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

Spencer Riley [1,2,¶], **Vicki Kelsey**[3], and **Kenneth Minschwaner**[2]

1 Now at Physics Department, Montana State University, USA **2** Physics Department, New Mexico Institute of Mining and Technology, USA **3** Atmospheric and Environmental Sciences Program, South Dakota School of Mines and Technology, USA **¶** Corresponding author

Summary

Precipitable water (PW) 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 measured at zenith (directly overhead) 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 The source code for the legacy edition of PMAT has been archived to Zenodo with linked DOI: (Riley & Kelsey, 2022).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 is important for setting initial conditions in weather forecasting models, and the accuracy and coverage (both temporal and spatial) of PW measurements greatly impact near term forecasting skill (Marcus et al., 2007). In addition, long term increases in PW have been shown to be associated with warming due to increasing greenhouse gas levels in the atmosphere, leading to a positive feedback on climate (Held & Soden, 2006). PMAT addresses the need for an easily integrated analysis workflow to quantitatively characterize the relationship between regional PW 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 using data on atmospheric brightness temperature obtained from a series of infrared thermometers in the Central New Mexico climate zone (Kelsey et al., 2022). The suite is well suited for community science endeavors designed to augment PW coverage.

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

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 the raw data file defined in Prerequisite (1) with data from National Weather Service balloon launches via siphon (May et al., 2014 - 2017) and surface dewpoint and relative humidity data from local ground stations via the University of Utah's 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:

$$\mathrm{PW} = Ae^{T_b B}. \tag{1}$$

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. 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

Held, I. M., & Soden, B. J. (2006). Robust responses of the hydrological cycle to global warming. *Journal of Climate*, *19*(21), 5686–5699. https://doi.org/10.1175/JCLI3990.1

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

Marcus, S., Kim, J., Chin, T., Danielson, D., & Laber, J. (2007). Influence of GPS precipitable water vapor retrievals on quantitative precipitation forecasting in southern california. *Journal of Applied Meteorology and Climatology*, *46*(11), 1828–1839. https://doi.org/10.1175/2007JAMC1502.1

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

82 Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., & Leisch, F. (2022). *e1071: Misc*
83 *functions of the department of statistics, probability theory group (formerly: E1071), TU*
84 *wien*. https://CRAN.R-project.org/package=e1071

85 Riley, S., & Kelsey, V. (2022). *Precipitable-water model analysis tool* (legacy) [Computer
86 software]. https://doi.org/10.5281/zenodo.4587475

87 Salby, M. L. (1996). *Fundamentals of atmospheric physics* (1st ed.). Elsevier.
88 ISBN: 0126151601