Package ‘aire.zmvm’

June 26, 2024

Type Package

Title Download Mexico City Pollution, Wind, and Temperature Data

Version 1.0.0

Description Tools for downloading hourly averages, daily maximums and minimums from each of the pollution, wind, and temperature measuring stations or geographic zones in the Mexico City metro area. The package also includes the locations of each of the stations and zones. See <http://aire.cdmx.gob.mx/> for more information.

License BSD_3_clause + file LICENSE

URL https://hoyodesmog.diegovalle.net/aire.zmvm/, https://github.com/diegovalle/aire.zmvm

BugReports https://github.com/diegovalle/aire.zmvm/issues

Depends R (>= 3.2.3)

Imports dplyr, httr, lubridate, progress, readr, readxl, rvest, sp, stats, stringr, tidyr, utils, xml2

Suggests ggplot2, knitr, rmarkdown, testthat

Encoding UTF-8

LazyData TRUE

RoxygenNote 7.3.1

NeedsCompilation no

Author Diego Valle-Jones [aut, cre]

Maintainer Diego Valle-Jones <diego@diegovalle.net>

Repository CRAN

Date/Publication 2024-06-26 12:20:02 UTC

Contents

`convert_to_imeca` *Convert pollution values to IMECA*

Description

This function converts pollution running averages in the original units (ppb, µg/m³, etc) to IMECA

Usage

```
convert_to_imeca(value, pollutant, showWarnings = TRUE)
```

Arguments

`value` a numeric vector of values to convert to IMECAs. Note that the concentration of pollutants can be measured in different ways, for NO2, and O3 a 1 hour average is used, for CO, an 8 hour average, and for SO2, PM10 and PM25 a 24 hour average is used.

`pollutant` type of pollutant. A vector of one or more of the following options:

- SO2 - Sulfur Dioxide - ppb (24 hour average)
- CO - Carbon Monoxide - ppm (8 hour average)
- NO2 - Nitrogen Dioxide - pbb (1 hour average)
- O3 - Ozone ppb (1 hour average)
- PM10 - Particulate matter 10 micrometers or less (24 hour average)
- PM25 - Particulate matter 2.5 micrometers or less (24 hour average)

`showWarnings` deprecated; you can use the function `suppressWarnings` instead.

Details

Air quality in Mexico City is reported in IMECAs (Índice Metropolitano de la Calidad del Aire), a dimensionless scale where all pollutants can be compared.

Note that each pollutant has different averaging periods (see the arguments section). Because of rounding error results may be off by a couple of points.

Value

A vector containing the converted value in IMECAs

See Also

For the formulas on how to convert visit: AVISO POR EL QUE SE DA A CONOCER EL PROYECTO DE NORMA AMBIENTAL PARA EL DISTRITO FEDERAL

Other convert functions: `convert_to_index()`

Examples

```
## IMECA is a dimensionless scale that allows for the comparison of
## different pollutants
convert_to_imeca(157, "O3")
convert_to_imeca(c(450, 350, 250), rep("NO2", 3))
## Since this is PM10 the 80 is supposed to be the 24 hour average
convert_to_imeca(80, "PM10")

## warning about recycling elements in a vector
convert_to_imeca(c(157, 200), c("O3", "O3"))

convert_to_imeca(67, "O3")
convert_to_imeca(77, "O3")
convert_to_imeca(205, "O3")
convert_to_imeca(72, "O3")
convert_to_imeca(98, "O3")
```

`convert_to_index` *Convert a pollutant concentration to its air quality category*

Description

This functions converts a pollutant value in its original units into one of the 5 categories used by the Mexican government to communicate to the public how polluted the air currently is and its health risks.

Usage

```
convert_to_index(value, pollutant)
```

Arguments

`value`	a numeric vector of values to convert to index
`pollutant`	type of pollutant. A vector of one or more of the following options: • SO2 - Sulfur Dioxide - ppb (24 hour average) • CO - Carbon Monoxide - ppm (8 hour average) • NO2 - Nitrogen Dioxide - pbb (1 hour average) • O3 - Ozone ppb (1 hour average) • PM10 - Particulate matter 10 micrometers or less (24 hour average) • PM25 - Particulate matter 2.5 micrometers or less (24 hour average)

Value

the IMECA value of the concentration indexed into 5 categories

- BUENA - Good: 0-50 minimal health risk
- REGULAR - Regular: 51-100 moderate health effects
- MALA - Bad: 101-150 sensitive groups may suffer adverse heatlh effects
- MUY MALA - Very Bad: 151-200 everyone can experience negative health effects
- EXTREMADAMENTE MALA - Extremely Bad: > 200 serious health issues

See Also

NADF-009-AIRE-2006

Other convert functions: `convert_to_imeca()`

Examples

```
convert_to_index(c(12.1, 215, 355), c("PM25", "PM10", "PM10"))
```

`download_24hr_average` *Download archives of the 24 hour averages of pollutants*

Description

Data comes from Promedios de 24 horas de partículas suspendidas(PM10 Y PM2.5) and Promedios de 24 horas de Dióxido azufre

Usage

```
download_24hr_average(type, year, progress = interactive())
```

Arguments

`type` type of data to download.

- SO2 - Sulfur Dioxide (parts per billion)
- PS - Suspended solids

`year` a numeric vector containing the years for which to download data (the earliest possible value is 1986 for SO2 and 1995 for PS)

`progress` whether to display a progress bar (TRUE or FALSE). By default it will only display in an interactive session.

Value

A data.frame with pollution data.

Examples

```
## Not run:
head(download_24hr_average("PS", 2017))

## End(Not run)
```

`download_deposition` *Download Acid Rain Measurements Archives*

Description

Download data on rainfall samples collected weekly during the rainy season, available at Depósito and Depósito

Usage

```
download_deposition(deposition, type)
```

Arguments

`deposition` type of deposition to download

- TOTAL - Total deposition (1988-2000)
- HUMEDO - Wet and dry deposition (1997-)

`type` type of ion measurement

- DEPOSITO - ion quantity deposition
- CONCENTRACION - ion concentration

Value

A data.frame with deposition data.

Examples

```
## Not run:
## Download rainfall in mm
df <- download_deposition(deposition = "HUMEDO", type = "CONCENTRACION") %>%
        filter(pollutant == "PP")
head(df)

## End(Not run)
```

download_lead *Download Lead Pollution Archives*

Description

Download data on lead pollution from the archives available at Plomo and Partículas suspendidas

Usage

```
download_lead(type)
```

Arguments

type type of data to download.

- PbPST
- PST, PM10, PM25

Value

A data.frame with pollution data.

Examples

```
## Not run:
head(download_lead("PbPST"))

## End(Not run)
```

download_meteorological *Download Meteorological Data Archives*

Description

Download the files available at Meteorología

Usage

```
download_meteorological(year, progress = interactive())
```

Arguments

year	a numeric vector containing the years for which to download data (the earliest possible value is 1986)
progress	whether to display a progress bar (TRUE or FALSE). By default it will only display in an interactive session.

Value

a data.frame with meterological information: "RH","TMP","WDR","WSP","PBa"

Examples

```
## Not run:
head(download_meteorological(2017))

## End(Not run)
```

download_pollution *Download Pollution Archives*

Description

Download the pollution files available at Contaminante

Usage

```
download_pollution(year, progress = interactive())
```

Arguments

year	a numeric vector containing the years for which to download data (the earliest possible value is 2009)
progress	whether to display a progress bar (TRUE or FALSE). By default it will only display in an interactive session.

Value

a data.frame with pollution information for the following pollutants "CO", "NO", "NO2", "NOX", "O3", "PM10", "SO2", "PM25", and "PMCO"

Examples

```
## Not run:
head(download_pollution(2017))

## End(Not run)
```

---

download_pressure *Download Atmospheric Pressure Archives*

Description

The data comes from Presión Atmosférica

Usage

```
download_pressure(year, progress = interactive())
```

Arguments

`year`	a numeric vector containing the years for which to download data (the earliest possible value is 2009)
`progress`	whether to display a progress bar (TRUE or FALSE). By default it will only display in an interactive session.

Value

A data.frame with atmospheric pressure data.

Examples

```
## Not run:
head(download_pressure(2017))

## End(Not run)
```

`download_radiation` *Download Ultraviolet Radiation Archives*

Description

Download data on UVA and UVB from the pollution archives available at Radiación Solar (UVA) and Radiación Solar (UVB)

Usage

```
download_radiation(type, year, progress = interactive())
```

Arguments

`type`	type of data to download. • UVA - long wave ultraviolet A • UVB - short wave ultraviolet B
`year`	a numeric vector containing the years for which to download data (the earliest possible value is 2000)
`progress`	whether to display a progress bar (TRUE or FALSE). By default it will only display in an interactive session.

Value

A data.frame with pollution data. The hours correspond to the *Etc/GMT+6* timezone, with no daylight saving time

Examples

```
## Not run:
head(download_radiation("UVA", 2017))

## End(Not run)
```

`get_latest_imeca` *Get the latest pollution values for each station*

Description

Download the latest hourly values for the pollutants with the highest values for each station as measured in IMECAs

Usage

```
get_latest_imeca()
```

Details

Note that in 2015 it was determined that the stations with codes ACO, AJU, INN, MON and MPA would no longer be taken into consideration when computing the pollution index because they didn't meet the objectives of monitoring air quality, and are no longer included in the index, even if they are still part of the SIMAT (Sistema de Monitoreo Atmosférico de la Ciudad de México). Thus, even if they are located inside a zone, they are not included in the pollution values for that zone.

Value

A data.frame with pollution values in IMECAs, the hour corresponds to the *America/Mexico_City* timezone (which changes with daylight saving time)

See Also

Reporte de calidad del aire

Other IMECA functions: `get_station_imeca()`, `get_zone_imeca()`

Examples

```
df <- get_latest_imeca()
head(df)
```

`get_station_data` *Download pollution data by station*

Description

Retrieve pollution data by station, in the original units, from the air quality server at Consulta de Concentraciones, or for earlier years use the archive files available from Contaminante, or Meteorología for meteorological data. There's a mistake in the 2016 wind speed data, so for this year, and only this year, the alternative Excel file was used.

Usage

```
get_station_data(criterion, pollutant, year, progress = interactive())
```

Arguments

`criterion` Type of data to download.

- HORARIOS - Hourly data
- MAXIMOS - Daily maximums
- MINIMOS - Daily minimums

`pollutant` The type of pollutant to download.

- SO2 - Sulfur Dioxide (parts per billion)

- CO - Carbon Monoxide (parts per million)
- NOX - Nitrogen Oxides (parts per billion)
- NO2 - Nitrogen Dioxide (parts per billion)
- NO - Nitric Oxide (parts per billion)
- O3 - Ozone (parts per billion)
- PM10 - Particulate matter 10 micrometers or less (micrograms per cubic meter)
- PM25 - Particulate matter 2.5 micrometers or less (micrograms per cubic meter)
- WSP - Wind velocity (meters per second)
- WDR - Wind direction (degrees)
- TMP - Temperature (degrees Celsius)
- RH - Relative humidity (percentage)

`year` a numeric vector containing the years for which to download data (the earliest possible value is 1986)

`progress` whether to display a progress bar (TRUE or FALSE). By default it will only display in an interactive session.

Details

Temperature (TMP) archive values are correct to one decimal place, but the most recent data is only available rounded to the nearest integer.

Value

A data.frame with pollution data. When downloading "HORARIOS" the hours correspond to the *Etc/GMT+6* timezone, with no daylight saving time

Warning

The data for the current month is in the process of being validated

See Also

`stations` for a data.frame with the location and names of all pollution measuring stations,

Other raw data functions: `get_station_month_data()`

Examples

```
## Not run:
## Download daily maximum PM10 data (particulate matter 10 micrometers or
## less in diameter) from 2015 to 2016
df <- get_station_data("MAXIMOS", "PM10", 2015:2016)
head(df)

## Download ozone concentration hourly data for 2016
df2 <- get_station_data("HORARIOS", "O3", 2016)
```

```
## Convert to local Mexico City time
df2$mxc_time <- format(as.POSIXct(paste0(df2$date, " ", df2$hour, ":00"),
                                  tz = "Etc/GMT+6"),
                       tz = "America/Mexico_City")
head(df2)

## End(Not run)
```

get_station_imeca *Download pollution data by station in IMECAs*

Description

Retrieve hourly averages of pollution data, by station, measured in IMECAs

Usage

```
get_station_imeca(pollutant, date, show_messages = TRUE)
```

Arguments

`pollutant` The type of pollutant to download

- SO2 - Sulfur Dioxide
- CO - Carbon Monoxide
- NO2 - Nitrogen Dioxide
- O3 - Ozone
- PM10 - Particulate matter 10 micrometers or less
- PM25 - Particulate matter 2.5 micrometers or less

`date` The date for which to download data in YYYY-MM-DD format (the earliest possible date is 2009-01-01).

`show_messages` show a message about issues with excluded stations

Details

Note that in 2015 it was determined that the stations with codes ACO, AJU, INN, MON and MPA would no longer be taken into consideration when computing the pollution index because they didn’t meet the objectives of monitoring air quality, and are no longer included in the index, even if they are still part of the SIMAT (Sistema de Monitoreo Atmosférico de la Ciudad de México). Thus, even if they are located inside a zone, they are not included in the pollution values for that zone.

Value

A data.frame with pollution data measured in IMECAs, by station. The hours correspond to the *Etc/GMT+6* timezone, with no daylight saving time

See Also

Índice de calidad del aire por estaciones

Other IMECA functions: `get_latest_imeca()`, `get_zone_imeca()`

Examples

```
## Not run:
## There was an ozone pollution emergency on May 15, 2017
df_o3 <- get_station_imeca("O3", "2017-05-15", show_messages = FALSE)

## Convert to local Mexico City time
df_o3$mxc_time <- format(as.POSIXct(paste0(df_o3$date,
                                           " ",
                                           df_o3$hour,
                                           ":00"),
                                    tz = "Etc/GMT+6"),
                         tz = "America/Mexico_City")
head(df_o3[order(-df_o3$value), ])

## End(Not run)
```

`get_station_month_data` *Download monthly pollution data*

Description

Retrieve hourly averages, daily maximums, or daily minimums of pollution data in the original units, by station, from the air quality server at Consulta de Concentraciones

Usage

```
get_station_month_data(criterion, pollutant, year, month)
```

Arguments

`criterion` Type of data to download.

- HORARIOS - Hourly data
- MAXIMOS" - Daily maximums
- MINIMOS - Daily minimums

`pollutant` The type of pollutant to download.

- SO2 - Sulfur Dioxide (parts per billion)
- CO - Carbon Monoxide (parts per million)
- NOX - Nitrogen Oxides (parts per billion)
- NO2 - Nitrogen Dioxide (parts per billion)
- NO - Nitric Oxide (parts per billion)

- O3 - Ozone (parts per billion)
- PM10 - Particulate matter 10 micrometers or less (micrograms per cubic meter)
- PM25 - Particulate matter 2.5 micrometers or less (micrograms per cubic meter)
- WSP - Wind velocity (meters per second)
- WDR - Wind direction (degrees)
- TMP - Temperature (degrees Celsius)
- RH - Relative humidity (percentage)

`year` an integer indicating the year for which to download data (the earliest possible value is 1986)

`month` month number to download

Details

Temperature (TMP) data was rounded to the nearest integer, but the `get_station_data` function allows you to download data accurate to one decimal point in some cases (i.e. for old data).

Value

A data.frame with pollution data, the hours correspond to the *Etc/GMT+6* timezone, with no daylight saving time

Warning

The data for the current month is in the process of being validated

See Also

`stations` for a data.frame with the location and names of all pollution measuring stations

Other raw data functions: `get_station_data()`

Examples

```
## Not run:
## Download daily hourly PM10 data (particulate matter 10 micrometers or
## less in diameter) from March 2016
df_pm10 <- get_station_month_data("HORARIOS", "PM10", 2016, 3)
head(df_pm10)

## Download daily hourly O3 data from October 2017
df_o3 <- get_station_month_data("HORARIOS", "O3", 2018, 1)
## Convert to local Mexico City time
df_o3$mxc_time <- format(as.POSIXct(paste0(df_o3$date,
                                           " ",
                                           df_o3$hour, ":00"),
                                    tz = "Etc/GMT+6"),
                         tz = "America/Mexico_City")
head(df_o3)
```

```
## End(Not run)
```

get_zone_imeca *Download pollution data by zone in IMECAs*

Description

Retrieve pollution data in IMECAs by geographic zone from the air quality server at Consultas

Usage

```
get_zone_imeca(
  criterion,
  pollutant,
  zone,
  start_date,
  end_date,
  showWarnings = TRUE,
  show_messages = TRUE
)
```

Arguments

`criterion` The type of data to download. One of the following options:

- HORARIOS - Hourly data
- MAXIMOS" - Daily maximums

`pollutant` The type of pollutant to download. One or more of the following options:

- SO2 - Sulfur Dioxide
- CO - Carbon Monoxide
- NO2 - Nitrogen Dioxide
- O3 - Ozone
- PM10 - Particulate matter 10 micrometers or less
- TC - All the pollutants

`zone` The geographic zone for which to download data. One or more of the following:

- NO - Noroeste
- NE - Noreste
- CE - Centro
- SO - Suroeste
- SE - Sureste
- TZ - All zones

`start_date` The start date in YYYY-MM-DD format (earliest possible value is 2008-01-01).

`end_date` The end date in YYYY-MM-DD format.

`showWarnings` deprecated; you can use the function `suppressWarnings` instead.

`show_messages` show a message about issues with performing the conversion

Details

Note that in 2015 it was determined that the stations with codes ACO, AJU, INN, MON and MPA would no longer be taken into consideration when computing the pollution index because they didn't meet the objectives of monitoring air quality. They are no longer included in the index, even if they are still part of the SIMAT (Sistema de Monitoreo Atmosférico de la Ciudad de México). Thus, even if they are located inside a zone, they are not included in the pollution values for that zone.

The different geographic zones were defined in the Gaceta Oficial de la Ciudad de México No. 230, 27 de Diciembre de 2016.

Zona Centro: Benito Juárez, Cuauhtémoc, Iztacalco and Venustiano Carranza.

Zona Noreste: Gustavo A. Madero, Coacalco de Berriozábal, Chicoloapan, Chimalhuacán, Ecatepec de Morelos, Ixtapaluca, La Paz, Nezahualcóyotl and Tecámac.

Zona Noroeste: Azcapotzalco, Miguel Hidalgo, Atizapán de Zaragoza, Cuautitlán, Cuautitlán Izcalli, Naucalpan de Juárez, Nicolás Romero, Tlalnepantla de Baz and Tultitlán.

Zona Sureste: Iztapalapa, Milpa Alta, Tláhuac, Xochimilco, Chalco and Valle de Chalco.

Zona Suroeste: Álvaro Obregón, Coyoacán, Cuajimalpa, Magdalena Contreras, Tlalpan and Huixquilucan.

Value

A data.frame with pollution data measured in IMECAs, by geographic zone. The hours correspond to the *Etc/GMT+6* timezone, with no daylight saving time

See Also

`zones` a data.frame containing the municipios belonging to each zone, and Índice de calidad del aire por zonas

Other IMECA functions: `get_latest_imeca()`, `get_station_imeca()`

Examples

```
## There was a regional (NE) PM10 pollution emergency on Jan 6, 2017
get_zone_imeca("MAXIMOS", "PM10", "NE", "2017-01-05", "2017-01-08",
               show_messages = FALSE)

## There was an ozone pollution emergency on May 15, 2017
get_zone_imeca("MAXIMOS", "O3", "TZ", "2017-05-15", "2017-05-15",
               show_messages = FALSE)


## Not run:
## Download daily maximum PM10 data (particulate matter 10 micrometers or
## less in diameter) from 2015-01-01 to 2016-03-20 for all geographic zones
df <- get_zone_imeca("MAXIMOS", "PM10", "TZ", "2015-01-01", "2016-03-20")
head(df)

## Download hourly O3 pollution data for May 15, 2017. Only the suroeste zone
df2 <- get_zone_imeca("HORARIOS", "O3", "SO", "2017-05-15", "2017-05-15")
```

```
## Convert to local Mexico City time
df2$mxc_time <- format(as.POSIXct(paste0(df2$date, " ", df2$hour, ":00"),
                                  tz = "Etc/GMT+6"),
                       tz = "America/Mexico_City")

head(df2)

## End(Not run)
```

`idw360` *Inverse Distance Weighting with Directional Data*

Description

Function for inverse distance weighted interpolation with directional data. Useful for when you are working with data whose unit of measurement is degrees (i.e. the average of 35 degrees and 355 degrees should be 15 degrees). It works by finding the shortest distance between two degree marks on a circle.

Usage

```
idw360(values, coords, grid, idp = 2)
```

Arguments

`values`	the dependent variable
`coords`	the spatial data locations where the values were measured. First column x/longitude, second y/latitude
`grid`	data frame or Spatial object with the locations to predict. First column x/longitude, second y/latitude
`idp`	The inverse distance weighting power

Value

data.frame with the interpolated values for each of the grid points

Examples

```
library("sp")
library("ggplot2")

## Could be wind direction values in degrees
values <- c(55, 355)

## Location of sensors. First column x/longitud, second y/latitude
locations <- data.frame(lon = c(1, 2), lat = c(1, 2))
```

```
coordinates(locations) <- ~lon+lat

## The grid for which to extrapolate values
grid <- data.frame(lon = c(1, 2, 1, 2), lat = c(1, 2, 2, 1))
coordinates(grid) <- ~lon+lat

## Perform the inverse distance weighted interpolation
res <- idw360(values, locations, grid)
head(res)

## Not run:
df <- cbind(res, as.data.frame(grid))
## The wind direction compass starts where the 90 degree mark is located
ggplot(df, aes(lon, lat)) +
  geom_point() +
  geom_spoke(aes(angle = ((90 - pred) %% 360) * pi / 180),
             radius = 1,
             arrow=arrow(length = unit(0.2, "npc")))

library("mapproj")
## Random values in each of the measuring stations
locations <- stations[, c("lon", "lat")]
coordinates(locations) <- ~lon+lat
crs_string <- "+proj=longlat +ellps=WGS84 +no_defs +towgs84=0,0,0"
proj4string(locations) <- CRS(crs_string)
values <- runif(length(locations), 0, 360)
pixels <- 10
grid <- expand.grid(lon = seq((min(coordinates(locations)[, 1]) - .1),
                              (max(coordinates(locations)[, 1]) + .1),
                              length.out = pixels),
                    lat = seq((min(coordinates(locations)[, 2]) - .1),
                              (max(coordinates(locations)[, 2]) + .1),
                              length.out = pixels))
grid <- SpatialPoints(grid)
proj4string(grid) <- CRS(crs_string)
## bind the extrapolated values for plotting
df <- cbind(idw360(values, locations, grid), as.data.frame(grid))
ggplot(df, aes(lon, lat)) +
  geom_point(size = .1) +
  geom_spoke(aes(angle = ((90 - pred) %% 360) * pi / 180),
             radius = .07,
             arrow=arrow(length = unit(0.2,"cm"))) +
  coord_map()

## End(Not run)
```

stations *Pollution measuring stations in Mexico City*

Description

This dataset contains all pollution measuring stations in Mexico City. The station with code SS1 was added manually since it was missing from the official source dataset (its location was found in the Audit of Ambient Air Monitoring Stations for the Sistema de Monitoreo Atmosférico de la Ciudad de México).

Usage

```
stations
```

Format

A data frame with 63 rows and 7 variables:

station_code abbreviation of the station

station_name name of the station

lon longitude of the station

lat latitude of the station

altitude altitude of the station

comment comment

station_id id of the station

Source

‘`http://148.243.232.112:8080/opendata/catalogos/cat_estacion.csv`’

Examples

```
head(stations)
```

`zones` *Pollution zones in Mexico City*

Description

This data set contains the municipios (counties) that make up the 5 geographic zones into which Mexico City was divided for the purpose of disseminating information about the IMECA.

Usage

```
zones
```

Format

A data frame with 36 rows and 6 variables:

region INEGI code of the region (state_code + municipio_code)

state_code INEGI code of the state

state_abbr state abbreviation

municipio_code INEGI code of the municipio

municipio_name name of the municipio

zone zone

Details

Note that in 2015 it was determined that the stations with codes ACO, AJU, INN, MON and MPA would no longer be taken into consideration when computing the pollution index because they didn't meet the objectives of monitoring air quality, and are no longer included in the index, even if they are still part of the SIMAT (Sistema de Monitoreo Atmosférico de la Ciudad de México). Thus, even if they are located inside a zone, they are not included in the pollution values for that zone.

A transparency request was used to determine the zone to which the municipios of Acolman, Texcoco and Atenco belong.

Source

Gaceta Oficial de la Ciudad de México No. 230, 27 de Diciembre de 2016, and *Solicitud de Información* FOLIO 0112000033818

Examples

```
head(zones)
```

Index