Global observation-based climatology of precipitation occurrence and peak intensity

Hylke Beck¹, Seth Westra², and Eric Wood¹

¹Department of Civil and Environmental Engineering, Princeton University, Princeton, New Jersey, USA
²School of Civil, Environmental and Mining Engineering, University of Adelaide, Adelaide, Australia

We introduce a unique set of global observation-based climatologies of daily precipitation (P) occurrence (related to the lower tail of the P distribution) and peak intensity (related to the upper tail of the P distribution). The climatologies were produced using Random Forest (RF) regression models trained with an unprecedented collection of daily P observations from 93,138 stations worldwide. Five-fold cross-validation was used to evaluate the generalizability of the approach and to quantify uncertainty globally. The RF models were found to provide highly satisfactory performance, yielding cross-validation coefficient of determination ($R^2$) values from 0.74 for the 15-year return-period daily P intensity to 0.86 for the >0.5 mm d⁻¹ daily P occurrence. The performance of the RF models was consistently superior to that of state-of-the-art reanalysis (ERA5) and satellite (IMERG) products. The highest P intensities over land were found along the western equatorial coast of Africa, in India, and along coastal areas of Southeast Asia. Using a 0.5 mm d⁻¹ threshold, P was estimated to occur 23.2 % of days on average over the global land surface (excluding Antarctica). The climatologies including uncertainty estimates will be released as the Precipitation DISTRIBUTion (PDIST) dataset via www.gloh2o.org/pdist. We expect the dataset to be useful for numerous purposes, such as the evaluation of climate models, the bias correction of gridded P datasets, and the design of hydraulic structures in poorly gauged regions.