

EGU23-13852, updated on 20 Dec 2023 https://doi.org/10.5194/egusphere-egu23-13852 EGU General Assembly 2023 © Author(s) 2023. This work is distributed under the Creative Commons Attribution 4.0 License.



Can machine learning help us to create *improved* and *trustworthy* satellite-based precipitation products?

Ioannis Tsoukalas¹, **Panagiotis Kossieris**¹, Luca Brocca², Silvia Barbetta², Hamidreza Mosaffa², and Christos Makropoulos¹

¹Department of Water Resources & Environmental Engineering, School of Civil engineering, National Technical University of Athens, Greece

²Research Institute for Geo-Hydrological Protection, National Research Council, Perugia, Italy

Key variable of earth observation (EO) systems is precipitation, as indicated by the wide spectrum of applications that is involved (e.g., water resources and early warning systems for flood/drought events). During the last decade, the EO community has put significant research efforts towards the development of satellite-based precipitation products (SPPs), however, their deployment in realworld applications has not yet reached the full potential, despite their ever-growing availability, spatiotemporal coverage and resolution. This may be associated with the reluctancy of end-users to employ SPPs, either worrying about uncertainty and biases inherited in SPPs or even due to the existence of multiple SPPs, whose performance fluctuates across the globe, and thus making it difficult to select the most *appropriate* SPP (some sort of a choice paradox). To address this issue, this work targets the development of an explainable machine learning approach capable of integrating multiple satellite-based precipitation (P) and soil moisture (SM) products into a single precipitation product. Hence, in principle, to create a new dataset that optimally combines the properties of each individual satellite dataset (used as predictors), better matching the groundbased observations (used as predictand, i.e., reference dataset). The proposed approach is showcased via a benchmark dataset consisted of 1009 cells/locations around the world (Europe, USA, Australia and India), highlighting its robustness as well as its application capability which are independent of specific climatic regimes and local peculiarities.