




# Precipitable-Water Model Analysis Tool: An open-source suite for estimating precipitable water with low-cost instrumentation.

Spencer Riley<sup>1</sup>, Vicki Kelsey<sup>2</sup>, and Kenneth Minschwaner<sup>1</sup>

<sup>1</sup> Physics Department, New Mexico Institute of Mining and Technology <sup>2</sup> Atmospheric and Environmental Sciences Program, South Dakota School of Mines and Technology  Corresponding author

DOI: [10.xxxxxx/draft](https://doi.org/10.xxxxxx/draft)

## Software

- [Review](#) 
- [Repository](#) 
- [Archive](#) 

Editor: [Open Journals](#) 

## Reviewers:

- [@openjournals](#)

Submitted: 01 January 1970

Published: unpublished

## License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#)).

## Summary

Precipitable water is the vertically integrated amount of water vapor in a column of air from the surface to the top of the atmosphere ([Salby, 1996](#)). The Precipitable-water Model Analysis Tool (PMAT) is an open-source software suite designed to analyze the correlation between atmospheric brightness temperature and precipitable water. The suite is built using R and Python programming languages and packaged via Docker. The source code for the legacy edition of PMAT has been archived to Zenodo with linked DOI: ([Riley & Kelsey, 2022](#)).

## Statement of Need

Precipitable water data can be used in weather forecasting to determine the amount of potential rainfall available and to study the dynamical evolution of convective storms. The lack of reliable and sufficient precipitable water data can have a significant impact on the quality of weather forecasts ([Smith & Toumi, 2008](#); [Yang & Smith, 2018](#)). PMAT addresses the need for an easily integrated analysis workflow to quantitatively characterize the relationship between regional PWAT and localized atmospheric brightness temperature observations for areas that lack high to moderate-resolution data collection infrastructure. The software suite has already been configured and deployed for use in the South-Central New Mexico climate zone using a series of handheld infrared thermometers as the source of atmospheric brightness temperature data ([Kelsey et al., 2022](#)). As we plan to expand the scope of atmospheric science research, we aim to incorporate community science endeavors in South Dakota.

## Software Architecture and Design

PMAT contains three primary elements: deployment, pre-processing, and analysis. The three components together can configure the container, collect and organize data, conduct primary and secondary analysis functions, then visualize the results.

The prerequisites for the full deployment of PMAT include:

- (1) A data file consisting of measurements that include date and time stamps, ground and sky temperatures, and visual condition of the sky (either “clear sky” or “overcast”)
- (2) A configuration file written in YAML that details sensor information, analysis parameters, and information required to retrieve data

\*co-first author

The deployment mechanism allows end-users to implement the software suite on local and cloud-based systems, which facilitates access to the generated data products and visualizations. We have configured a template repository that illustrates the required directory structure.

The pre-processing module involves aggregating atmospheric brightness temperature data collected in the field with data from radiosondes and local ground stations. The core of this subprocess is a Python script that will read in the raw data and collect atmospheric data from the University of Wyoming Upper-Air database via *siphon* (May et al., 2014 - 2017) in addition to the University of Utah MesoWest database. PMAT supports the usage of external data sources, such that local data files that contain precipitable water and relative humidity data may be processed. This process also includes a series of three primary filters to organize and refine the data. These filters

- separate the data into structures based on the assigned label (“clear sky” or “overcast”)
- removes entries that are manually labeled Do-Not-Analyze or DNA.
- compares the standard deviation of the individual days with the average standard deviation of the dataset

The threshold of the standard deviation filter is defined by the end user’s configuration file.

The analysis module has primary and secondary functionality. First and foremost, the primary analysis function is an iterative regression algorithm that defines a correlation profile between the atmospheric brightness temperature and the precipitable water vapor content. The relationship is characterized by a two-parameter exponential equation that takes the form:

$$\text{PWAT} = Ae^{T_b B}. \quad (1)$$

The result of this process is two output files. One is for the results at each step, and the other is the final averaged regression statistics. The plots associated with the analysis process will utilize the averaged coefficients and statistics.

The secondary analysis functions consist of climatological modeling, time series analysis, and a linear support vector machine. The climatological analysis computes the average value on the same day of the year. The time series analysis includes the computation of regression models for sinusoidal data. This feature is specifically applied to the atmospheric brightness temperature observations as a means to extrapolate the corresponding precipitable water vapor content. Finally, the included linear support vector machine employs the *e1071* R package (Meyer et al., 2022) to classify atmospheric brightness temperature and precipitable water data into the aforementioned label scheme.

At the conclusion of this workflow, the end-users have access to a collection of generated data files, plot sets, and analysis results.

## Acknowledgements

## References

- Kelsey, V., Riley, S., & Minschwaner, K. (2022). Atmospheric precipitable water vapor and its correlation with clear-sky infrared temperature observations. *Atmospheric Measurement Techniques*, 15(5), 1563–1576. <https://doi.org/10.5194/amt-15-1563-2022>
- May, R., Arms, S., Leeman, J., & Chastang, J. (2014 - 2017). *Siphon: A collection of Python utilities for accessing remote atmospheric and oceanic datasets* (Version 0.5.1) [Computer software]. Unidata. <https://doi.org/10.5065/D6CN72NW>
- Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., & Leisch, F. (2022). *e1071: Misc functions of the department of statistics, probability theory group (formerly: E1071), TU wien*. <https://CRAN.R-project.org/package=e1071>

- 80 Riley, S., & Kelsey, V. (2022). *Precipitable-water model analysis tool* (legacy) [Computer  
81 software]. <https://doi.org/10.5281/zenodo.4587475>
- 82 Salby, M. L. (1996). *Fundamentals of atmospheric physics* (1st ed.). Elsevier.  
83 ISBN: 0126151601
- 84 Smith, S., & Toumi, R. (2008). Measuring Cloud Cover and Brightness Temperature with a  
85 Ground-Based Thermal Infrared Camera. *Jamc*, 47(2), 683–693. [https://doi.org/10.1175/](https://doi.org/10.1175/2007JAMC1615.1)  
86 [2007JAMC1615.1](https://doi.org/10.1175/2007JAMC1615.1)
- 87 Yang, L., & Smith, J. (2018). Sensitivity of extreme rainfall to atmospheric moisture content  
88 in the arid/semiarid southwestern united states: Implications for probable maximum  
89 precipitation estimates. *Jgr: Atmospheres*, 123(3), 1638–1656. [https://doi.org/10.1002/](https://doi.org/10.1002/2017JD027850)  
90 [2017JD027850](https://doi.org/10.1002/2017JD027850)

DRAFT