Version 2.2 documentation

Hylke Beck (hylkeb@princeton.edu)

Princeton University
Department of Civil and Environmental Engineering
Princeton, NJ, United States

www.gloh2o.org

1 Introduction

This is the documentation for Multi-Source Weighted-Ensemble Precipitation (MSWEP) V2.2, a fully global precipitation dataset (1979–2017) with a 3-hourly temporal and 0.1° spatial resolution. MSWEP takes advantage of the complementary strengths of gauge-, satellite-, and reanalysis-based data to provide reliable precipitation estimates over the entire globe. MSWEP is available for download via `www.gloh2o.org`. The dataset has been comprehensively evaluated byBeck et al. (2017) globally and by Beck et al. (2018) for the conterminous US.

2 Data format

MSWEP V2 is provided in the widely used netCDF-4 format at 3-hourly, daily, and monthly temporal resolution, and 0.1° and 0.5° spatial resolution. The 3-hourly and daily data are supplied as monthly files, while the monthly data are supplied as single files. The actual precipitation estimates are stored in the `precipitation` netCDF field (dimensions 1800 × 3600 for the 0.1° data and dimensions 360 × 720 for the 0.5° data) in mm/3-hour, mm/day, and mm/month units for the 3-hourly, daily, and monthly data, respectively.

3 Reading data

Precipitation data for a particular day (e.g., April 25, 2010) can be read using MATLAB as follows:

```
global_precip = ncread('201004.nc','precipitation',[1 1 25],[Inf Inf 1])';
```

The transposition accent is necessary because MATLAB incorrectly reorders the dimensions when opening netCDF files. The same data are read using Python as follows:

```
from netCDF4 import Dataset
dataset = Dataset('201004.nc', 'r')
global_precip = dataset.variables['precipitation'][24,:,:]
dataset.close()
```

Note that 24 instead of 25 is used for the slicing because Python starts indexing at zero.

4 Version history

Version 2.2 (March 2, 2018)

Initial release corresponding exactly to the description in Beck et al. (2019). This version includes three changes:

1. The dataset was extended to October 31, 2017 (reflecting the availability of ERA-Interim).

2. We conducted some additional quality control of the GridSat infrared archive and now incorporate GridSat-based rainfall estimates from 1980 (rather than 1983) onwards.

3. In version 2.1, the long-term precipitation trends were amplified due to the CDF corrections, particularly in the tropics during the pre-TRMM era. This issue was resolved in version 2.2 by rescaling the estimates to match the trends of the non-CDF corrected data.

Version 2.1 (November 20, 2017)

Two changes. First, NCEP-CFSR precipitation data were removed due to the presence of spurious trends (NCEP-CFSR was added in version 2.0). Second, corrections using the monthly gauge-based GPCC precipitation dataset were reintroduced (these were removed in version 2.0).

Version 2.01 (October 8, 2017)

Fixed a mistake in the netCDF metadata. For all netCDF files, the units for the time variable were changed from “days since 1900-01-01 00:00:00” to “days since

1899-12-31 00:00:00". The actual time data have not been changed. In the case of the 3-hourly data, a (fictional) time value of, for example, 1.125 would indicate that the data represent the period 03:00Z to 05:59Z on January 1, 1900.

Version 2.0 (July 23, 2017)

Major upgrade containing numerous changes in the data sources as well as the merging algorithm, and as a result markedly different precipitation estimates, especially at the daily and 3-hourly time scales (notably less drizzle and higher peaks). The most important changes in version 2 include:

1. The correction of distributional precipitation biases to account for the spurious drizzle and attenuated peaks evident in version 1.
2. Increasing the spatial resolution from 0.25° to 0.1° to increase the local relevance of the precipitation estimates.
3. The inclusion of ocean areas, to enable oceanic studies and avoid missing data in coastal areas.
4. The addition of precipitation estimates derived from GridSat thermal infrared imagery for the pre-TRMM era to supplement the reanalysis and gauge data.
5. Daily gauge corrections for each grid cell based on the five nearest gauges, to replace the coarse 0.5° CPC Unified.
6. The use of a daily gauge correction scheme that accounts for differences in gauge reporting times, to minimize timing mismatches when merging the daily gauge estimates with the satellite and reanalysis data.
7. Extension of the data record to 2017.

Version 1.2 (November 28, 2016)

Despite the many changes in version 1.2, the terrestrial precipitation estimates have not changed considerably since the previous version. The full list of changes is as follows:

1. MSWEP now also provides experimental precipitation estimates for ocean areas. When using MSWEP ocean data it should be kept in mind that: (i) the weights used for the temporal dynamics are almost entirely based on land stations which are not necessarily representative of ocean areas; (ii) the gauge-based data sources (CPC Unified and GPCC) are unavailable over ocean areas; and (iii) the estimates have not been validated (this is work in progress).

   For determining the long-term mean over ocean areas we could not use CHPclim as it only covers land areas. The long-term mean over ocean areas was therefore derived by weighting the long-term means of CMORPH, TMPA 3B42RT, GSMaP-MVK, ERA-Interim, and JRA-55. The weights for CMORPH,

TMPA 3B42RT, and GSMaP-MVK were set to 1 for latitudes $< 25°$ and to 0 for latitudes $> 35°$. The weights decrease linearly from 1 at 25° to 0 at 35°. The weights for ERA-Interim and JRA-55 were set to 1. In the future, we intend to refine the weight estimates over ocean areas to obtain more reliable long-term means.

2. The record has been extended from 2014 to 2015.

3. ERA-Interim data were mistakenly offset by +3 hours in previous versions.

4. The threshold temperature for the inclusion of satellite data has been increased to 5°C for all time scales, to minimize the probability of incorporating potentially erroneous satellite data.

5. Satellite data prior to the year 2000 have been excluded. Among the satellite sources, only CMORPH provides data prior to 2000. However, the data were in lesser agreement with the gauge and reanalysis estimates and have therefore been excluded.

6. For generating the weight maps used for determining the temporal dynamics, the GHCN-D and GSOD station data are now normalized prior to the computation of grid-cell average time series.

7. The weight maps have been produced at 0.25° rather than 0.5°.

8. For each grid cell, rather than normalizing the satellite data, we rescaled the satellite data to match the reanalysis data for the period of overlap, to ensure retainment of the long-term trends.

Version 1.1 (August 2, 2016)

No changes to the actual data, the only changes are in the netCDF formatting. First, we changed the order of the variables from "lon, time, lat" to "time, lat, lon", which should solve some of the problems people have had with reading the data. Second, we corrected the time variable for the daily data, which was mistakenly offset by 1 day.

Version 1.0 (May 30, 2016)

Initial release corresponding exactly to the description in Beck et al. (2017).

5 Acknowledgements

By using MSWEP in any publication you agree to cite Beck et al. (2019). The dataset is being developed by Hylke Beck (Princeton University, Princeton, USA) in collaboration with Ming Pan, Eric Wood (Princeton University, Princeton, USA), Albert van Dijk (ANU, Canberra, Australia), Ad de Roo (JRC, Ispra, Italy), Vincenzo Levizzani (CNR-ISAC, Bologna, Italy), Jaap Schellekens (Deltares, Delft, The Netherlands),

Diego Miralles (VU University Amsterdam, The Netherlands), and Brecht Martens (Ghent University, Ghent, Belgium). We gratefully acknowledge the precipitation dataset developers for producing and making available their datasets. The work was supported through IPA support for Hylke Beck from the U.S. Army Corps of Engineers' International Center for Integrated Water Resources Management (ICIWaRM), under the auspices of UNESCO, to further develop a Latin America and Caribbean Drought Monitor.

References

Beck, H. E., M. Pan, T. Roy, G. P. Weedon, F. Pappenberger, A. I. J. M. van Dijk, G. J. Huffman, R. F. Adler, and E. F. Wood (2018). Daily evaluation of 26 precipitation datasets using Stage-IV gauge-radar data for the CONUS. *Hydrology and Earth System Sciences Discussions*, 1–23.

Beck, H. E., A. I. J. M. van Dijk, V. Levizzani, J. Schellekens, D. G. Miralles, B. Martens, and A. de Roo (2017). MSWEP: 3-hourly 0.25° global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data. *Hydrology and Earth System Sciences 21*(1), 589–615.

Beck, H. E., N. Vergopolan, M. Pan, V. Levizzani, A. I. J. M. van Dijk, G. P. Weedon, L. Brocca, F. Pappenberger, G. J. Huffman, and E. F. Wood (2017). Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling. *Hydrology and Earth System Sciences 21*(12), 6201–6217.

Beck, H. E., E. F. Wood, M. Pan, C. K. Fisher, D. M. Miralles, A. I. J. M. van Dijk, T. R. McVicar, and R. F. Adler (2019). MSWEP v2 global 3-hourly 0.1° precipitation: methodology and quantitative assessment. *Bulletin of the American Meteorological Society in press*.