SoftwareX 30 (2025) 102185

Contents lists available at ScienceDirect

SoftwareX

journal homepage: www.elsevier.com/locate/softx

DataMetProcess: An open-source package and Shiny application for the acquisition and processing of meteorological data from INMET

Wagner Martins dos Santos [a], Hoi Leong Lee [b,*], Edimir Xavier Leal Ferraz [a], Abelardo Antônio de Assunção Montenegro [a], Lady Daiane Costa de Sousa Martins [a], Alan Cézar Bezerra [c], Ênio Farias de França e Silva [a], Thieres George Freire da Silva [a,c], João L.M.P. de Lima [d], Xuguang Tang [e], Alexandre Maniçoba da Rosa Ferraz Jardim [c,f,*]

[a] Department of Agricultural Engineering, Federal Rural University of Pernambuco, Recife 52171-900, Pernambuco, Brazil
[b] Faculty of Electronic Engineering & Technology, Universiti Malaysia Perlis, Arau, 02600 Perlis, Malaysia
[c] Academic Unit of Serra Talhada, Federal Rural University of Pernambuco, Serra Talhada, 56909-535, Pernambuco, Brazil
[d] MARE—Marine and Environmental Sciences Centre, ARNET—Aquatic Research Network, Department of Civil Engineering, Faculty of Sciences and Technology, University of Coimbra, Rua Luís Reis Santos, Pólo II - Universidade de Coimbra, Coimbra, 3030-788, Portugal
[e] Institute of Remote Sensing and Geosciences, Hangzhou Normal University, Hangzhou, 311121, China
[f] Department of Biodiversity, Institute of Biosciences, São Paulo State University—UNESP, Rio Claro, 13506-900, São Paulo, Brazil

ARTICLE INFO

Keywords:
Climate analysis
Evapotranspiration
Intuitive tool
Software R

ABSTRACT

The acquisition and understanding of meteorological data are crucial for management in various sectors of society. Thus, the objective was to develop a package in the R programming environment to download and process data provided by Brazilian National Institute of Meteorology (INMET) and to create a free web-based tool using Shiny (interactive web apps) to simplify the use of the functions provided in the package. The DataMetProcess package was developed, offering four essential functions to extract information from the database, correct time zones, change the time scale, and calculate reference evapotranspiration according to the Penman-Monteith model. The Shiny web application developed is available in three versions: (1) the code; (2) portable; and (3) a web version. The DataMetProcess package and the developed web application are effective in processing the data provided by INMET, democratizing access to information and facilitating analysis through the interactive web application. Finally, being open-source, it can be easily adapted and used worldwide. We encourage the research community to adopt this tool and provide feedback to further refine its usability and efficiency.

Metadata

Nr	Code metadata description	*Metadata*
C1	Current code version	*v1.0.3*
C2	Permanent link to code/ repository used for this code version	https://github.com/wagnerms97/DataMetProcess/tree/main
C3	Permanent link to reproducible capsule	
C4	Legal code license	GPL-3
C5	Code versioning system used	*Git*
C6	Software code languages, tools and services used	R, JavaScript
C7	Compilation requirements, operating environments and dependencies	*R packages: dplyr, tidyverse, Rlang, utils, shiny, shinydashboard, shinyjs, shinybusy, shinyalert, shinyWidgets, reactable, plotly, ggplot2, zoo, janitor*
C8	If available, link to developer documentation/manual	https://github.com/wagnerms97/DataMetProcess/blob/main/Readme.Rmd
C9	Support email for questions	wagnnerms97@gmail.com

* Corresponding authors.
E-mail addresses: wagner.martins@ufrpe.br (W.M. Santos), hoileong@unimap.edu.my (H.L. Lee), edimir.ferraz@ufrpe.br (E.X.L. Ferraz), abelardo.montenegro@ufrpe.br (A.A.A. Montenegro), lady.daiane@ufrpe.br (L.D.C.S. Martins), alan.bezerra@ufrpe.br (A.C. Bezerra), enio.fsilva@ufrpe.br (Ê.F. França e Silva), thieres.silva@ufrpe.br (T.G.F. Silva), plima@dec.uc.pt (J.L.M.P. de Lima), xgtang@hznu.edu.cn (X. Tang), alexandremrfj@gmail.com (A.M.R.F. Jardim).

https://doi.org/10.1016/j.softx.2025.102185
Received 25 February 2025; Received in revised form 6 April 2025; Accepted 28 April 2025
Available online 8 May 2025

1. Motivation and significance

The acquisition and processing of meteorological data are fundamental for managing various sectors of society, such as agriculture, urban planning, and industry [1,2]. Monitoring variables such as air temperature, relative humidity, precipitation, wind speed, solar radiation, and evapotranspiration enables more efficient planning based on reliable information. However, these data are often extensive and require either labor-intensive manual processing or programming expertise for proper handling and analysis [3].

Among the various free databases available, the National Institute of Meteorology (INMET), affiliated with the Brazilian Ministry of Agriculture and Livestock, is one of the primary sources used in projects conducted in Brazil that involve meteorological information [4–6]. INMET provides hourly measurements from 568 stations distributed across the national territory, including 144 meteorological stations in the Northeast, 87 in the North, 96 in the Central-West, 147 in the Southeast, and 94 in the South. This highlights the importance of processing this database in the development of meteorological software. However, since these data are collected on an hourly scale, many applications require conversion to larger temporal scales, such as daily, monthly, and annual. Additionally, graphical analyses are essential to facilitate data interpretation and support decision-making.

In this context, computational tools, which can be developed in different programming languages, streamline data processing and analysis while offering various visual resources to aid in information comprehension [7,8]. However, they require a certain level of programming knowledge, and depending on the complexity of the required analyses, their use can be challenging. In this regard, the development of reusable solutions that reduce the need for coding, along with open-source initiatives such as The Comprehensive R Archive Network (CRAN) for the R language, provides a wide range of resources for the research community.

R is a programming language and environment for statistical computing and graphics, known for its extensive range of techniques and its ease of extension and modification. Its open-source nature has led to a large number of active contributors on CRAN, making it a programming environment that facilitates the seamless integration of different packages [9]. Enhancing this language, the Shiny package enables the straightforward development of interactive web applications, leveraging several auxiliary packages for widget creation and interactivity within the application. Additionally, services such as shinyapps.io, provided by Posit (Posit Software, PBC, Boston, MA, USA), allow free sharing of Shiny applications on the web, making this tool both powerful and essential for the development of free and open-source solutions.

Although R packages such as **climatol** [10], **metR** [11], **rnoaa** [12] and **BrazilMet** [13] provide solutions for handling and analyzing meteorological data, they have limitations in terms of accessibility for users without programming experience and integration with intuitive interfaces. Furthermore, the solutions provided by these packages differ from what is proposed for **DataMetProcess**, despite some similarities (Table 1). For example, **climatol** primarily focuses on data homogenization and gap filling. Although it provides summaries of climatological series, it lacks functions for batch processing of variables and requires data conversion into specific formats used by the tool. In turn, **metR** is more focused on graphical visualization of climatic data, offering various integrations with **ggplot** [14]. Similarly, **rnoaa** shares some characteristics with **DataMetProcess**, as it provides access to data made available by NOAA (National Oceanic and Atmospheric Administration). However, it does not directly support data manipulation from other sources or for different temporal scales, given that NOAA's API already provides data at multiple scales. Among these, **BrazilMet** is the closest package to the proposed one; however, despite also enabling the download of INMET stations, its focus is on estimating reference evapotranspiration (ET_0) and potential evapotranspiration (PET). It simplifies access to essential data for estimating these variables by using functions that allow the separate calculation of different input variables required in this process, such as relative humidity estimation, radiation conversion, shortwave solar radiation, and longwave radiation, although **DataMetProcess** also calculates and returns some variables related to ET_0, its primary focus is not evapotranspiration estimation. **BrazilMet**, on the other hand, is more comprehensive in this regard, as it provides specific functions for estimating both ET_0 and PET, along with detailed methods for calculating intermediate variables.

Table 1
Functionality comparison between R packages for climate data analysis.

Functionality	climatol	rnoaa	metR	BrazilMet	DataMetProcess
Quality control	✓	×	×	×	×
Data homogenization	✓	×	×	×	×
Gap filling	✓	×	×	×	×
Direct access to NOAA data	×	✓	✓	×	×
Conversion between multiple temporal scales	×	×	×	×	✓
Time zone adjustment	×	×	×	×	✓
Graphical visualization	✓	✓	✓	×	✓
Batch processing of climate variables	×	×	×	×	✓
Integration with an intuitive web interface	×	×	×	×	✓
Direct support for INMET data	×	×	×	✓	✓

Note: ✓ - the item is offered by the package; × - not included.

In contrast, **DataMetProcess** is a more general package designed for meteorological data processing, allowing the conversion of data from hourly intervals into daily, monthly, and annual formats. This tool was developed to bridge this gap, offering an open-source R package with optimized functionalities for meteorological data processing. Therefore, **DataMetProcess** can even be seen as a complement to existing packages, adding more functionality and possibilities for different temporal scales of climatic data. Although designed to facilitate the download and processing of meteorological data from Brazil's INMET, its other functions are generalized to be applicable to other databases. Additionally, a free web-based tool was developed using the Shiny framework, enabling users to visually explore and analyze data without requiring extensive programming knowledge. Below, we delve into the development process of the tool and present application examples to illustrate its functionalities.

2. Software description

2.1. Software architecture

The **DataMetProcess** package was developed to simplify and expedite the initial processing of meteorological data, offering four essential functions that can be easily utilized by the end user: **list_inmet(), adjustDate(), calculateDMY()**, and **calculateETrefPM()**. These functions encompass the download and listing of data from the Brazilian INMET database, time zone adjustment, temporal scale adjustments, and reference evapotranspiration calculations, respectively. The **DataMetProcess** was developed in R version 4.3.2, and it utilized basic support functions that are provided as defaults in R, as well as various functionalities from packages available on CRAN, which were crucial for the package development (Table 2).

The package can be installed in its stable version through CRAN (https://cran.r-project.org/web/packages/DataMetProcess/index.html) or in its development version through GitHub (https://github.

Table 2
Packages used in the development of DataMetProcess.

Package	Version	Reference	Functionality
base	4.3.2	[9]	Basic support functions
dplyr	1.1.4	[15]	Data manipulation
tidyr	1.3.0	[16]	Data manipulation
lubridate	1.9.3	[17]	Date and time manipulation
rlang	1.1.2	[18]	Collection of frameworks and APIs
utils	4.3.2	[9]	Set of functions

com/wagnnerms97/DataMetProcess) (Source Code 1), working on Windows, Linux, or Mac environments, as tested for the package submission to the CRAN repository. Here, we present an example of installing and using the **DataMetProcess** R package in two parts:

Source Code 1. Install and load package.

```
#1 Installing the package
#1.1. CRAN:
install.packages("DataMetProcess")
#1.2. GitHub (R package devtools)
devtools::install_github("https://github.com/wagnnerms97/DataMetProcess")
#2. Load package
library(DataMetProcess)
```

2.2. Software functionalities

2.2.1. list_inmet()

A the **list_inmet()** function allows users to extract information from the database of INMET's automatic meteorological stations (Source Code 2). Data extraction is performed by directly downloading compressed (".zip") files hosted on the INMET website. To do this, the user must specify the desired year and, optionally, the storage location for the downloaded file. If no storage location is defined, the file will be temporarily saved in the system's default directory. The extracted data includes a list containing: (1) A table with the file addresses within the zip archive, enabling later extraction using the **unzip()** function from the **utils** package; (2) A structured table with information extracted from the file names, including city, station code, year, start and end dates of data collection; (3) The path of the saved zip file. The extracted data is processed using functions from the **tidyr** and **dplyr** packages, which structure and organize the information. The separation of elements contained in the file names is based on predefined patterns using slashes ("/") and underscores ("_") as delimiters. After processing, the data is sorted by city to facilitate querying (Algorithm 1).

The **unzip()** function from the **utils** package allows extracting only specific data from the regions of interest. Thus, if the main file containing the entire database is stored in a temporary file, as per the default of the **list_inmet()** function, only the specified files will be retained on the user's system. The **unzip()** function takes as its main parameters from **list_inmet()**: the address of the file returned in the third item of the list, and one or more download addresses contained in the first table of that same list (Source Code 2).

Source Code 2. Using the **list_inmet()** and **unzip()** function.

```
#1. List of data available at INMET by year
infoInmet <- list_inmet("2000")
#2. File addresses in the main zip file
infoInmet$Adresses
#3. Data frame with database information
infoInmet$Details
#4. File address zip file
infoInmet$Saved
#5. Applying to utils::unzip()
unzip.file <- utils::unzip(zipfile = infoInmet$Saved, exdir = "./", files = infoInmet
$Adresses$Name[2])
```

2.2.2. adjustDate()

The **adjustDate()** function corrects the time zone of the data, an optional step depending on the data source and location. The function requires three inputs: (1) a column containing date records, (2) a column with time values, and (3) the desired time zone for correction. It assumes that the provided times are in UTC (Coordinated Universal Time) to facilitate processing and converts them accordingly. The time zone adjustment process consists of the following steps: (1) Extracting the hour component from the time column using **substr()** (considering only the first two characters when available); (2) Merging the date and time columns into a single date-time column using **tidyr**; (3) Converting the date-time column to POSIXct format using **strptime()**; (4) Applying the specified time zone correction with **as.POSIXct()**, ensuring the proper conversion from UTC to the target time zone (Algorithm 2).

2.2.3. calculateDMY()

The **calculateDMY()** function was developed to perform the conversion of hourly data to daily, from daily to monthly, and from monthly to annual data. Direct conversion from hourly to monthly or annual data is possible; however, attention must be paid to the grouping of variables that may change according to the temporal scale between maximum, minimum, averages or sums.

In the present study, the **calculateDMY()** function was developed to convert time series data across different temporal scales: from hourly to daily, from daily to monthly, and from monthly to annual. Additionally, direct conversion from hourly data to monthly or annual scales is supported. However, careful consideration is required when aggregating variables, as the appropriate method (e.g., sum, mean, maximum, or minimum) may vary depending on the selected temporal scale. The

Algorithm 1
list_inmet().

```
Input:
year → Year of the data to be downloaded
filename → Path and name of the zip file (optional)
Initialization:
IF filename is not provided THEN
Create a temporary file
File download:
Construct the URL of the zip file using the provided year
Download the zip file
Information extraction:
Get the list of files inside the zip
Extract information (Origin, Region, State, Id, City, Dates) from the database using the filename as a separator
Output:
Return a list containing:
Addresses: list of files inside the zip
Details: structured table with metadata
Saved: path of the saved zip file
End
```

Algorithm 2
adjustDate().

```
Input:
  data → Data frame containing the data
  col_date → Name of the column with the dates
  col_hour → Name of the column with the time (format "hh", "hh:mm", or "hh:mm:ss"; only "hh" will be used)
  fuso → Time zone for correction
  Processing:
  Extract only the hours ("hh") from the col_hour column and overwrite the column in the data frame
  Merge the col_date and col_hour columns into a new column called "Date_Hour" in datetime format:
  Interpret "Date_Hour" as GMT-0
  Adjust it to the time zone specified in fuso
  Output:
  Return the updated data frame with the "Date_Hour" column corrected to the desired time zone
  End
```

function processes data based on user-defined parameters, applying appropriate transformations to summarize the dataset with the help of the **dplyr** and **rlang** packages. The function follows these steps: (1) Data Selection: The function extracts the specified columns from the input data frame, including the date column and the variables to be summarized. (2) Temporal Aggregation: The data is aggregated according to the selected scale ("Daily", "Monthly" or "Annual"). (3) Grouping and Summarization: The function groups data by the transformed date column according to the selected temporal scale and calculates the sum, mean, maximum, or minimum based on user specifications, although the function includes a default specification based on INMET. By incorporating these steps, **calculateDMY()** ensures a precise and flexible approach to temporal data aggregation, accommodating different statistical measures based on user requirements (Algorithm 3).

To demonstrate the practical application of the **calculateDMY()** function, the example code snippet shows its usage in meteorological data from a file provided by default in **DataMetProcess** (Source code 3).

Source Code 3. Using the **adjustDate()** and **calculateDMY()** functions.

```
#1. Loading and read example file (default in the package)
address <- base::system.file("extdata", "ex1_inmet.CSV", package =
"DataMetProcess")
df <- read.table(address, h = TRUE, sep = ";", dec = ",", skip = 8, na.strings =
−9999, check.names = FALSE)
#2. Default date format
df$Data <- as.Date(df$Data, format = "%d/%m/%Y")
#3. Fix the time zone
df <- adjustDate(data = df, col_date = "Data", col_hour = "Hora", fuso = "America/
Bahia")
#4. Calculation of daily scale
df.d <- calculateDMY(data = df, col_date = "Date_Hour", col_sum = colnames(df)[c
(2,6)], col_mean = colnames(df)[c(3,7,8,15,16,18)], col_max = colnames(df)[c
(4,9,11,13,17)], col_min = colnames(df)[c(5,10,12,14)], type = "Daily")
```

2.2.4. calculateETrefPM()

In this function, we calculate the reference evapotranspiration (ET_0) by the Food and Agriculture Organization (FAO) Penman-Monteith (PM) method [19] (Eq. (1)). The crop evapotranspiration (ET_c) can be calculated by the product of the ET_0 and a crop coefficient (Kc).

$$ET_0 = \frac{0.408\Delta(R_n - G) + \gamma \frac{900}{Ta + 273} WS(e_s - e_a)}{\Delta + \gamma(1 + 0.34WS)} \tag{1}$$

where ET_0 - reference evapotranspiration (mm day^{-1}), R_n - net radiation at the crop surface (MJ m^{-2} day^{-1}), G - soil heat flux (MJ m^{-2} day^{-1}), Ta - air temperature at 2 m height (°C), WS - wind speed at 2 m height (m s^{-1}), e_s - saturation vapor pressure (kPa), e_a - actual vapor pressure (kPa), Δ - slope of saturation vapor pressure versus air temperature curve (kPa °C^{-1}), and γ - psychrometric constant (kPa °C^{-1}).

The function takes the parameters: latitude (lat, decimal degrees), altitude (alt, m), anemometer height (za, m), days after planting (DAP, for the first date), and the records of date, temperature (°C), relative humidity (RH, %), soil heat flux (G, MJ m^{-2} day^{-1}), global solar radiation (R_g, MJ m^{-2} day^{-1}), atmospheric pressure (AP, hPa), wind speed (WS, m s^{-1}), and Kc values. The implementation of the Penman-Monteith method follows these steps: (1) Calculation of R_n: Determined based on R_g and crop surface properties; (2) Determination of G: Considered negligible at daily scales or empirically estimated; (3) Calculation of e_s and e_a: Derived from relative humidity and air temperature; (4) Calculation of vapor pressure deficit ($e_s - e_a$): The difference between saturation vapor pressure and actual vapor pressure; (5)

Algorithm 3
calculateDMY().

```
Input:
  data → Data frame containing the data
  col_date → Name of the column with dates
  col_sum → Columns to apply summation
  col_mean → Columns to calculate mean
  col_max → Columns to get maximum
  col_min → Columns to get minimum
  n.round → Number of decimal places for rounding
  type → Aggregation type: "Daily", "Monthly", or "Yearly" (default: "Daily")
  Processing:
  Select only the columns specified in col_date, col_sum, col_mean, col_max, and col_min from the data frame
  IF type = "Monthly" THEN
  Convert col_date to the first day of the month ("%Y-%m-01″)
  ELSE IF type = "Yearly" THEN
  Convert col_date to year only ("%Y")
  Group the data by the col_date column
  Calculate aggregations: col_sum, col_mean, col_max, and col_min
  Round all numeric columns to n.round decimal places
  Output:
  Return the aggregated data frame according to the selected scale
  End
```

Determination of the γ and the slope of the Δ: Derived from air temperature and altitude; (6) Application of the Penman-Monteith equation: ET_0 is calculated by combining climatic and aerodynamic factors; (7) Calculation of ET_c: ET_0 is multiplied by Kc corresponding to each phenological stage (Algorithm 4).

To demonstrate the practical application of the **calculateETrefPM()** function, the example code snippet shows its usage in calculating ET_0 using meteorological data from a file provided by default in **DataMetProcess** (Source code 4).

Source Code 4. Using the **calculateETrefPM()** function.

```
#1. Loading and read example file (default in the package)
address <- base::system.file("extdata","ex2_daily.CSV", package =
"DataMetProcess")
df <- read.table(address, h = TRUE, sep = ";")
#2 Converting to megajoules per square meter per day
df$solar_radiation_kj_m <- df$solar_radiation_kj_m/1000
colnames(df)[3] <- "solar_radiation_mj_m"
#3 ETref calculation
df.Eto <- calculateETrefPM(data = df, lat = −21.980353, alt = 859.29, za = 10,
DAP = 1, date = colnames(df)[1], Ta = colnames(df)[7], G = NULL, RH = colnames
(df)[15], Rg = colnames(df)[3], AP = colnames(df)[4], WS = colnames(df)[18], Kc
= NULL)
```

2.2.5. *Web application framework - shiny*

The developed Shiny web application is organized into four main sections: About, Download Inmet, Data Processing, and Evapotranspiration. Both the Shiny app and the source code are available on GitHub (https://github.com/wagnnerms97/DataMetProcess), allowing users to download, edit, and reuse the code (or portions of it) as a valuable resource for future applications. The application is available in three versions: (1) the standard version that requires the installation of R and all package dependencies (Table 3), (2) a portable version that requires the installation and execution of the application, which is advantageous for maintaining development versions of the application, avoiding errors due to updates or removals of dependencies from CRAN, and (3) a web version on shinyapps.io (https://wagnnermds.shinyapps.io/datametprocess_shiny) that does not require installation and is a convenient alternative for initial users. However, there are limitations in processing due to the limited availability of the server.

The application is divided into several sections to facilitate access to various features and available information. Each section is designed to meet a specific need, ensuring an intuitive and efficient user experience.

1) About: The "About" section was developed to provide users with information about the developed Shiny application, including general information such as a brief description, authors, citation

Algorithm 4
calculateETrefPM().

Input:
data → Data frame containing meteorological data
lat → Latitude (decimal degrees)
alt → Altitude (meters)
za → Anemometer height (meters)
DAP → Days after planting (default: 1)
date → Name of the column with dates
Ta → Name of the column with air temperature (°C)
G → Name of the column with soil heat flux (optional, default: 0)
RH → Name of the column with relative humidity (%)
Rg → Name of the column with global radiation ($MJ/m^2/day$)
AP → Name of the column with atmospheric pressure (hPa)
WS → Name of the column with wind speed (m/s)
Kc → Name of the column with crop coefficient (optional)
Processing:
IF G is not provided **THEN** create a new column $G = 0$
Generate auxiliary columns:
Day of the year extracted from date
Create a growing sequence starting from the initial value
Calculate climatic parameters:
Saturation vapor pressure (kPa)
Slope of the vapor pressure curve (kPa/ °C)
Actual vapor pressure (kPa)
Atmospheric pressure (atm) converted to kPa
Psychrometric constant (kPa/ °C)
Wind speed adjusted to 2 m (m/s)
Calculate astronomical parameters:
Relative Earth-Sun distance
Solar declination
Sunset hour angle
Extraterrestrial radiation
Calculate radiation:
Clear-sky radiation
Net shortwave radiation
Net longwave radiation
Total net radiation
Calculate Reference Evapotranspiration (ET_0) using the Penman-Monteith equation
IF Kc is provided **THEN** calculate Crop Evapotranspiration (ET_c) as $ET_c = ET_0 * Kc$
Output:
Return a data frame with the following columns:
date → Date
etref → Reference evapotranspiration (mm/day)
JD → Day of the year, i.e., Julian Day (JD)
DAP → Days after planting
es, ea, delta, y, rn → Intermediate parameters
etc → Crop evapotranspiration (line 22 applicable)
End

Table 3
Shiny app dependencies.

Package	Version	Reference	Functionality
shiny	1.8.0	[20]	Web Application Framework for R
shinydashboard	0.7.2	[21]	Create Dashboards
shinyjs	2.1.0	[22]	Perform JavaScript operations
shinybusy	0.3.2	[23]	Busy Indicators and Notifications
shinyalert	3.0.0	[24]	Create Popup Messages (Modals)
shinyWidgets	0.8.0	[25]	Custom Inputs Widgets
reactable	0.4.4	[26]	Interactive Data Tables
plotly	4.10.3	[27]	Create Interactive Web Graphics via 'plotly.js'
ggplot2	3.4.4	[14]	Create Data Visualizations
tidyverse	2.0.0	[28]	Set of packages for data science
zoo	1.8–12	[29]	Manipulation and analysis of time series data
janitor	2.2.0	[30]	Simple Tools for Examining and Cleaning Dirty Data

recommendation, contact, acknowledgments, and useful links (e.g., CRAN, GitHub, article DOI) (Fig. 1).

2) Download Inmet: The "Download Inmet" section is responsible for implementing the **list_inmet()** function through a selector containing the available years. It returns a table on the Shiny screen containing the data available in the INMET database for the year selected by the user, allowing data download through the selection made directly in the table. The section is designed to allow only one selection at a time, facilitating user interaction and preserving the autonomy of INMET (Fig. 2).
3) Data Processing: The "Data Processing" section was developed for the implementation of the **adjustDate()** and **calculateDMY()** functions, divided into two subsections: Load File and Process. The Load File is intended for reading the file, containing various settings to ensure proper file reading by the internal R system (e.g., separator, missing values, adjustment of column names, and date formats), including the time zone correction by **adjustDate()** (Fig. 3).

The process subsection is responsible for applying the **calculateDMY()** function, containing a navigation panel to access the specific area for each scale (daily, monthly, and annual), where each level unlocks as each step is completed, starting with daily data. To allow the user to determine which columns will be grouped by sum and average, two selection areas are built from the columns of the file loaded by the user, ensuring that the column selected for sum cannot be selected for average, and vice versa, ensuring the proper functioning of the application (Fig. 4).

After processing both, the plots area is released on the panel above (Results), intended for building interactive graphs with various design settings, allowing interaction, customization, and downloading of the graphs (Fig. 5). For example, it is possible to select between the three available data scales, define a variable for each axis, switch between types of graphs (i.e., line, point, point and line), and define colors for the elements.

4) Evapotranspiration

The evapotranspiration section was developed similarly to the previous section, containing the subsections Load File and ETref calculation. The Load File subsection is similar to that of the Data Processing section (Fig. 6), while the "ETref calculation" subsection is responsible for implementing the **calculateETrefPM()** function, receiving all the necessary parameters for the function's operation through various selectors.

3. Illustrative examples

To demonstrate the major functionalities of the developed code, we present the use case of the **list_inmet()** function, designed to facilitate interaction with meteorological datasets from the Brazilian National Institute of Meteorology (INMET). Upon processing the input data, the function outputs a list containing three key elements: File Addresses Table: A structured table providing the addresses of individual files for subsequent extraction using the **unzip()** function from the utils package. Metadata Table: A table that organizes metadata extracted from the file names, such as city, station code, year, start date, and end date of data collection. Main File Address: The address of a consolidated file that encompasses all available data (Table 4). The function operates

Fig. 1. Initial screen of the Shiny *DataMetProcess* application.

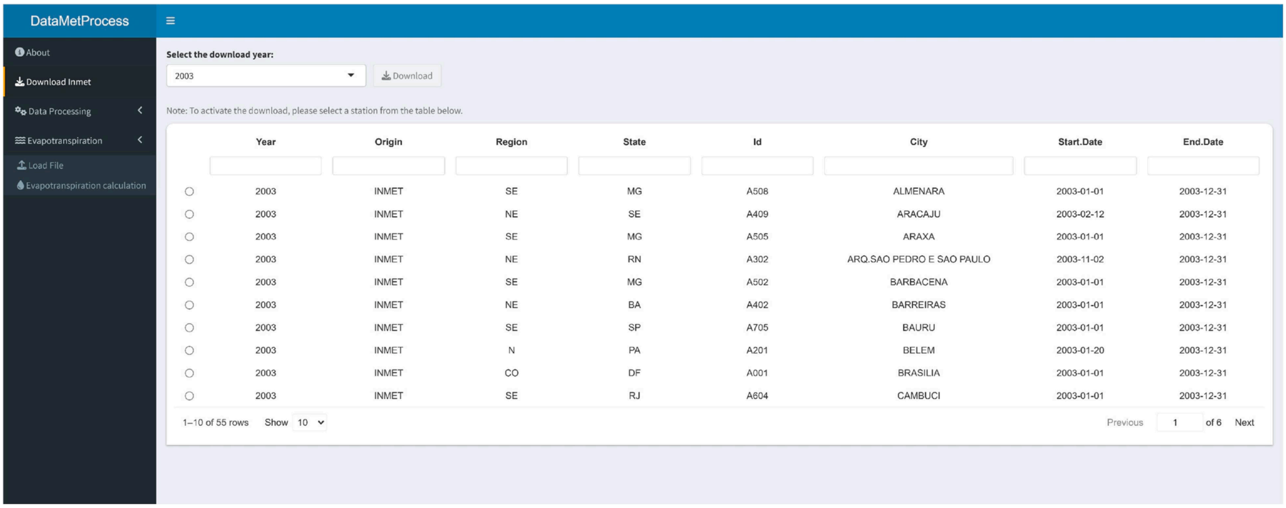

Fig. 2. Download section of INMET data in the Shiny *DataMetProcess* application.

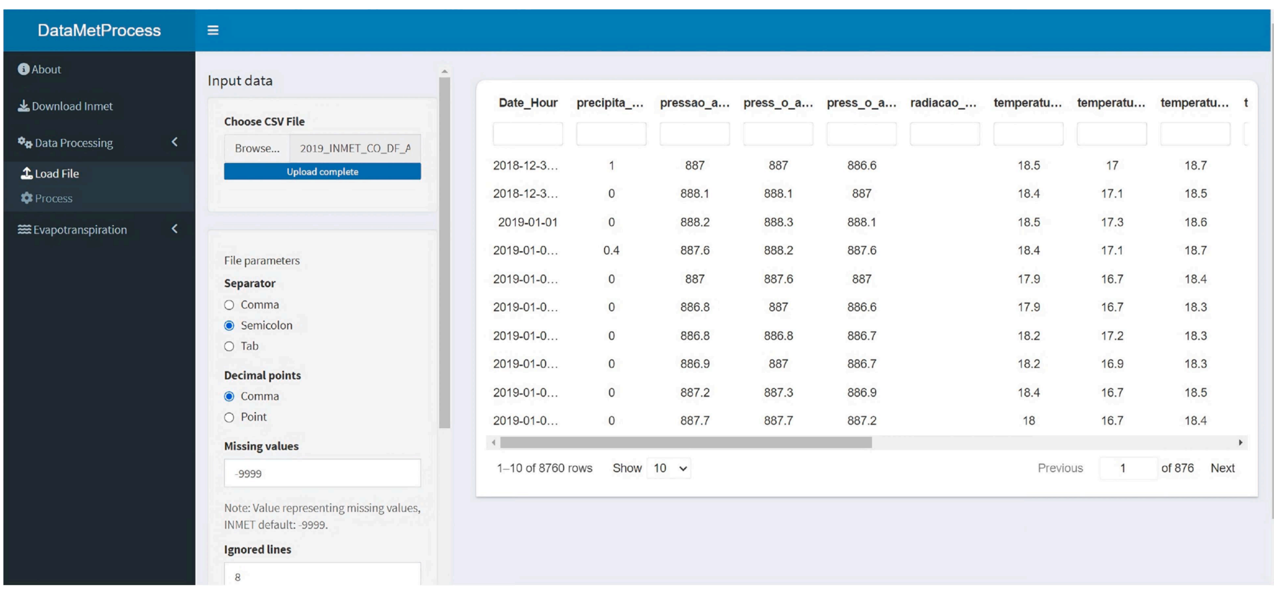

Fig. 3. Data loading section for processing in the Shiny *DataMetProcess* application.

effectively within the context of INMET's datasets, adhering to strict file naming conventions. These conventions include a prescribed order of attributes: origin, region, state, station ID, city, start date, and end date. This dependency on consistent naming patterns renders the function less adaptable to alternative databases where such conventions are absent or differ significantly.

The application of the **adjustDate()** and **calculateDMY()** functions provides an efficient workflow for processing meteorological data. These functions were utilized to analyze daily temperature data, enabling a comparative study of different Brazilian biomes during the year 2019. The selected biomes and corresponding representative locations include Pantanal: Corumbá; Amazon: Manaus; Caatinga: Petrolina; Pampa: Porto Alegre; Cerrado: Uberaba and Atlantic Forest: Vitória. This approach offers a comprehensive visualization of the temperature dynamics within each biome, while facilitating inter-regional comparisons. Such analyses are instrumental in advancing climate and environmental research, providing insights into biome-specific temperature patterns and their broader implications (Fig. 7). The entire data processing procedure can be consulted in the https://github.com/wagnnerms97/DataMetProcess/tree/main/examples.

The developed functions are easily adaptable to other databases, such as those provided by CLIMA.AML (Metropolitan Meteorological Monitoring and Alert Network) in Portugal [31] and MeteoNet, a set of meteorological data provided by Météo-France [32]. For this study, data from two meteorological stations located in Lisbon (Portugal) and Bernières (France) were used. The data were collected at sub-hourly intervals, every 30 min for Portugal and every 6 min for France, and subsequently processed on a daily scale using the **calculateDMY()** function (Fig. 8). The entire data processing procedure can be consulted in the https://github.com/wagnnerms97/DataMetProcess/tree/main/examples.

A performance evaluation study of the **adjustDate()** and **calculateDMY()** functions was conducted to assess the time required to process different time intervals due to the complexity of the processing steps involved in these functions. The time required for data reading, time zone correction, and calculations for daily, monthly, and annual data was recorded. The test was conducted on a notebook with an Intel® Core™ i5–10210 U processor (2.11 GHz), 8 GB DDR4–2666 MHz RAM, and an NVIDIA GeForce MX330 2GB graphics card, running Windows 11 Pro version 22H2. The total number of years was defined as 30,

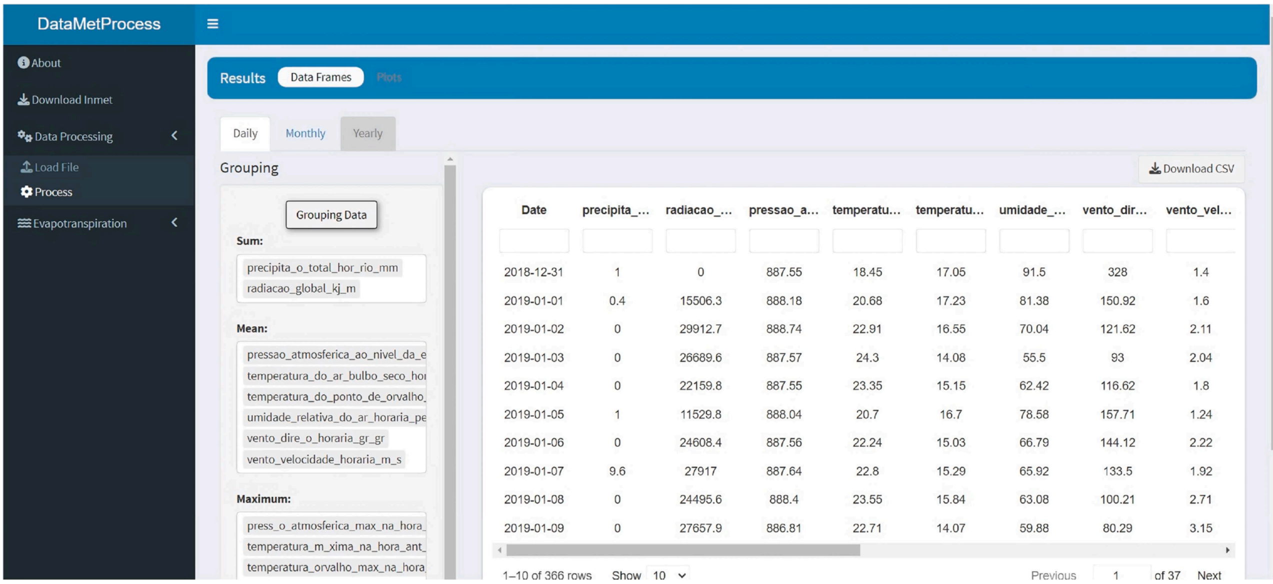

Fig. 4. Data processing section for daily, monthly, or annual grouping in the Shiny *DataMetProcess* application.

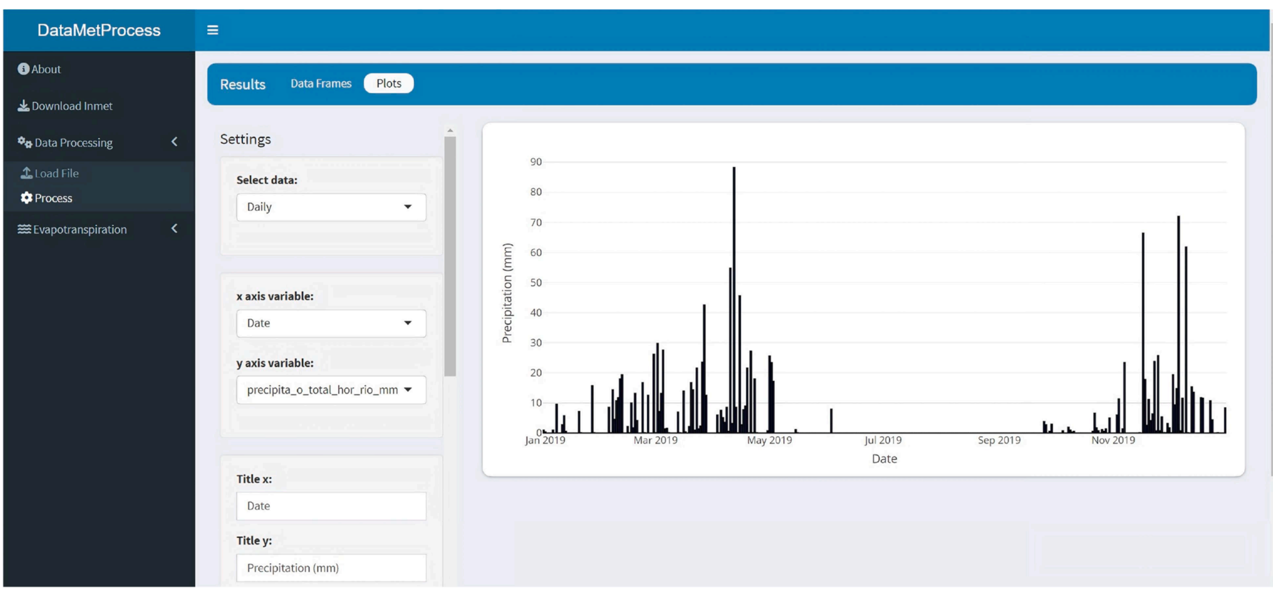

Fig. 5. Section for plotting daily, monthly, or annual data in the Shiny *DataMetProcess* application.

corresponding to the period necessary to determine the climatological normal. Due to the lack of available datasets with the complete number of years for testing, simulated databases were used. These were evaluated regarding the processing time required for periods of 5, 10, 15, 20, 25, and 30 years, in 10 repetitions. Two types of datasets were considered (1) Multiple files: a set of 30 years of data distributed across individual files, i.e., 30 files; (2) Grouped files: a set of data grouped into six files corresponding to the total number of years evaluated. As expected, the processing time for individual data (Multiple files) increased with the number of years, reaching a maximum of 10.5 ± 0.7 s due to the time required to read the files separately. Overall, the functions demonstrated efficiency across different data volumes, particularly in determining the climatological normal (Fig. 9).

4. Impact

The **DataMetProcess** package stands out by offering a comprehensive and intuitive solution for processing meteorological data, eliminating the need for manual conversions and reducing the complexity of analysis. Unlike other available tools, it allows for the direct conversion of data across different temporal scales, time zone adjustments, and integration with a user-friendly web interface based on Shiny. This enables researchers and professionals to focus on data interpretation and strategic decision-making without having to invest significant time in programming or raw data processing. Moreover, its open-source nature allows for customization and the expansion of functionalities, making it a flexible and adaptable choice for various projects.

DataMetProcess was developed in the R environment for acquiring and processing meteorological data provided by the National Institute of Meteorology (INMET - Brazil). However, its structure allows for adaptation to other databases, provided that the necessary instructions are supplied. The web interface, built with Shiny, further enhances accessibility, enabling individuals without programming experience to

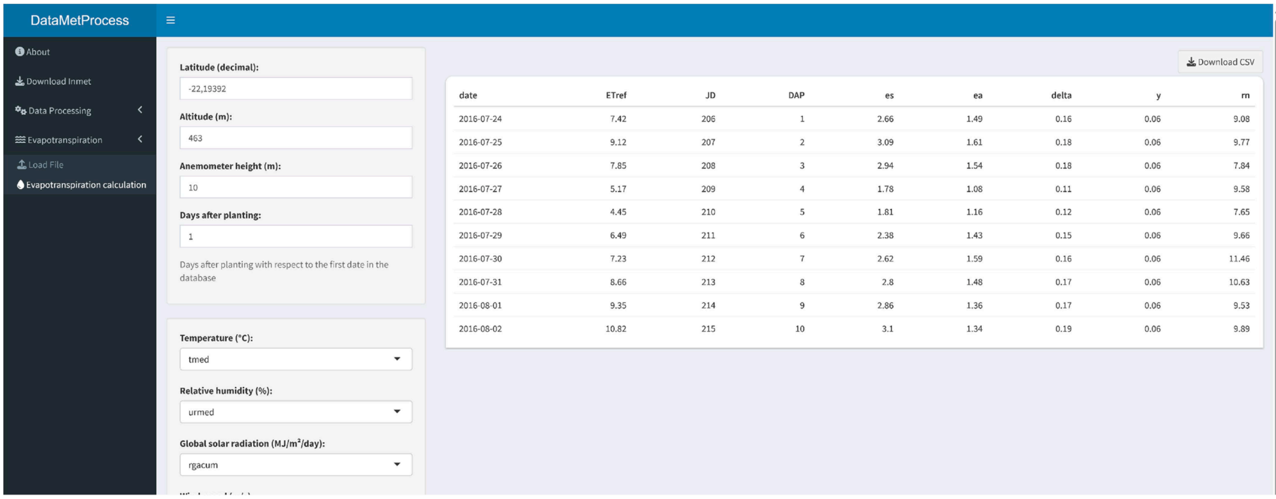

Fig. 6. Section for calculating reference or crop evapotranspiration in the Shiny *DataMetProcess* application.

Table 4
Summary example of the result from **list_inmet()** for the year 2019 in Brazilian INMET.

Item 1
Name
2019/
2019/INMET_CO_DF_A001_BRASILIA_01–01–2019_A_31–12–2019.CSV
2019/INMET_CO_DF_A042_BRAZLANDIA_01–01–2019_A_31–12–2019.CSV
2019/INMET_CO_DF_A045_AGUAS EMENDADAS_01–01–2019_A_31–12–2019.CSV
(…)
Item 2

Year	Origin	Region	State	Id	City	Start. Date	End. Date
2019	INMET	NE	BA	A422	ABROLHOS	01/01/2019	31/12/2019
2019	INMET	NE	CE	A360	ACARAU	01/01/2019	31/12/2019
2019	INMET	SE	ES	A657	AFONSO CLAUDIO	01/01/2019	31/12/2019

(…)
Item 3
C:\Users\{user}\AppData\Local\Temp\RtmpUZp29P\file3d444fa63977

process data interactively and intuitively from any internet-connected device, without requiring installation.

The automation of essential processes, such as time zone corrections and evapotranspiration calculations using the Penman-Monteith model, significantly reduces the time and effort required for data preparation, allowing users to focus on more in-depth analyses. As an open-source project, **DataMetProcess** is not a static product; its evolution will depend on contributions from the scientific community and the developers themselves, with new features being incorporated as needed. With this approach, it has the potential to reach a broad audience, including academics, students, and professionals in environmental sciences, establishing itself as an essential tool for meteorological analyses and strategic decision-making.

5. Conclusions

The development of the **DataMetProcess** package and the Shiny web application represents a significant advancement in the accessibility and analysis of meteorological data. By integrating functions for data extraction, time zone correction, temporal scale adjustment, and evapotranspiration calculation, combined with an open-source interface, it not only democratizes access to information but also facilitates analysis by users with varying levels of technical expertise. The open-

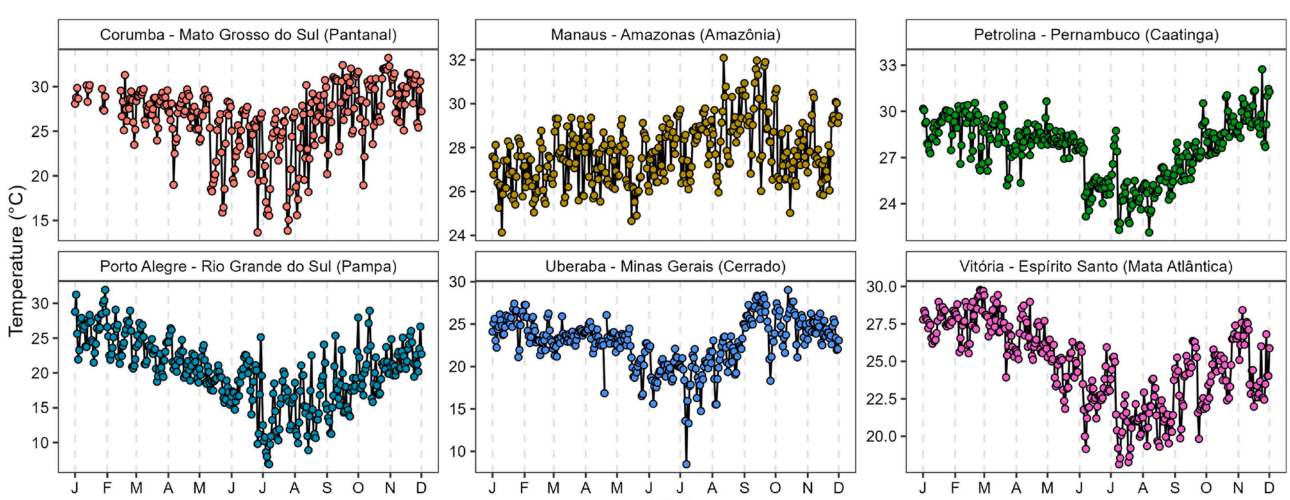

Fig. 7. Visualization of daily air temperature variation in different Brazilian biomes throughout 2019 after processing hourly data into daily data using the functions **adjustDate()** and **calculateDMY()**.

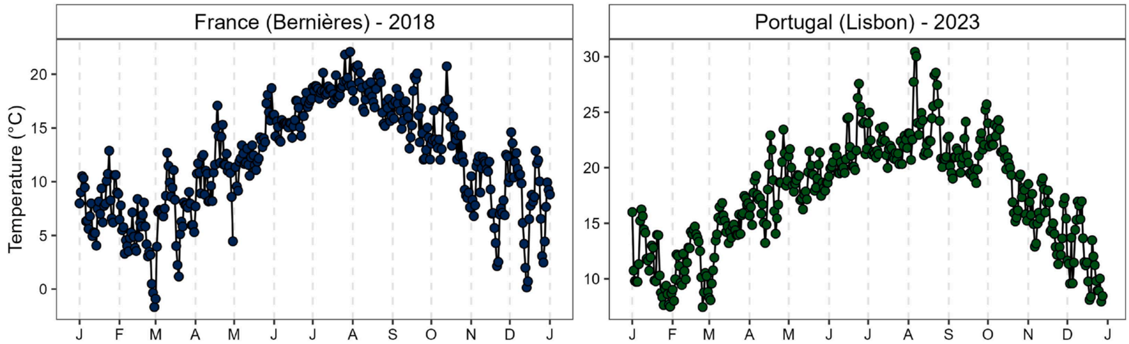

Fig. 8. Visualization of daily air temperature variation in Bernières (France) and Lisbon (Portugal) throughout 2018 and 2023, respectively, after processing hourly data into daily data using the function **calculateDMY()**.

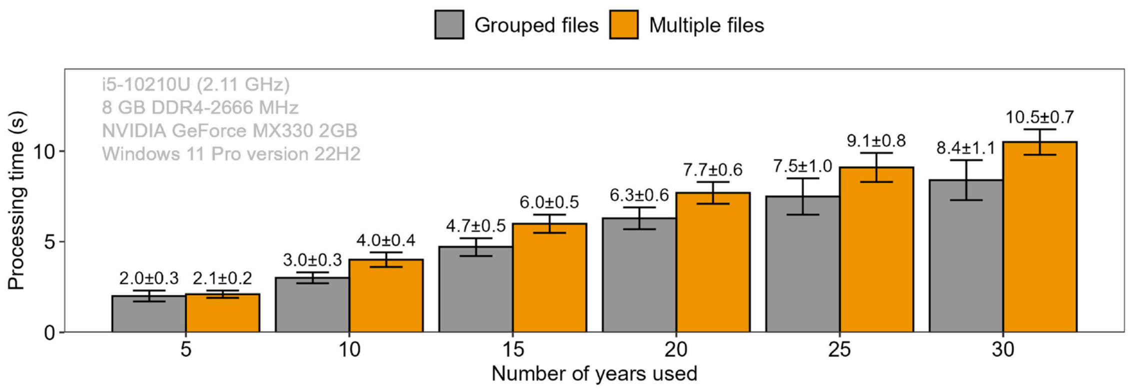

Fig. 9. Processing time for files containing different numbers of years. Multiple files: a set of 30 years of data distributed across individual files, i.e., 30 files. Grouped files: a set of data grouped into six files corresponding to the total number of years evaluated.

source nature of **DataMetProcess** enables the scientific community to adapt and expand its functionalities, fostering a more collaborative and innovative research ecosystem. Future work may explore the integration of data from other meteorological sources and the application of machine learning techniques to predict climate patterns, further enhancing the impact of this software. This work underscores the importance of open and collaborative solutions for environmental research and management.

CRediT authorship contribution statement

Wagner Martins dos Santos: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Hoi Leong Lee:** Writing – review & editing, Writing – original draft, Visualization, Software, Resources, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Edimir Xavier Leal Ferraz:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Abelardo Antônio de Assunção Montenegro:** Writing – review & editing, Methodology, Investigation, Conceptualization. **Lady Daiane Costa de Sousa Martins:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Alan Cézar Bezerra:** Writing – review & editing, Supervision, Methodology, Investigation. **Ênio Farias de França e Silva:** Writing – review & editing, Supervision, Methodology, Investigation. **Thieres George Freire da Silva:** Writing – review & editing, Supervision, Resources, Project administration. **João L.M.P. de Lima:** Writing – review & editing, Supervision, Resources, Project administration. **Xuguang Tang:** Writing – review & editing, Writing – original draft, Visualization, Software, Resources, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Alexandre Maniçoba da Rosa Ferraz Jardim:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This study was supported by the 2025 Universiti Malaysia Perlis (UniMAP) Scientific Publication Grant. The authors would like to thank the Postgraduate Program in Agricultural Engineering (PGEA) and the Federal Rural University of Pernambuco (UFRPE), Brazil, for their support and provision of equipment for the development of this research. We acknowledge the Coordination for the Improvement of Higher Education Personnel (CAPES - Finance Code 001), Brazil, the Portuguese Foundation for Science and Technology (FCT), the University of Coimbra, Portugal, the National Natural Science Foundation of China, and the São Paulo Research Foundation—FAPESP (grant number 2023/05323-4), Brazil, for fellowships and financial support for this study. We also thank the Editor-in-Chief and all anonymous reviewers for their comments and suggestions on how to improve the manuscript.

References

[1] Nouri M, Homaee M, Pereira LS, Bybordi M. Water management dilemma in the agricultural sector of Iran: a review focusing on water governance. Agric Water Manag 2023;288:108480.
[2] Szewrański S, Świąder M, Kazak JK, Tokarczyk-Dorociak K, van Hoof J. Socio-environmental vulnerability mapping for environmental and flood resilience assessment: the case of ageing and poverty in the City of Wrocław, Poland. Integr Environ Assess Manag 2018;14:592–7.
[3] Czernecki B, Głogowski A, Nowosad J. Climate: an R package to access free in-situ meteorological and hydrological datasets for environmental assessment. Sustainability 2020;12:394.
[4] Assis DMS, Franco VS, Dias TSS, Sodré GRC, Tavares-Martins ACC, Godoy BS. Local perceptions do not follow rainfall trends: a case study in traditional Marajo island communities (eastern para state, BR). Heliyon 2023;9:e15497. -e15497.
[5] Júnior JC, da S, Medeiros V, Garrozi C, Montenegro A, Gonçalves GE. Random forest techniques for spatial interpolation of evapotranspiration data from Brazilian's Northeast. Comput Electron Agric 2019;166:105017.
[6] Teixeira de Aguiar J, Lobo M. Reliability and discrepancies of rainfall and temperatures from remote sensing and Brazilian ground weather stations. Remote Sens Appl Soc Environ 2020;18:100301.
[7] Atkins JW, Stovall AEL, Alberto Silva C. Open-Source tools in R for forestry and forest ecology. For Ecol Manage 2022;503:119813.
[8] Liu C, Cui Y, Li X, Yao M. microeco: an R package for data mining in microbial community ecology. FEMS Microbiol Ecol 2021:97.
[9] R Core Team. R: A language and environment for statistical computing. 2024. Available: https://www.r-project.org/.
[10] Guijarro JA. climatol: climate tools (Series homogenization and derived products). 2024. https://doi.org/10.32614/CRAN.package.climatol.
[11] Campitelli E. metR: tools for easier analysis of meteorological fields. 2021. https://doi.org/10.5281/zenodo.2593516.
[12] Chamberlain S, Hocking D. rnoaa: "NOAA" weather data from R. 2025. Available: https://docs.ropensci.org/rnoaa/.
[13] Filgueiras R, Venancio LP, Aleman CC, da Cunha FF. BrazilMet: download and processing of automatic weather stations (AWS) data of INMET-Brazil. 2025. Available: https://cran.r-project.org/package=BrazilMet.
[14] Wickham H. ggplot2: elegant graphics for data analysis. 3rd ed. New York: Springer-Verlag; 2016 Available: https://ggplot2.tidyverse.org.
[15] Wickham H, François R, Henry L, Müller K, Vaughan D. dplyr: A grammar of data manipulation. 2023. Available: https://cran.r-project.org/package=dplyr.
[16] Wickham H, Vaughan D, Girlich M. tidyr: tidy messy data. 2024. Available: https://cran.r-project.org/package=tidyr.
[17] Grolemund G, Wickham H. Dates and times made easy with lubridate. J Stat Softw 2011;40:1–25. https://doi.org/10.18637/jss.v040.i03.
[18] Henry L, Wickham H. rlang: functions for base types and core R and "Tidyverse" features. 2024. Available: https://cran.r-project.org/package=rlang.
[19] Allen RG, Pereira LS, Raes D, Smith M. Crop evapotranspiration - guidelines for computing crop water requirements. editor. FAO irrigation and dranaige. Roma, Itália: FAO; 1998. p. 56.
[20] Chang W, Cheng J, Allaire JJ, Sievert C, Schloerke B, Xie Y, Allen J, McPherson J, Dipert A, Borges B. shiny: web application framework for R. 2024. Available: https://cran.r-project.org/package=shiny.
[21] Chang W, Borges Ribeiro B. shinydashboard: create dashboards with "Shiny". 2021. Available: https://cran.r-project.org/package=shinydashboard.
[22] Attali D. shinyjs: easily improve the user experience of your shiny apps in seconds. 2021. Available: https://cran.r-project.org/package=shinyjs.
[23] Meyer F, Perrier V. shinybusy: busy indicators and notifications for "Shiny" applications. 2024. Available: https://cran.r-project.org/package=shinybusy.
[24] Attali D, Edwards T. shinyalert: easily create pretty popup messages (Modals) in "Shiny". 2024. Available: https://cran.r-project.org/package=shinyalert.
[25] Perrier V, Meyer F, Granjon D. shinyWidgets: custom inputs widgets for shiny. 2024. Available: https://cran.r-project.org/package=shinyWidgets.
[26] Lin G. reactable: interactive data tables for R. 2023. Available: https://cran.r-project.org/package=reactable.
[27] Sievert C. Interactive web-based data visualization with R, plotly, and shiny. Chapman and Hall/CRC; 2020. Available: https://plotly-r.com.
[28] Wickham H, Averick M, Bryan J, Chang W, D' L, Mcgowan A, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Lin Pedersen T, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H. Welcome to the Tidyverse. J Open Source Softw 2019;4:1686. https://doi.org/10.21105/JOSS.01686.
[29] Zeileis A, Grothendieck G. Zoo: S3 infrastructure for regular and irregular time series. J Stat Softw 2005;14:1–27. https://doi.org/10.18637/JSS.V014.I06.
[30] Firke S. janitor: simple tools for examining and cleaning dirty data. 2023. Available: https://cran.r-project.org/package=janitor.
[31] AML Meteorologia. n.d. Available: https://clima.aml.pt/page/publico. Accessed 17 Jun 2024.
[32] Larvor G, Berthomier L, Chabot V, Le Pape B, Pradel B, Perez L. MeteoNet: an open reference weather dataset for AI by météo-france. 2021. Available: https://ams.confex.com/ams/101ANNUAL/meetingapp.cgi/Paper/382915. Accessed 17 Jun 2024.