A New Monthly Precipitation Climatology for the Global Land Areas for the Period 1951 to 2000

Ch. Beck, J. Grieser and Bruno Rudolf

1 INTRODUCTION

Globally gridded precipitation-data sets are an essential base for various applications in the geosciences and especially in climate research, as for instance global and regional studies on the hydrological cycle and on climate variability, verification and calibration of satellite based climate data or the evaluation of global circulation models (GCMs). As all applications require reliable high quality precipitation fields the underlying station data have to meet high demands concerning the quality of the observed precipitation data as well as the correctness of station meta data and also with respect to sufficient spatial station density and distribution. Concerning the use of globally gridded climate data for analyses of long-term climate variability it has to be ensured that station data used for gridding are as continous and homogeneous as possible.

In recent years various globally gridded data-sets of monthly terrestrial precipitation observations have been developed for example by Dai et al. (1997) and New et al. (2000). Several gridded datasets utilising gauge data arose from the global precipitation climatology project (GPCP) (Rudolf et al. 1994, Xie et al. 1996, Huffman et al. 1997, Chen et al. 2002, Adler et al., 2003).

Within the framework of the DEKLIM (German Climate Research Programme)-funded research project VASClimO (Variability Analysis of Surface Climate Observations) a new gridded monthly precipitation dataset for the period 1951 to 2000 covering the global land areas with a spatial resolution of 0.5° x 0.5° is developed on the basis of the most comprehensive data-base of monthly observed precipitation data world-wide that resides with the GPCC (Global Precipitation Climatology Centre).

Prior to gridding, all available station-data are subjected to a multi-stage quality control of observed values as well as of station-meta data. Only station time series with a minimum of 90% data availability during the analysed period (1951 – 2000) are used for interpolation to a regular 0.5° x 0.5° grid in order to minimise the risk of generating temporal inhomogeneities in the gridded data due to varying station densities. The interpolation is done by Ordinary Kriging (Krige, 1962) which proved to result in the least interpolation error of several methods tested. Thus, the resulting gridded dataset is highly suitable for the application in studies concerning long-term aspects of climate variability.

2 DATA SOURCES AND THEIR AVAILABILITY

The creation of gridded monthly precipitation data is based on station data from a number of data-sources available at the GPCC. Differences between these data-sources have to be considered. They can be attributed to varying methods used for aggregation and compilation of the data and for their transfer to the GPCC.

Three essential data-sources are the historical data-bases of the Food and Agriculture Organization of the UN (FAO, 13,500 stations) of the Climatic Research Unit (CRU, 9,500 stations) and of the Global Historical Climatology Network (GHCN, 22,600 stations). Unfortunately many of the time series originating from these sources and often starting early in the past do not last into most recent years.

Another three data-sources comprise data transferred via the Global Telecommunication System (GTS). GPCC-Synop and CPC-Synop include monthly data that have been aggregated on the basis of near-real time transferred data with high temporal resolution by the Global Precipitation Climatology Centre (GPCC) and the Climate Prediction Centre (CPC) respectively. These two sources exhibit significant differences because the near-real time transferred synoptic data contain numerous missing values and are not quality controlled. Processing of these data differs between GPCC and CPC. Also transferred via GTS the Climat-network contains quality controlled monthly precipitation for currently around 2000 stations.

In addition to these six data-sources the GPCC disposes of supplementary data derived directly from the various national meteorological and hydrological services (National) as well as from other research institutions or from research projects (Regional). Data from these sources cover varying subperiods of the analysed period and are updated irregularly.

The current availability of monthly precipitation data for the period 1951 to 2000 differentiated between the various data-sources is depicted in Fig. 1.

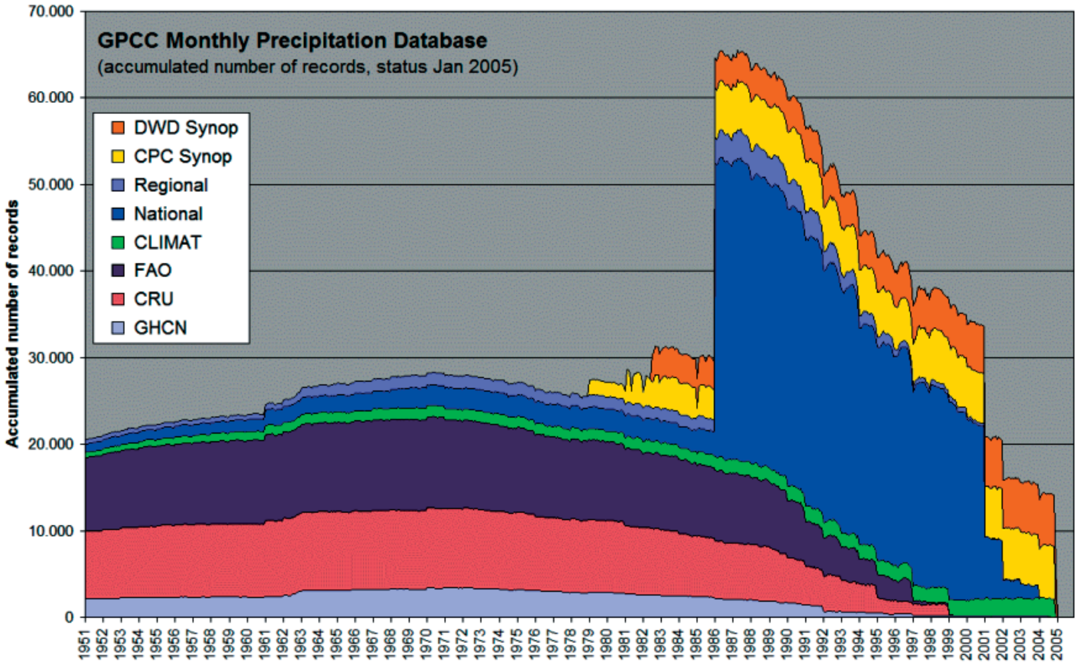

Figure 1 Temporal evolution of the availability of monthly precipitation data since 1951 at the GPCC (Global Precipitation Climatology Centre) differentiated between various data sources.

Most striking is the maximum of the data availability around 1986. This corresponds with the beginning of the GPCC's primary analysing period. However, several historical data sources (CRU, FAO, GHCN) allow for an extension of the analysed period further into the past. As data from the GHCN at the time are only partially integrated into the data-base of the GPCC the availability of historical data is currently restricted to FAO and CRU. The decline in data availability in the most recent period that may be attributed to the delayed update of most data sources can be partially compensated through the use of near real-time transferred GTS-data (Climat, GPCC-Synop, CPC-Synop).

3 QUALITY CONTROL OF STATION META DATA

Different data sources contain partially concurrent data for the same stations. In order to avoid duplicates and false assignment of data records it is necessary to conduct a thorough quality control of station-meta data. This is a substantial step towards the compilation of long station time series from different data sources.

The geographical position, station height, station name and additional parameters like WMO code are used to identify stations. Additionally, precipitation data for overlapping periods are compared in order to correctly assign stations.

A mostly automated procedure succeeds in roughly half of all station assignments. The remaining cases exhibit specific problems that require further often time consuming examination.

The relevance of this quality control step with respect to subsequent analyses is illustrated in Fig. 2. Errors in station meta-data may cause the assignment of stations to wrong climate regions. It is essential for the reliability of the gridded climatology that such errors are detected and if possible corrected by quality control mechanisms.

4 COMPILATION OF LONG TIME SERIES FOR THE PERIOD 1951 TO 2000

Long time series of monthly precipitation totals for the period 1951 to 2000 are compiled on the basis of the various available data sources. For each month of each series with concurrent availability of data from more than one data source it has to be decided which observation from which data source to use for constructing the series.

This decision turns out to be delicate because as it is shown in Table 1 there are partially striking differences between data for the same station for the same month that originate from different sources. Whereas the historical data sources (FAO, GHCN, CRU) correspond well with each other and also with data from national suppliers (National) especially the two data sources originating from aggregated synoptic observations (GPCC-Synop, CPC-Synop) show only limited agreement with each other but also with all other data sources.

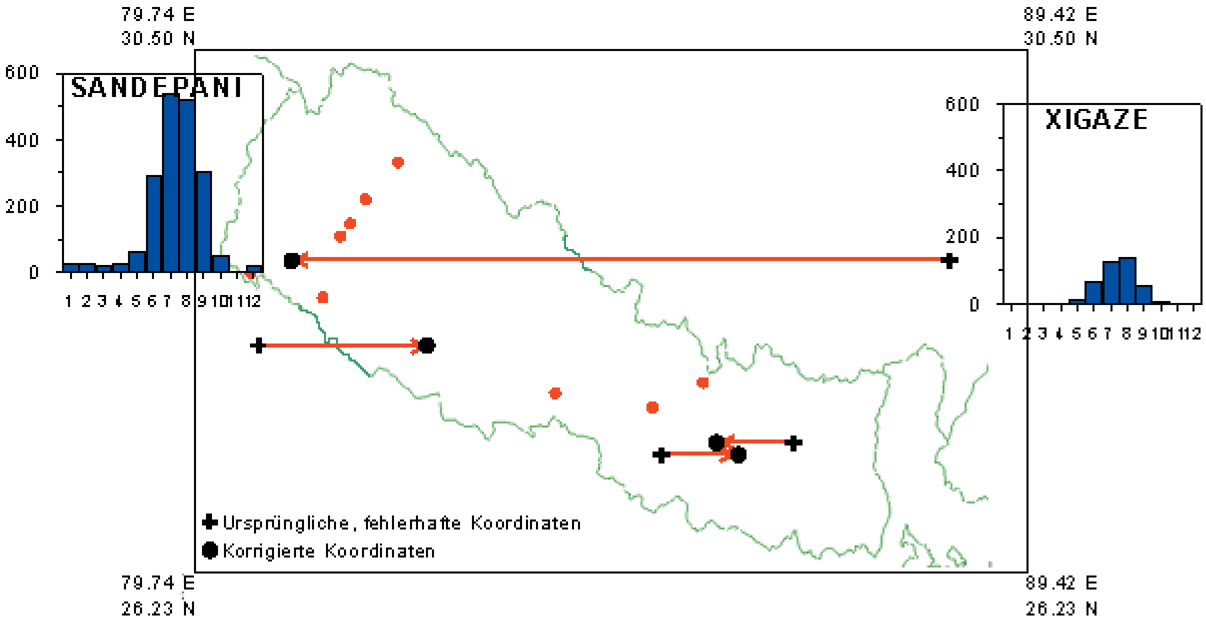

Figure 2 Originally wrong and during quality control corrected geographical coordinates of several Nepalesian stations. The seasonal variation of precipitation for Sandepani is shown for its correct and its originally wrong location (neighbouring station Xigaze) illustrating the climatological relevance of quality control of station-metadata.

In order to minimise the risk of generating temporal inhomogeneities due to changes between data from different sources a strategy for the compilation of long time series is developed that ensures that the minimum possible number of different data sources is used to construct each series. Additionally if possible only data from sources that exhibit a high degree of similarity are combined into one series. And finally data from GPCC-Synop and CPC-Synop - the two data sources that show the most distinct differences to all other sources - are only incorporated if all other sources have no data for the specific month.

Table 1 Mean number (in %) of equal monthly data (±1mm) available for corresponding months and stations from different data-sources. Averaged over 9343 Stations for the period 1951-2000.

	National	FAO	GHCN	CRU	Climat	Regional	GPCC	CPC
National	**-**	**91.3**	**88.4**	**93.1**	**85.4**	**85.6**	**40.2**	**45.2**
FAO	**91.3**	**-**	**94.7**	**90.5**	**83.6**	**89.7**	**38.0**	**56.2**
GHCN	**88.4**	**94.7**	**-**	**98.8**	**84.3**	**74.2**	**43.1**	**56.9**
CRU	**93.1**	**90.5**	**98.8**	**-**	**87.3**	**87.7**	**39.0**	**53.6**
Climat	**85.4**	**83.6**	**84.3**	**87.3**	**-**	**74.6**	**42.2**	**56.3**
Regional	**85.6**	**89.7**	**74.2**	**87.7**	**74.6**	**-**	**41.6**	**39.8**
GPCC	**40.2**	**38.0**	**43.1**	**39.0**	**42.2**	**41.6**	**-**	**38.6**
CPC	**45.2**	**56.2**	**56.9**	**53.6**	**56.3**	**39.8**	**38.6**	**-**

According to this strategy finally 9343 long station time series with a minimum data availability of 90% during the period 1951 to 2000 are compiled. The spatial distribution of these time series is depicted in Fig. 3.

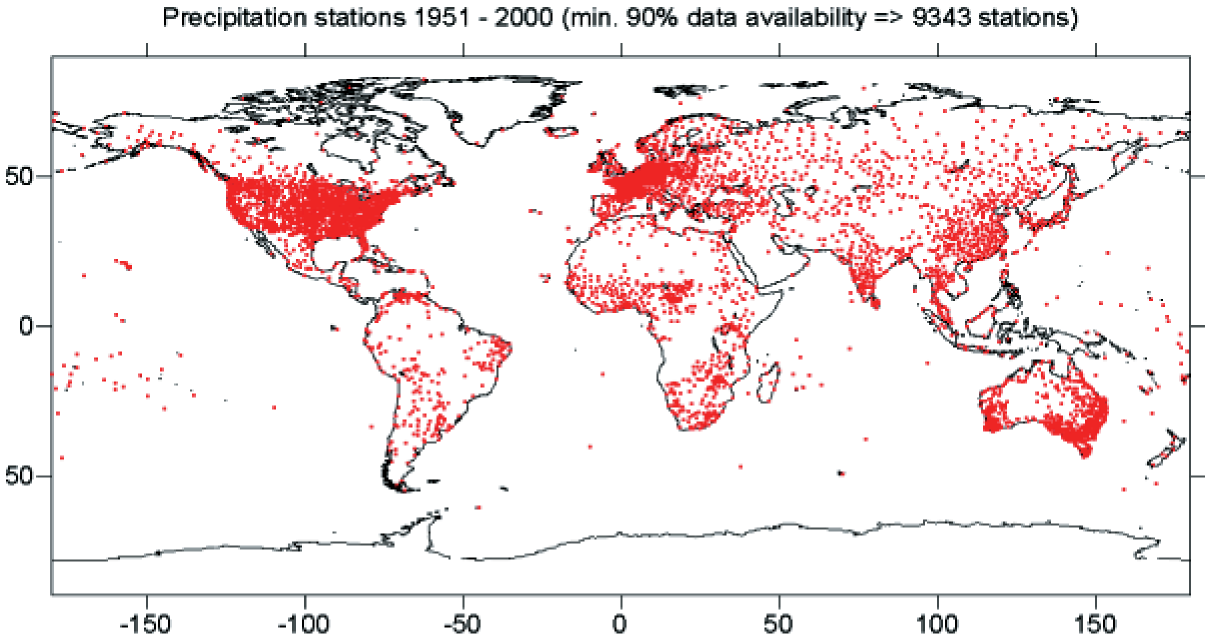
Precipitation stations 1951 - 2000 (min. 90% data availability => 9343 stations)
50
0
50
-150
-100
-50
0
50
100
150

Figure 3 Spatial distribution of 9343 compiled monthly precipitation time series with a minimum of 90% data availability during the period 1951 to 2000.

Fig. 4 depicts the temporal variations of the contributions of each data source to the compiled series in the period 1951 to 2000.

The largest fraction of all compiled time series accounting for about 49% of all monthly data originates from national data sources. Another 47% stem from the three historical data sources (CRU, FAO, GHCN). The data sources Regional and Climat together account for roughly 2% and less than 2% of all data are derived from GPCC-Synop and CPC-Synop respectively. However it becomes clear from Fig. 4 that the latter two data sources are indispensable for the compilation of long time series as they are the only available data source for a number of stations during the two most recent decades. Since the beginning of the 1990's there is also an increase of missing data - reaching its absolute maximum of roughly 15% in 2000 - to be registered that may be attributed to the delay concerning the update of data sources.

The majority (about 55%) of all station time series comprise data of more than one data source but only about one fifth of all stations incorporate data from GPCC-Synop and CPC-Synop, the two data sources with greatest deviations from all other sources. Therefore it can be assumed that the number of temporal inhomogeneities generated by the compilation process is minimized.

5 QUALITY CONTROL OF LONG TIME SERIES

With respect to planned climate-variation studies based on the gridded data it is of great importance to detect and if necessary eliminate temporal as well as spatial outliers in the compiled station time series and – in a subsequent step - to detect temporal inhomogeneities within the series.

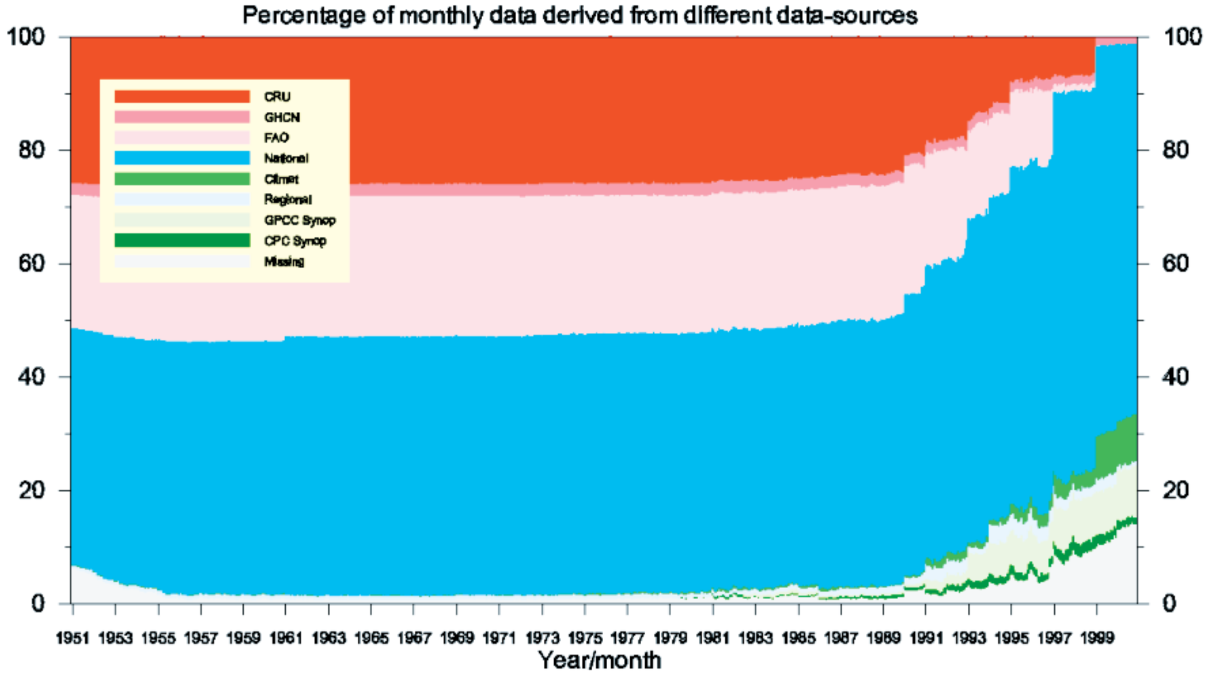

Figure 4 Temporal variations of the contributions of different data-sources to the compiled series in the period 1951 to 2000.

Temporal checks for outliers are performed utilising the sample distribution of each calendar month for each station. Values are flagged as potential outliers if

$$X_i - q50 \; > f \; IR$$

where *Xi* is the monthly value of year *i*, q^{50} is the 50th percentile, *IR* is the interquartile range (75th percentile minus the 25th percentile) and *f* is the multiplication factor. We used a value of 4.00 as multiplication factor. This factor turned out to be adequate for the detection of temporal outliers in monthly precipitation data (Eischeid et al. 1995).

Checks for spatial outliers use nearby stations to estimate a monthly value for a specific time series in a specific month (Eischeid et al. 1995, Peterson and Vose 1997). Several methods may be used to interpolate from surrounding stations to a target station (see for example Eischeid et al. 1995). We used a simple averaging algorithm that estimates each monthly value of each station from its surrounding stations as

$$X_i = \left[\sum_{j=1}^{k} p_j^2 R_{ji} \overline{X} / \overline{R_j} \right] / \sum_{j=1}^{k} p_j^2$$

where X_j is the target stations estimated monthly value of year *i*, p_j is the pearson correlation coefficient between the target station and surrounding station *j*, R_{ji} is the monthly value of surrounding station *j* of year *i*, $\overline{X}$ is the long-term monthly mean of the target station and $\overline{R}_{ij}$ is the long-term monthly mean of surrounding station *j*. Only data from surrounding series that have a pearson correlation of 0.7 or greater with the target station time-series are used for this estimate.

By analysing the difference series between the target stations observed and estimated monthly series (reference series) it is determined for each potential temporal outlier if the observed value is also an outlier in a regional context. Although for the estimation of reference series a much more comprehensive data-base as shown in Fig. 3 (additionally containing time series with up to 70% missing data in the period 1951 to 2000) is used, the creation of reference series wasn't possible for all stations. Thus the spatial outlier check could not be performed for all temporal outliers. Finally all data points that failed both outlier tests were removed prior to interpolation.

The reference series that are determined as described above according to Peterson and Easterling (1994) are also the basis of the relative homogeneity test according to Alexandersson (1986) which is applied to the time series of ratios between compiled station time series and reference series. In order to ensure that only time series without significant inhomogeneities are incorporated into the reference series used for homogeneity testing the following test strategy according to Rapp and Schönwiese (1995) is applied. In a first step all time series are used to create reference series and the homogeneity of all available time series is tested. In a second step only time series that can be assumed as homogenous according to the results of the first application of the homogeneity test are used to create reference series for the second and final run of the homogeneity test. As the Alexandersson test only detects one, the most significant inhomogeneity in one run each time series is homogenized if a significant inhomogeneity was detected and is tested again. This procedure continues until no further significant inhomogeneity is detected. Finally all detected significant inhomogeneities are homogenized (starting with the most recent one) by utilizing the mean quotients between target series and reference series calculated for homogeneous subintervals before and after each break year. In order to avoid unreasonable monthly precipitation adjustments the annual course of the monthly calculated adjustment factors is smoothed by a 7-point Gaussian low-pass filter.

6 INTERPOLATION AND GRIDDING

Many methods are known for the interpolation of climate data. Therefore a jackknifing procedure is used in order to compare a variety of these methods with respect to monthly precipitation sums and the station density given.

The Jackknife Error is the difference between the interpolated and the observed value at a station, leaving the station observation out for the interpolation. It therefore measures how different the interpolation result would be at a given station location if there was no station. The higher the Jackknife Error the larger is the spatial variability compared to the station density. Reasons may either be erroneous values, or rather local information as it is the case with high mountain stations. Thus, a high Jackknife Error may indicate sparse or erroneous data as well as specific climate characteristics of the station under consideration.

With respect to different errors (relative, absolute, root-mean-square error) the interpolation method by Krige (1962) applied to relative deviations reveals to be the best with respect to the underlying data, closely followed however by the method of Shepard (Shepard, 1968). Though Shepards Method is standard with the Global Precipitation Climatology Centre we used Kriging in this application.

Monthly precipitation totals reveal pronounced spatial structures. Within Germany about 50% of the long term spatial variability can be attributed to altitude. Despite this knowledge no deterministic spatial structures are used for the interpolation. The low station density in some areas of the world does not allow to reliably utilise this information worldwide. Experiments revealed that though the results could be optimised for some regions of the world the incorporation of deterministic structures with respect to the given data would lead to worse results for other regions.

The observations of the 9343 stations are interpolated on a 0.5° x 0.5° global land surface grid. Greenland and Antarctica are left out because of lack of data. This led to about 71.000 grid points, i.e. 8 times the number of stations. According to the inhomogeneous station distribution the reliability of the results varies from region to region. This should be kept in mind when interpreting the results. In order to estimate grid averages of the total precipitation the interpolated values at the four corner points of a 0.5° x 0.5° grid cell are averaged. Global maps with 1° and 2.5° resolution are produced by area-weighted averaging of the 0.5° grids.

7 TREND INVESTIGATIONS

The gridded precipitation climatology offers a basis for climate trend investigations. From the variety of properties that may be analysed (e.g. absolute and relative linear trends, trend-to-noise ratios) only two examples are provided here. Firstly, this is the temporal variability of the annually averaged mean monthly precipitation total averaged over the global land surface (Fig. 5). No significant trend may be detected thus indicating that no evidence of a globally enhanced hydrological cycle can be concluded from the data.

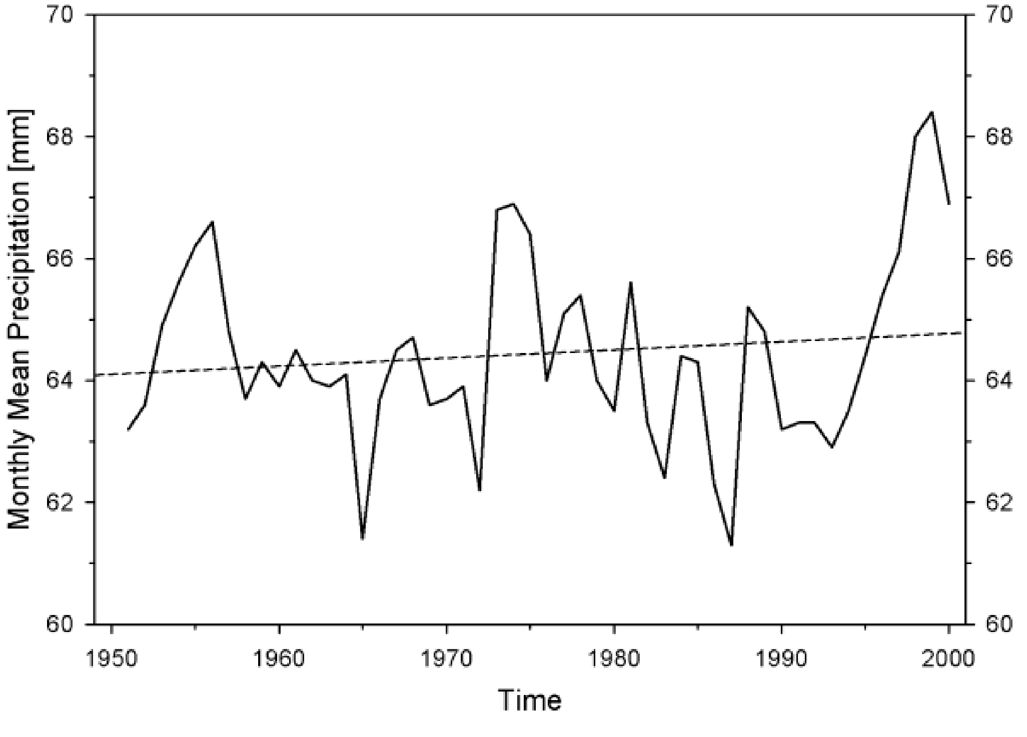

Figure 5 Time series of the annual mean of monthly precipitation totals averaged over the global land surface together with the linear trend from 1951 to 2000.

Additionally the Mann-Kendall-trend test is applied to each of the grid points and for each calendar month. This allows to distinguish regions where no non-random trends are visible from those with significant positive or negative ones. As an example Fig. 6 shows the local trend significance of annual precipitation totals.

8 OUTLOOK

A first version of a globally gridded homogeneous monthly precipitation climatology from 1951 to 2000 is produced and available at http://www.dwd.de/vasclimo

With an increasing amount of available information it is going to be updated irregularly. Different versions will be produced in order to optimally meet different needs, i.e. for trend investigations extreme precipitation should be removed, whereas they are crucial for the investigation of temporal variability.

The interpolated data are going to be compared to the products of other groups (i.e. New et al. 2000). Finally the grid data will be used for climate variability investigation within the project VASClimO.

ACKNOWLEDGEMENTS

The joint project VASClimO (Variability Analyses of Surface Climate Observations) is part of the German climate research programme (DEKLIM) and is supported financially by the Bundesministerium für Bildung und Forschung under DEKLIM project-No. 33 11 0307.

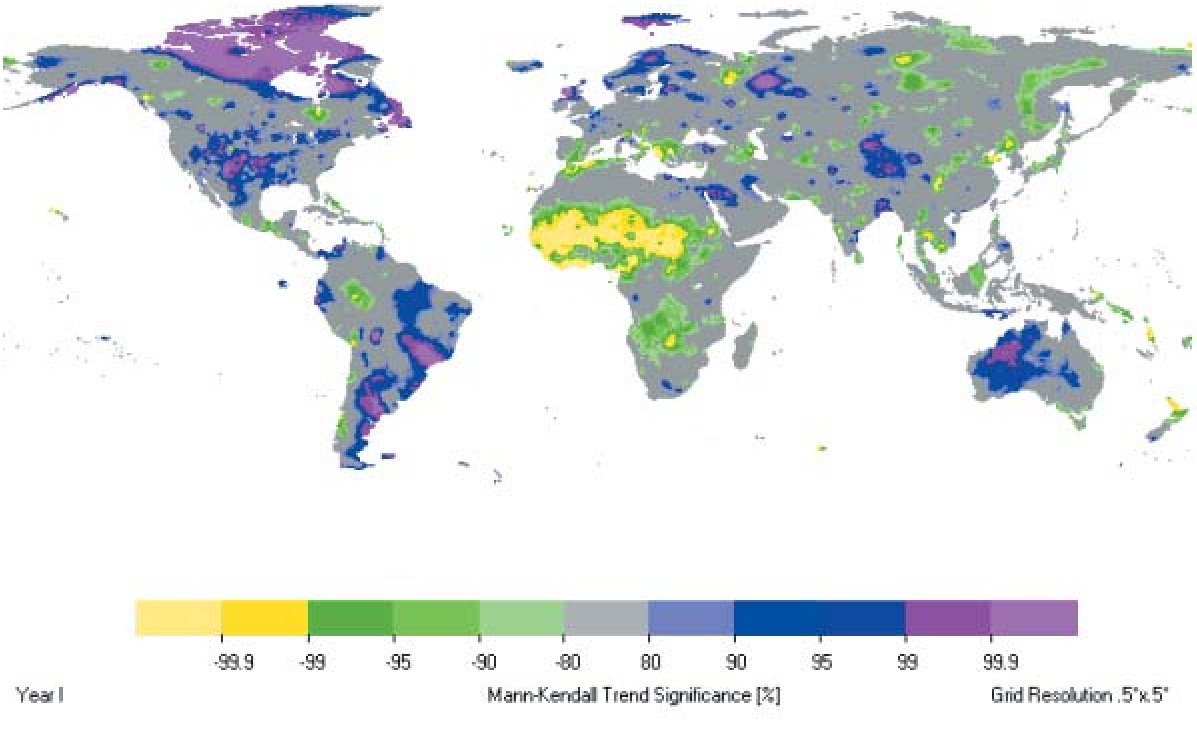

Figure 6 Significance of local trends in annual precipitation totals within the period from 1951 to 2000 according to the Mann-Kendall-Trend test. Significance of regions with negative trends are given in green and yellow, for positive trends the regions are given in blue and red.

References

Adler, R.F., G.J. Huffman, A. Chang, R. Ferraro, P. Xie, J. Janowiak, B. Rudolf, U. Schneider, S. Curtis, D. Bolvin, A. Gruber, J. Suesskind, P. Arkin und E. Nelkin (2003). The Version-2 Global Precipitation Climatology Project (GPCP) monthly precipitation data analysis (1979-present), Journal of Hydrometeorology, 4, 1147-1167.

Alexandersson, H. (1986). A homogeneity test applied to precipitation data, J. Climatol., 6, 661-675.

Chen, M., P.Xie, J. E. Janowiak und P. A. Arkin (2002). Global Land Precipitation: A 50-yr monthly analysis based on gauge observations. Journal of Hydrometeorology, 3, 249-266.

Dai, A. und A. D. Del Genio (1997). Surface observed global land precipitation variations during 1900-1988, J. Climate, 10, 2943-2962.

Eischeid, J. K., C. B. Baker, T. R. Karl and H. F. Diaz (1995): The quality control of long-term climatological data using objective data analysis, J. Appl. Met. 34: 2787-2795.

Huffman, G. J., R. F. Adler, P. Arkin, A. Chang, R. Ferraro, A. Gruber, J. Janowiak, A. McNab, B. Rudolf und U. Schneider (1997).

The Global Precipitation Climatology Project (GPCP) combined precipitation dataset, Bulletin of the American Meteorological Society, 78, 5-20.

Krige, D.G. (1962). Statistical applications in mine valuation. J. Inst. Min. Survey. S. Afr., 12(2), 45-84, 12(3), 95-136.

New, M. G., M. Hulme und P. D. Jones (2000). Representing twentieth-century space-time climate variability. Part II: development of 1901–1996 monthly grids of terrestrial surface climate, Journal of Climate, 13, 2217-2238.

Peterson, T. C. and D. R. Easterling (1994): Creation of homogeneous composite climatological reference series, Int. J. Climatol. 14: 671-679.

Peterson, T. C. and R. S. Vose (1997): An overview of the global historical climatology network temperature database, Bull. Of the American Met. Soc. 78/12: 2837-2849.

Peterson, T. C., R. S. Vose, R. Schmoyer and R. Razuvaev (1998): Global historical climatology network (GHCN) quality control of monthly temperature data, Int. J. Climatol. 18: 1169-1179.

Rapp, J. and C.-D. Schönwiese (1995): Atlas der Niederschlags- und Temperaturtrends in Deutschland 1891-1990.- Frankfurter Geowiss. Arb., Serie B, Band 5.

Rudolf, B., H. Hauschild, W. Rueth und U. Schneider (1994). Terrestrial precipitation analysis: Operational method and required density of point measurements, in Global Precipitation and Climate Change, Desbois, M. und F. Desalmand (eds.) (Springer, Berlin), 173-186.

Shepard, D. (1968). A two-dimensional interpolation function for irregularly spaced data. 23rd ACM National Conference. Brandon Syst. Press: Princeton, USA, 517-524.

Xie, P., B. Rudolf, U. Schneider und P.A. Arkin (1996). Gauge-based monthly analysis of global land precipitation from 1971 to 1994. Journal of Geophys. Research, 101, 19023-19034.