycd.output=system.file("extdata", "dysim.nc", package = "dycdtools"),
                        var.extract=c("TEMP"))

 for(i in 1:length(var.values)){
   expres<-paste0(names(var.values)[i],"<-data.frame(var.values[[",i,"]])")
   eval(parse(text=expres))
  }

# interpolate temperature for depths from 0 to 13 m at increment of 0.5 m
  temp.interpolated<-interpol(layerHeights = dyresmLAYER_HTS_Var,
                              var = dyresmTEMPTURE_Var,
                              min.dept = 0,max.dept = 13,by.value = 0.5)

 data(obs_temp)
# time series plot of temperature sim and obs
 p <- plot_ts(sim = temp.interpolated,
         obs = obs_temp,
         target.depth=c(1,6),
         sim.start="2017-06-06",
         sim.end="2017-06-15",
         plot.start="2017-06-06",
         plot.end="2017-06-15",
         ylabel="Temperature \u00B0C",
         min.depth=0,
         max.depth=13,
         by.value=0.5)

 p
```

`run_iteration` *Internal function to provide parallel processing support to the calibration assistant function.*

Description

Internal function to provide parallel processing support to the calibration assistant function.

Usage

```
run_iteration(this.sim, dycd.wd)
```

Arguments

`this.sim`	a numeric denoting which parameter combination to be tried.
`dycd.wd`	working directory where input files (including the bat file) to DYRESM-CAEDYM are stored.

Index