Precipitation Analysis Based on Gauge Measurements

Bruno Rudolf, Tobias Fuchs, Joerg Rapp and Udo Schneider

Deutscher Wetterdienst, Offenbach a.M., Germany,
Global Precipitation Climatology Centre (GPCC)

Email: bruno.rudolf@dwd.de

Abstract: Precipitation analyses based on raingauge observations are required to validate satellite-based precipitation algorithms, adjust operational satellite precipitation products to a ground truth, and to verify model precipitation output. The Global Precipitation Climatology Centre (GPCC) provides the international climate research community with global gridded precipitation data sets derived from raingauge observations. The problems to deal with are data availability, quality-control, and error assessment (systematic measuring errors of raingauges, sampling errors). On global scale, it seems to be possible to achieve a mostly sufficient spatial and temporal coverage of monthly precipitation data. The accuracy of error-corrected monthly analyses can be described by a random error of 10 % to 20 % The international availability of daily precipitation data is more restricted due to the large amount of data, which would be required, and the resulting processing problems. Therefor, it seems to be the only realistic and practicable way, to limit high resolution precipitation analyses to selected periods and regions to be defined with respect of the scale concerned.

1. Introduction

In situ observed precipitation data are the basis for verification and validation of precipitation estimates derived from remotely sensed information or precipitation forecasts provided by NWP models. However, raingauge-based precipitation data and analyses have to be examined carefully for errors resulting from mis-recording (i.e. typing errors), systematic undercatch or low sampling rates. This is the area of activity of the Global Precipitation Climatology Centre (GPCC).

The GPCC has been established in 1989 on invitation of the World Meteorological Organization (WMO). The Centre is operated by the Deutscher Wetterdienst (DWD), Offenbach, Germany. It contributes to the World Climate Research Programme (WCRP) and the Global Climate Observing System (GCOS). The scientific and technical functions of the GPCC are defined by the Implementation and Data Management Plan for the Global Precipitation Climatology Project (GPCP) (WMO/TD No. 367) and comprise:

- Collection of in situ observed precipitation data from global and national networks;
- Quality-control of the data and correction of errors;
- Calculation of monthly gridded area-mean precipitation for the Earth's land surface;
- Error assessment for the calculated precipitation for each individual gridbox and month;
- Combination of the gridded results from surface-based observations and satellite data in co-operation with other GPCP-participants (NASA/GSFC, NOAA/NCEP, Universities);
- Climatological studies based on the analysis results.

The GPCC also co-operates with the regional projects Baltic Sea Experiment (BALTEX) and the Mesoscale Alpine Programme (MAP). High-resolution precipitation data collections of these projects are a well suited data base to verify the satellite-based one degree daily precipitation product provided by the GPCP since the year 1997. First comparison studies have been performed at the GPCC (Rubel and Rudolf, 2000).

2. The Observational Database

2.1 Availability of precipitation data

Conventionally measured data from raingauge networks are still the most reliable information to obtain area-averaged precipitation for the landsurface. Satellite-based estimates are subject to larger biases and stochastic errors and need to be adjusted to in-situ observations (Barrett et al. 1994, Rudolf et al. 1996).

A first meteorological database for precipitation can be obtained from synoptically observed weather reports (at least with a daily resolution) and monthly climatic data, which are distributed worldwide as "SYNOP" and "CLIMAT" reports via the World Weather Watch Global Telecommunication System (GTS). GPCC regularly collects monthly precipitation totals from these sources for nearly 7,000 stations worldwide. These data being available near real-time are the basis for monthly monitoring of the global precipitation, resp. the "Monitoring Product" of the GPCC.

With respect to the high spatial and temporal variability of precipitation, station networks of high density are necessary also for global scale analyses. In order to derive monthly area-mean precipitation on a grid of 2.5° latitude by longitude with an error of most probably less than 10%, observed data from 2 to 10 raingauges per area of 10,000 km² are required, depending on the regional variability of precipitation. A number of 2 stations per 10,000 km² already results in the need of data from 40,000 equally distributed stations for the total global landsurface (Rudolf et al. 1994). But for spatial (e.g. 1° x 1°) and/or temporal (e.g. daily) higher resoluted analyses, much more stations per 10,000 km2 are necessary to remain below the 10 % sampling error goal (Rudolf 1995, after US Weather Bureau 1947 and WMO 1985).

In order to achieve the desired database for GPCC's analyses, WMO has sent out requests for additional data by letters to all Members in 1992 and 1994. The Executive Council of WMO "reaffirmed the importance of the task undertaken by the GPCC to collect raingauge data worldwide and produce a global monthly-mean climatology of precipitation". The Council "urged the Members to assist the Centre and make available precipitation records in a timely fashion." (WMO 1993). The data collection period as defined by the GPCP Implementation and Data Management Plan (WCRP 1990) starts with the year 1986.

So far, national institutes from about 150 countries have supplied additional data on a voluntary basis, following the WMO requests and bilateral negotiation with GPCC. The entire GPCC database includes now monthly precipitation totals of about 48,000 stations (GPCC's full data set). The data period begins with 1986, the time series are largely complemented by climatological means for the normal period 1961-1990. The year with the best data coverage is 1987 with monthly precipitation data for about 38,000 stations (Figure 1).

A gradual decrease of the number of stations after 1987 down to 7,000 stations for 1999 (i.e. GPCC's GTS data) is caused by the delay of the delivery of additional data and by the time required by the full data set still needs spatial and temporal completions, as well as retrospective temporal extension (a precipitation re-analysis from 1961 onwards is required with regard to CLIVAR) and continuation of update deliveries to GPCC by the countries in future.

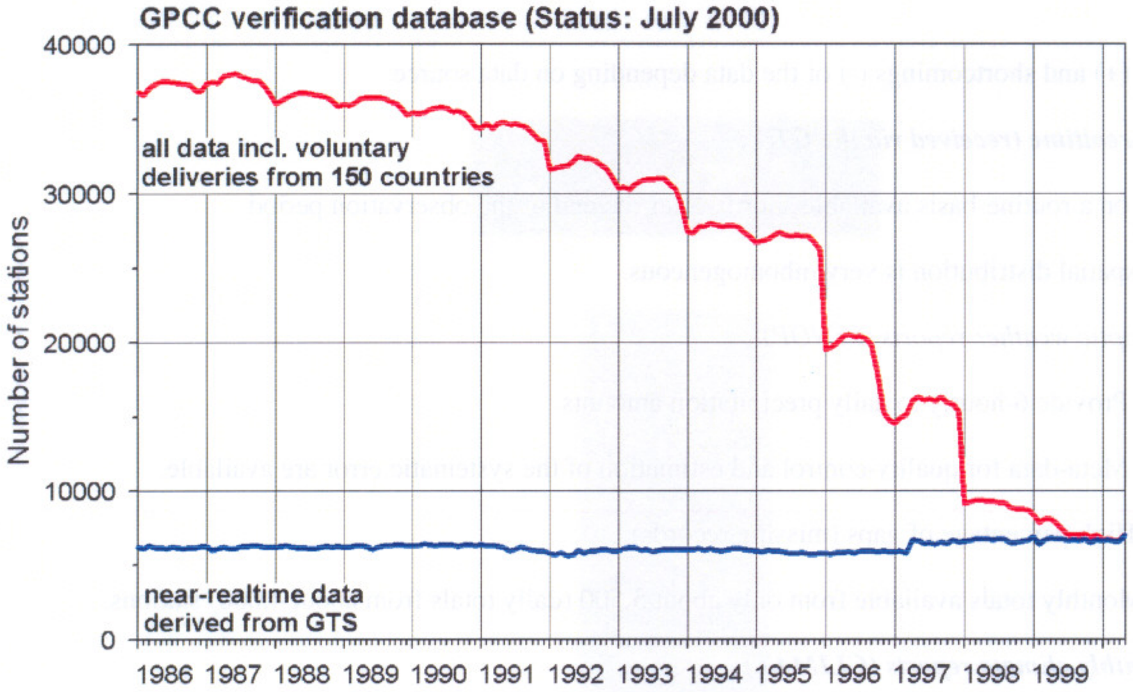

Fig. 1: Number of monthly precipitation data in the GPCC data base collected from various sources, July 2000.

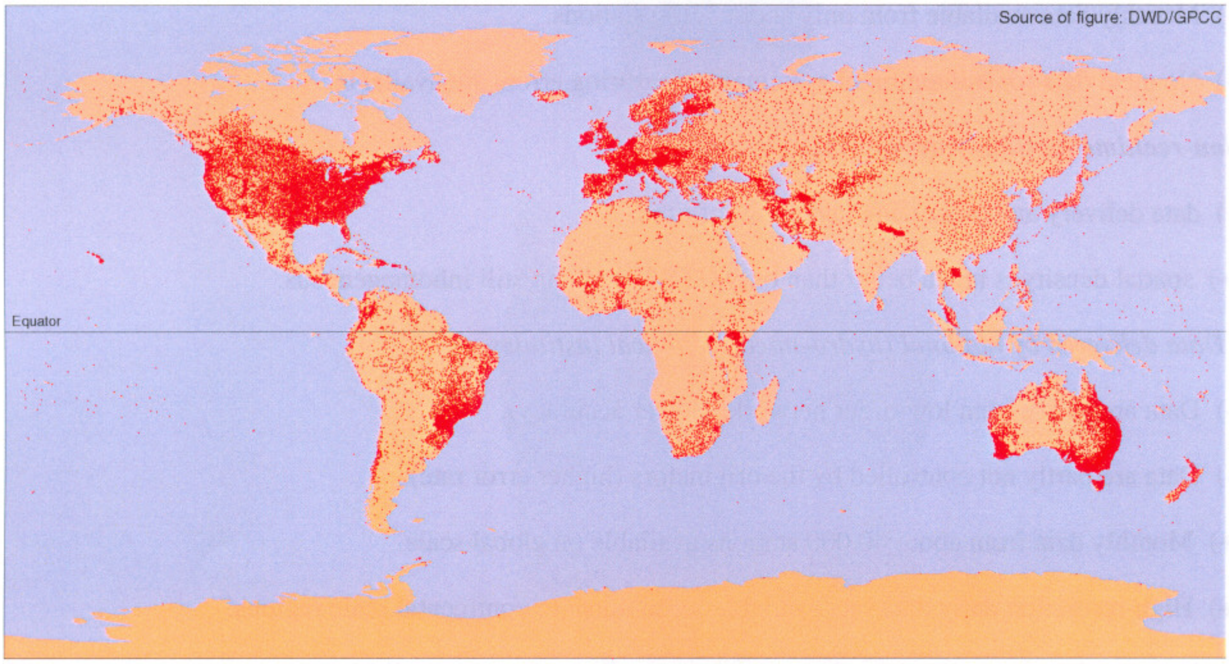

Fig. 2: Distribution of stations with monthly precipitation in the GPCC data base for July 1987.

The GPCC is using a Relational Data Base Management System (RDBMS) to archive and handle the precipitation data and station-related meta information. The data from the different sources are stored separately in the RDBMS.

Advantages (+) and shortcomings (-) of the data depending on data source

a) ***Near-realtime (received via the GTS):***

(+) on a routine-basis available shortly after the end of the observation period.

(-) spatial distribution is very inhomogeneous.

Synoptic weather reports (SYNOP)

(+) Provide 6-hourly to daily precipitation amounts.

(+) Meta-data for quality-control and estimation of the systematic error are available.

(-) High percentage of gaps (missing records).

(-) Monthly totals available from only about 5,500 (daily totals from about 4,000) stations.

Monthly climate reports (CLIMAT)

(+) Reports are mostly pre-controlled by the originator.

(+) Meta-data for additional quality-control are available.

(-) Monthly data available from only about 1,700 stations.

(-) No meta data for estimation of systematic measuring errors are available.

b) ***Non-realtime (various ways of transmission):***

(-) data delivery and processing takes a lot of time.

(+) spatial density is much better than only GTS-based, but still inhomogeneous.

Data delivered by national (hydro-)meteorological institutes

(-) Data are partly from low-order networks (lower accuracy).

(-) Data are partly not controlled by the originators (higher error rate).

(+) Monthly data from about 40,000 stations available on global scale.

(+) High-resolution daily data sets available for national to continental scale regions.

(-) No meta data for estimation of systematic measuring errors are available.

2.2 Problems of analyses of daily precipitation based on GTS data on a global scale

There is a strong demand from the international research community for analyses of daily precipitation. Just recently a satellite-based product of daily global precipitation on a 1° by 1° grid has become operational (http://rsd.gsfc.nasa.gov/912/gpcp/gpcp_daily_comb.html). Validation studies regarding this data set (see. Section 7.4) show, that it needs to be adjusted to raingauge data on a daily timescale. But until now it has not been possible to provide a global analysis of daily precipitation based on raingauge data, because of 2 problems:

- Availability of raingauge data on a daily time resolution

  The number of stations for which SYNOP reports are received via GTS at DWD varies from day to day from about 5300 up to 5800 stations. Some of these stations do not report precipitation data at all. Other stations are not covered with precipitation data during a whole analysis day due to problems in measuring and data transfer, as well as (part-time) automation of stations. Calculated daily precipitation totals need to be based on 100 % data coverage per day to receive reliable precipitation totals.

  In test months between 3800 and 4300 GTS-stations were usable for deriving daily precipitation totals, if not regarding the summation time for precipitation totals. A quite sufficient regional station density for daily precipitation analyses is only reached (see Fig. 3) in Europe, partly also in Eastern Asia, North America, Brazil and South Africa. But the station density in other regions, especially in most parts of Africa, South-East Asia and Oceania, and in high northern latitudes, is low. The high spatial variability of daily precipitation derived from routinely available GTS data obstructs the production of a gridded daily precipitation product covering the whole landsurface of the earth.

  Only additional gauge data from international projects (BALTEX, MAP, LBA, CEOP, etc.) can improve the available database. But regarding the required number of stations per grid to meet the high sampling error criterion on a daily basis, it is still more advisable to provide gauge-based daily precipitation assessments only for the regions with sufficient data coverage mentioned above.

- Definition of analysis days

  Precipitation is measured at specific times for specific time intervals. Most of the SYNOP-stations calculate and report precipitation totals on a 6-hourly or multiple 6-hourly (e.g. 12, 18, 24 hours) basis. The climatological summation day for precipitation totals mostly starts within a time zone at a specific time in the morning of the analysis day and continues until the same time on the following day.

  One of the main problems in calculating daily totals of precipitation exceeding one time zone is the definition of the analysis day. 24-hour observations are usually fixed on local time (mostly in the morning of a specific time zone). But the summation time of precipitation totals at an analysis day can differ with time zone, region, country, sometimes also in a country itself (see Fig. 3). An adjustment of a gauge-based precipitation day to the day used in satellite-based precipitation analyses (00 UTC-00 UTC next day) is not possible due to the lack of intermediate data from gauges. As satellite-based precipitation products provide very high (e.g. 3-hourly) time resolution, the best way to use gauge-based data for satellite-adjustment is to adjust the 'satellite precipitation day' to the 'gauge observation day'.

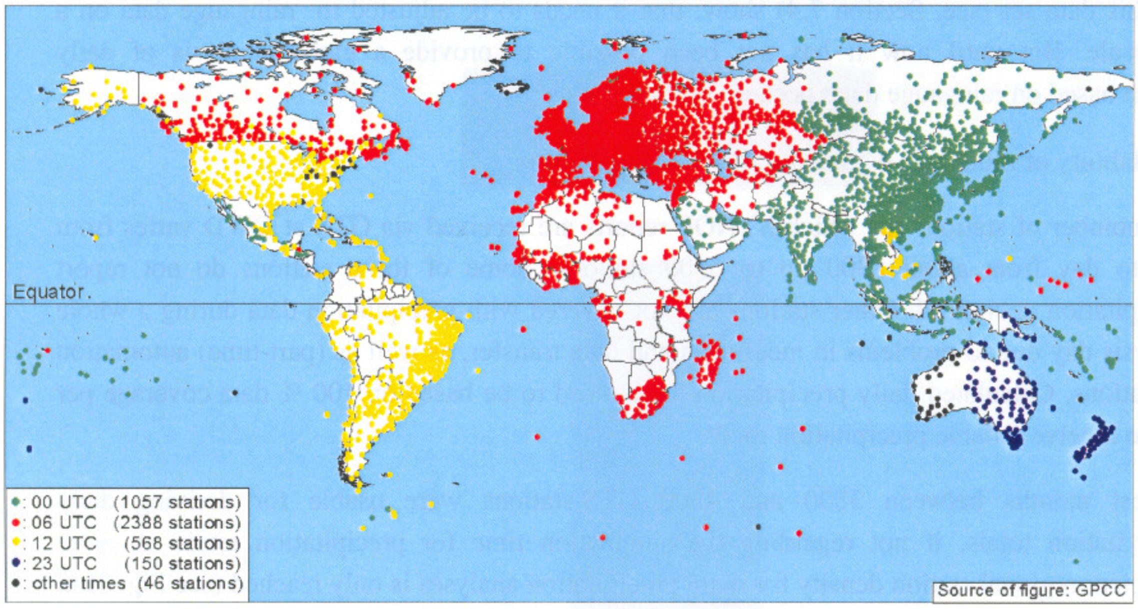

Fig 3: Summation time for calculating daily precipitation totals based on SYNOP-reports from GTS. Day: Jan 23rd 1999. Total number of usable stations: 4209.

3. The GPCC Quality-Control Procedures

The necessity of an effective quality-control is evidenced by any experience in processing of observational meteorological data. Conditions complicating the quality-control of precipitation data are:

- The temporal and spatial variability of precipitation is very large.
- A large number of the statistical outliers are true extreme data.
- Meta-information to be used as objective quality criteria is mostly not available
- for monthly data from the hydrometeorological raingauge networks.

Data which arrived at GPCC are partly affected by typing or coding errors and other modifications happening during the transmission from the originator to the archive. In many questionable cases it is not possible to get replies from the data originator. The following problems occur:

- Delivered datasets are irregularly formatted.
- Important meta data (station identifications, geogr. location, format descriptions) are missing and have to be procured.
- Delivered station coordinates are erroneous (occurs frequently).

- Doublettes of stations have to be eliminated.
- Missing precipitation is not clearly indicated in the data.
- Recorded precipitation depths are affected by coding/decoding errors.
- Temporal misplacement of data in time-series.

First of all the station meta data (identification, geographical location) have to be checked, corrected (geogr. coordinates are partly erroneous) or complemented. If monthly precipitation data at a station are available from more than one source, then an "optimal" value is selected automatically based on statistically predefined random errors assigned to the data from the different sources and intercomparisons between them.

The quality-control of the precipitation data at the GPCC is semi-automatic. In the automatic part the precipitation data at a station are first checked against the climatological normal and for spatial homogeneity and questionable data are flagged. In addition to that the data from the different (being stored separately in the RDBMS) are checked against each other. A statistical check is performed on the basis of frequency distributions derived from time series of monthly precipitation (full GPCC data base, GHCN data set), too. According to a suitable combination of these different criteria, part of the data (less than 10%) are classified as questionable and flagged.

The automatic part of the control-procedure has not been designed to correct data, but to reduce the number of data for which a visual control is necessary. The problem with a fully automatic quality-control is that it would eliminate all questionable data and, with regard to the high variability of precipitation, also remove a large amount of true data, in particular extreme values. These data, however, are very important to describe the real structure of the spatial distribution and the variability within the gridded precipitation analysis. In order to keep the true extrema and also to remove obviously wrong data from the analysis system, a visual check of the questionable data, although very time-consuming, seems to be inalienable.

Data marked as questionable in the automatic quality-control process subsequently can be manually reviewed by a trained expert using an interactive programme on a graphics workstation. This software is showing all relevant information of the station being checked, as well as the precipitation data of the neighbouring stations and background fields such as gridded climatologies or a 3d-orography (the data source is displayed by symbol, stations with data flagged as questionable are marked by colour). Obvious errors in the precipitation data are corrected, if possible. Otherwise incorrect data are set to the code for missing values or, if available, the monthly precipitation from another source can be selected for the analysis. If a station is misplaced, its geographical location can also be corrected. Original as well as corrected data are archived in the RDBMS of DWD.

The quality-control procedure at GPCC includes the following components (A - E):

A **Pre-control of recorded locations and identifiers of the stations**

B **Separate pre-control of precipitation data from:**

- GTS CLIMATs (DWD Offenbach): ca. 1,700 stations;

processed at DWD; monthly precipitation;

statistical/visual quality-control using quintiles,

quality-flags preset = high

- GTS SYNOPs (DWD Offenbach): ca. 5,500 stations;

  processed at DWD;

  automatic correction of typical coding errors using the full SYNOP information,

  quality-flags preset = f(coverage)

- GTS SYNOPs (CPC Washington): ca. 6,000 stations;

  processed at NCEP; autom. correction and filling of gaps;

  delivered to GPCC incl. meta-information on the monthly data coverage,

  quality-flags preset = f(coverage)

- Additional national data sets from 150 countries: ca. 40,000 stations;

  processed at GPCC; delivered by mail or email on individual request;

  overall visual quality-control for the single delivered datasets,

  quality-flags preset = still open

C **Storage of these data in the RDBMS "MIRAKEL" of GPCC**

D **Automatic quality-control of monthly precipitation data**

1) Selection of one value for each station based on quality flags.

2) Calculation of gridded area-means from the selected data (first guess).

3) Quality checks and setting of quality flags:

   o Constistancy check (comparison of monthly data from different sources);

   o Plausibility check in the temporal structure (comparison of monthly data and frequency distributions derived from the data-series of the station); Plausibility check in the spatial structure (comparison of monthly data and monthly area-mean).

4) Supplementing of quality-flags to the data.

E **Visual quality-control of monthly precipitation data flagged as questionable**

5) Visualization of data and background information (questionable data marked).

6) Manual confirmation or elimination or correction of precipitation records, or modification of quality-flags or of data selections.

→ **Calculation of gridded area-means from the checked and corrected data.**

4. Calculation of Area-Means on the Grid

The final analysis of gridded area-averaged precipitation is performed on the basis of the quality-controlled data by using the objective analysis method SPHEREMAP (Shepard 1968; Willmott et al. 1985), which is based on an inverse distance and directional weighting scheme. The major steps of the procedure are:

- Calculation of 0.5°-gridpoint values by weighted interpolation of nearby gauge data (weighting scheme after Shepard, 1968);
- use of a limited number of nearest stations only;
- weighting coefficients inverse proportional to the gauge gridpoint distance, corrected for directionally clustered stations;
- calculation of the area-mean value of each 0.5° x 0.5°-gridbox as the arithmetic mean of the interpolated gridpoint values at the box-corners (land surface only);
- calculation of the area-mean value on a 2.5° x 2.5°-gridbox or a 1.0° x 1.0°-gridbox as a weighted average of the means of all 0.5° x 0.5°-gridboxes inside; the 0.5° means are weighted by the portion of landsurface within the 0.5°-gridbox.

5. GPCC raingauge-based analyses of global land surface precipitation

The GPCC products, gridded data sets based on raingauge observations, are available in two resolutions, 2.5° by 2.5° and 1.0° by 1.0° geographical latitude and longitude, and with two different databases, i.e. near real-time with GTS data only and non real-time including complemented GTS data and additionally the data, which are delivered later from national institutions to the GPCC.

GPCC Monitoring Product: The near real-time 'Monitoring Product' for the land surface is based on GTS data from about 6,000 to 7,000 stations located over the continents and ocean islands. This product is the in-situ basis of the GPCP Combined Products and has also been used at the GPCC for special near real-time analyses (e.g. for river flood events). All data used are fully quality-controlled. Re-analyses were produced back to 1986 using the GTS data base. The product is available on a routine basis with a delay of about two months after observation.

GPCC Full Data Product: The non real-time 'Full Data Product' for the land surface is based on observations from (up to) 38,000 stations including the additional data from national collections provided by 150 countries so far. The sampling error of this products is lower than it is for the

Monitoring Product. A first version has been produced for the period from 1986 to 1995, but has not yet been published, since some more quality control has to be done on the additional data.

Variables which are supplied with both products on the grid are:

- Monthly precipitation amount (mm/month) (Figure 4);
- Mean monthly precipitation for the normal period 1961-90 (mm/month);
- Monthly precipitation deviation from normal 1961-90 (mm/month) (Figure 5);
- Monthly precipitation anomaly (percentage of normal 1961-90);
- Number of raingauges per gridcell for estimation of the sampling error (Figure 6);

Mean correction factors for the systematic gauge-measuring error.

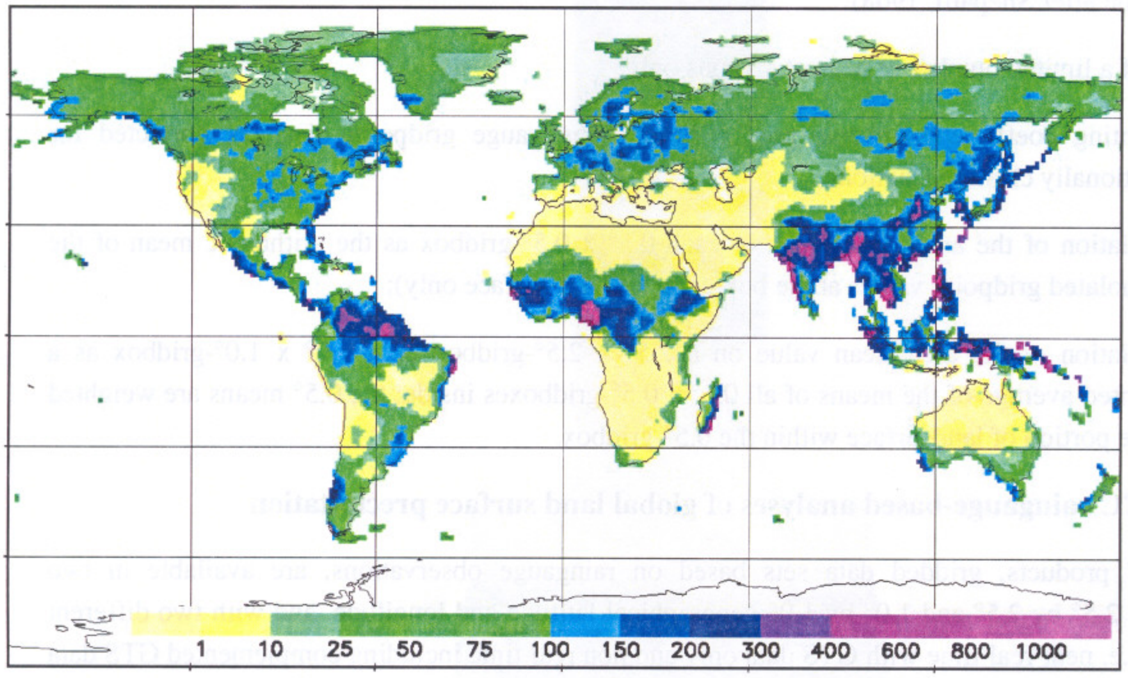

Fig 4: Monthly precipitation in mm/month for July 2000, based on precipitation data received via GTS, controlled and corrected, from about 7,000 meteorological stations.

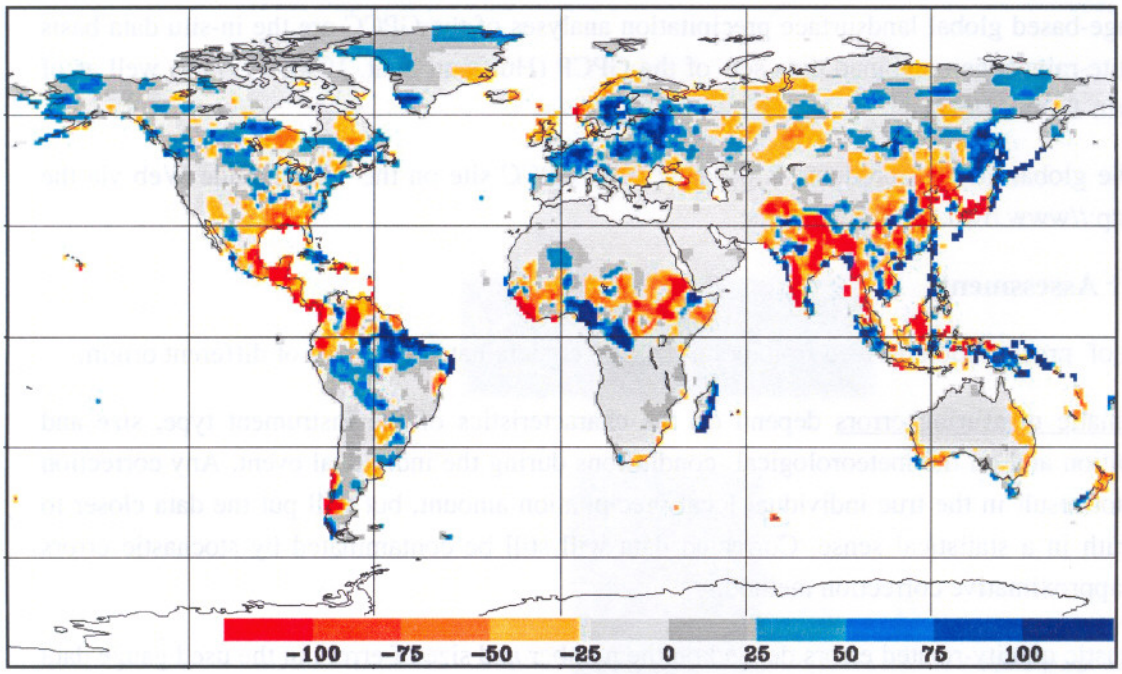

Fig 5: Precipitation anomalies in mm/month (deviation from July mean 1961–90) for the month July 2000.

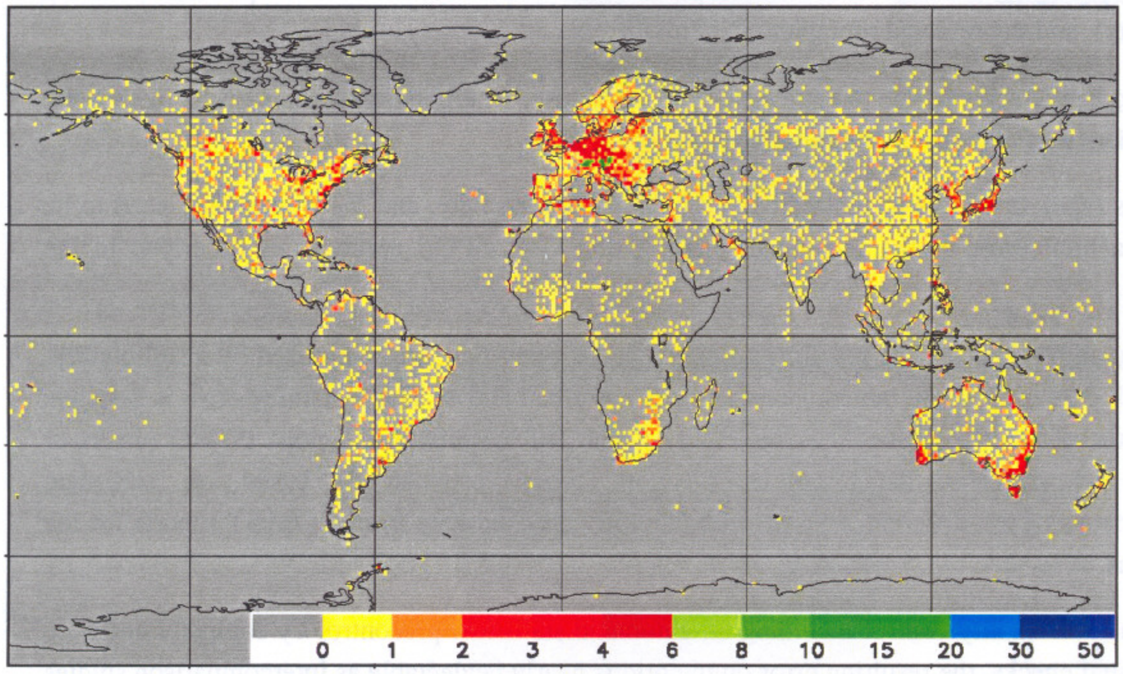

Fig 6: Number of stations per grid for the precipitation analysis shown in Figure 4.

The raingauge-based global landsurface precipitation analyses of the GPCC are the in-situ data basis of the satellite-raingauge combined data sets of the GPCP (Huffman et al. 1995, 1997) as well as of CMAP (Arkin & Xie, 1997).

Access to the global gridded products is possible from GPCC site on the World Wide Web via the address: http://www.dwd.de/research/gpcc

6. Error Assessment

Area-means of precipitation derived from point data are contaminated by errors of different origin:

- Systematic measuring errors depend on the characteristics of the instrument type, size and exposition and on the meteorological condititions during the individual event. Any correction will not result in the true individual local precipitation amount, but will put the data closer to the truth in a statistical sense. Corrected data will still be contaminated by stochastic errors from approximative correction method.

- Stochastic quality-related errors depend on the number and size of errors in the used gauge data set. Data transmitted from the observer to the archive are partly affected by typing or coding errors and similar modifications which are assumed to be random. The size of the resulting errors on the grid strongly depends on the level of quality-control.

- Sampling errors are area-related and depend on the number of observations per area (network density) and the regional precipitation variability.

- Methodical errors are due to approximations of the used interpolation scheme or the method for calculation of area-mean precipitation.

These errors types first have to be treated and quantified separately, and the results then need to be merged to a total error of the area-mean precipitation. The GPCC approach is described in the following:

Stochastic quality-related errors resulting from erroneous input data are minimized by a full quality-control of all data used in the raingauge analysis as described above. The remaining error is assumed to have a stochastic distribution. It is separately quantified for different regions (WMO blocks or countries) and different data sources (network types, originators), based on statistical intercomparisons of the different datasets with carefully checked data from CLIMAT reports for the same stations.

With a high-level quality-control including an automatic statistical precontrol by computer software and final visual checks, the resulting error on the grid is nearly neglectable as intercomparison studies of controlled and uncontrolled data versus a reference have shown. On the one hand, an automatic-only control, which removes all questionable data, would also eliminate numerous true extreme values, and reduce the variability in the gridded products. On the other hand, single erroneous outliers can strongly affect gridded results. Both can cause errors which may exceed the size of the sampling error. Therefore the GPCC performs the full quality-control (automatic precontrol plus visual control) of all data used in the Monitoring Product. The critical point is the realization of the time-consuming visual quality-control month by month for a dataset of 40,000 stations.

of all data used in the Monitoring Product. The critical point is the realization of the time-consuming visual quality-control month by month for a dataset of 40,000 stations.

Systematic measuring errors are trendly compensated using long-term mean correction factors, which were derived by Legates (1987), who corrected the monthly long-term-mean precipitation for about 25,000 stations, using Sevruk's correction formulas and approximated meta data. The monthly mean correction factors, which are mostly in the range of 1.0 to 3.0, describe the mean annual cycle but not the year-to-year variation. From this and from the limited accuracy of the bulk correction in principle results, that after the correction with regard to systematic measuring errors a stochastic error component is remaining. The GPCC quantifies this error component by 40% of the systematic correction term. This method is not yet sufficient, since the systematic error strongly depends on the meteorological conditions during the individual rain- or snowfall event. GPCC is working on an advanced method using wind-speed and precipitation-type data from synoptic data for correction on a daily basis (Fuchs et al. 2000).

The sampling error has been investigated by GPCC using data from dense networks of Australia, Canada, Finland, Germany and USA. Based on statistical experiments performed for 322 test cases, relations were derived between the sampling error and the number of observations and spatial variability of precipitation in these regions. The relative sampling error with a number of 5 raingauges per 2.5° x 2.5°-gridcell is between 7% and 40% of the true area-mean of monthly precipitation, with 10 stations an error between 5% and 20% can be expected (Rudolf et al. 1994).

The methodical error is much smaller than the sampling errors, as intercomparison studies performed at GPCC for various analysis methods have shown, and is neglected for large-scale GPCC analyses.

The total stochastic error on the grid is calculated from the individual error components, after systematic errors have been eliminated: The station-related stochastic errors are transferred to the grid-area, and on the grid they are combined with the sampling error using error progression theory. The size of the resulting total error of the calculated gridded precipitation ranges from a few up to some tens in percent of the assumed true area-mean. Errors resulting from an insuffient correction of systematic measuring errors dominate in regions with snowfall and high windspeeds, the sampling errors are major for data-poor regions, especially where precipitation is highly variable.

7. Research Activities

GPCC is going to prepare products of higher resolution and to develop advanced methods for quality-control, error assessment and spatial analysis.

7.1 Improved quality-control of monthly precipitation data

Frequency distributions of monthly precipitation data have been derived on the basis of the extended GPCC data base and the GHCN data collection. The implementation of an additional statistical check of the monthly precipitation totals for the Monitoring Product has reduced the number of data flagged as questionable in the automatic QC process, so that the number of stations to be checked manually has been reduced. The percentage of stations flagged as questionable, which actually have to be corrected or removed because of obvious errors, has increased from 10-20% (earlier) to between 40 and 50% in the improved QC version. The new QC procedure has been operationally used for all GPCC Monitoring Products from December 1999 onwards.

7.2 Improved method for correction of the systematic gauge-measuring error

Implementation of a new operational method for correction of daily gauge data due to systematic measuring errors has been developed at the University of Vienna (F. Rubel) in the framework of BALTEX (Rubel and Hantel 1999). The new method for on-event correction is based on synoptical data and has been extended for applications on a global scale in cooperation between F. Rubel and GPCC and has been implemented at the GPCC. The method takes wind effects and evaporation losses into account and is using information about raingauge type and installation specifics from a WMO report and weather information during precipitation events in order to determine the precipitation phase, which is very important to estimate the systematic gauge-measuring error (Fuchs et al. 2000).

Intercomparison studies between Legates' climatological correction and the on-event correction method have been carried out for the regions of the GEWEX Continental Scale Experiments (Ungersböck et al. 2000). Overall the new method gives slightly lower correction factors than Legates' bulk correction, mainly because Legates' method has a tendency to overestimate the correction in the case of snowfall.

7.3 Evaluation of daily raingauge data for EuroTRMM

The goal of EuroTRMM is to demonstrate the feasibility of the assimilation of TRMM data for numerical weather prediction. Contributors to this EU-Project are ESA/ESTEC, ECMWF, the MPI for Meteorology and the GPCC. The GPCC has started analysing synoptic and daily non-GTS raingauge data for this project. The MPI (K. Arpe et al.) will test the improvement of ECMWF forecasts by using TRMM precipitation data for assimilation. See Section 8.

7.4 Validation of GPCP products (GPCP-1DD)

The GPCC is performing a validation study for the GPCP-1DD product based on high resolution raingauge data sets for the areas of BALTEX (ca. 4,200 stations) and of the European Alps (ca. 3,100 stations). First comparison results have been presented at several meetings (GEWEX Hydrology Panel Meeting, Geesthacht, Germany, Sep. 1999, MAP Workshop on Alpine Climatology in Interlaken/ Switzerland, March 2000, EGS General Assembly, Nice, France, April 2000, EUMETSAT Users Conference, Bologna, Italy, April 2000, GPCP WGDM meeting, Greenbelt/MD., USA, Aug. 2000; see e.g. Skomorowski et al. 2000, Rudolf and Rubel 2000).

The correlation between GPCP-1DD and raingauge analyses shows large variations between 0.3 and 0.8. For individual days the structure of precipitation events is generally correctly reproduced by the GPCP-1DD product, but quantitative applications are somewhat problematic. Problems still occuring are that GPCP-1DD is based on only 2 SSM/I images per day (quick moving precipitation events can not be determined properly) and that there is a time shift between the raingauge observation day and the satellite evaluation day. A report on the results for the BALTEX area is available on internet: http://www-med-physik.vu-wien.ac.at/staff/rub/pro/GPCC/applications.htm

The intercomparison studies on the GPCP-1DD product versus regional high-resolution daily raingauge analyses will be extended to longer periods. This study is jointly done by GPCC and F.

7.5 Planned projects

The GPCC applied for funds of the German Polar Research Programme with regard to ACSYS in order to continue operation of the Arctic Precipitation Data Archive (APDA), to collect and analyse snow depth data and to develop an improved Arctic precipitation climatology. A project proposal with regard to CLIVAR has been jointly prepared by Prof. C.-D. Schönwiese (Univ. of Frankfurt) and B. Rudolf (GPCC). Goals are the compilation of a quality-controlled and homogenized global climatic dataset covering the total observational period and its statistical analysis. This proposal was accepted by the German Climate Research Programme to receive the opportunity for application of national funding.

The GPCC projected a large-scale hydrological study jointly with the GRDC and scientists from the Universities of New Hampshire (Vorosmarty) and of Tokyo (Taikan Oki). Project goal is to assess the reliability of observed precipitation and run-off data based on comparison of both using hydrological models.

7.6 Other current and planned Developments at the GPCC

- Final development of the operational procedure to assess .the total error and its components on each individual gridbox and month of the GPCC products.
- Calculation of the area-mean monthly precipitation for selected gridboxes using orographical data.
- Analysis of snow cover, snow depth and liquid water equivalent in order to complement information on total precipitation (within ACSYS).
- Development of the GCOS Surface Network Precipitation Data Set (longterm time-series for selected climate stations).

8. Evaluation of temporal high-resolution raingauge data for EuroTRMM

8.1 Compilation of in-situ data for EuroTRMM

Only a brief overview of the compilation of in-situ data is given here. The final report of the project part will comprise an expanded description.

Data acquisition

Required data are SYNOP reports comprising 6, 12, or 24-hourly precipitation totals on the one hand. These data are transmitted from the GTS (Global Telecommunication System) to the German Meteorological Service (DWD) regularly. On the other hand, the project part tried to get additional data sets from various national sources, like national meteorological or hydrological services, research institutions, etc. Unfortunately, the result of this inquiries was not satisfied, because the temporal resolution of the data was too large.

SYNOP data

Now, the global SYNOP availability described in Chapter 2 was put in concrete form in case of the TRMM area. Figure 7 gives the number of SYNOP reports with derived daily precipitation totals at

September, 1st 1998. The number of stations was calculated for 2.5° grids. More than 5 stations are located only within a few of grids, especially in China and Southern Africa near the 30th parallel. The number of stations per grid decreases, if the grid become smaller. For example, only four 0.5° grids exist between the mentioned latitudes with more than two SYNOP stations in it.

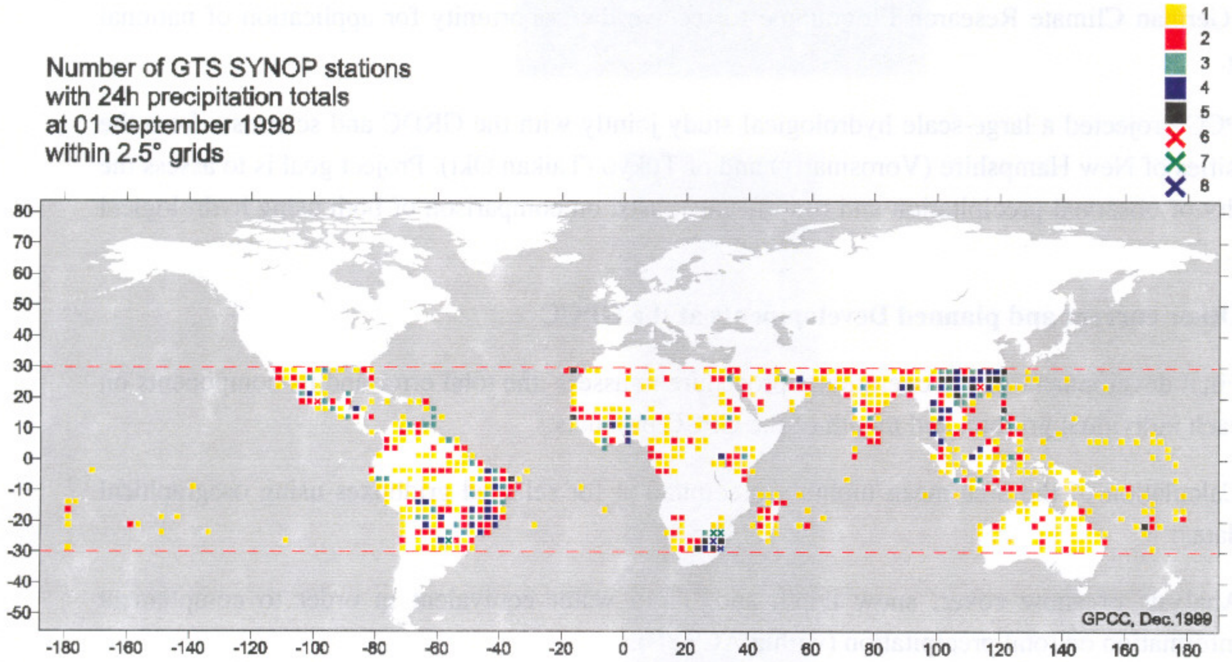

Fig 7: Number of SYNOP reports with derived daily precipitation totals at September, 1st 1998.

Some statistical characteristics of the data

The mean spatial representativeness gives an impression how similar time series of neighbouring stations fluctuate. The representativeness is defined as the distance in which the correlation coefficient reachs a certain value, e.g. 0.5. The higher the mean distance, the better the representativeness of a time series. In case of 6-hourly precipitation Figure 8 shows the mean spatial representativeness for 20° x 20° boxes within the TRMM area. It appears that the representativeness is better in higher than in tropical latitudes, which could be caused by more convective precipitation.

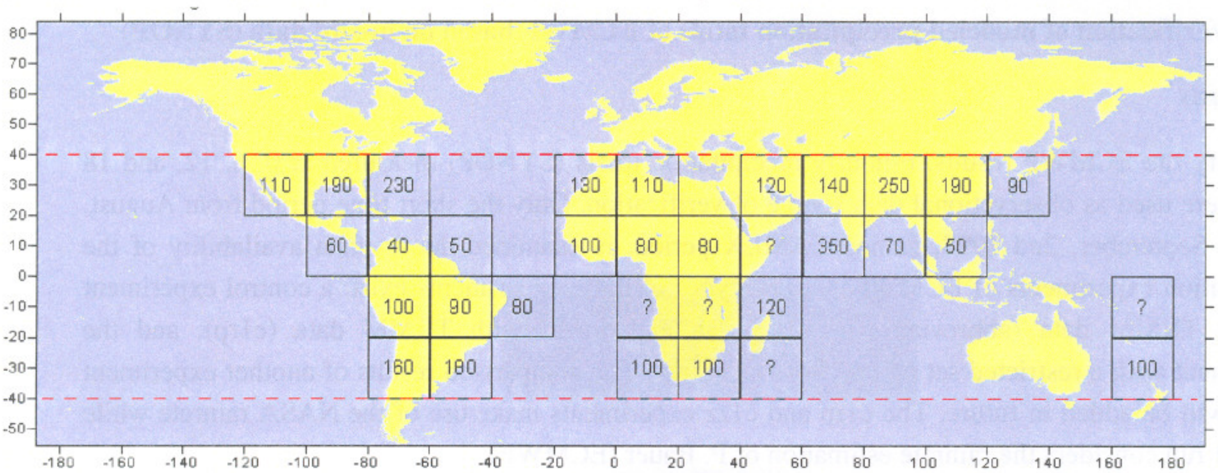

Fig 8: Mean spatial representativeness (in km) of 6-hourly precipitation totals for 20° x 20° boxes; January 1998 - December 1999

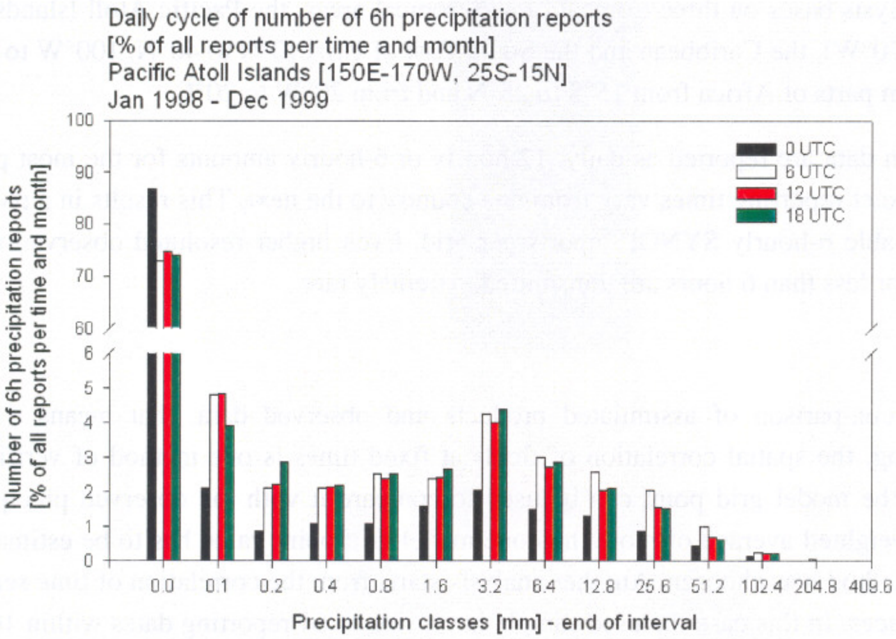

Fig 9: Daily cycle of number of 6-hourly precipitation reports for the Pacific Atoll Islands.

Anyway, an example of the daily cycle of the number of 6-hourly precipitation reports is shown in Figure 9 for the Pacific Atoll Islands. The time of day with low precipitation amounts is between 18 and 0 UTC, that means in the morning of this region. The local maximum of the 1.6-3.2 mm class is an artefact due to the typical form of the SYNOP report.

For data error discussion see Section 6.

8.2 Verification of modeled precipitation totals of ECMWF based on in-situ data (SYNOP)

Data basis

To verify the ECMWF results, 6-hourly precipitation totals (SYNOP) measured at 0, 6, 12, and 18 UTC were used as observational data source of verification. Only the short time period from August, 18th to September, 2nd 1998, named CAMEX period, is examined due to data availability of the assimilation experiments of ECMWF. Several types of these experiments exist: a control experiment without TRMM data (abbreviated as e1ko), an experiment with TRMM data (e1rp), and the experiment with a restricted set of TRMM data (e1zz). The comparison results of another experiment (e3j1) will be added in future. The e1rp and e1zz experiments make use of the NASA rainrate while the e3j1 run considers the rainrate estimation of P. Bauer (ECMWF).

he model results are stored as 3-hourly and 6-hourly amounts on a Gaussian grid. The grid point distance is c. 0.56° x 0.56° in case of the equatorial region and c. 0.70° x 0.56° in case of a latitude of 40°, respectively. The precipitation data of the CAMEX experiments were calculated as sums of the assimilated convective and large scale precipitation.

The verification analysis bases on three tropical or sub-tropical areas, the Pacific Atoll Islands (25°S to 10°N, 150°E to 170°W), the Caribbean and the South-East of USA (0°N to 40°N, 100°W to 60°W) as well as the western parts of Africa from 15°S to 25°N and from 20°W to 20°E.

SYNOP precipitation data are reported as daily, 12-hourly or 6-hourly amounts for the most parts of the world. But the exact reporting times vary from one country to the next. This results in a reduction of the number of usable 6-hourly SYNOP reports per grid. Even higher resoluted observations like precipitation totals for less than 6 hours are transmitted extremely rare.

Methods

Besides the visual comparison of assimilated products and observed data, that means a purely qualitative proceeding, the spatial correlation of fields at fixed times is one method of verification. Either the value of the model grid point can be used to compare it with the observed precipitation amount or an area-weighted average of more than one model grid point value has to be estimated. In this study the first method was choosen. Another analysis starts from the correlation of time series for fixed observation places. In this case the used sample has a size of 64 reporting dates within 16 days. A modification of these estimations is the correlation depending on topography, which is distinguished between mountainous, hilly, and flat or between coastal and interior area, respectively.

As verification quantities serve

- the rank correlation coefficient after Spearman as well as Kendall's tau;
- contingency tables based on typical frequency classes, which themselves depend on the total sample size,
- categorical quality quantities, like PEC (Percent Correct), TSS (True Skill Statistics), and others,

- continuous quality quantities, like BIAS, RMSE (Root Mean Square Error), RV (Reduction of Variance), and others.

In this context it should be emphasized that not all mentioned quality quantities are suitable for verification of precipitation totals.

Results

Table 1 summarizes the results of the field verification for fixed times. The ECMWF control experiments shows values of the RMSE fluctuating between 4.1 and 8.1 mm and a BIAS from -0.6 to 0.5 mm. While the RMSE is lower in case of the Pacific Atolls, the averaged BIAS of the precipitation totals is negative in the West Africa area, that means, the observed values are higher than the modeled ones. Isolated minimum values of BIAS resulting from observed but not predicted precipitation maxima are the reason for this spreading BIAS values.

Verification ECMWF – OBS (SYNOP)
6 h precipitation totals
18 Aug 1998 – 2 Sep 1998

RMSE [mm]	Pacific Atoll Islands	West Africa/ Ivory coast	Caribbean Sea/SE USA
Control experiment [e1ko]	4.06	8.07	7.38
Experiment with TRMM data [e1rp]	3.90	8.20	7.41
Experiment with a restricted set of TRMM data [e1zz]	4.04	8.16	7.39

BIAS [mm]	Pacific Atoll Islands	West Africa/ Ivory coast	Caribbean Sea/SE USA
Control experiment [e1ko]	0.54	-0.55	0.24
Experiment with TRMM data [e1rp]	0.50	-0.51	0.24
Experiment with a restricted set of TRMM data [e1zz]	0.50	-0.49	0.23

Table 1: Verification results for various ECMWF assimilation experiments regarding three special defined areas.

The verification analyses on the basis of experiments with TRMM data don't have clear results. The calculation for the Atoll Islands indicates slightly decreasing errors, whereas the estimation for West Africa provides a slightly worsening quality. The experiment with the restricted set of TRMM data doesn't improve this situation significantly. It should be noticed that it is necessary to increase the sample number, especially to use longer time intervals. The two weeks of the CAMEX interval reflect only a small part of the seasonal characteristics of the chosen climate zones.

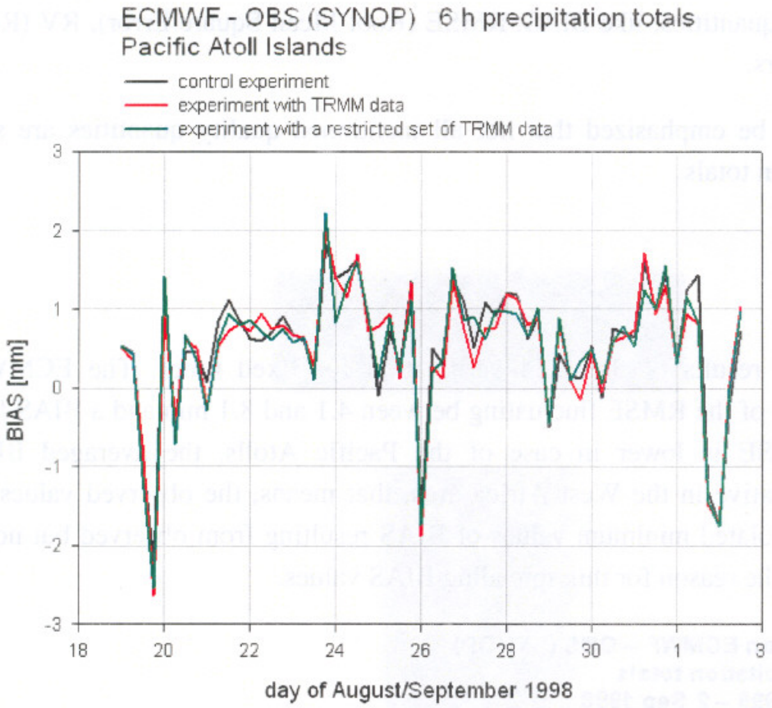

Fig 10: Time series of BIAS of modeled vs. observed data for the Pacific Atoll area.

Verification ECMWF – OBS (SYNOP)
6 h Precipitation totals, <u>all data</u>
18 Aug 1998 – 2 Sep 1998
Spatial correlation coefficients

r (Spearman)	Pacific Atoll Islands	West Africa/ Ivory coast	Caribbean Sea/SE USA
Mean number of combinations grid point - obs	*44.8*	*81.4*	*138.6*
Control experiment [e1ko]	0.370	0.270	0.282
Experiment with TRMM data [e1rp]	0.381	0.264	0.286
Experiment with a restricted set of TRMM data [e1zz]	0.365	0.274	0.282

Kendall's τ	Pacific Atoll Islands	West Africa/ Ivory coast	Caribbean Sea/SE USA
Control experiment [e1ko]	0.299	0.224	0.238
Experiment with TRMM data [e1rp]	0.308	0.219	0.242
Experiment with a restricted set of TRMM data [e1zz]	0.295	0.228	0.237

Pacific Atoll Islands: 25 S – 10 N, 150 E – 170 W
West Africa/Ivory coast: 15 S – 25 N, 20 W – 20 E
Caribbean Sea/SE USA: 0 N – 40 N, 100 W – 60 W

Verification ECMWF – OBS (SYNOP)
6 h Precipitation totals, <u>without obs data = 0 mm and > 20 mm</u>
18 Aug 1998 – 2 Sep 1998
Spatial correlation coefficients

r (Spearman)	Pacific Atoll Islands	West Africa/ Ivory coast	Caribbean Sea/SE USA
Mean number of combinations grid point - obs	*10.7*	*11.9*	*20.4*
Control experiment [e1ko]	0.237	0.151	0.216
Experiment with TRMM data [e1rp]	0.257	0.127	0.215
Experiment with a restricted set of TRMM data [e1zz]	0.235	0.137	0.203

Kendall's τ	Pacific Atoll Islands	West Africa/ Ivory coast	Caribbean Sea/SE USA
Control experiment [e1ko]	0.183	0.119	0.163
Experiment with TRMM data [e1rp]	0.188	0.098	0.164
Experiment with a restricted set of TRMM data [e1zz]	0.177	0.098	0.154

Pacific Atoll Islands: 25 S – 10 N, 150 E – 170 W
West Africa/Ivory coast: 15 S – 25 N, 20 W – 20 E
Caribbean Sea/SE USA: 0 N – 40 N, 100 W – 60 W

Table 2: Verification results for various ECMWF assimilation experiments regarding three special defined areas, left side with all in-situ data, right side without observed precipitation events = 0 mm and > 20 mm.

Nevertheless, further alternatives of verification were used. If the mean BIAS for each station of the defined areas is plotted on a map, a strong spatial BIAS variation in a very confined space becomes obvious due to singular precipitation events. Figure 10 illustrates these peaks reaching negative values of BIAS. Therefore the rank correlation of modeled and measured precipitation totals (without values of exactly 0.0 mm or greater than 20 mm!) produces only weak significant coefficients (Table 2), whereas the correlation with all data shows plainly higher coefficients. But the differences between the results based on the control experiment and the experiments with TRMM data are again small. With the exception of Western Africa, the improvements using TRMM data for assimilation experiments seem to be insignificant. The separation of the SYNOP stations with respect to specific location criteria leads to the knowledge that in case of mountainous or hilly located stations the RMSE is smaller that in case of stations located on the flat (Table 3). One reason for this behaviour could be an increased quality of the model output due to convective precipitation induced by topography.

Caribbean Sea / SE USA
18 Aug – 2 Sep 1998
ECMWF-OBS (SYNOP)
6 h precipitation totals
RMSE [mm]

	Number of stations	e1ko	e1rp	e1zz
MV (mountainous)	27	2.09	2.04	2.04
HI (hilly)	41	2.14	2.09	2.11
FL (flat)	65	3.13	3.11	3.14

	Number of stations	e1ko	e1rp	e1zz
CO (coast)	45	3.04	3.01	3.07
NO	83	2.46	2.43	2.43

General topography around the station:
FL flat
HI hilly
MV mountainous valley or at least not on the top of a mountain
Station location based on specific criteria:
CO station within 30 km from the coast
LA station next to a large (> 25 km**2) lake
NO if none of the above
Station may be all three but only labeled with one with the priority CO, then LA.

Table 3: Verification results for various ECMWF assimilation experiments regarding three special defined areas, depending on general topography criteria.

Expanded verification analyses are very necessary to achieve more reliable results. The final project report will be more detailed. It will include verification results regarding the e3j1 experiment of ECMWF.

References:

Barrett, E.C., J. Doodge, M. Goodman, J. Janowiak, E. Smith & C. Kidd (1994): The First WetNet Precipitation Intercomparison Project (PIP-1). Remote Sensing Review, Vol. 11 (1-4), 49 - 60.

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

Fuchs, T., J. Rapp, F. Rubel and B. Rudolf (2000): Correction of synoptic precipitation observations due to systematic measuring errors with special regard to precipitation phases: submitted to Phys. Chem. Earth.

Huffman, G.J., R.F. Adler, B. Rudolf, U. Schneider, P.R. Kehn (1995): Global Precipitation Estimates Based on a Technique for Combining Satellite-Based Estimates, Raingauge Analyses and NWP Model Information. Journal of Climate 8(5), p. 1284 - 1295.

Huffman, G.J., R.F. Adler, P.A. Arkin, A. Chang, R. Ferraro, A. Gruber, J. Janowiak, A. McNab, B. Rudolf, U. Schneider (1997): The Global Precipitation Climatology Project (GPCP) Combined Precipitation Dataset. Bull. Americ. Meteor. Soc. 78(1), 5-20.

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

Morrissey, M.L., J.A. Maliekal, J.S. Greene and J. Wang (1995): The uncertainty of simple spatial averages using rain gauge networks. Water Resources Res. 31, 2011-2017.

Rubel, F. and M. Hantel (1999): Correction of daily rain gauge measurements in the Baltic Sea drainage basin. Nordic Hydrology, 30(3), 191-208.

Rudolf, B., H. Hauschild, M. Reiss, U. Schneider (1992): Beiträge zum Weltzentrum für Niederschlagsklimatologie. Meteorologische Zeitschrift N.F. 1(1), 7-84 (in German language).

Rudolf, B. (1993): Management and Analysis of Precipitation Data on a Routine Basis. Proceedings Int. Symp. on Precipitation and Evaporation (Ed. B. Sevruk and M. Lapin), ETH Zuerich, Vol. 1, p. 69 - 76.

Rudolf, B., H. Hauschild, W. Rueth, U. Schneider (1994): Terrestrial Precipitation Analysis: Operational Method and Required Density of Point Measurements. NATO ASI I/26, Global Precipitations and Climate Change (Ed. M. Desbois and F. Desalmand), Springer Verlag Berlin, p. 173 - 186.

Rudolf, B. (1995): Die Bestimmung der zeitlich-räumlichen Struktur des globalen Niederschlags. Berichte des Deutschen Wetterdienstes, Offenbach am Main, 153 Seiten (in German language).

Rudolf, B., H. Hauschild, W. Rueth, U. Schneider (1996): Comparison of Raingauge Analyses, Satellite-Based Precipitation Estimates and Forecast Model Results. Advances in Space Research, 18(7), p. (7)53 - (7)62.

Rudolf, B., T. Fuchs, W. Rüth, U. Schneider (1998): Precipitation Data for Verification of NWP Model Re-Analyses: The Accuracy of Observational Results. Proceedings First WCRP International Conference on Reanalyses (Washington, DC, USA, 27-31 Oct 1997), WMO/TD-No. 876, 215-218.

Rudolf, B. and F. Rubel (2000): Regional Validation of Satellite-Based Global Precipitation Estimates. EUMETSAT Meteorological Satellite Data Users' Conference, Bologna, Italy, 29 May to 2 June 2000. Submitted to Conference Proceedings.

Schneider, U. (1993): The GPCC quality-control system for gauge-measured precipitation data. Report of a GEWEX workshop on Analysis Methods of Precipitation on a Global Scale, Koblenz, 14-17 Sept. 1992, WCRP-81, WMO/TD-No. 558, A5-A7.

Schneider, U., D. Henning, H. Hauschild, M. Reiss and B. Rudolf (1992): Zur Berechnung monatlicher Niederschlagshöhen aus synoptischen Meldungen. Meteorol. Zeitschr.N.F. 1, 22-31 (in German language).

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

Skomorowski, P., F. Rubel and B. Rudolf (2000): Verification of GPCP-1DD global satellite precipitation products using MAP surface observations. Submitted to Phys. Chem. Earth.

Ungersböck, M., F. Rubel, T. Fuchs and B. Rudolf (2000): Bias correction of global daily rain gauge measurements. Submitted to Phys. Chem. Earth.

US-Weather Bureau (1947):Thunderstorm rainfall. Hydromet. Rep. 5.

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. The American Cartographer 12 (1), 5-16.

Willmott, C.J., S.M. Robeson and J.J. Feddema (1994): Estimating continental and terrestrial precipitation averages from rain-gauge networks. Int. J. Climatol. 14, 403-414.

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

WMO (1985): Review of requirements for area-averaged precipitation data, surface-based and space-based estimation techniques, space and time sampling, accuracy and error, data exchange. WCP-100, WMO/TD-No. 115.

WMO (1993): Forty-Fifth Session of the Executive Council, Geneva, 8-18 June 1993, Abbridged Report with Resolutions. General Summary No. 4.5.8, WMO/No. 794.

Visit GPCC's Homepage: http://www.dwd.de/research/gpcc