Package 'nowcaster'

November 13, 2026

Title Statistical Models for Notification Delay Correction of Epidemiological Data

Version 1.0.1

Description Statistical models for correcting notification delays of epidemiological data. The package wraps statistical models to estimate not yet reported data using empirical delay distribution either from individual-level data or from the differences of aggregated time series of cases. For more details see Bastos et al. (2019) <doi:10.1002/sim.8303>.

License GPL (>= 3)

Encoding UTF-8

RoxygenNote 7.3.2

URL https://covid19br.github.io/nowcaster/

BugReports https://github.com/covid19br/nowcaster/issues

Suggests INLA (>= 22.5.3), ggplot2, knitr, rmarkdown, vroom, parallel, lubridate, lme4, sn, stringr

Depends R (>= 4.1.0)

LazyData true

Additional_repositories https://inla.r-inla-download.org/R/stable/

Imports dplyr, tidyr, tibble, mgcv

Config/testthat/edition 3

NeedsCompilation no

Author Rafael Lopes [aut] (ORCID: <https://orcid.org/0000-0002-9416-6145>), Tatiana Portella [aut] (ORCID: <https://orcid.org/0000-0001-8364-1978>), Theo Economou [aut] (ORCID: <https://orcid.org/0000-0001-8697-1518>), Marcelo Gomes [aut] (ORCID: <https://orcid.org/0000-0003-4693-5402>), Leo Bastos [aut, cre] (ORCID: <https://orcid.org/0000-0002-1406-0122>)

Maintainer Leo Bastos <leonardo.bastos@fiocruz.br>

Repository CRAN

Date/Publication 2026-11-13 16:10:15 UTC

Contents

data.w 2
data.w_no_age 3
nowcasting.summary 4
nowcasting_age 4
nowcasting_age_mgcv 5
nowcasting_diff_inla 6
nowcasting_inla 7
nowcasting_mgcv 9
nowcasting_no_age 11
nowcasting_no_age_mgcv 12
slope.estimate.quant 12
sragBH 13

Index **15**

`data.w` *data.w*

Description

Function to put the data into the format proper to the nowcasting estimation, has to use the maximum of data present on the last week parsed to the estimation. Makes the account of maximum days per week to be used.

Usage

```
data.w(
  dataset,
  trim.data,
  bins_age = c("SI-PNI", "10 years", "5 years", bins_age),
  date_onset,
  date_report,
  age_col,
  use.epiweek = FALSE,
  K = 0,
  silent = FALSE
)
```

Arguments

`dataset`	dataset to be formatted as data by week
`trim.data`	How much to trim of the data?
`bins_age`	Bins of age to cu the data, parsing from nowcasting_inla
`date_onset`	Column of dates of onset of the events, normally date of onset of first symptoms of cases

date_report	Column of dates of report of the event, normally date of digitation of the notification of cases
age_col	Age column to be where to cut the data into age classes
use.epiweek	If TRUE, it uses the CDC epiweek definition where the week starts on Sunday, if FALSE it the week ends at the weekday of the last record date.
K	How much weeks to forecast ahead? The default for K is 0, no forecasting ahead
silent	Deprecated. To be removed.

Value

Data in weeks format, with the maximum dates for the last week used

---

`data.w_no_age` *data.w_no_age*

Description

Function to put the data into the format proper to the nowcasting estimation, has to use the maximum of data present on the last week parsed to the estimation. Makes the account of maximum days per week to be used. With no age structure.

Usage

```
data.w_no_age(
  dataset,
  trim.data,
  date_onset,
  date_report,
  use.epiweek = FALSE,
  K = 0,
  silent = FALSE
)
```

Arguments

dataset	dataset to be formatted as data by week
trim.data	How much to trim of the data?
date_onset	Column of dates of onset of the events, normally date of onset of first symptoms of cases
date_report	Column of dates of report of the event, normally date of digitation of the notification of cases
use.epiweek	If TRUE, it uses the CDC epiweek definition where the week starts on Sunday, if FALSE it the week ends at the weekday of the last record date.
K	How much weeks to forecast ahead? **Default** K is 0, no forecasting ahead
silent	DEPRECATED to be removed.

Value

Data in weeks format, with the maximum dates for the last week used

---

`nowcasting.summary` *nowcasting.summary*

Description

Function to summarize the output from the posteriors generated from any of the of the workhorse functions, nowcasting_no_age and nowcasting_age

Usage

```
nowcasting.summary(trajectory, age = FALSE)
```

Arguments

`trajectory`	Trajectories to be summarized, trajectory should have sample, Time, dt_event and Y columns.
`age`	Is by age data? Default FALSE, If its TRUE it will have columns for age class, fx_etaria e fx_etaria.num

Value

A list with 2 elements summarized and grouped by, with the 'Median', 'LS' upper limit, 95% CI 'LI' lower limit, 95% CI 'LSb' upper limit, 50% CI 'LIb' lower limit, 50% CI

`nowcasting_age` *nowcasting_age*

Description

Run INLA model on structured data by age-class data has to be in the format of delay-triangle

Usage

```
nowcasting_age(
  dataset,
  zero_inflated = FALSE,
  timeREmodel = "rw2",
  INLAoutput = FALSE,
  INLAoutputOnly = FALSE,
  WAIC = FALSE,
  DIC = FALSE
)
```

Arguments

dataset	data pre formatted in to age classes and delays by week for each cases, delay triangle format
zero_inflated	Experimental! It deals with zero-inflated data by using a zeroinflatednbinomial2 model.
timeREmodel	Latent model for time random effects. The default is a second-order random walk model.
INLAoutput	return the INLA output. Default is FALSE.
INLAoutputOnly	return the only the INLA output. Default is FALSE.
WAIC	return the WAIC. Default is FALSE.
DIC	return the DIC. Default is FALSE.

Value

Trajectories from the inner 'INLA' model

`nowcasting_age_mgcv` *nowcasting_age_mgcv*

Description

Run INLA model on structured data by age-class data has to be in the format of delay-triangle

Usage

```
nowcasting_age_mgcv(dataset, method = "fs")
```

Arguments

dataset	data pre formatted in to age classes and delays by week for each cases, delay triangle format
method	method used grouped gam, "by" where the smooth function will be s(., by= age-group) or "fs" (factor smooth) where the smooth function uses s(., age_group, ,bs="fs")

Value

Trajectories from the inner 'INLA' model

nowcasting_diff_inla *nowcasting_diff_inla*

Description

Function to estimate the number of events that have already occurred but have not yet been reported, in situations where the notification date is unavailable. The function calculates delayed notifications by comparing case counts for the same event date across successive database versions. nowcasting_diff_inla, calculates reporting delays based on differences between database versions and fits a statistical distribution to the empirical delay.

Usage

```
nowcasting_diff_inla(
  dataset,
  Dmax = 10,
  wdw = 10,
  date_start,
  date_release,
  cases,
  age_col,
  silent = FALSE,
  K = 0,
  trajectories = FALSE,
  zero_inflated = FALSE,
  timeREmodel = "rw2",
  INLAoutput = FALSE,
  INLAoutputOnly = FALSE,
  WAIC = FALSE,
  DIC = FALSE,
  ...
)
```

Arguments

dataset	Dataset with at least three columns: event start date, database release date, and number of cases. Optionally, a fourth column with age strata can be included. Data must must be agregated by week
Dmax	(in weeks) Window of dates the estimation will act, i.e., till how many past weeks the nowcasting will estimate. Default is 10 weeks.
wdw	(in weeks) Until which maximum amount of weeks the Nowcasting will use to the estimation. Default is 10 weeks.
date_start	Column containing the dates when the events ocurred. Data must must be agregated by week
date_release	Column containing the dates when the databases were released. Data must must be agregated by week

cases	Number of reported cases.
age_col	Column for ages
silent	Deprecated. Should be the warnings turned off? . The default is TRUE.
K	(in weeks) How much weeks to forecast ahead? . The default is K = 0, no forecasting ahead
trajectories	Returns the trajectories estimated from the inner 'INLA' model . The default is FALSE.
zero_inflated	Experimental! In non-structured models, fit a model that deals with zero-inflated data. The default is FALSE. If the age_col is not missing this flag is ignored.
timeREmodel	Latent model for time random effects. . The default is a second-order random walk model.
INLAoutput	return the INLA output. Default is FALSE.
INLAoutputOnly	return the only the INLA output. Default is FALSE.
WAIC	return the WAIC. The default is FALSE.
DIC	return the DIC.The default is FALSE
...	list parameters to other functions

Value

a list of 2 elements, each element with a data.frame with nowcasting estimation, 'Total', 'data' with the time-series out of wdw . If 'age_col' is parsed, add a third element with by age estimation 'age' . If 'trajectories' = TRUE, add a forth element with the returned trajectories from 'inla'. If 'INLAoutput' = TRUE, the INLA output is returned as a list object named 'output'. If 'INLAoutputOnly' = TRUE, just the INLA output is returned in a list object named 'output'. If 'WAIC' = TRUE or 'DIC' = TRUE, then 'INLAoutput' is forced to be TRUE returning the INLA output and a list object named waic or dic are also returned.

`nowcasting_inla` *nowcasting_inla*

Description

function to estimate amount of events already started by not yet notified. The main use is to estimate how many cases in a outbreak has already started their onset date of symptons but has not yet notified. nowcasting_inla, fits a statistical distribution to the empirical distribution of time of delay between onset date and report date.

Usage

```
nowcasting_inla(
  dataset,
  bins_age = "SI-PNI",
  trim.data = 0,
  Dmax = 15,
```

```
  wdw = 30,
  use.epiweek = TRUE,
  age_col,
  date_onset,
  date_report,
  data.by.week = FALSE,
  return.age = NULL,
  silent = FALSE,
  K = 0,
  trajectories = FALSE,
  zero_inflated = FALSE,
  timeREmodel = "rw2",
  INLAoutput = FALSE,
  INLAoutputOnly = FALSE,
  WAIC = FALSE,
  DIC = FALSE,
  ...
)
```

Arguments

dataset	Dataset with at least 2 columns, date of onset, date of report. It can be a dataset with 3 columns, two dates columns as before said and a another one being an stratum column, in which data will stratified, usually being age.
bins_age	Age bins to do the nowcasting, it receive a vector of age bins, or options between, "SI-PNI", "10 years", "5 years". The default is "SI-PNI".
trim.data	(in weeks) Date to be trimmed out from the data base, in days. Default is 0 days.
Dmax	(in weeks) Window of dates the estimation will act, i.e., till how many past weeks the nowcasting will estimate. Default is 15 weeks.
wdw	(in weeks) Until which maximum amount of weeks the Nowcasting will use to the estimation. Default is 30 weeks.
use.epiweek	If TRUE, it uses the CDC epiweek definition where the week starts on Sunday, if FALSE it the week ends at the weekday of the last record date. Default is TRUE
age_col	Column for ages
date_onset	Column of dates of onset of the events, normally date of onset of first symptoms of cases
date_report	Column of dates of report of the event, normally date of digitation of the notification of cases
data.by.week	If it has to be returned the whole time-series data. Default is FALSE.
return.age	Deprecated. If the estimate by Age should be returned. Default is TRUE.
silent	Deprecated. Should be the warnings turned off? . The default is FALSE
K	(in weeks) How much weeks to forecast ahead? . The default is K = 0, no forecasting ahead

trajectories	Returns the trajectories estimated from the inner 'INLA' model . The default is FALSE.
zero_inflated	Experimental! In non-structured models, fit a model that deals with zero-inflated data. The default is FALSE. If the age_col is not missing this flag is ignored.
timeREmodel	Latent model for time random effects. . The default is a second-order random walk model.
INLAoutput	return the INLA output. Default is FALSE.
INLAoutputOnly	return the only the INLA output. Default is FALSE.
WAIC	return the WAIC. The default is FALSE.
DIC	return the DIC.The default is FALSE.
...	list parameters to other functions

Value

a list of 2 elements, each element with a data.frame with nowcasting estimation, 'Total', 'data' with the time-series out of wdw . If 'age_col' is parsed, add a third element with by age estimation 'age' . If 'trajectories' = TRUE, add a forth element with the returned trajectories from 'inla'. If 'INLAoutput' = TRUE, the INLA output is returned as a list object named 'output'. If 'INLAoutputOnly' = TRUE, just the INLA output is returned in a list object named 'output'. If 'WAIC' = TRUE or 'DIC' = TRUE, then 'INLAoutput' is forced to be TRUE returning the INLA output and a list object named waic or dic are also returned.

Examples

```
## Not run:
# Loading Belo Horizonte SARI dataset
data(sragBH)
output <- nowcasting_inla(dataset = sragBH,
                date_onset = DT_SIN_PRI,
                date_report = DT_DIGITA,
                silent = T)

## End(Not run)
```

---

nowcasting_mgcv *nowcasting_mgcv*

Description

function to estimate amount of events already started by not yet notified. The main use is to estimate how many cases in a outbreak has already started their onset date of symptons but has not yet notified. nowcasting_mgcv, fits a statistical distribution to the empirical distribution of time of delay between onset date and report date.

Usage

```
nowcasting_mgcv(
  dataset,
  bins_age = "SI-PNI",
  trim.data = 0,
  Dmax = 15,
  wdw = 30,
  use.epiweek = TRUE,
  age_col,
  date_onset,
  date_report,
  data.by.week = FALSE,
  return.age = NULL,
  silent = FALSE,
  method.group.gam = "fs",
  K = 0,
  trajectories = FALSE,
  ...
)
```

Arguments

dataset	Dataset with at least 2 columns, date of onset, date of report. It can be a dataset with 3 columns, two dates columns as before said and a another one being an stratum column, in which data will stratified, usually being age.
bins_age	Age bins to do the nowcasting, it receive a vector of age bins, or options between, "SI-PNI", "10 years", "5 years". The default is "SI-PNI".
trim.data	(in weeks) Date to be trimmed out from the data base, in days. Default is 0 days.
Dmax	(in weeks) Window of dates the estimation will act, i.e., till how many past weeks the nowcasting will estimate. Default is 15 weeks.
wdw	(in weeks) Until which maximum amount of weeks the Nowcasting will use to the estimation. Default is 30 weeks.
use.epiweek	If TRUE, it uses the CDC epiweek definition where the week starts on Sunday, if FALSE it the week ends at the weekday of the last record date. Default is TRUE
age_col	Column for ages
date_onset	Column of dates of onset of the events, normally date of onset of first symptoms of cases
date_report	Column of dates of report of the event, normally date of digitation of the notification of cases
data.by.week	If it has to be returned the whole time-series data. Default is FALSE.
return.age	Deprecated. If the estimate by Age should be returned. Default is TRUE.
silent	Deprecated. Should be the warnings turned off? . The default is FALSE
method.group.gam	Method for aggregated smooth function in gam. Default is "fs" factor smooth, another alternative is "by".

K	(in weeks) How much weeks to forecast ahead? . The default is K = 0, no forecasting ahead
trajectories	Returns the predictive trajectories. The default is FALSE.
...	list parameters to other functions

Value

a list of 2 elements, each element with a data.frame with nowcasting estimation, 'Total', 'data' with the time-series out of wdw . If 'age_col' is parsed, add a third element with by age estimation 'age' . If 'trajectories' = TRUE, add a forth element with the returned trajectories from 'inla'.

Examples

```
# Loading Belo Horizonte SARI dataset
data(sragBH)
output <- nowcasting_mgcv(dataset = sragBH,
                date_onset = DT_SIN_PRI,
                date_report = DT_DIGITA)
```

`nowcasting_no_age` *nowcasting_no_age*

Description

Run INLA model on non-structured data, data has to be in the format of delay-triangle

Usage

```
nowcasting_no_age(
  dataset,
  zero_inflated = FALSE,
  timeREmodel = "rw2",
  INLAoutput = FALSE,
  INLAoutputOnly = FALSE,
  WAIC = FALSE,
  DIC = FALSE
)
```

Arguments

dataset	data pre formatted in to age classes and delays by week for each cases, delay triangle format
zero_inflated	zero-inflated model. Default is FALSE.
timeREmodel	Latent model for time random effects. The default is a second-order random walk model.

INLAoutput	return the INLA output. Default is FALSE.
INLAoutputOnly	return the only the INLA output. Default is FALSE.
WAIC	return the WAIC. Default is FALSE.
DIC	return the DIC. Default is FALSE.

Value

Trajectories from the inner 'INLA' model

`nowcasting_no_age_mgcv` *nowcasting_no_age_mgcv*

Description

Run INLA model on non-structured data, data has to be in the format of delay-triangle

Usage

```
nowcasting_no_age_mgcv(dataset)
```

Arguments

dataset	data pre formatted in to age classes and delays by week for each cases, delay triangle format

Value

Trajectories from the Negative Binomial GAM

`slope.estimate.quant` *Slope.estimate.quant*

Description

Fits a linear model to trajectories returned from 'nowcasting_inla()' within a given window. The default is 3 weeks. If 'end.week' is missing uses the maximum date in 'trajectories'.

Usage

```
slope.estimate.quant(end.week, trajectories, window = 3)
```

Arguments

end.week	(in weeks) The end of the week wanted to the slope estimate. Default: max. date in 'trajectories'.
trajectories	Data.frame with the predicted or nowcasted estimate
window	(in weeks) Window of how much time will be used to calculate the slope estimate. The default is 3 weeks.

Value

The numerical value of the slope of the estimate

Examples

```
# Loading Belo Horizonte SARI dataset
data(sragBH)
now <- nowcasting_mgcv(dataset = sragBH,
                date_onset = DT_SIN_PRI,
                date_report = DT_DIGITA,
                trajectories = TRUE,
                silent = TRUE)
slope.estimate.quant(trajectories = now$trajectories)
```

---

sragBH *SARI data from Belo Horizonte*

Description

An anonymized dataset containing cases of severe acute respiratory illness (SARI) from a Brazilian municipality, Belo Horizonte, with symptoms onset varying from 2019-12-29 to 2022-03-27.

Usage

```
sragBH
```

Format

A data frame with 65404 rows and 7 variables:

DT_SIN_PRI Date of onset symptoms.

DT_DIGITA Date of recording.

CLASSI_FIN Final classification of the case. (Not used for nowcasting)

EVOLUCAO Case evolution. (Not used for nowcasting)

CO_MUN_RES IBGE municipality code. Belo Horizonte is 310620.

Idade Age in years.

fx_etaria Age brackets. ...

Source

https://opendatasus.saude.gov.br/

Index

∗ **datasets**
sragBH, 13

data.w, 2
data.w_no_age, 3

nowcasting.summary, 4
nowcasting_age, *4*, 4
nowcasting_age_mgcv, 5
nowcasting_diff_inla, 6
nowcasting_inla, 7
nowcasting_mgcv, 9
nowcasting_no_age, *4*, 11
nowcasting_no_age_mgcv, 12

slope.estimate.quant, 12
sragBH, 13