Package ‘postGGIR’

July 14, 2021

Version 2.4.0

Title Data Processing after Running 'GGIR' for Accelerometer Data

Description
Generate all necessary R/Rmd/shell files for data processing after running 'GGIR' (v2.4.0) for accelerometer data. In part 1, all csv files in the GGIR output directory were read, transformed and then merged. In part 2, the GGIR output files were checked and summarized in one excel sheet. In part 3, the merged data was cleaned according to the number of valid hours on each night and the number of valid days for each subject. In part 4, the cleaned activity data was imputed by the average Euclidean norm minus one (ENMO) over all the valid days for each subject. Finally, a comprehensive report of data processing was created using Rmarkdown, and the report includes few exploratory plots and multiple commonly used features extracted from minute level actigraphy data.

Maintainer Wei Guo <wei.guo3@nih.gov>

License GPL-3

Imports refund, denseFLMM, dplyr, xlsx, survival, stats, tidyr, zoo, ineq, cosinor, cosinor2, abind, accelerometry, ActCR, ActFrag, minpack.lm, kableExtra, GGIR

Depends R (>= 3.6.0)

Suggests knitr, rmarkdown

Encoding UTF-8

ByteCompile true

Type Package

VignetteBuilder knitr

RoxygenNote 7.1.1

NeedsCompilation no

Author Wei Guo [aut, cre], Vadim Zipunnikov [aut], Kathleen Merikangas [aut]

Repository CRAN

Date/Publication 2021-07-14 16:50:02 UTC
\section*{R topics documented:}

ActCosinor2 .................................................. 2
ActCosinor_long2 ........................................... 3
ActExtendCosinor2 .......................................... 4
ActExtendCosinor_long2 .................................... 5
aftergir ....................................................... 7
bin2_data2 ..................................................... 9
create.postGGIR ............................................. 9
data.imputation ............................................... 10
DataShrink .................................................... 11
fragmentation2 ............................................... 13
fragmentation_long2 ....................................... 15
ggir.datatransform ......................................... 16
ggir.summary ................................................ 18
IS2 ............................................................. 19
IS_long2 ...................................................... 20
IV2 ............................................................... 21
IV_long2 ...................................................... 21
jive.predict2 ................................................ 22
PAfun ........................................................... 23
pheno.plot ..................................................... 23
RA2 ............................................................. 24
RA_long2 ...................................................... 25
SVDmiss2 ...................................................... 26
Time2 ........................................................... 26
Time_long2 .................................................... 27
Tvol2 ............................................................ 28
wear_flag ...................................................... 29

\section*{Index} 30

\begin{tabular}{ll}
 ActCosinor2 & \textit{Cosinor Model for Circadian Rhythmicity} \\
\end{tabular}

\section*{Description}

A parametric approach to study circadian rhythmicity assuming cosinor shape.

\section*{Usage}

\texttt{ActCosinor2(x, window = 1, n1440 = 1440)}
**Arguments**

- **x**: vector of dimension n*1440 which represents n days of 1440 minute activity data.
- **window**: The calculation needs the window size of the data. E.g. window = 1 means each epoch is in one-minute window.
- **n1440**: the number of points of a day. Default is 1440 for the minute-level data.

**Value**

A list with elements

- **mes**: MESOR which is short for midline statistics of rhythm, which is a rhythm adjusted mean. This represents mean activity level.
- **amp**: amplitude, a measure of half the extend of predictable variation within a cycle. This represents the highest activity one can achieve.
- **acro**: acrophase, a measure of the time of the overall high values recurring in each cycle. Here it has a unit of radian. This represents time to reach the peak.
- **acrotile**: acrophase in the unit of the time (hours)
- **ndays**: Number of days modeled

**References**


---

**ActCosinor_long2**

*Cosinor Model for Circadian Rhythmicity for the Whole Dataset*

**Description**

A parametric approach to study circadian rhythmicity assuming cosinor shape. This function is a whole dataset wrapper for ActCosinor.

**Usage**

`ActCosinor_long2(count.data, window = 1)`

**Arguments**

- **count.data**: data.frame of dimension n * (p+2) containing the p dimensional activity data for all n subject days. The first two columns have to be ID and Day. ID can be either character or numeric. Day has to be numeric indicating the sequence of days within each subject.
- **window**: The calculation needs the window size of the data. E.g. window = 1 means each epoch is in one-minute window.
Value
A data.frame with the following 5 columns

<table>
<thead>
<tr>
<th>ID</th>
<th>ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>ndays</td>
<td>number of days</td>
</tr>
<tr>
<td>mes</td>
<td>MESRO, which is short for midline statistics of rhythm, which is a rhythm adjusted mean. This represents mean activity level.</td>
</tr>
<tr>
<td>amp</td>
<td>amplitude, a measure of half the extend of predictable variation within a cycle. This represents the highest activity one can achieve.</td>
</tr>
<tr>
<td>acro</td>
<td>acrophase, a measure of the time of the overall high values recurring in each cycle. Here it has a unit of radian. This represents time to reach the peak.</td>
</tr>
<tr>
<td>acrotimes</td>
<td>acrophase in the unit of the time (hours)</td>
</tr>
<tr>
<td>ndays</td>
<td>Number of days modeled</td>
</tr>
</tbody>
</table>

ActExtendCosinor2
Extended Cosinor Model for Circadian Rhythmicity

Description
Extended cosinor model based on sigmoidally transformed cosine curve using anti-logistic transformation

Usage
ActExtendCosinor2(
  x,
  window = 1,
  lower = c(0, 0, -1, 0, -3),
  upper = c(Inf, Inf, 1, Inf, 27),
  n1440 = 1440
)

Arguments

- **x**: vector vector of dimension n*1440 which represents n days of 1440 minute activity data
- **window**: The calculation needs the window size of the data. E.g window = 1 means each epoch is in one-minute window.
- **lower**: A numeric vector of lower bounds on each of the five parameters (in the order of minimum, amplitude, alpha, beta, acrophase) for the NLS. If not given, the default lower bound for each parameter is set to -Inf.
- **upper**: A numeric vector of upper bounds on each of the five parameters (in the order of minimum, amplitude, alpha, beta, acrophase) for the NLS. If not given, the default lower bound for each parameter is set to Inf.
- **n1440**: the number of points of a day. Default is 1440 for the minute-level data.
**Value**

A list with elements

- **minimum**: Minimum value of the function.
- **amp**: Amplitude, a measure of half the extend of predictable variation within a cycle. This represents the highest activity one can achieve.
- **alpha**: It determines whether the peaks of the curve are wider than the troughs: when alpha is small, the troughs are narrow and the peaks are wide; when alpha is large, the troughs are wide and the peaks are narrow.
- **beta**: It determines whether the transformed function rises and falls more steeply than the cosine curve: large values of beta produce curves that are nearly square waves.
- **acrotimes**: Acrophase is the time of day of the peak in the unit of the time (hours)
- **F_pseudo**: Measure the improvement of the fit obtained by the non-linear estimation of the transformed cosine model
- **UpMesor**: Time of day of switch from low to high activity. Represents the timing of the rest-activity rhythm. Lower (earlier) values indicate increase in activity earlier in the day and suggest a more advanced circadian phase.
- **DownMesor**: Time of day of switch from high to low activity. Represents the timing of the rest-activity rhythm. Lower (earlier) values indicate decline in activity earlier in the day, suggesting a more advanced circadian phase.
- **MESOR**: A measure analogous to the MESOR of the cosine model (or half the deflection of the curve) can be obtained from mes=\text{min}+\text{amp}/2. However, it goes through the middle of the peak, and is therefore not equal to the MESOR of the cosine model, which is the mean of the data.
- **ndays**: Number of days modeled.

**References**

Usage

ActExtendCosinor_long2(
  count.data,  
  window = 1,  
  lower = c(0, 0, -1, 0, -3),  
  upper = c(Inf, Inf, 1, Inf, 27)
)

Arguments

count.data  data.frame of dimension n * (p+2) containing the p dimensional activity data for all n subject days. The first two columns have to be ID and Day. ID can be either character or numeric. Day has to be numeric indicating the sequence of days within each subject.

window  The calculation needs the window size of the data. E.g window = 1 means each epoch is in one-minute window. window size as an argument.

lower  A numeric vector of lower bounds on each of the five parameters (in the order of minimum, amplitude, alpha, beta, acrophase) for the NLS. If not given, the default lower bound for each parameter is set to -Inf.

upper  A numeric vector of upper bounds on each of the five parameters (in the order of minimum, amplitude, alpha, beta, acrophase) for the NLS. If not given, the default lower bound for each parameter is set to Inf.

Value

A data.frame with the following 5 columns

<table>
<thead>
<tr>
<th>ID</th>
<th>ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>ndays</td>
<td>number of days</td>
</tr>
<tr>
<td>minimum</td>
<td>Minimum value of the of the function.</td>
</tr>
<tr>
<td>amp</td>
<td>amplitude, a measure of half the extend of predictable variation within a cycle. This represents the highest activity one can achieve.</td>
</tr>
<tr>
<td>alpha</td>
<td>It determines whether the peaks of the curve are wider than the troughs: when alpha is small, the troughs are narrow and the peaks are wide; when alpha is large, the troughs are wide and the peaks are narrow.</td>
</tr>
<tr>
<td>beta</td>
<td>It determines whether the transformed function rises and falls more steeply than the cosine curve: large values of beta produce curves that are nearly square waves.</td>
</tr>
<tr>
<td>acrotime</td>
<td>acrophase is the time of day of the peak in the unit of the time (hours)</td>
</tr>
<tr>
<td>F_pseudo</td>
<td>Measure the improvement of the fit obtained by the non-linear estimation of the transformed cosine model</td>
</tr>
<tr>
<td>UpMesor</td>
<td>Time of day of switch from low to high activity. Represents the timing of the rest- activity rhythm. Lower (earlier) values indicate increase in activity earlier in the day and suggest a more advanced circadian phase.</td>
</tr>
</tbody>
</table>
DownMesor

Time of day of switch from high to low activity. Represents the timing of the rest-activity rhythm. Lower (earlier) values indicate decline in activity earlier in the day, suggesting a more advanced circadian phase.

MESOR

A measure analogous to the MESOR of the cosine model (or half the deflection of the curve) can be obtained from mes = min + amp/2. However, it goes through the middle of the peak, and is therefore not equal to the MESOR of the cosine model, which is the mean of the data.

Description

This R script will generate all necessary R/Rmd/shell files for data processing after running GGIR for accelerometer data.

Usage

```r
afterggir(
  mode,
  useIDs.FN = NULL,
  currentdir,
  studyname,
  bindir = NULL,
  outputdir,
  epochIn = 5,
  epochOut = 5,
  flag.epochOut = 60,
  log.multiplier = 9250,
  use.cluster = TRUE,
  QCdays.alpha = 7,
  QChours.alpha = 16,
  QCnights.feature.alpha = c(0, 0),
  Rversion = "R",
  filename2id = NULL,
  PA.threshold = c(50, 100, 400),
  desiredtz = "US/Eastern",
  RemoveDaySleeper = FALSE,
  trace = FALSE
)
```

Arguments

- **mode** number Specify which of the five parts need to be run, e.g. mode = 0 makes that all R/Rmd/sh files are generated for other parts. When mode = 1, all csv files in the GGIR output directory were read, transformed and then merged. When
mode = 2, the GGIR output files were checked and summarized in one excel sheet. When mode = 3, the merged data was cleaned according to the number of valid hours on each night and the number of valid days for each subject. When mode = 4, the cleaned data was imputed.

useIDs.FN character Filename with or without directory for sample information in CSV format, including "filename" and "duplicate" in the headlines at least. If duplicate="remove", the accelerometer files will not be used in the data analysis of part 5-7. Default is NULL, which makes all accelerometer files will be used in part 5-7.

currentdir character Directory where the output needs to be stored. Note that this directory must exist.

studyname character Specify the study name that used in the output file names

bindir character Directory where the accelerometer files are stored or list

outputdir character Directory where the GGIR output was stored.

epochIn number Epoch size to which acceleration was averaged (seconds) in GGIR output. Default is 5 seconds.

epochOut number Epoch size to which acceleration was averaged (seconds) in part1. Default is 5 seconds.

flag.epochOut number Epoch size to which acceleration was averaged (seconds) in part 3. Default is 60 seconds.

log.multiplier number The coefficient used in the log transformation of the ENMO data, i.e. log( log.multiplier * ENMO + 1), which have been used in part 5-7. Default is 9250.

use.cluster logical Specify if part1 will be done by parallel computing. Default is TRUE, and the CSV file in GGIR output will be merged for every 20 files first, and then combined for all.

QCdays.alpha number Minimum required number of valid days in subject specific analysis as a quality control step in part2. Default is 7 days.

QChours.alpha number Minimum required number of valid hours in day specific analysis as a quality control step in part2. Default is 16 hours.

QCnights.feature.alpha number Minimum required number of valid nights in day specific mean and SD analysis as a quality control step in the JIVE analysis. Default is c(0,0), i.e. no additional data cleaning in this step.

Rversion character R version, eg. "R/3.6.3". Default is "R".

filename2id R function User defined function for converting filename to sample IDs. Default is NULL.

PA.threshold number Threshold for light, moderate and vigorous physical activity. Default is c(50,100,400).

desiredtz character desired timezone: see also http://en.wikipedia.org/wiki/Zone.tab. Used in g.inspectfile(). Default is "US/Eastern".

RemoveDaySleeper logical Specify if the daysleeper nights are removed from the calculation of number of valid days for each subject. Default is FALSE.
trace logical  Specify if the intermediate results is printed when the function was executed. Default is FALSE.

Value
See postGGIR manual for details.

bin_data2  Bin data into longer windows

Description
Bin minute level data into different time resolutions

Usage
bin_data2(x = x, window = 1, method = c("average", "sum"))

Arguments
x  vector of activity data.
window  window size used to bin the original 1440 dimensional data into. Window size should be an integer factor of 1440
method  character of "sum" or "average", function used to bin the data

Value
a vector of binned data

create.postGGIR  Create a template shell script of postGGIR

Description
Create a template shell script of postGGIR, named as STUDYNAME_part0.maincall.R.

Usage
create.postGGIR()

Value
The function will create a template shell script of postGGIR in the current directory, names as STUDYNAME_part0.maincall.R
Data imputation for the cleaned data with annotation

**Description**

Data imputation for the merged ENMO data with annotation. The missing values were imputated by the average ENMO over all the valid days for each subject.

**Usage**

```r
data.imputation(workdir, csvInput)
```

**Arguments**

- `workdir` character: Directory where the output needs to be stored. Note that this directory must exist.
- `csvInput` character: File name with or without directory for sample information in CSV format. The ENMO data will be read through `read.csv(csvInput, header=1)` command, and the missing values were imputated by the average ENMO over all the valid days for each subject at each time point. In this package, `csvInput = flag_All_studyname_ENMO.data.Xs.csv`.

**Value**

Files were written to the specified sub-directory, named as `impu.flag_All_studyname_ENMO.data.Xs.csv`, which Xs is the epoch size to which acceleration was averaged (seconds) in GGIR output. This excel file includes the following columns,

- **filename** accelerometer file name
- **Date** date recorded from the GGIR part2.summary file
- **id** IDs recorded from the GGIR part2.summary file
- **calendar_date** date in the format of yyyy-mm-dd
- **N.valid.hours** number of hours with valid data recorded from the part2_daysummary.csv file in the GGIR output
- **N.hours** number of hours of measurement recorded from the part2_daysummary.csv file in the GGIR output
- **weekday** day of the week-Day of the week
- **measurementday** day of measurement-Day number relative to start of the measurement
- **newID** new IDs defined as the user-defined function of `filename2id()`, e.g. substrings of the filename
- **Nmiss_c9_c31** number of NAs from the 9th to 31th column in the part2_daysummary.csv file in the GGIR output
DataShrink

Annotating the merged data for all accelerometer files in the GGIR output

**Description**

Annotating the merged ENMO/ANGLEZ data by adding some descriptive variables such as number of valid days and missing pattern.

**Usage**

DataShrink(
    studyname,
    outputdir,
    workdir,
    QCdays.alpha = 7,
    QChours.alpha = 16,
    summaryFN = "../summary/part24daysummary.info.csv",
    epochIn = 5,
    epochOut = 60,
    useIDs.FN = NULL,
    RemoveDaySleeper = FALSE,
    trace = FALSE,
    Step = 1
)
**Arguments**

- **studyname** character Specify the study name that used in the output file names
- **outputdir** character Directory where the GGIR output was stored.
- **workdir** character Directory where the output needs to be stored. Note that this directory must exist.
- **QCdays.alpha** number Minimum required number of valid days in subject specific analysis as a quality control step in part2. Default is 7 days.
- **QChours.alpha** number Minimum required number of valid hours in day specific analysis as a quality control step in part2. Default is 16 hours.
- **summaryFN** character Filename with or without directory for sample information in CSV format, which includes summary description of each accelerometer file. Some description will be extracted and merged into the ENMO/ANGLEZ data.
- **epochIn** number Epoch size to which acceleration was averaged (seconds) in GGIR output. Default is 5 seconds.
- **epochOut** number Epoch size to which acceleration was averaged (seconds) in part1. Default is 60 seconds.
- **useIDs.FN** character Filename with or without directory for sample information in CSV format, which includes "filename" and "duplicate" in the headlines at least. If duplicate="remove", the accelerometer files will not be used in the data analysis of part 5-7. Default is NULL, which makes all accelerometer files will be used in part 5-7.
- **RemoveDaySleeper** logical Specify if the daysleeper nights are removed from the calculation of number of valid days for each subject. Default is FALSE.
- **trace** logical Specify if the intermediate results is printed when the function was executed. Default is FALSE.
- **Step** number Specify which of the variable need to be cleaned. For example, Step = 1 for the "anglez" variable, and Step = 2 for the "enmo" variable.

**Value**

Files were written to the specified sub-directory, named as flag_ALL_studyname_ENMO.data.Xs.csv and flag_ALL_studyname_ANGLEZ.data.Xs.csv, which Xs is the epoch size to which acceleration was averaged (seconds) in GGIR output. This excel file includes the following columns,

- **filename** accelerometer file name
- **Date** date recored from the GGIR part2.summary file
- **id** IDs recored from the GGIR part2.summary file
- **calender_date** date in the format of yyyy-mm-dd
- **N.valid.hours** number of hours with valid data recored from the part2_daysummary.csv file in the GGIR output
- **N.hours** number of hours of measurement recored from the part2_daysummary.csv file in the GGIR output
weekday
measurementday
day of the week-Day of the week

day of measurement-Day number relative to start of the measurement

newID
new IDs defined as the user-defined function of filename2id(), e.g. substrings of
the filename

Nmiss_c9_c31
number of NAs from the 9th to 31th column in the part2_daysummary.csv file
in the GGIR output

missing
"M" indicates missing for an invalid day, and "C" indicates completeness for a
valid day

Ndays
number of days of measurement

ith_day
rank of the measurementday, for example, the value is 1,2,3,4,-3,-2,-1 for mea-
surementday = 1,...,7

Nmiss
number of missing (invalid) days

Nnonmiss
number of non-missing (valid) days

misspattern
indicators of missing/nonmissing for all measurement days at the subject level

RowNonWear
number of columns in the non-wearing matrix

NonWearMin
number of minutes of non-wearing

Nvalid.day
number of valid days with/without removing daysleeper nights; It is equal to
Nnonmiss when RemoveDaySleeper=FALSE.

daysleeper
If 0 then the person is a nightsleeper (sleep period did not overlap with noon) if
value=1 then the person is a daysleeper (sleep period did overlap with noon).

remove16h7day
indicator of a key quality control output. If remove16h7day=1, the day need to
be removed. If remove16h7day=0, the day need to be kept.

duplicate
If duplicate="remove", the accelerometer files will not be used in the data anal-
ysis of part5-7.

---

**fragmentation2**

**Fragmentation Metrics**

**Description**

Fragmentation methods to study the transition between two states, e.g. sedentary v.s. active.

**Usage**

```r
fragmentation2(
  x,
  w,
  thresh,
  bout.length = 1,
  metrics = c("mean_bout", "TP", "Gini", "power", "hazard", "all")
)
```
Arguments

- **x**: integer vector of activity data.
- **w**: vector of wear flag data with same dimension as x.
- **thresh**: threshold to binarize the data.
- **bout.length**: minimum duration of defining an active bout; defaults to 1.
- **metrics**: What is the fragmentation metrics to extract. Can be "mean_bout","TP","Gini","power","hazard", or all the above metrics "all".

Details

Metrics include mean_bout (mean bout duration), TP (between states transition probability), Gini (gini index), power (alpha parameter for power law distribution) hazard (average hazard function)

Value

A list with elements

- **mean_r**: mean sedentary bout duration
- **mean_a**: mean active bout duration
- **SATP**: sedentary to active transition probability
- **ASTP**: active to sedentary transition probability
- **Gini_r**: Gini index for active bout
- **Gini_a**: Gini index for sedentary bout
- **h_r**: hazard function for sedentary bout
- **h_a**: hazard function for active bout
- **alpha_r**: power law parameter for sedentary bout
- **alpha_a**: power law parameter for active bout

References

Junrui Di, Andrew Leroux, Jacek Urbanek, Ravi Varadhan, Adam P. Spira, Jennifer Schrack, Vadim Zipunnikov. Patterns of sedentary and active time accumulation are associated with mortality in US adults: The NHANES study. bioRxiv 182337; doi: https://doi.org/10.1101/182337
Description

Fragmentation methods to study the transition between two states, e.g. sedentary v.s. active. This function is a whole dataset wrapper for `fragmentation`.

Usage

```r
fragmentation_long2(
  count.data,  # data.frame of dimension n*1442 containing the 1440 minutes of activity data for all n subject days. The first two columns have to be ID and Day. ID can be either character or numeric. Day has to be numeric indicating the sequence of days within each subject.
  weartime,    # data.frame with dimension of `count.data`. The first two columns have to be ID and Day. ID can be either character or numeric. Day has to be numeric indicating the sequence of days within each subject.
  thresh,      # threshold to define the two states.
  bout.length = 1,  # minimum duration of defining an active bout; defaults to 1.
  metrics = c("mean_bout", "TP", "Gini", "power", "hazard", "all"),  # What is the fragmentation metrics to extract. Can be "mean_bout","TP","Gini","power","hazard", or all the above metrics "all".
  by = c("day", "subject")  # Determine whether fragmentation is calculated by day or by subjects (i.e. aggregate bouts across days). by-subject is recommended to gain more power.
)
```

Arguments

- `count.data`: data.frame of dimension n*1442 containing the 1440 minutes of activity data for all n subject days. The first two columns have to be ID and Day. ID can be either character or numeric. Day has to be numeric indicating the sequence of days within each subject.
- `weartime`: data.frame with dimension of `count.data`. The first two columns have to be ID and Day. ID can be either character or numeric. Day has to be numeric indicating the sequence of days within each subject.
- `thresh`: threshold to define the two states.
- `bout.length`: minimum duration of defining an active bout; defaults to 1.
- `metrics`: What is the fragmentation metrics to extract. Can be "mean_bout","TP","Gini","power","hazard", or all the above metrics "all".
- `by`: Determine whether fragmentation is calculated by day or by subjects (i.e. aggregate bouts across days). by-subject is recommended to gain more power.

Details

Metrics include mean_bout (mean bout duration), TP (between states transition probability), Gini (gini index), power (alpha parameter for power law distribution) hazard (average hazard function).

Value

A dataframe with some of the following columns

- **ID**: identifier of the person
- **Day**: numeric vector indicating the sequence of days within each subject.
mean_r  mean sedentary bout duration
mean_a  mean active bout duration
SATP    sedentary to active transition probability
ASTP    active to sedentary transition probability
Gini_r  Gini index for active bout
Gini_a  Gini index for sedentary bout
h_r     hazard function for sedentary bout
h_a     hazard function for active bout
alpha_r power law parameter for sedentary bout
alpha_a power law parameter for active bout

---

**ggir.datatransform**

Transform the data and merge all accelerometer files in the GGIR output

---

**Description**

An accelerometer file was transformed into wide data matrix, in which the rows represent available
days and the columns including all timestamps for 24 hours. Further, the wide data was merged
together.

**Usage**

```r
ggir.datatransform(
  outputdir,
  subdir,
  studyname,
  numericID = FALSE,
  sortByid = "newID",
  f0 = 1,
  f1 = 1e+06,
  epochIn = 5,
  epochOut = 600,
  mergeVar = 1
)
```

**Arguments**

- `outputdir` character Directory where the GGIR output was stored.
- `subdir` character Sub-directory where the summary output was stored under the current
directory. Default is "data".
- `studyname` character Specify the study name that used in the output file names
- `numericID` logical Specify if the ID is numeric when checking ID errors in part2. Default is FALSE.
sortByid character Specify the name of "ID" for each accelerometer file in the report of part5. The value could be "newID","id" and "filename". Default is "filename".

f0 number File index to start with (default = 1). Index refers to the filenames sorted in increasing order.

f1 number File index to finish with. Note that file ends with the minimum of f1 and the number of files available. Default = 1000000.

epochIn number Epoch size to which acceleration was averaged (seconds) in GGIR output. Default is 5 seconds.

epochOut number Epoch size to which acceleration was averaged (seconds) in part1. Default is 600 seconds.

mergeVar number Specify which of the varaible need to be processed and merged. For example, mergeVar = 1 makes that the M$metalong varialbes were read from R data on the directory of /meta/basic under GGIR output directory, which includes "nonwearscore","clippingscore","lightmean","lightpeak","temperaturemean" and "EN". When mergeVar = 2, makes that the "enmo" and "anglez" varialbes were read from csv data on the directory of /meta/csv under GGIR output directory.

Value

mergeVar = 1 Six files were written to the specified sub-directory as follows,

nonwearscore_studyname_f0_f1_Xs.xlsx
Data matrix of nonwearscore, where f0 and f1 are the file index to start and finish with and Xs is the epoch size to which acceleration was averaged (seconds) in GGIR output.

clippingscore_studyname_f0_f1_Xs.xlsx
Data matrix of clippingscore

lightmean_studyname_f0_f1_Xs.xlsx
Data matrix of lightmean

lightpeak_studyname_f0_f1_Xs.xlsx
Data matrix of lightpeak

temperaturemean_studyname_f0_f1_Xs.xlsx
Data matrix of temperaturemean

EN_studyname_f0_f1_Xs.xlsx
Data matrix of EN

mergeVar = 2 Two files were written to the specified sub-directory as follows,

studyname_ENMO.dataf0_f1_Xs.xlsx
Data matrix of ENMO, where f0 and f1 are the file index to start and finish with and Xs is the epoch size to which acceleration was averaged (seconds) in GGIR output.

studyname_ANGLEZ.dataf0_f1_Xs.xlsx
Data matrix of ANGLEZ
ggir.summary  

Description of all accelerometer files in the GGIR output

Description

Description of all accelerometer files in the GGIR output and this script was executed when mode=2 in the main call.

Usage

```r
ggit.summary(
  bindir = NULL,
  outputdir,
  studyname,
  numericID = FALSE,
  sortByid = "filename",
  subdir = "summary",
  part5FN = "WW_L50M125V500_T5A5",
  QChours.alpha = 16,
  filename2id = NULL,
  desiredtz = "US/Eastern",
  trace = FALSE
)
```

Arguments

- **bindir** character Directory where the accelerometer files are stored or list for the purpose of extracting the bin file list. Default=NULL when it is not available and therefore the bin file list is extracted from the /meta/basic folder of the GGIR output.
- **outputdir** character Directory where the GGIR output was stored.
- **studyname** character Specify the study name that used in the output file names
- **numericID** logical Specify if the ID is numeric when checking ID errors in part2. Default is FALSE.
- **sortByid** character Specify the name of "ID" for each accelerometer file in the report of part2. The value could be "newID","id" and "filename". Default is "filename".
- **subdir** character Sub-directory where the summary output was stored under the current directory. Default is "summary".
- **part5FN** character Specify which output is used in the GGIR part5 results. Default is "WW_L50M125V500_T5A5", which means that part5_daysummary_WW_L50M125V500_T5A5.csv and part5_personsummary_WW_L50M125V500_T5A5.csv are used in the analysis.
- **QChours.alpha** number Minimum required number of valid hours in day specific analysis as a quality control step in part2. Default is 16 hours.
**filename2id**  
*R* function User defined function for converting filename to sample IDs. Default is NULL.

**desiredtz**  
character desired timezone: see also http://en.wikipedia.org/wiki/Zone.tab. Used in `g.inspectfile()`. Default is "US/Eastern".

**trace**  
logical Specify if the intermediate results is printed when the function was executed. Default is FALSE.

---

**Value**

Four files were written to the specified sub-directory

**studyname_ggir_output_summary.xlsx**

This excel file includes 9 pages as follows,

- page 1 List of files in the GGIR output
- page 2 Summary of files
- page 3 List of duplicate IDs
- page 4 ID errors
- page 5 Number of valid days
- page 6 Table of number of valid/missing days
- page 7 Missing pattern
- page 8 Frequency of the missing pattern
- page 9 Description of all accelerometer files
- page 10 Inspects accelerometer file for key information, including: monitor brand, sample frequency and file header

**studyname_ggir_output_summary_plot.pdf**

Some plots such as the number of valid days, which were included in the part2a_studyname_postGGIR.report.html file as well.

**part24daysummary.info.csv**

Intermediate results for description of each accelerometer file.

**studyname_samples_remove_temp.csv**

Create studyname_samples_remove.csv file by filling "remove" in the "duplicate" column in this template. If duplicate="remove", the accelerometer files will not be used in the data analysis of part 5-7.

---

**IS2**

*Interdaily Statbility*

---

**Description**

This function calculate interdaily stability, a nonparametric metric of circadian rhythm.

**Usage**

```
IS2(x)
```
Arguments

- `x` data.frame of dimension n* by p, where p is the dimension of the data.

References


---

IS_long2

Interdaily Stability for the Whole Dataset

Description

This function calculates interdaily stability, a nonparametric metric of circadian rhythmicity. This function is a whole dataset wrapper for IS

Usage

`IS_long2(count.data, window = 1, method = c("average", "sum"))`

Arguments

- `count.data` data.frame of dimension n*(1440+2) containing the 1440 dimensional activity data for all n subject days. The first two columns have to be ID and Day. ID can be either character or numeric. Day has to be numeric indicating the sequency of days within each subject.
- `window` an integer indicating what is the window to bin the data before the function can be applied to the dataset. For details, see `bin_data`.
- `method` character of "sum" or "average", function used to bin the data

Value

A data.frame with the following 2 columns

- `ID` ID
- `IS` IS

References

Intradaily Variability

Description
This function calculates intradaily variability, a nonparametric metric representing fragmentation of circadian rhythmicity.

Usage
IV2(x)

Arguments
x vector of activity data

Value
IV

References

Intradaily Variability for the Whole Dataset

Description
This function calculates intradaily variability, a nonparametric metric representing fragmentation of circadian rhythmicity. This function is a whole dataset wrapper for IV.

Usage
IV_long2(count.data, window = 1, method = c("average", "sum"))

Arguments
count.data data.frame of dimension n * (1440+2) containing the 1440 dimensional activity data for all n subject days. The first two columns have to be ID and Day. ID can be either character or numeric. Day has to be numeric indicating the sequency of days within each subject.
window an integer indicating what is the window to bin the data before the function can be applied to the dataset. For details, see bin_data.
method character of "sum" or "average", function used to bin the data.
**Value**

A `data.frame` with the following 5 columns:

- ID
- Day
- IV

**References**


---

**jive.predict2**

*Modified jive.predict function (package: r.jive)*

**Description**

Replace SVDmiss by SVDmiss2 in the function.

**Usage**

```r
jive.predict2(data.new, jive.output)
```

**Arguments**

- `data.new` - A list of two or more linked data matrices on which to estimate JIVE scores. These matrices must have the same column dimension \( N \), which is assumed to be common.
- `jive.output` - An object of class "jive", with row dimensions matching those for `data.new`.

**Details**

See `jive.predict(package:r.jive)` for details.

**Value**

See `r.jive:: jive.predict` for details.
Description

This function is a whole dataset wrapper for Time

Usage

PAfun(count.data, weartime, PA.threshold = c(50, 100, 400))

Arguments

count.data data.frame of dimension n*1442 containing the 1440 minute activity data for all n subject days. The first two columns have to be ID and Day.
weartime data.frame with dimension of count.data. The first two columns have to be ID and Day.
PA.threshold threshold to calculate the time in minutes of sedentary, light, moderate and vigorous activity the data.

Value

A dataframe with some of the following columns

ID identifier of the person
Day indicator of which day of activity it is, can be a numeric vector of sequence 1,2,... or a string of date
time time of certain state

Description

This R script will generate plot for each variable and write description to a log file.

Usage

pheno.plot(
    inputFN,
    outFN = paste("plot_", inputFN, ".pdf", sep = ""),
    csv = TRUE,
    sep = " ",
    start = 3,
    read = TRUE,
logFN = NULL,
    track = TRUE
)

Arguments

inputFN    character Input file name or input data
outFN      character Output pdf file name for the plots
csv        logical Specify if input file is a CSV file. Default is TRUE.
sep        character Separator between columns. Default is space. If csv=TRUE, this will not be used.
start      number The location of the first phenotype variable starts in the input file.
read       logical Specify if inputFN is a file name or a data. Default is TRUE when inputFN is a file name.
logFN      character File name of the log file. Default is NULL, while logFN=paste(inputFN,".log",sep="") in the function.
track      logical Specify if the intermediate results is printed when the function was executed. Default is TRUE.

Value

Files were written to the current directory. One is .pdf file for plots and the other is .log file for variable description.

---

RA2  Relative Amplitude

Description

This function calculate relative amplitude, a nonparametric metric representting fragmentation of circadian rhythmicity

Usage

RA2(x, window = 1, method = c("average", "sum"))

Arguments

x          vector vector of activity data
window     since the calculation of M10 and L5 depends on the dimension of data, we need to include window size as an argument.
method     character of "sum" or "average", function used to bin the data

Value

RA
References


RA_long2

Relative Amplitude for the Whole Datset

Description

This function calculates relative amplitude, a nonparametric metric of circadian rhythmicity. This function is a whole dataset wrapper for RA.

Usage

RA_long2(count.data, window = 1, method = c("average", "sum"))

Arguments

  count.data data.frame of dimension n * (p+2) containing the p dimensional activity data for all n subject days. The first two columns have to be ID and Day. ID can be either character or numeric. Day has to be numeric indicating the sequency of days within each subject.

  window since the calculation of M10 and L5 depends on the dimension of data, we need to include window size as an argument. This function is a whole dataset wrapper for RA.

  method character of "sum" or "average", function used to bin the data

Value

A data.frame with the following 3 columns

<table>
<thead>
<tr>
<th>ID</th>
<th>Day</th>
<th>RA</th>
</tr>
</thead>
</table>
SVDmiss2

Modified SVDmiss function (package SpatioTemporal)

Description

Modify ncomp = min(ncol(X),nrow(X),ncomp) for the matrix with nrow(X)<ncol(X)

Usage

SVDmiss2(X, niter = 200, ncomp = dim(X)[2], conv.reldiff = 0.001)

Arguments

X X Data matrix, with missing values marked by 'NA'.
niter Maximum number of iterations to run before exiting, 'Inf' will run until the 'conv.reldiff' criteria is met.
ncomp Number of SVD components to use in the reconstruction (>0).
conv.reldiff Assume the iterative procedure has converged when the relative difference between two consecutive iterations is less than 'conv.reldiff'.

Details

See SVDmiss(package:SpatioTemporal) for details.

Value

See SpatioTemporal:: SVDmiss for details

Time2

Time of A Certain activity State

Description

Calculate the total time of being in certain state, e.g. sedentary, active, MVPA, etc.

Usage

Time2(x, w, thresh, smallerthan = TRUE, bout.length = 1)

Arguments

x vector of activity data.
w vector of wear flag data with same dimension as x.
thresh threshold to binarize the data.
smallerthan Find a state that is smaller than a threshold, or greater than or equal to.
bout.length minimum duration of defining an active bout; defaults to 1.
Time

Value

Time

Timne Metrics for Whole Dataset

Description

This function is a whole dataset wrapper for Time

Usage

Time_long2(count.data, weartime, thresh, smallerthan = TRUE, bout.length = 1)

Arguments

count.data  data.frame of dimension n*1442 containing the 1440 minute activity data for all n subject days. The first two columns have to be ID and Day.

weartime  data.frame with dimension of count.data. The first two columns have to be ID and Day.

thresh  threshold to binarize the data.

smallerthan  Find a state that is smaller than a threshold, or greater than or equal to.

bout.length  minimum duration of defining an active bout; defaults to 1.

Value

A dataframe with some of the following columns

ID  identifier of the person

Day  indicator of which day of activity it is, can be a numeric vector of sequence 1,2,... or a string of date

time  time of certain state
**Tvol2** *Total Volumen of Activity for Whole Dataset*

**Description**

Calculate total volume of activity level, which includes TLAC (total log transformed activity counts), TAC (total activity counts).

**Usage**

```r
tvol2(count.data, weartime, logtransform = FALSE, log.multiplier = 9250)
```

**Arguments**

- `count.data` data.frame of dimension n*1442 containing the 1440 minute activity data for all n subject days. The first two columns have to be ID and Day.
- `weartime` data.frame with dimension of `count.data`. The first two columns have to be ID and Day.
- `logtransform` if TRUE, then calcualte TLAC. Or calculate TAC.
- `log.multiplier` number The coefficient used in the log transformation of the ENMO data, i.e. \( \log(\log(\text{multiplier} \times \text{ENMO} + 1)) \). Default is 9250.

**Details**

log transormation is defined as \( \log(\text{x}+1) \).

**Value**

A dataframe with some of the following columns

- **ID** identifier of the person
- **Day** indicator of which day of activity it is, can be a numeric vector of sequence 1,2,... or a string of date
- **TAC** total activity count
- **TLAC** total log activity count
Description

Determine during which time period, subject should wear the device. It is preferable that user provide their own wear/non wear flag which should has the same dimension as the activity data. This function provide wear/non wear flag based on time of day.

Usage

wear_flag(count.data, start = "05:00", end = "23:00")

Arguments

count.data  data.frame of dimension n*1442 containing the 1440 minute activity data for all n subject days. The first two columns have to be ID and Day.

start   start time, a string in the format of 24hr, e.g. "05:00"; defaults to "05:00".

date.end   end time, a string in the format of 24hr, e.g. "23:00"; defaults to "23:00"

Details

Fragmentation metrics are usually defined when subject is awake. The weartime provide time periods on which those features should be extracted. This can be also used as indication of wake/sleep.

Value

A data.frame with same dimension and column name as the count.data, with 0/1 as the elements representing wear, nonwear respectively.
Index

ActCosinor2, 2
ActCosinor_long2, 3
ActExtendCosinor2, 4
ActExtendCosinor_long2, 5
afterggir, 7

bin_data2, 9

create.postGGIR, 9

data.imputation, 10
DataShrink, 11

fragmentation2, 13
fragmentation_long2, 15

ggir.datatransform, 16
ggir.summary, 18

IS2, 19
IS_long2, 20
IV2, 21
IV_long2, 21

jive.predict2, 22

PAfun, 23
pheno.plot, 23

RA2, 24
RA_long2, 25

SVDmiss2, 26

Time2, 26
Time_long2, 27
Tvol2, 28

wear_flag, 29

30