



## Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling

Hylke E. Beck (1), Noemi Vergopolan (1), Ming Pan (1), Vincenzo Levizzani (2), Albert I.J.M. van Dijk (3), Graham P. Weedon (4), Luca Brocca (5), Florian Pappenberger (6), George J. Huffman (7), and Eric F. Wood (1)  
(1) Department of Civil and Environmental Engineering, Princeton University, Princeton, New Jersey, USA, (2) National Research Council of Italy, Institute of Atmospheric Sciences and Climate (CNR-ISAC), Bologna, Italy, (3) Fenner School of Environment & Society, The Australian National University, Canberra, Australia, (4) Met Office, Joint Centre for Hydro-Meteorological Research, Wallingford, UK, (5) Research Institute for Geo-Hydrological Protection, National Research Council, Perugia, Italy, (6) European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, UK, (7) Mesoscale Atmospheric Processes Laboratory, NASA Goddard Space Flight Center, Greenbelt, Maryland, USA

We undertook a comprehensive evaluation of 22 gridded (quasi-)global (sub-)daily precipitation ( $P$ ) datasets for the period 2000–2016. Thirteen non-gauge-corrected  $P$  datasets were evaluated using daily  $P$  gauge observations from 76 086 gauges worldwide. Another nine gauge-corrected datasets were evaluated using hydrological modeling, by calibrating the conceptual model HBV against streamflow records for each of 9053 small to medium-sized ( $< 50\,000\text{ km}^2$ ) catchments worldwide, and comparing the resulting performance. Marked differences in spatio-temporal patterns and accuracy were found among the datasets. Among the uncorrected  $P$  datasets, the satellite- and reanalysis-based MSWEP-ng V1.2 and V2.0 datasets generally showed the best temporal correlations with the gauge observations, followed by the reanalyses (ERA-Interim, JRA-55, and NCEP-CFSR) and the the satellite- and reanalysis-based CHIRP V2.0 dataset, the estimates based primarily on passive microwave remote sensing of rainfall (CMORPH V1.0, GSMaP V5/6, and TMPA 3B42RT V7) or near-surface soil moisture (SM2RAIN-ASCAT), and finally, estimates based primarily on thermal infrared imagery (GridSat V1.0, PERSIANN, and PERSIANN-CCS). Two of the three reanalyses (ERA-Interim and JRA-55) unexpectedly obtained lower trend errors than the satellite datasets. Among the corrected  $P$  datasets, the ones directly incorporating daily gauge data (CPC Unified and MSWEP V1.2 and V2.0) generally provided the best calibration scores, although the good performance of the fully gauge-based CPC Unified is unlikely to translate to sparsely or ungauged regions. Next best results were obtained with  $P$  estimates directly incorporating temporally coarser gauge data (CHIRPS V2.0, GPCP-1DD V1.2, TMPA 3B42 V7, and WFDEI-CRU), which in turn outperformed the one indirectly incorporating gauge data through another multi-source dataset (PERSIANN-CDR V1R1). Our results highlight large differences in estimation accuracy, and hence, the importance of  $P$  dataset selection in both research and operational applications. The good performance of MSWEP emphasizes that careful data merging can exploit the complementary strengths of gauge-, satellite- and reanalysis-based  $P$  estimates.