wcswatin: An R Package for Processing Multi-Source Weather and Climate Data for SWAT Hydrological Modeling

Reginal Exavier [1]¶ **and Peter Zeilhofer** [2]

1 University of São Paulo (USP), Postgraduate Program in Physical Geography, São Paulo - SP, Brazil.
2 Federal University of Mato Grosso (UFMT), Postgraduate Program in Geography, Cuiabá - MT, Brazil.
¶ Corresponding author

Summary

Weather and climate variables constitute essential boundary conditions for hydrological modeling frameworks; however, the acquisition, harmonization, preprocessing, and quality control of heterogeneous datasets originating from multiple sources remain computationally intensive and susceptible to systematic and random errors. wcswatin is an R package developed to standardize and automate the preprocessing of multi-source weather and climate datasets, with a specific emphasis on generating rigorously formatted meteorological inputs for the Soil and Water Assessment Tool (SWAT), a globally established, process-based ecohydrological model.

The package provides integrated tools for extracting, cleaning, validating, and formatting Global Weather and Climate Grids (GWCG), as well as for processing measured station data through spatial interpolation. By supporting common data formats including NetCDF, GeoTIFF, and CSV, wcswatin enables consistent and reproducible preparation of weather and climate inputs for SWAT modeling. Figure 1 illustrates the main processing pipelines implemented in the package.
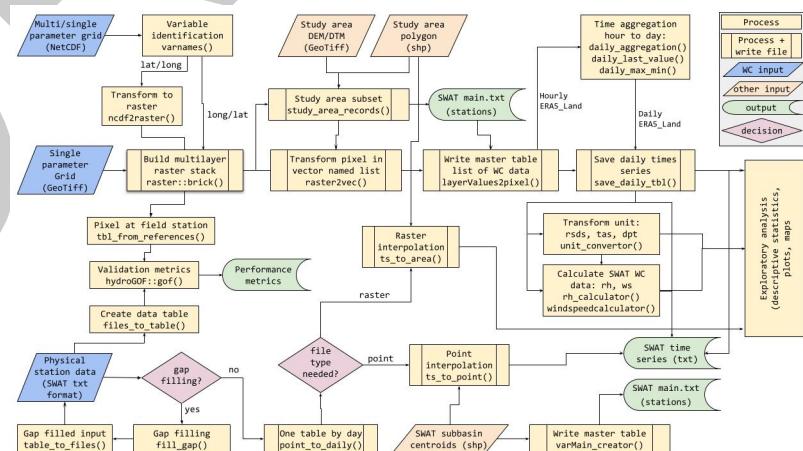
Figure 1: Processing pipelines implemented in the wcswatin R package for the handling of multi-source weather and climate datasets, exemplified using hourly ERA5-Land reanalysis data (Muñoz Sabater 2019).

Through the implementation of a unified preprocessing workflow and integrated validation procedures for the systematic comparison of gridded datasets with in situ observations, wcswatin

24 lowers methodological and technical barriers for hydrologists and environmental scientists,
25 particularly in applications requiring the integration of multiple climate data sources.

Statement of need

27 The Soil and Water Assessment Tool (SWAT) (Arnold et al., 2012) is widely used for watershed-
28 scale hydrological modeling worldwide (Tan et al., 2020) and requires accurate weather and
29 climate inputs, including precipitation, temperature, solar radiation, wind speed, and relative
30 humidity. While Global Weather and Climate Grids (GWCG) derived from reanalysis products,
31 satellite missions, and other modeled datasets provide increasingly high-resolution alternatives
32 to sparse station networks, their preprocessing for SWAT remains technically demanding.

33 These datasets typically require multiple processing steps prior to model application, including
34 temporal aggregation, unit conversion, spatial subsetting, validation against in situ observations,
35 and conversion into model-specific input formats. Performing these steps manually or using
36 multiple disconnected tools is time-consuming and increases the risk of errors, particularly
37 when comparing alternative climate products or working with large spatio-temporal datasets.

38 wcswatin addresses these challenges by providing a generalized and integrated workflow for
39 preprocessing multisource weather and climate data for SWAT modeling. The package supports
40 the processing of gridded datasets in common formats, offers specialized functionality for
41 compiling complete SWAT-compatible weather datasets from ERA5-Land, and includes tools for
42 gap-filling, spatial interpolation, and validation using measured station data. By consolidating
43 these tasks within a single R-based environment, wcswatin facilitates reproducible and efficient
44 preparation of weather and climate inputs for hydrological and ecohydrological studies

Functionality

46 The wcswatin package supports the extraction, preprocessing, exploratory analysis, spatial
47 interpolation, and validation of weather and climate data from multiple sources, with a focus
48 on generating SWAT-compatible inputs.

49 For gridded datasets, the package provides functions to process Global Weather and Climate
50 Grids supplied in NetCDF or GeoTIFF formats. This includes automated handling of temporal
51 resolution, aggregation of hourly data to daily values, unit conversion, and derivation of
52 additional variables required by SWAT, such as wind speed and relative humidity. These
53 capabilities have been applied to multi-parameter reanalysis products including ERA5-Land.

54 In addition to gridded data, wcswatin supports the use of measured in situ station data. Station
55 observations can be imported in standard SWAT format, screened for missing values, gap-filled
56 using regression-based approaches, and spatially interpolated using trend surface analysis.
57 The resulting interpolated datasets can be generated as point-based or gridded outputs and
58 subsequently converted into SWAT-compatible weather inputs. Figure 2 demonstrates an
59 example of the spatial interpolation capabilities, showing a third-order trend surface interpolation
60 for daily precipitation data.
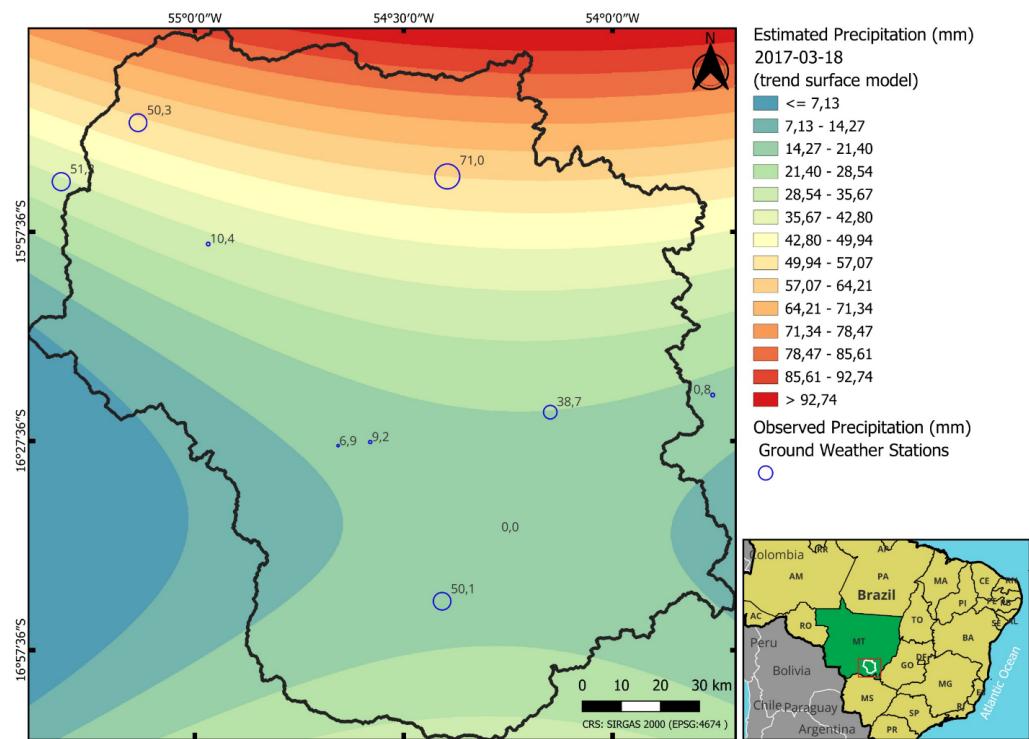
Figure 2: Example of third-order trend surface interpolation for precipitation data (March 18, 2017) using wcswatin's spatial interpolation function.

The package also includes tools for validating gridded climate products against station measurements through statistical metrics and visual inspection, enabling users to compare alternative weather and climate datasets prior to hydrological modeling. Performance optimizations, including parallel processing and reduced memory overhead, allow wcswatin to handle large spatio-temporal datasets efficiently on standard computing systems.

Related work

Several existing tools partially support the preparation of weather and climate data for hydrological modeling. The NASAaccess R package (Ibrahim, 2023) provides functionality for downloading and reformatting selected NASA datasets for use in models such as SWAT, but is limited to a small number of variables and data sources. The pRecipe package (Godoy & Markonis, 2023) focuses on global precipitation datasets and visualization but does not offer comprehensive workflows for generating complete SWAT weather inputs.

Web-based tools like the NCEP Climate Forecast System Reanalysis portal (https://global-weather.tamu.edu/) are restricted to specific time periods (1979-2014) and have demonstrated weak performance in discharge simulations (Senent-Aparicio et al., 2021). In contrast, wcswatin emphasizes integrated preprocessing, validation, and formatting of multisource weather and climate data, supporting both gridded and station-based inputs within a unified and extensible workflow tailored to SWAT modeling.

Acknowledgements

References

Arnold, J. G., Moriasi, D. N., Gassman, P. W., Abbaspour, K. C., White, M. J., Srinivasan, R., Santhi, C., Harmel, R. D., Griensven, A. van, Van Liew, M. W., Kannan, N., & Jha, M. K. (2012). SWAT: Model use, calibration, and validation. *Transactions of the ASABE*, *55*(4), 1491–1508. https://doi.org/10.13031/2013.42256

Godoy, M. R. V., & Markonis, Y. (2023). pRecipe: A global precipitation climatology toolbox and database. *Environmental Modelling & Software*, *165*, 105711. https://doi.org/10.1016/j.envsoft.2023.105711

Ibrahim, M. (2023). *NASAaccess: Downloading and reformatting tool for NASA earth observation data products*. National Aeronautics; Space Administration, Goddard Space Flight Center. https://github.com/nasa/NASAaccess

Senent-Aparicio, J., Blanco-Gómez, P., López-Ballesteros, A., Jimeno-Sáez, P., & Pérez-Sánchez, J. (2021). Evaluating the potential of GloFAS-ERA5 river discharge reanalysis data for calibrating the SWAT model in the grande san miguel river basin (el salvador). *Remote Sensing*, *66*(16: 3299). https://doi.org/https://doi.org/10.3390/rs13163299

Tan, M. L., Gassman, P. W., Yang, X., & Haywood, J. (2020). A review of SWAT applications, performance and future needs for simulation of hydro-climatic extremes. *Advances in Water Resources*, *143*, 103662. https://doi.org/10.1016/j.advwatres.2020.103662