**hoboR: An R package to summarize and manipulate weather station data.**

Alcalá Briseño R.I., Carson A.R., Lan Y.-H., Peterson E., Grünwald, N.J., and LeBoldus, J.M.

**Summary**

Meteorological records captured by weather stations and data loggers can accumulate large amounts of digital information. Local and regional efforts to create weather station networks, as well as projects oriented to the study of microclimates and smart agriculture, have increased the use of weather stations (Estévez et al., 2011; Lembrechts et al., 2021; Hachimi et al., 2022). These large and complex datasets can be difficult to analyze for non-expert users. HOBO data (ONSET, United Kingdom) loggers are among the most popular weather stations because they are relatively inexpensive and easy to use. A HOBO graphic user interface exists but is incompatible with post-process data analysis and large data sets. As a result, we developed the hoboR, an R package to read, combine, and manipulate CSV files. These algorithms support the removal of redundant data, create meteorological summaries by time and date, support the identification of sensor failures and out-of-range values, and the calculation of summary statistics (i.e., max, min, mean, and standard deviation).

**Statement of need**

HoboR is an R package (R Core Team, 2024) for efficiently processing large datasets, automating these tasks by loading the CSV files into a data frame structure specific to HOBO dataloggers, significantly reducing the time and effort required for data handling and management, increasing the accuracy and reproducibility of the analysis. HoboR functions remove duplicate entries, summarize the data by time intervals (minutes, hours, and/or days), and subset files by user-determined ranges. The package can also identify and address common data quality and accuracy issues related to sensor failures, out-of-range entries, time zone discrepancies, and data formats.

Developing software to automate data processing collected by weather stations and data loggers can facilitate the analysis of local weather and microclimate patterns. Typically, this data is used to correlate meteorological measurements with weather depending on biological processes, species compositions, and a variety of agricultural applications (Hachimi et al., 2022; Dahl et al., 2023; Nikolauo et al., 2023; Wu et al., 2023). Traditional spreadsheet interfaces pose an inherent challenge when handling data from large and complex studies, making managing and curating these datasets time-consuming and error-prone. In many cases, the large file sizes make spreadsheet-based interfaces unable to handle an entire dataset and/or very difficult to remove redundant data. The integration of advanced algorithms and user-friendly software makes hoboR accessible to researchers and programmers with differing levels of experience, facilitating the implementation of the analysis of meteorological data to ecology, agriculture and other meteorological science improving data wrangling. Other packages to analyze weather data exist in Turbo Pascal (Pickering et al., 1994), and R packages have been developed to analyze weather data captured by satellites, including NASA Power and rnoa (Sparks 2018, Chamberlain and Hocking, 2023). To our knowledge, no R packages are available for collecting and analyzing large meteorological data sets collected from HOBO weather stations.

Weather station data can be logged at various time intervals for different types of sensors, including rain gauges, temperature, relative humidity (RH), leaf wetness, and solar radiation, among others. The main functions of hoboR implement dynamic interpretation programming, which enables the processing of independent spreadsheets with any number of sensors given by the HOBO weather station or data logger, allowing the data to be arranged into an initial column structure. The package output includes the summary statistics of the minimum, maximum, mean, and standard deviation values and can be rounded to the nearest minute, hour, or day. Other functions were developed to help summarize the data by time intervals and different ranges of dates. Among the challenges in recording meteorological data are the various errors that occur during data collection. These errors could include damage to the logger, debris blocking the sensors, battery replacement, and malfunctioning sensors or loggers. These issues can result in multiple entries that might be challenging and time-consuming to detect, correct, and curate in tabular format. Hfunctions theseIn addition, we provide a which is important for studying microclimatesA schematic representation of the processing pipeline is illustrated in Fig. 1.

A diagram of a computer

Description automatically generated

Fig. 1) A workflow for the hoboR package, describing the recommended steps for effective data analysis using HOBO weather stations and data loggers. The process initiates by aggregating CSV files from a directory corresponding to each location, followed by summarizing duplicate entries and parsing the data chronologically. Subsequent steps include direct analysis of the data or further manipulation of the data by time and range, complemented by a pre-processing stage for quality assessment to detect impossible values. Optional calibration steps are also illustrated to enhance data accuracy. Solid lines indicate the standard workflow for HOBO data analysis, while discontinuous lines represent the optional calibration process.

**Example**

A test dataset is provided with the hoboR package. This data set was collected in Brookings, Oregon, between August and December 2021 (Fig. 2), as part of the Carson (2022) thesis at the Botany and Plant Pathology Department at Oregon State University. We tested the package using partial datasets from different weather stations and data loggers. A full dataset consists of millions of entries. The code is reproduced below.

```R

*# load the library*

library(hoboR)

*# Analysis*

*# Add the PATH to your sites for weather data (from HOBO)*

path **=** paste0("~/Desktop/Adam/site\_12\_date\_adj2/")

*# make sure the path to your CSV files exists*

file.exists(path) *# this will return a logical value TRUE*

*# loading all hobo files*

hobofiles **<-** hobinder(path, skip **=** 1)

hobocleaned **<-** hobocleaner(hobofiles, format **=** "yymd")

head(hobocleaned)

*# getting hobo mean summary by time*

hobot5 **<-** hobotime(hobocleaned, summariseby **=** "30 mins", na.rm **=** T)

hobomeans5 **<-** meanhobo(hobot5, summariseby **=** "1 day", na.rm **=** T)

head(hobomeans5)

*# getting hobo means by date*

hobomeans **<-** meanhobo(hobocleaned, summariseby **=** "24 h", na.rm **=** T)

head(hobomeans)

*# Specify a window range*

timerange **<-** hoborange(hobocleaned, start**=**"2022-08-08", end**=**"2022-12-12")

head(timerange)

*# Snapshot of a time interval*

a **<-** timestamp(hobocleaned, stamp **=** "2022-08-05 00:01", by **=** "24 hours",

days **=** 100, na.rm **=** **FALSE**, plot **=** F, var **=** "Temp")

a**$**Group **<-** **rep**("night", nrow(a))

b **<-** timestamp(hobocleaned, stamp **=** "2022-08-05 12:01", by **=** "24 hours",

days **=** 100, na.rm **=** **FALSE**, plot **=** F, var **=** "Temp")

b**$**Group **<-** **rep**("day", nrow(b))

daynight **<-** rbind(a, b)

ggplot(daynight, aes(x **=** Date, y **=** Temp, group **=** Group, color **=** Group)) **+**

geom\_line() **+**

scale\_x\_datetime() **+**

scale\_y\_continuous(limits **=** **c**(0, 30)) **+**

scale\_color\_manual(values **=** **c**("orange", "black")) **+**

labs(color **=** "Source") **+**

theme\_minimal()

```

A graph showing a line of orange and black lines

Description automatically generated

Figure 2: Temperature trends from August to December in Brookings, Oregon, processed using the hoboR package. This graph displays daily, and nightly temperatures recorded by HOBO weather stations, with the orange line representing daytime temperatures and the black line representing nighttime temperatures, demonstrating the diurnal temperature variations.

This package requires R version 4.3.0 or later. It also requires the following packages:

data.table, lubridate, dplyr, plyr, reshape, and ggplot2. These dependencies should be installed automatically when dependencies = TRUE is set in the command used to install the

package.

```R

> if (!require("devtools"))

> install.packages("devtools")

> devtools::install\_github("leboldus\_lab/hoboR", dependencies = TRUE)

```

**Authors contributions**

Ricardo I. Alcalá Briseño developed and maintained the package, wrote the documentation, debugged the code, and wrote the manuscript. Adam R. Carson collected the data, wrote code implemented in the package, and debugged the code. Sky Lan collected the data, wrote code implemented in the package, and assisted in the user-functionality of the code functions. Ebba Peterson assisted in best practices for post-processing weather stations and data loggers. Niklaus J. Grunwald participated in manuscript preparation and funding. Jared M. LeBoldus supervised the project, manuscript preparation, and funding.

**Acknowledgements**

United States Department of Agriculture Co-operative agreement number: 58-2072-1-039

**Disclosure**

Authors declare no conflict of interest.

**References**

@article{Estevez2011,

title = {Guidelines on validation procedures for meteorological data from automatic weather stations},

journal = {Journal of Hydrology},

volume = {402},

number = {1},

pages = {144-154},

year = {2011},

issn = {0022-1694},

doi = {https://doi.org/10.1016/j.jhydrol.2011.02.031},

url = {https://www.sciencedirect.com/science/article/pii/S0022169411001594},

author = {J. Estévez and P. Gavilán and J.V. Giráldez},

keywords = {Quality control, Validation, Agrometeorology, Weather network},

}

@article{Lembrechts2021,

author = {Lembrechts, Jonas J. and Lenoir, Jonathan and R. Scheffers, Brett and De Frenne, Pieter},

title = {Designing countrywide and regional microclimate networks},

journal = {Global Ecology and Biogeography},

volume = {30},

number = {6},

pages = {1168-1174},

keywords = {biogeography, climate change, data loggers, location selection, microclimate, microweather, sensors, temperature},

doi = {https://doi.org/10.1111/geb.13290},

url = {https://onlinelibrary.wiley.com/doi/abs/10.1111/geb.13290},

eprint = {https://onlinelibrary.wiley.com/doi/pdf/10.1111/geb.13290},

year = {2021}

}

@Article{Hachimi2023,

Author = {Hachimi, Chouaib El and Belaqziz, Salwa and Khabba, Saïd and Sebbar, Badreddine and Dhiba, Driss and Chehbouni, Abdelghani},

Title = {Smart Weather Data Management Based on Artificial Intelligence and Big Data Analytics for Precision Agriculture},

Journal = {Agriculture},

Volume = {13},

Year = {2023},

Number = {1},

Article-Number = {95},

url = {https://www.mdpi.com/2077-0472/13/1/95},

issn = {2077-0472},

doi = {10.3390/agriculture13010095}

}

@Manual{R2024,

title = {R: A Language and Environment for Statistical Computing},

author = {{R Core Team}},

organization = {R Foundation for Statistical Computing},

address = {Vienna, Austria},

year = {2024},

url = {https://www.R-project.org/},

}

@article{Pickering1994,

author = {Pickering, N. B. and Hansen, J. W. and Jones, J. W. and Wells, C. M. and Chan, V. K. and Godwin, D. C.},

title = {WeatherMan: A Utility for Managing and Generating Daily Weather Data},

journal = {Agronomy Journal},

volume = {86},

number = {2},

pages = {332-337},

doi = {https://doi.org/10.2134/agronj1994.00021962008600020023x},

url = {https://acsess.onlinelibrary.wiley.com/doi/abs/10.2134/agronj1994.00021962008600020023x},

eprint = {https://acsess.onlinelibrary.wiley.com/doi/pdf/10.2134/agronj1994.00021962008600020023x},

}

@article{Sparks2018,

doi = {10.21105/joss.01035},

url = {https://doi.org/10.21105/joss.01035},

year = {2018},

publisher = {The Open Journal},

volume = {3}, number = {30},

pages = {1035},

author = {Adam H. Sparks},

title = {nasapower: A NASA POWER Global Meteorology, Surface Solar Energy and Climatology Data Client for R},

journal = {Journal of Open Source Software} }

@Manual{Chamberlain2024,

title = {rnoaa: 'NOAA' Weather Data from R},

author = {Scott Chamberlain and Daniel Hocking},

year = {2024},

note = {R package version 1.4.0,

https://github.com/ropensci/rnoaa (devel)},

url = {https://docs.ropensci.org/rnoaa/ (docs)},

}

@article{Dahl2023,

author = {Dahl, Mathilde Borg and Kreyling, Juergen and Petters, Sebastian and Wang, Haitao and Mortensen, Martin Steen and Maccario, Lorrie and Sørensen, Søren J. and Urich, Tim and Weigel, Robert},

title = {Warmer winters result in reshaping of the European beech forest soil microbiome (bacteria, archaea and fungi)—With potential implications for ecosystem functioning},

journal = {Environmental Microbiology},

volume = {25},

number = {6},

pages = {1118-1135},

doi = {https://doi.org/10.1111/1462-2920.16347},

url = {https://enviromicro-journals.onlinelibrary.wiley.com/doi/abs/10.1111/1462-2920.16347},

eprint = {https://enviromicro-journals.onlinelibrary.wiley.com/doi/pdf/10.1111/1462-2920.16347},

year = {2023}

}

@article{Nikolaou2023,

title = {Improved daily estimates of relative humidity at high resolution across Germany: A random forest approach},

journal = {Environmental Research},

volume = {238},

pages = {117173},

year = {2023},

issn = {0013-9351},

doi = {https://doi.org/10.1016/j.envres.2023.117173},

url = {https://www.sciencedirect.com/science/article/pii/S0013935123019771},

author = {Nikolaos Nikolaou and Laurens M. Bouwer and Marco Dallavalle and Mahyar Valizadeh and Massimo Stafoggia and Annette Peters and Kathrin Wolf and Alexandra Schneider},

keywords = {Relative humidity, Spatiotemporal modeling, Machine learning, External validation, Exposure assessment, Environmental epidemiology},

}

@article{Wu2023,

author = {Wu, Carrie and Powers, John M and Hopp, David Z and Campbell, Diane R},

title = "{Effects of experimental warming on floral scent, display and rewards in two subalpine herbs}",

journal = {Annals of Botany},

pages = {mcad195},

year = {2023},

month = {12},

issn = {0305-7364},

doi = {10.1093/aob/mcad195},

url = {https://doi.org/10.1093/aob/mcad195},

eprint = {https://academic.oup.com/aob/advance-article-pdf/doi/10.1093/aob/mcad195/55122486/mcad195.pdf},

}