



GPCC's data base and quality-control for the new gridded global precipitation analyses

Udo Schneider, Meyer-Christoffer Anja, Ziese Markus, Finger Peter, Becker Andreas, and Rudolf Bruno
German Weather Service (DWD), Hydrometeorology, Offenbach a.M., Germany (udo.schneider@dwd.de)

Since its start in 1989 the Global Precipitation Climatology Centre (GPCC) performs global analyses of monthly precipitation for the earth's land-surface on the basis of in-situ measurements. Meanwhile, the data set has continuously grown both in temporal coverage (original start of the evaluation period was 1986), as well as extent and quality of the underlying data base. Due to the high spatio-temporal variability of precipitation, even its global analysis requires a high density of measurement data.

The GPCC receives the daily SYNOP and the monthly climate (CLIMAT) messages in near real-time via the WMO GTS. Core data source of the GPCC analyses are the data from station networks operated by the National Meteorological/Hydrological Services worldwide; data deliveries have been received from ca. 190 countries. The GPCC integrates also other global precipitation data collections (i.e. FAO, CRU and GHCN), as well as regional data sets. Recently the Africa data set from S. Nicholson (Univ. Tallahassee) has been integrated. As a result of these efforts the GPCC holds the worldwide largest and most comprehensive collection of precipitation data, which is continuously updated and extended.

The acquired data sets are pre-checked, reformatted and then imported into a relational data base, where they are archived separately in source specific slots, thus allowing a cross-comparison of data from the different sources. Any time new data sets are imported to the data base the metadata in the input data set are compared to those already available in the data base. In case of discrepancies (e.g. deviating coordinates), external geographical sources of information are utilized to decide whether a correction of the metadata in the data base is required or not, thus resulting in a perpetual improvement of the station metadata.

Since the beginning of 2009 the precipitation data to be imported is compared against a back-ground statistic. Exceptional values are checked and either confirmed, corrected if possible, or excluded from the analyses. This approach requires a high level of manual interaction, due to the complexity of the error analysis. In preparation of GPCC's new analyses the entire data base has been checked statistically, for homogeneity over time and for spatial consistency.

Typical errors identified in the QC processing are factor-10, factor 2.54 or also factor 25.4 errors due to erroneous inch to mm conversions, typing errors as switching of digits, sometimes shifts of data by 1 or more months or even a year. Besides the corrections applied by the GPCC, the original data is also kept, allowing backtracking of the corrections.

In comparative analyses the quality of the data from the different sources was evaluated and ranked. If data from different sources are available for a given station and month/year, the data with the highest quality are selected for the analyses according to this priority scheme.

There is a broad spectrum of requirements coming from different kinds of users, for which different analysis products have been designed and implemented at the GPCC. Since the data base has substantially grown compared to the previous reanalysis in 2008 the GPCC has put intensive efforts into its updates of the publicly available (<http://gpcc.dwd.de>) gridded precipitation climatology and re-analyses data sets, accomplished in December 2010.