

EGU22-13345

<https://doi.org/10.5194/egusphere-egu22-13345>

EGU General Assembly 2022

© Author(s) 2022. This work is distributed under the Creative Commons Attribution 4.0 License.



A novel approach to systematically analyze the error structure of precipitation datasets using decision trees

Xinxin Sui¹, Zhi Li², Guoqiang Tang³, Zong-Liang Yang⁴, and Dev Niyogi⁴

¹Department of Civil, Architectural and Environmental Engineering, Cockrell School of Engineering, The University of Texas at Austin, Austin, TX 78712, USA

²School of Civil Engineering and Environmental Science, University of Oklahoma, Norman, OK 73072, USA

³Centre for Hydrology, University of Saskatchewan, Canmore, Alberta, Canada

⁴Jackson School of Geosciences, The University of Texas at Austin, Austin, TX 78712, USA

Multiple environmental factors influence the error structure of precipitation datasets. The conventional precipitation evaluation method over-simply analyzes how the statistical indicators vary with one or two factors via dimensionality reduction. As a result, the compound influences of multiple factors are superposed rather than disassembled. To overcome this deficiency, this study presents a novel approach to systematically and objectively analyze the error structure within precipitation products using decision trees. This data-driven method can analyze multiple factors simultaneously and extract the compound effects of various influencers. By interpreting the decision tree structures, the error characteristics of precipitation products are investigated. Three types of precipitation products (two satellite-based: 'top-down' IMERG and 'bottom-up' SM2RAIN-ASCAT, and one reanalysis: ERA5-Land) are evaluated across CONUS. The study period is from 2010 to 2019, and the ground-based Stage IV precipitation dataset is used as the ground truth. By data mining 60 binary decision trees, the spatiotemporal pattern of errors and the land surface influences are analyzed.

Results indicate that IMERG and ERA5-Land perform better than SM2RAIN-ASCAT with higher accuracy and more stable interannual patterns for the ten years of data analyzed. The conventional bias evaluation finds that ERA5-Land and SM2RAIN-ASCAT underestimate in summer and winter, respectively. The decision tree method cross-assesses three spatiotemporal factors and finds that underestimation of ERA5-Land occurs in the eastern part of the rocky mountains, and SM2RAIN-ASCAT underestimates precipitation over high latitudes, especially in winter. Additionally, the decision tree method ascribes system errors to nine physical variables, of which the distance to the coast, soil type, and DEM are the three dominant features. On the other hand, the land cover classification and the topography position index are two relatively weak factors.