Global Precipitation Climatology Centre

Deutscher Wetterdienst

Operational processing, quality-control and analysis of precipitation data at the GPCC

Udo Schneider and Bruno Rudolf, Deutscher Wetterdienst, Offenbach a.M., Germany

1 Background and objectives

The main task of the GEWEX Global Precipitation Climatology Project (GPCP) is the compilation of global gridded precipitation data sets on the basis of the available observing systems, i.e. conventional surface networks and remote sensing data. The products are required for verification of global climate model simulations, the investigation of climate variability and special phenomena such as El Niño/Southern Oscillation, as well as the determination of the Earth's water balance and budgets (WCRP, 1990).

The GPCC (Global Precipitation Climatology Centre) is the in-situ component of the GPCP. Its function within GPCP is the analysis of global land-surface precipitation by interpolation of raingauge data on an operational basis. The specific requirements are:
- High accuracy of the gridded results (desired error of monthly precipitation totals < 10%),
- Gridded analysis results are accompanied by information on analysis quality/error estimates.

The major problems in this context are:
- Errors in the reported precipitation data and station meta information,
- Sampling errors (depending on the variability in the precipitation field and the density of the station network),
- Systematic errors in raingauge-measurements.

Dealing with the difficulties mentioned above the following points are very important:
- A thorough quality-control (QC) of precipitation data <u>and</u> station meta information,
- A high density and a good spatial coverage of the station network,
- Analyses have to be corrected for the systematic gauge-measuring error.

2 GPCC data base

In-situ raingauge measurements still provide the most reliable information to analyse area-mean precipitation for the land-surface. All station meta data, as well as the precipitation data are archived in an Oracle-based relational data base management system (RDBMS). In merging the data from different sources the quality-control (QC) and harmonization of the station meta information is crucial. A thorough QC is necessary to detect errors in the station meta data (especially geogr. coordinates) and a harmonization is required to ensure consistency of time series and to avoid dublettes, as far as possible, in the merged data set.

2.1 Near real-time GTS data

Via the WMO World Weather Watch Global Telecommunication System (GTS) data can be obtained near real-time from synoptic weather reports (at least with a daily resolution) and monthly climate reports. Monthly precipitation data are routinely obtained at GPCC from 3 sources: (1) monthly totals calculated at GPCC from SYNOP reports received at DWD, Offenbach, (2) monthly CLIMAT reports received also at DWD, Offenbach, and (3) monthly totals calculated at CPC/NCEP from SYNOP reports received at NCEP, Washington DC.
The data from these sources, which are checked and merged by GPCC in order to improve the spatial coverage and data quality, form the basis of its monitoring of global monthly precipitation (see section 5). The total number of stations with monthly precipitation data available via GTS has somewhat increased over time and has reached ca. 7,000 stations during recent years (Fig. 1, left).

2.2 GPCC's Full data base

Owing to the large variability of precipitation in space and time the GPCC is acquiring additional precipitation data from national weather services, hydrological institutes etc. to enlarge the data base. So far, institutes from about 160 countries have supplied additional data on a voluntary basis, following WMO requests and bilateral contacts of GPCC. GPCC's full data base includes monthly precipitation totals of more than 50,000 stations for which any monthly precipitation data are available in the period since 1986. According to the GPCP Implementation and Data Management Plan (WCRP, 1990) 1986 originally was defined as the starting year for the evaluation period.

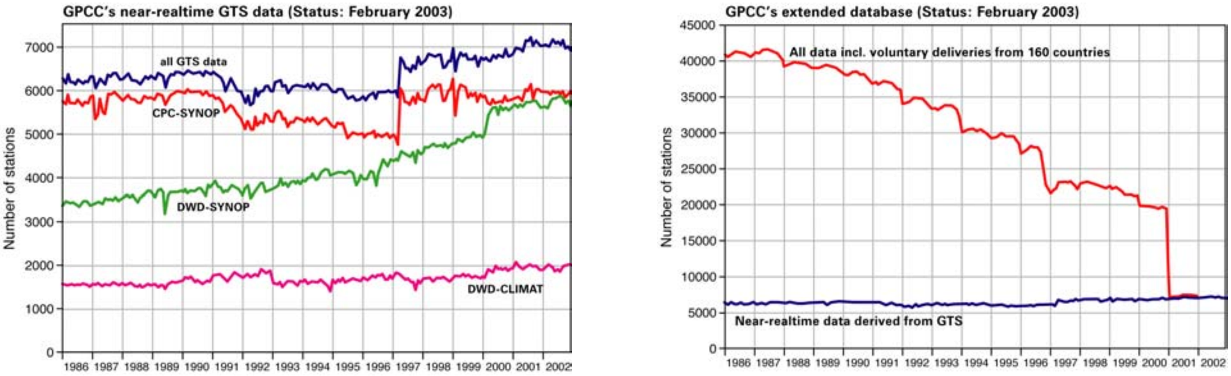

Fig. 1: Availability of precipitation data as a function of time since 1986 for (left) the data received via GTS and (right) GPCC's full data base including the delivered national/regional data collections.

The year with the best data coverage is 1987 with data of ca. 41,000 stations available. The decrease during the last years (ca. 28,000 stations in 1995, less thereafter) is caused by the delay of the data delivery and the time required by the GPCC to process and check these data sets before loading them into the RDBMS.

2.3 Historical extension of GPCC's data base

Following recommendations from GCOS, GEWEX and CLIVAR the GPCC is working on the historical extension of its data base. Therefore GPCC, in co-operation with the University of Frankfurt a.M., in Oct. 2001 started a research project*[1] „Variability Analysis of Surface Climate Observations“ (VASClimO), which is aiming at the following main goals:

- the compilation of a comprehensive global climate data base for precipitation, snow cover, surface air temperature (average and extremes) and mean sea level pressure, including existing historical data collections and additional data to be acquired (performed at GPCC) and
- a detailed statistical analysis of the data set (mainly performed at Univ. Frankfurt a.M.).

The subproject at GPCC can build upon the data base available at the GPCC consisting of precipitation data and meta data of more than 50,000 stations world-wide. However, most of the time series start as late as 1986. Some large data collections already available at GPCC, such as from CRU (Univ. East Anglia, Norwich, UK), GHCN (NCDC, Asheville, NC) and FAO (Food and Agriculture Organization of UN, Rome, Italy) are including a large amount of historical climate data. These data collections will be integrated into the GPCC data base to cover also the historical period (Fig. 2). Some time series are even extending back to the early 18th century.

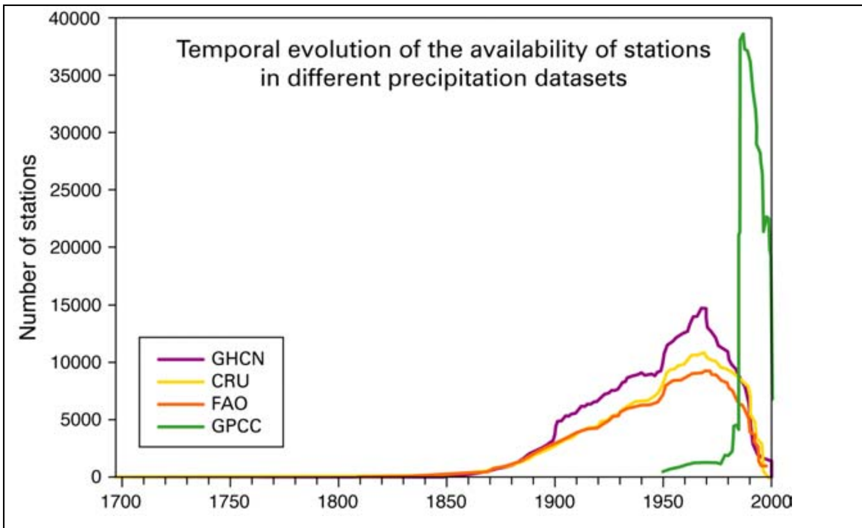

Fig. 2: Availability of precipitation data (number of stations with full year of data) as a function of time for the GHCN, FAO and CRU precipitation data set, as well as for the GPCC data base.

*[1] *The project is sponsored within the German Climate Programme (DEKLIM) by the Federal Ministry for Science and Education of Germany.*

These data sets are currently quality-controlled and loaded into the database (11,868 precipitation time series from CRU, 22,654 precipitation and 7,280 temperature time series of the GHCN, as well as 13,530 precipitation and 5,996 temperature time series of the FAO). Checking the historical data collection from CRU and loading into the GPCC data base has almost been completed. Several significant errors in the station meta data have been detected and could be corrected (examples are given in section 3 and in the presentation). Checking the station meta information of the other historical data collections and loading them into GPCC's data base is in progress.

3 Quality-control and harmonization of station meta information

The GPCC has developed a quality-control (QC) system consisting of different levels. The focus here is lying on the QC of the station meta data. The first step is a pre-control of the station coordinates performed during the pre-processing and reformatting of the individual national/regional data sets and updates delivered to the GPCC using the software CLIDAVIS displaying the station locations and country borders. Station coordinates of the global data collections can be controlled using a polygon check method testing whether the stations of a country are lying within the polygons representing the border of the corresponding country or not.

The central element in checking the meta information (especially the geogr. coordinates) of the stations is the intercomparison of the meta data in the input data files with the station catalogue (more than 50.000 stations) included in the RDBMS during the loading process. The procedure for this will be presented at the meeting and is briefly outlined in the following.

First each station record in the input data file undergoes a check of the validity of the geogr. coordinates, elevation and the country code (Alpha3-code). For each station record in the input data file similar stations are selected from the data base (DB) system. If the station meta data (WMO-ID, Nat. No., geogr. coordinates, name) is identical and the elevation of the DB station is within +-10 m of the station elevation in the data file being loaded, the data are assigned automatically to the corresponding DB station.

Alternatively the selected similar stations are grouped according to the degree of similarity in an “Inner Pool” (stations with good agreement) and an “Outer Pool” (stations with less good agreement). These data pools are evaluated further. In case of a very close matching with the DB stations (i.e. in the geogr. coordinates a difference of 0.01° is tolerated) the data are assigned to the corresponding DB station.

If there is no station with similar meta information in the DB system and no station in the vicinity, a new station record is inserted into the DB.

However, in many cases only part of the station meta information in the input data set is identical to that of DB stations, whereas part of the information differs (for example due to typing errors or rounding effects in the geogr. coordinates, a different spelling of station names etc.). These unclear cases station (“UCS”) have to be clarified manually. The procedure displays all stations with similar meta data selected from the DB on a screen. Then an expert has to decide whether the data can be assigned to a DB station or if a new station record has to be inserted. In some cases the meta information in the data base has to be corrected; however in most cases where discrepancies occur these can be attributed to erroneous meta data in the input data sets (see Fig. 3). All manually clarified cases with an assignment to a DB station are stored in the UCS library, so that the system is learning and exactly the same case hasn't to be treated again.

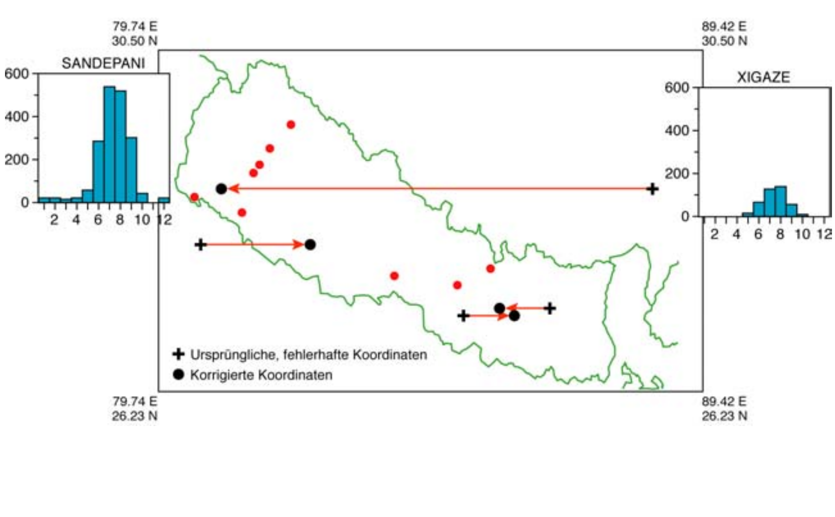

Fig. 3: Example for erroneous station meta data (geogr. coordinates) in Eurasia part of the CRU data set (Nepal).

4 Method used for operational analysis of precipitation at GPCC

GPCC is using a spherical adaptation of SPHEREMAP (Shepard, 1968) for applications on a global scale (after Willmott et al., 1985) for the interpolation of the irregularly distributed raingauge measurements to a regular grid (0.5° resolution). The interpolation method is based on an inverse distance weighting scheme, taking the directional distribution (clustering) of stations also into account.

Following external studies (Legates, 1987, 1991; Bussieres and Hogg, 1989) and internal intercomparison studies (Rudolf et al., 1992, 1994) the SPHEREMAP method was selected and implemented at GPCC in 1991 for operational objective analysis of global precipitation. These studies indicated the SPHEREMAP method being particularly suitable in analysis of global precipitation climatologies. In an intercomparison study of 4 different interpolation schemes (Bussieres and Hogg, 1989) it was the best of the empirical schemes and did a job almost as well as Optimum Interpolation.

The construction of gridded fields of area-average precipitation consists of 2 major steps: (1) interpolation of the irregularly distributed raingauge observations onto points of a regular grid and (2) to convert the grid point values to area-averages for each grid box. The area-average precipitation on a 0.5° grid is calculated by averaging the grid point values at its 4 corners. Area-average precipitation on a coarser grid (1° or 2.5°) is then calculated as the area-weighted average over the corresponding 0.5° grids.

5 GPCC products

Near real-time “Monitoring Product”

The “Monitoring Product” is based on data received via the WMO GTS (ca. 7,000 stations, details see section 2.1). These global analyses of monthly precipitation are generated on a 1° and 2.5° grid and form the in-situ basis of the combined satellite-raingauge data sets of the GPCP (Huffman et al. 1997). Generally the analysis procedure (including a high-level quality-control) is continuing on a routine monthly basis and the monitoring product is available about 1 to 2 months after the observation period. The “Monitoring Product” is now covering the period January 1986 up to Dec. 2002 and can be downloaded via Internet using the redesigned GPCC-Visualizer:
http://gpcc.dwd.de/visu_gpcc.html

Non real-time “Full Data Product”

A re-analysis based upon the full data base („Full Data Product“) has been carried out for the period January 1986 up to December 1995 (ca. 28,000 to 40,000 stations). The results have been calculated on a 0.5°-grid (and 1°-grid) and have been provided to NASA/GSFC for publication on the ISLSCP-II initiative CD-ROM. After future expansions of the data base the re-analysis will be repeated from time to time. A comparative analysis to the Monitoring Product has been performed (section 6).

Both products up to now are accompanied by bulk correction factors to compensate for the systematic gauge-measuring errors, that have been derived by Legates (1987) for climatological conditions. An improved method for the correction of the systematic gauge-measuring error taking into account the weather conditions during the given month is in work.

6 Intercomparison of “Monitoring Product” and “Full Data Product”

The precipitation differences between both products are large where the precipitation itself or the gradient is large, and where the “Full Data Product” is based on significantly more rain-gauges than the “Monitoring Product”.

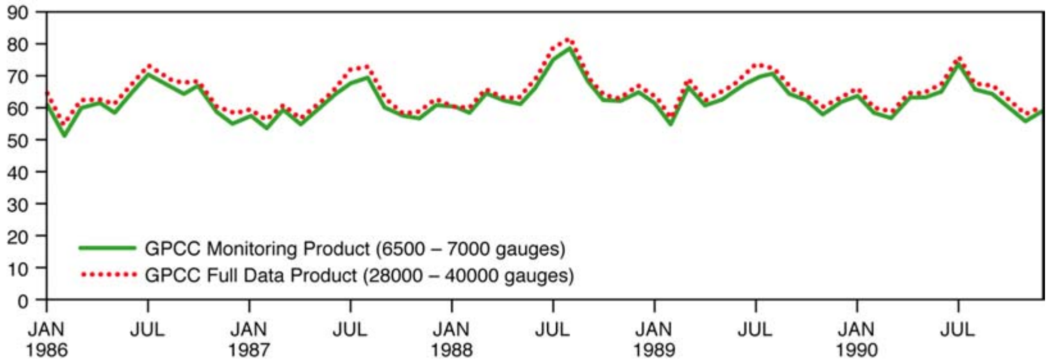

Fig. 4: Area-average monthly precipitation for the earth’s land-surface for the “Monitoring Product” and the “Full Data Product” over the period Jan. 1986 to Dec. 1990.

In Fig. 4 area-average monthly precipitation for the earth's land-surface is shown for the "Monitoring Product" and the "Full Data Product" over the period Jan. 1986 to Dec. 1990. The area-average monthly precipitation for the "Full Data Product" is always slightly higher (up to ca. 5%) than for the "Monitoring Product". One reason may be that the heavy rainfall events, which generally are quite local in extent, are depicted better by a dense station network. Regional differences vary largely depending on the station density (number of raingauges per grid) and the location. This emphasizes the importance of a high density of the station network, too.

7. Conclusions

Experience of GPCC with the processing and evaluation of the many national or regional data sets received, as well as of the global data collections (i.e. CRU, FAO, GHCN) shows that a *thorough QC not only of the precipitation (climate) data, but also of the station meta data is crucial*, because of frequently occuring errors.

Very important in this context is also the *harmonization of station meta data in merging the data sets from different sources* to ensure consistency of time series and to avoid dublettes, as far as possible, to get the best possible data base.

Compensation of the systematic gauge-measuring error is important in conventional measurements.

The *analysis method* SPHEREMAP is in operational use at GPCC since ca. 10 years. The GPCC is planning to implement a new analysis method; the selection of a new analysis technique has to be based upon statistical investigations and intercomparison studies. The meeting will be very helpful in this process. However, a new analysis technique has to be suited for operational application and the analyses for the "Monitoring Product" and the "Full Data Product" will have to be repeated then using the new methodology to create consistent data sets.

References

Bussieres, N. and W. Hogg (1989): The objective analysis of daily rainfall by distance weighting schemes on a mesoscale grid. *Atmos. Oceans*, **27**, 521-541.

Huffman, G.J., R.F. Adler, P.A. Arkin, A. Chang, R. Ferraro, A. Gruber, J. Janowiak, A. McNab, B. Rudolf and U. Schneider (1997): The Global Precipitation Climatology Project (GPCP) Combined Precipitation Dataset. *Bull. American Meteor. Soc*., **78**, 5-20.

Legates, D.R. (1987): A climatology of global precipitation. *Publ. in Climatology* **40** (1), Newark, DW, 85 pp.

Legates, D.R. (1991): Personal communication.

Rudolf, B., H. Hauschild, M. Reiss and U. Schneider (1992): Beiträge zum Weltzentrum für Niederschlags-klimatologie – Contributions to the Global Precipitation Climatology Centre. *Meteorol. Zeitschrift N.F.*, **1**, 7-84.

Rudolf, B., H. Hauschild, W. Rueth and U. Schneider (1994): Terrestrial precipitation analysis: Operational method and required density of point measurements. NATO ASI series, **I 26**, 173-186.

Shepard, D. (1968): A two-dimensional interpolation function for irregularly spaced data. *Proc. 23rd ACM Nat. Conf.*, Brandon/Systems Press, Princeton, NJ, 517-524.

WCRP (1990): The Global Precipitation Climatology Project - Implementation and Data Management Plan. WMO/TD-No. 367.

Willmott, C.J., C.M. Rowe and W.D. Philpot (1985): Small-scale climate maps: A sensitivity analysis of some common assumptions associated with grid-point interpolation and contouring. *Amer. Cartogr*., **12**, 5-16.

More information about GPCC: **http://gpcc.dwd.de**