# gridwxcomp: A Python package to evaluate and interpolate biases between station and gridded weather data

**John M. Volk**[1¶], **Christian Dunkerly**[1], **Christopher Pearson**[1], **Charles G. Morton**[1], and **Justin L. Huntington**[1]

**1** Desert Research Institute, Reno, USA ¶ Corresponding author

**DOI:** 10.21105/joss.07178

**Software**

- Review
- Repository
- Archive

**Editor:** Hugo Ledoux

**Reviewers:**

- @ArcticSnow
- @dvalters

**Submitted:** 22 July 2024
**Published:** 20 January 2025



## Introduction

Gridded weather data have become increasingly accessible and accurate over the recent decades, and such data enable a variety of applications and research that require spatially continuous and high spatio-temporal resolution data (Muñoz-Sabater et al., 2021; Rasmussen et al., 2023; Thornton et al., 2021). Gridded weather data are often developed from an assimilation of data production and measurement techniques, including the incorporation of *in-situ* observational data networks, land surface modeling techniques, and remote sensing techniques. Although well-curated *in-situ* measurements of weather variables are typically more accurate than gridded products, they provide data at a single point in space, and they involve difficulties and expenses related to deployment and sensor calibration and maintenance, resulting in incomplete spatial and temporal coverage. Gaps in spatial and temporal coverage in *in-situ* weather data are often filled by gridded data products. The increased coverage provided by data filling in gridded datasets comes with the tradeoff of uncertainty and potential for bias (Blankenau et al., 2020) that are introduced by gridded data development and assimilation techniques.

## Statement of Need

Commonly, *in-situ* measurements of weather data are used to validate and assess the bias and uncertainty in their gridded counterparts. Point biases can be interpolated to investigate spatial biases given sufficient density of measurement stations. Maps of spatial bias can subsequently be used to adjust the gridded weather data for the observed bias. `gridwxcomp` was developed to streamline these objectives in a reproducible Python framework.

This package has the functionality to download point data from a variety of gridded meteorological datasets that are hosted on Google Earth Engine (e.g., NLDAS, ERA5, gridMET) and pair those with station data. It also has functionality to make comparison plots, calculate monthly bias ratios and metrics, and interpolate those data to make spatially complete georeferenced raster images of bias between the gridded and station data using multiple interpolation techniques such as inverse distance weighting and linear interpolation. As far as the authors know, this is the only open-source software that accomplishes these tasks.

## Design and Features

`gridwxcomp` is a Python package that consists of five core submodules and two utility submodules (Figure 1). `gridwxcomp` can process the following meteorological variables: air temperature (minimum and maximum), dew point temperature, shortwave radiation, wind speed, vapor pressure, relative humidity (minimum, maximum, and average), and grass (short) and alfalfa (tall) reference evapotranspiration (ET). Daily gridded weather datasets hosted on Google Earth Engine can be accessed and compared against station data using `gridwxcomp` as long as the user has access to the dataset collection. Example public datasets include: CONUS404 (Rasmussen et al., 2023), ERA5-Land (Muñoz-Sabater et al., 2021), gridMET (Abatzoglou, 2013), NLDAS (Mitchell et al., 2004), RTMA (De Pondeca et al., 2011), and spatial CIMIS (Hart et al., 2009).

The Journal of Open Source Software

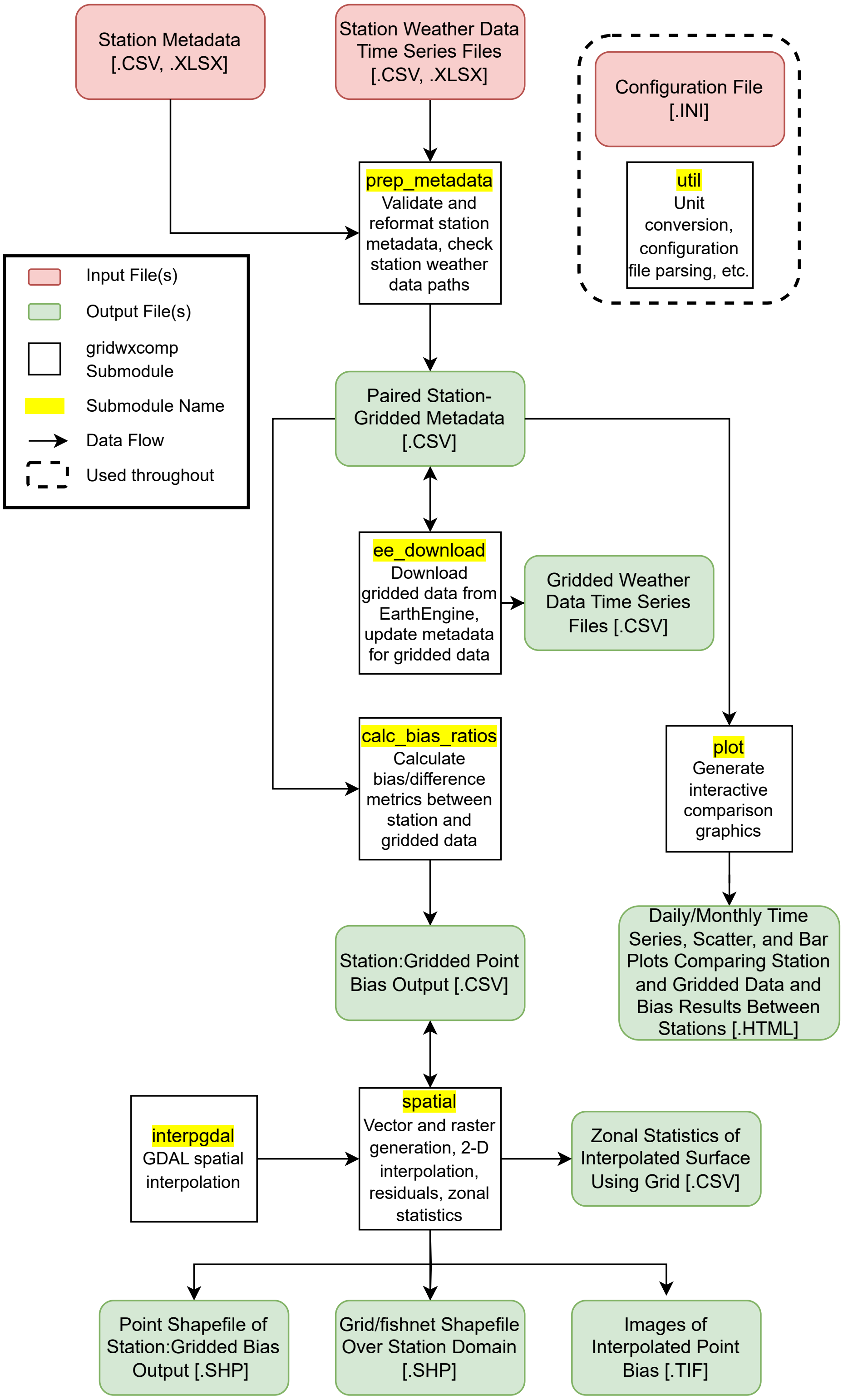


**Figure 1:** Flowchart diagram of submodules and data processing pipeline of `gridwxcomp`.

The prep_metadata submodule parses metadata of meteorological stations, including reprojection of station coordinates. The output file from prep_metadata is used by the ee_download submodule which queries Google Earth Engine for the gridded weather data specified by the user and downloads time series data at the corresponding station coordinates.

The calc_bias_ratios submodule pairs the station and gridded daily data, performs unit conversions, and computes average monthly, seasonal, and annual bias ratios or differences (for temperature variables) between the station and gridded data for a specified variable. In addition to computing station-to-gridded biases, the calc_bias_ratios routine calculates the interannual variability (standard deviation and coefficient of variation) of those metrics and the number of paired data used in each metric.

After computing station-gridded bias metrics, the spatial submodule offers tools to conduct spatial interpolation of the point biases and variability metrics and other spatial mapping functions. The main function of the spatial submodule is the interpolation of the point bias results; options include the algorithms from GDAL's gdal_grid tool. The coordinate reference system and resolution used for spatial interpolation can be specified by the user. In addition, methods for computing zonal statistics using a fishnet grid, generation of a point shapefile, and calculation of point residuals between the bias ratio results at station locations and the overlapping interpolated surfaces are provided in the spatial submodule.

The plot submodule in gridwxcomp provides tools for generating interactive diagnostic plots of both daily and monthly data, pairing the station data alongside its gridded counterpart.

# Research Enabled by gridwxcomp

The most significant application of gridwxcomp was the development of bias correction surfaces that are applied to gridded reference evapotranspiration (ETo) data which are key inputs to some of the remote sensing ET models that comprise the OpenET platform (Melton et al., 2021; Volk et al., 2024). Daily data from approximately 800 weather stations located in irrigated agricultural sites were curated, and the American Society of Civil Engineers (ASCE) standardized Penman-Monteith reference ET equation (Allen et al., 2005) was used to estimate ETo at the stations. Then gridwxcomp was used to pair these data with the nearest ETo data from the gridMET (Abatzoglou, 2013) dataset over temporally consistent periods. The long-term average monthly ratios for station ETo relative to the gridded ETo were calculated for each point and saved as georeferenced data by gridwxcomp and were subsequently spatially interpolated using a kriging approach. The interpolated monthly surfaces are used within the OpenET platform to correct gridMET ETo data before it is used by most of the remote sensing ET models as a major scaling flux.

# Co-author Roles

gridwxcomp was developed through the following efforts:

- John M. Volk: Conceptualization, Software, Validation, Writing & Editing
- Christian Dunkerly: Conceptualization, Software, Validation
- Christopher Pearson: Conceptualization, Software, Validation
- Charles Morton: Conceptualization, Software
- Justin L. Huntington: Conceptualization, Funding acquisition, Resources

# Acknowledgments

We would like to thank the Bureau of Reclamation, NASA Applied Sciences Program, and the Western States Water Use Program at the Desert Research Institute for providing funding for the development of this software. We also thank Sayantan Majumdar for his edits and suggestions.

## References

Abatzoglou, J. T. (2013). Development of gridded surface meteorological data for ecological applications and modelling. *International Journal of Climatology*, *33*, 121–131. https://doi.org/10.1002/joc.3413

Allen, R. G., Walter, I. A., Elliott, R. L., Howell, T. A., Itenfisu, D., Jensen, M. E., & Snyder, R. L. (2005). *The ASCE Standardized Reference Evapotranspiration Equation*. American Society of Civil Engineers. https://doi.org/10.1061/9780784408056

Blankenau, P. A., Kilic, A., & Allen, R. (2020). An evaluation of gridded weather data sets for the purpose of estimating reference evapotranspiration in the United States. *Agric. Water Manag.*, *242*(106376), 106376. https://doi.org/10.1016/j.agwat.2020.106376

De Pondeca, M. S. F. V., Manikin, G. S., DiMego, G., Benjamin, S. G., Parrish, D. F., Purser, R. J., Wu, W.-S., Horel, J. D., Myrick, D. T., Lin, Y., Aune, R. M., Keyser, D., Colman, B., Mann, G., & Vavra, J. (2011). The real-time mesoscale analysis at NOAA's national centers for environmental prediction: Current status and development. *Weather and Forecasting*, *26*(5), 593–612. https://doi.org/10.1175/waf-d-10-05037.1

Hart, Q. J., Brugnach, M., Temesgen, B., Rueda, C., Ustin, S. L., & Frame, K. (2009). Daily reference evapotranspiration for California using satellite imagery and weather station measurement interpolation. *Civil Engineering and Environmental Systems*, *26*(1), 19–33. https://doi.org/10.1080/10286600802003500

Melton, F., Huntington, J., Grimm, R., Herring, J., Hall, M., Rollison, D., Erickson, T., Allen, R., Anderson, M., Fisher, J. B., Kilic, A., Senay, G. B., Volk, J., Hain, C., Johnson, L., Ruhoff, A., Blankenau, P., Bromley, M., Carrara, W., … Anderson, R. G. (2021). OpenET: Filling a critical data gap in water management for the western United States. *JAWRA Journal of the American Water Resources Association*. https://doi.org/10.1111/1752-1688.12956

Mitchell, K. E., Lohmann, D., Houser, P. R., Wood, E. F., Schaake, J. C., Robock, A., Cosgrove, B. A., Sheffield, J., Duan, Q., Luo, L., Higgins, R. W., Pinker, R. T., Tarpley, J. D., Lettenmaier, D. P., Marshall, C. H., Entin, J. K., Pan, M., Shi, W., Koren, V., … Bailey, A. A. (2004). The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system. *Journal of Geophysical Research: Atmospheres*, *109*(D7). https://doi.org/10.1029/2003jd003823

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D. G., Piles, M., Rodríguez-Fernández, N. J., Zsoter, E., Buontempo, C., & Thépaut, J.-N. (2021). ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. *Earth System Science Data*, *13*(9), 4349–4383. https://doi.org/10.5194/essd-13-4349-2021

Rasmussen, R. M., Chen, F., Liu, C. H., Ikeda, K., Prein, A., Kim, J., Schneider, T., Dai, A., Gochis, D., Dugger, A., Zhang, Y., Jaye, A., Dudhia, J., He, C., Harrold, M., Xue, L., Chen, S., Newman, A., Dougherty, E., … Miguez-Macho, G. (2023). CONUS404: The NCAR–USGS 4-km long-term regional hydroclimate reanalysis over the CONUS. *Bulletin of the American Meteorological Society*, *104*, E1382–E1408. https://doi.org/10.1175/BAMS-D-21-0326.1

Thornton, P. E., Shrestha, R., Thornton, M., Kao, S.-C., Wei, Y., & Wilson, B. E. (2021). Gridded daily weather data for North America with comprehensive uncertainty quantification. *Sci. Data*, *8*(1), 190. https://doi.org/10.1038/s41597-021-00973-0

Volk, J. M., Huntington, J. L., Melton, F. S., Allen, R., Anderson, M., Fisher, J. B., Kilic, A., Ruhoff, A., Senay, G. B., Minor, B., Morton, C., Ott, T., Johnson, L., Comini de Andrade, B., Carrara, W., Doherty, C. T., Dunkerly, C., Friedrichs, M., Guzman, A., … Yang, Y. (2024). Assessing the accuracy of OpenET satellite-based evapotranspiration data to support water resource and land management applications. *Nature Water*, *2*(2), 193–205. https://doi.org/10.1038/s44221-023-00181-7