Package ‘EnvExpInd’

October 23, 2020

Type Package

Title Environmental Exposure on the Individual Level

Imports gstat, RCurl, dplyr, stringi, sp, maptools, zoo

Version 0.1.0

Depends R(>= 3.5.0)

Description Tools for the assessment of the environmental exposure. The package provides three methods (nearest monitoring site, inverse distance weighted as described in Li Wu (2017) <doi:10.1016/j.envint.2016.11.013>, and ordinary kriging) to calculate the environmental exposure (e.g. air pollution) on the individual level.

URL https://github.com/Spatial-R/EnvExpInd

License GPL-3

Encoding UTF-8

LazyData true

RoxygenNote 7.1.0

Suggests knitr, rmarkdown

VignetteBuilder knitr

NeedsCompilation no

Author Bing Zhang [aut, cre]

Maintainer Bing Zhang <zhangb269@mail2.sysu.edu.cn>

Repository CRAN

Date/Publication 2020-10-23 15:50:02 UTC

R topics documented:

  exposure_estimate_idw ............................................. 2
  exposure_estimate_krige ............................................ 4
  expoure_estimate_simple ........................................... 6
  get_lation_china .................................................. 7
  get_refrence_id_simple ............................................. 8
  individual_data ................................................... 10
exposure_estimate_idw

pollutant_data ......................................................... 10
site_data ............................................................... 11
timeseries_imput ....................................................... 11

Index 13

exposure_estimate_idw  Estimate the pollutant exposure using the inverse distance weighting method

Description

Used the pollutant concentration in the individual location as the reference point to estimate the environmental exposure. The pollutant concentration at the reference point was calculated based on the inverse distance weighting method.

Usage

exposure_estimate_idw(
  individual_data,
  individual_id,
  exposure_date,
  individual_lat,
  individual_lon,
  pollutant_data,
  pollutant_date = "date",
  pollutant_site_lat,
  pollutant_site_lon,
  pollutant_name = c("pm10", "so2"),
  estimate_interval = c(0:30)
)

Arguments

individual_data  data.frame, contains the reference id, individual_id and exposure_date
individual_id  character, variable name in individual_data, represents the unique id for each individual
exposure_date  character, variable name in individual_data, which represents the start date to estimate the environment exposure
individual_lat  character, variable name in individual_data, represents the latitude information of each individual
individual_lon  character, variable name in individual_data, represents the longitude information of each individual
pollutant_data  data.frame, contains the pollutant and site information. One column represents the site information and other columns represent the concentration of pollutants
exposure_estimate_idw

pollutant_date character, variable name represents the date information for the air pollutant dataset

pollutant_site_lat character, variable name in pollutant_data, includes the latitude information of each monitoring site

pollutant_site_lon character, variable name in pollutant_data, includes the longitude information of each monitoring site

pollutant_name vector, pollutant name in the pollutant_data, which represent the name of the target pollutants to be estimated

estimate_interval continue numeric vector, the estimation period, for example: 0:30, for each individual we estimate the environment exposure ranging from the exposure_date to exposure_date + 30 days

Value

A list. For each element in the list, there is a dataframe with the first column representing the individual id, the remaining columns represent the exposure estimation in different time points.

Author(s)

Bing Zhang, https://github.com/Spatial-R/EnvExpInd

Examples

library(EnvExpInd)
individual_data$date <- as.Date(individual_data$date)
pollutant_data$date <- as.Date(pollutant_data$date)
pollutant_data_full <- timeseries_imput(data = pollutant_data,date_var = "date",
  site_var = "site.name",imput_col = 3:8)
pollutant_data_tem <- merge(pollutant_data_full,site_data,by.x = "site.name",by.y = "site")

exposure_estimate_idw(
  individual_data = individual_data,
  individual_id = "id",
  exposure_date ="date",
  individual_lat ="lat",
  individual_lon ="lon",
  pollutant_data = pollutant_data_tem,
  pollutant_date = "date",
  pollutant_site_lat = "lat",
  pollutant_site_lon = "lon",
  pollutant_name = c("PM10","PM2.5"),
  estimate_interval = c(0:10))
exposure_estimate_krige

Assess the environmental exposure using the krige method

Description

Based on the krige method, the pollutant exposure in each individual location was estimated and then assess the total pollutant exposure through the estimate_interval

Usage

```r
exposure_estimate_krige(
  individual_data,
  individual_id,
  exposure_date,
  individual_lat,
  individual_lon,
  pollutant_data,
  pollutant_date = "date",
  pollutant_site_lat,
  pollutant_site_lon,
  pollutant_name = c("pm10", "so2"),
  estimate_interval = c(0:30),
  krige_model,
  nmax = 7,
  krige_method = "med"
)
```

Arguments

- `individual_data` data.frame, contains the reference id, individual_id and exposure_date
- `individual_id` character, variable name in individual_data, represents the unique id for each individual
- `exposure_date` character, variable name in individual_data, which represents the start date to estimate the environment exposure
- `individual_lat` character, variable name in individual_data, represents the latitude information of each individual
- `individual_lon` character, variable name in individual_data, represents the longitude information of each individual
- `pollutant_data` data.frame, contains the pollutant and site information. One column represents the site information and other columns represent the concentration of pollutants
- `pollutant_date` character, variable name represents the date information for the air pollutant dataset
- `pollutant_site_lat`, `pollutant_site_lon` character, variable name represents the location information for the air pollutant dataset
- `pollutant_name` character vector, represents the names of the pollutants (e.g., "pm10", "so2")
- `estimate_interval` numeric vector, represents the time interval to estimate the exposure
- `krige_model` character, method used for kriging (e.g., "med")
- `nmax` numeric, maximum number of interpolation points
- `krige_method` character, method used for kriging (e.g., "med")

Example Usage

```r
# Example data
individual_data <- data.frame(individual_id = c(1, 2, 3), exposure_date = c("2020-01-01", "2020-01-02", "2020-01-03"))
pollutant_data <- data.frame(site = c("Site1", "Site2", "Site3"), pm10 = c(10, 20, 30), so2 = c(5, 15, 25))

# Calculate exposure
exposure_estimate_krige(individual_data, individual_id = "id", exposure_date = "date", individual_lat = "lat", individual_lon = "lon", pollutant_data, pollutant_date = "date", pollutant_site_lat = "lat", pollutant_site_lon = "lon", pollutant_name = c("pm10", "so2"), estimate_interval = c(0:30), krige_model, nmax = 7, krige_method = "med")
```
pollutant_site_lat
character, variable name in pollutant_data, includes the latitude information of each monitoring site

pollutant_site_lon
character, variable name in pollutant_data, includes the longitude information of each monitoring site

pollutant_name
vector, pollutant name in the pollutant_data need to be estimated

estimate_interval
continue numeric vector, the estimation period, for example: 0:30, for each individual we estimate the environment exposure ranging from the exposure_date to exposure_date + 30 days

krige_model
?krige
mmax
?krige
krige_method
?krige

Value
A list. For each element in the list, there is a dataframe with the first column representing the individual id, the remaining columns represent the exposure estimation in different time points.

Author(s)
Bing Zhang, https://github.com/Spatial-R/EnvExpInd

Examples
## Not run:
library(EnvExpInd)
library(maptools)
library(gstat)
individual_data$date <- as.Date(individual_data$date)
pollutant_data$date <- as.Date(pollutant_data$date)
pollutant_data_full <- timeseries_imput(data= pollutant_data,date_var = "date",
site_var = "site.name",imput_col = 3:8)
pollutant_data_tem <- merge(pollutant_data_full,site_data,by.x = "site.name",by.y = "site")
test.pollutant <- pollutant_data_tem[pollutant_data_tem$date == "2014-09-20",]
coordinates(test.pollutant) = ~lat + lon
########## please define the variogram in a right way ####################
m <- fit.variogram(variogram(PM10~1, test.pollutant), vgm(1, "Sph", 200, 1))
exposure_estimate_krige(
  individual_data = individual_data,
  individual_id = "id",
  exposure_date ="date",
  individual_lat ="lat",
  individual_lon ="lon",
  pollutant_data = pollutant_data_tem,
  pollutant_date = "date",
  pollutant_site_lat = "lat",
  pollutant_site_lon = "lon",
  pollutant_name = c("PM10","PM2.5"),
knige_model = m,
krige_method = "med",
estimate_interval = c(0:10))

## End(Not run)

expoure_estimate_simple

Assess the environmental exposure using the simplest method: nearest monitoring site method

Description

Using the nearest surveillance site as the reference site to estimate the pollutant exposure.

Usage

expoure_estimate_simple(
  individual_data,
  individual_id,
  reference_id,
  exposure_date,
  pollutant_data,
  pollutant_site = "site",
  pollutant_date = "date",
  pollutant_name = c("pm10", "so2"),
  estimate_interval
)

Arguments

  individual_data data.frame, includes the reference id, individual_id and exposure_date
  individual_id character, variable name in the individual_data, which represents the unique id for each individual
  reference_id character, variable name in the individual_data, which represents the nearest surveillance site for each individual
  exposure_date character, variable name in the individual_data, which represents the start date to estimate the environment exposure
  pollutant_data data.frame, contains the pollutant and site information. One column represents the site information and other columns represent the concentration of pollutants
  pollutant_site character, variable name in the pollutant_data, which represents the monitoring site information
  pollutant_date character, variable name in the pollutant_data, which represents the surveillance date for pollutant concentration
pollutant_name: vector, variable names in the pollutant_data, which represent the name of the target pollutants to be estimated.

estimate_interval: continue numeric vector, the estimation period, for example: 0:30, for each individual we estimate the environment exposure ranging from the exposure_date to exposure_date + 30 days.

**Value**

A list. For each element in the list, there is a dataframe with the first column representing the individual id, the remaining columns represent the exposure estimation in different time points.

**Author(s)**

Bing Zhang, [https://github.com/Spatial-R/EnvExpInd](https://github.com/Spatial-R/EnvExpInd)

**Examples**

```r
library(EnvExpInd)
individual_data$date <- as.Date(individual_data$date)
pollutant_data$date <- as.Date(pollutant_data$date)
pollutant_data_full <- timeseries_imput(data= pollutant_data, 
    date_var = "date", site_var = "site.name", imput_col = 3:8)
pollutant_data_tem <- merge(pollutant_data_full, site_data, by.x = "site.name", by.y = "site")
individual_data$refrence_id <- get_refrence_id_simple(
    individual_data = individual_data, 
    individual_lat = "lat", 
    individual_lon = "lon", 
    individual_id = "id", 
    site_data = site_data, 
    site_lon = "lon", 
    site_lat = "lat", 
    site_id = "site")
expoure_estimate_simple(
    individual_data = individual_data, 
    individual_id = "id", 
    refrence_id = "refrence_id", 
    exposure_date = "date", 
    pollutant_data = pollutant_data_tem, 
    pollutant_site = "site.name", 
    pollutant_date = "date", 
    pollutant_name = c("PM10","PM2.5"), 
    estimate_interval = c(0:10))
```

**get_latlon_china**  
transform the address information into the longitude and latitude.
get_refrence_id_simple

Description

Match the nearing monitoring site for each individual

Usage

get_refrence_id_simple(
  individual_data,
  individual_lat,
  individual_lon,
  individual_id,
  site_data,
  site_lat,
get_refrence_id_simple

    site_lon,
    site_id
  )

Arguments

individual_data
data.frame, including three variables (individual_lat, individual_lon and individual_id)

individual_lat character, variable name in individual_data, includes the latitude information of each individual

individual_lon character, variable name in individual_data, includes the longitude information of each individual

individual_id character, variable name in individual_data, includes the unique id for each individual

site_data data.frame, including three variables (site_lat, site_lon and site_id)

site_lat character variable includes the latitude value of the site

site_lon character variable includes the longitude value of the site

site_id character variable includes the id for each site

Value

A vector, including the reference_id for each individual

Author(s)

Bing Zhang, https://github.com/Spatial-R/EnvExpInd

Examples

get_refrence_id_simple(
  individual_data = individual_data,
  individual_lat = "lat",
  individual_lon = "lon",
  individual_id = "id",
  site_data = site_data,
  site_lon = "lon",
  site_lat = "lat",
  site_id = "site")
individual_data

*The detailed information for each individual.*

**Description**

A dataset containing the detailed information for each individual

**Usage**

`individual_data`

**Format**

A data frame with 21 rows and 3 variables:

- **id** id number for each individual
- **date** the monitoring time point
- **lat** the latitude for each individual
- **lon** the longitude for each individual ...

pollutant_data

*The concentration of air pollutant at each time point.*

**Description**

A dataset containing the concentration of air pollutant at each time point

**Usage**

`pollutant_data`

**Format**

A data frame with 11090 rows and 8 variables:

- **date** the monitoring time point
- **site.name** the names of the monitoring site
- **SO2** the concentration of SO2
- **NO2** the concentration of NO2
- **PM10** the concentration of PM10
- **CO** the concentration of CO
- **O3** the concentration of O3
- **PM2.5** the concentration of PM2.5 ...
site_data  

Monitoring sites.

Description
A dataset containing the information of the monitoring sites

Usage
site_data

Format
A data frame with 10 rows and 2 variables:
- site  the name of monitoring sites
- lat  the latitude for each monitoring site
- lon  the longitude for each monitoring site ...

---

timeseries_imput  

Impute the missing value for the timeseries using the linear interpolation

Description
Complete the time series using the linear interpolation

Usage
timeseries_imput(data, date_var, site_var, imput_col)

Arguments
- data  data.frame, contains the reference id, individual_id and exposure_date
- date_var  character, variable name in data, represents the monitoring date.
- site_var  character, variable name in data, represents the name of monitoring site.
- imput_col  numeric, the column position of the target variables need to be imputed

Value
a data.frame

Author(s)
Bing Zhang. https://github.com/Spatial-R/EnvExpInd
Examples

library(EnvExpInd)
pollutant_data_com <- timeseries_imput(data = pollutant_data, date_var = "date",
                                       site_var = "site.name", imput_col = 3:8)
**Index**

* datasets
  individual_data, 10
  pollutant_data, 10
  site_data, 11

exposure_estimate_idw, 2
exposure_estimate_krige, 4
exposure_estimate_simple, 6

get_latlon_china, 7
get_reference_id_simple, 8

individual_data, 10
pollutant_data, 10
site_data, 11

timeseries_imput, 11