Pergamon

Phys. Chem. Earth (B), Vol. 26, No. 5-6, pp. 403–409, 2001

1464-1909/01/$ - see front matter

PII: S1464-1909(01)00026-0

Verification of GPCP-1DD Global Satellite Precipitation Products Using MAP Surface Observations

P. Skomorowski[1], F. Rubel[1] and B. Rudolf[2]

[1]WG Biometeorology, Institute of Medical Physics and Biostatistics, VUW, Vienna, Austria
[2]Global Precipitation Climatology Centre, DWD, Offenbach, Germany

Received 21 June 2000; accepted 9 October 2000

Abstract. Daily satellite precipitation estimates have become available for the first time by the Global Precipitation Climatology Project. The experimental version of the GPCP-1DD product, recently launched and available from NASA, provides the research community with 1 degree daily precipitation fields exclusively computed from satellite measurements. These comprise IR estimates from geostationary satellites and SSM/I data in the tropics as well as TOVS data from polar orbiting satellites at higher latitudes. The purpose of the present study is to verify the TOVS based GPCP-1DD rain rates over Central Europe with the surface precipitation measured with the dense rain gauge network operated during *The Mesoscale Alpine Programme* (MAP). 3 100 high quality daily rain gauge measurements during June/July 1997 have been objectively analyzed on the grid of the GPCP-1DD products. Verification results are given here in terms of difference fields (mean error= -0.59 mm/day) , correlation coefficients (mean monthly value =0.57) as well as accuracy scores (probability of detection=0.6, false alarm ratio=0.2) and skill scores (true skill statistics=0.36). These scores are slightly inferior than those known from NWP model verifications.

1 Introduction

Continuous global precipitation observation is a challenging task. It can be done only by the use of satellite techniques, because over wide areas of the globe no rain gauge or radar measurements are available. Here our interest is to verify the new experimental precipitation product of the Global Precipitation Climatology Project, the GPCP-1DD (resolution: 1 degree, 1 day) data. These global daily precipitation fields are based on a combination of different satellite estimates. In the 40 ° N-S belt IR estimates from geostationary satellites calibrated with SSM/I data have been used. Outside this region the rain estimates are based on TOVS data from the polar orbiting satellites NOAA-12 and NOAA-14. These two satellites, with 0130/1330 and 0730/1930 local time equator crossing times, are flying simultaneously and provide global precipitation estimates based on an empirical relationship between rain gauge observations and a function of the cloud-top pressure, fractional cloud cover and relative humidity profile (Susskind et al., 1997). The GPCP 1 degree daily precipitation fields over the verification region of the European Alps (Fig. 1) are therefore mainly based on TOVS-data, expect the calibration with monthly GPCP products (see e.g. Huffman et al. (1997), Rudolf et al. (2000)).

For the verification of this precipitation product regions with very dense rain gauge networks are needed. One such region are the European Alps. The rain gauge data from this region are stored in The Mesoscale Alpine Programme (MAP) data archive and fulfil all requirements to be used as ground truth data for the verification of satellite estimates or model predictions. Further comprehensive information on the precipitation climatology of the Alps is available from Frei and Schär (1998).

To prove the reliability of the GPCP-1DD products, the verification measures commonly used for model verification (Stanski et al., 1989) have been applied. Results in terms of scores are directly comparable with corresponding scores from NWP or climate model verifications.

Over the Alps currently no comprehensive model verifications have been done, but various case studies are documented in the literature. These comprise for example studies on the application of different parametrization schemes of the MM5 model (Cherubini et al., 1999) or on the accuracy of DM forecasts (Kreil and Volkert, 1999).

Nevertheless, comprehensive results from model verifications are available from other regions and give an

---

Correspondence to: F. Rubel

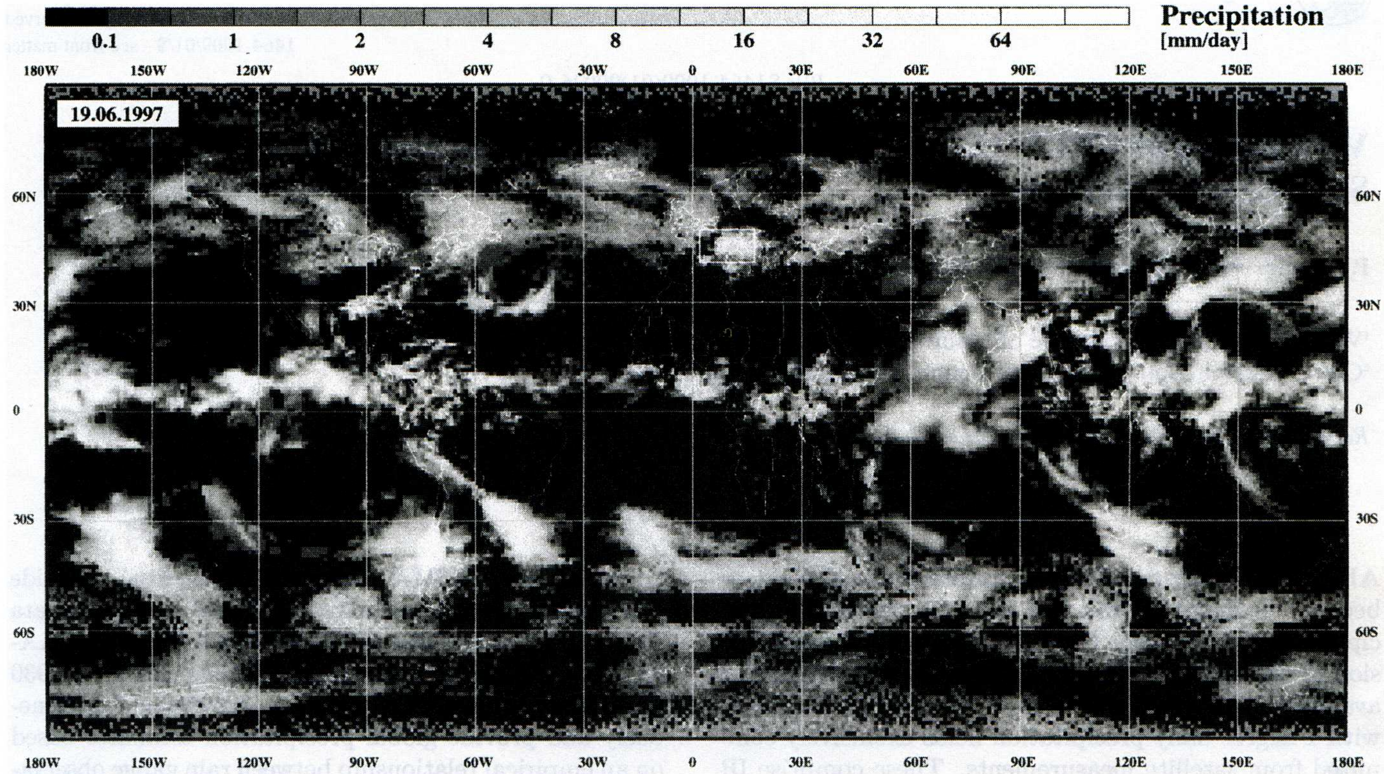

Fig. 1. Global satellite precipitation estimates for 19 June 1997 with the MAP verification region of 15 x 9 grid boxes (longitude: 2 ° E to 17 ° E, latitude 41 ° N to 50 ° N). Units mm/day.

idea of the accuracy and skill of modern NWP models. For example, the well documented verifications of the Australian LAPS model (Ebert et al., 1997), comprising scores from various climatic regions, could serve as preliminary model reference.

2 Data and Method

About 3 100 stations with rain gauge measurements, archived in the MAP data centre, have been used to verify the GPCP precipitation fields. The distribution of the stations has been documented by Frei and Schär (1998). However, during the verification period of June to July 1997, only for parts of the Alps the total number of rain gauges available has been collected. This is true for France, Switzerland and Slovenia, were the maximum coverage with stations is 10 times denser than the density of the synoptic network. For the other regions, comprising Austria and parts of Germany, Croatia and Italy only rain gauges from the synoptic network were available.

First of all areal precipitation amounts using the grid of the GPCP-1DD products have been calculated from the MAP rain gauge observations. For that an optimum interpolation method was applied. This method (see e.g. Rubel and Hantel, 2000) considers both the inhomogeneous distribution of the stations as well as the spatial

Observed
rain gauge analysis

Predicted
satellite estimates

	yes	no	
yes	h h...hits	f f...false	h+f
no	m m ... misses	z z ... zero	m+z
	h+m	f+z	n=h+f+m+z

Fig. 2. Contingency table for the calculation of accuracy and skill scores.

structure of the precipitation process at the scale considered (Rubel, 1996). Additionally to each areal precipitation estimate the interpolation error is known. Note that this interpolation error depends on the station density, the spatial auto-correlation function and the size of the grid boxes. The verification of the GPCP-1DD products is then performed by using only grid boxes with low interpolation error.

Having MAP and GPCP precipitation fields comparable on the same grid, a first visual comparison has been done by viewing difference fields (not shown). To quantify the verification results continuous and categorical statistics have been used. The measures of the continuous statistic comprises mean error, mean absolute error, RMS error and correlation coefficient. Each of these measures was calculated comparing daily grid values from observed (MAP) and estimated (GPCP) precipitation fields over the MAP model domain (Fig. 1).

The categorical statistics are based on two dimensional contingency tables (Fig. 2); one distinguishes between accuracy and skill scores (Wilks, 1995).

2.1 Accuracy Measures

Accuracy measures sum up the quality of a set of forecasts by comparing individual pairs of forecasts and observations. Several scalar measures of the accuracy are known. The simplest measure is known as accuracy (ACC) or hit rate (HR) and is the ratio of correct estimates to the total number of estimates. It can be computed from the contingency table (Fig. 2) as

$$ACC = \frac{\text{correct estimates}}{\text{total estimates}} = \frac{z+h}{n} \qquad (1)$$

Here we focus on probability of detection (POD) and false alarm ratio (FAR). POD is the fraction of those occasions where the estimation event occurred when it was also observed.

$$POD = \frac{\text{correct rain estimates}}{\text{rain observations}} = \frac{h}{h+m} \qquad (2)$$

and ranges from one for perfect estimates to zero. FAR is the fractional number of times that the event was predicted to occur but it did not occur. FAR is computed as

$$FAR = \frac{\text{false alarms}}{\text{rain forecasts}} = \frac{f}{f+h} \qquad (3)$$

The best FAR value is zero and the worst one.

2.2 Skill scores

With skill scores the improvement of the GPCP estimates over some reference estimates such as random chance, persistence or climatology can be measured. In general a skill score (SS) is defined as

$$SS = \frac{ACC_{estimate} - ACC_{reference}}{ACC_{perfect} - ACC_{reference}} \qquad (4)$$

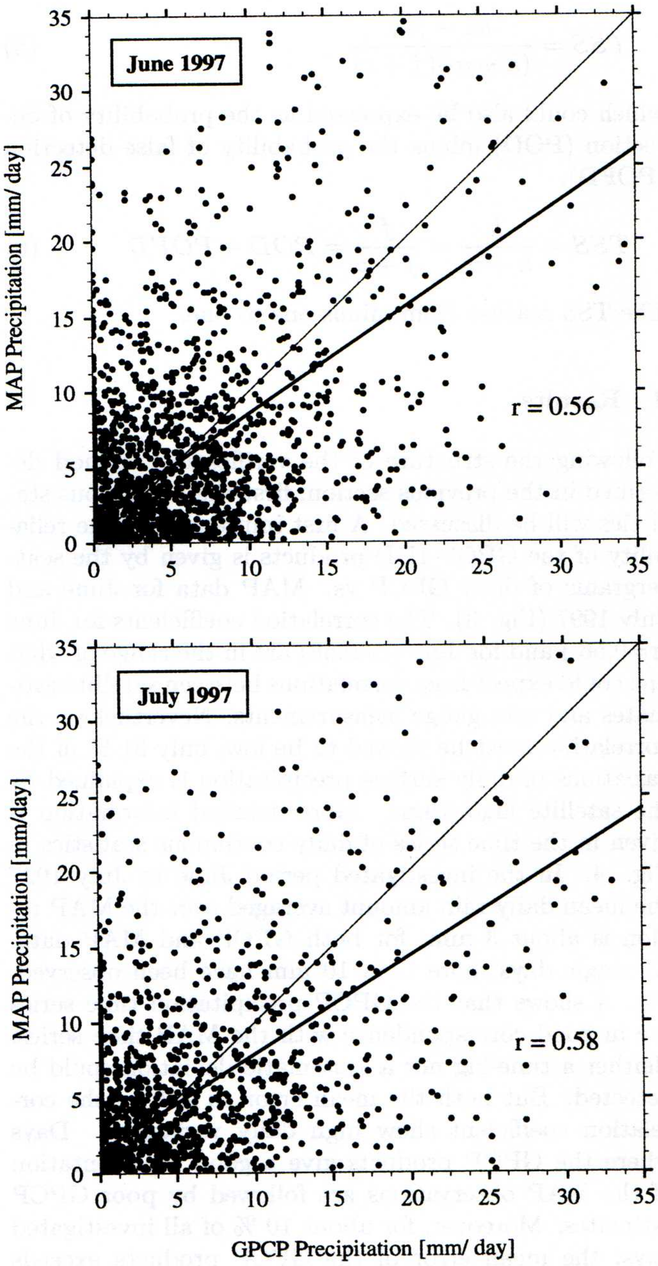

Fig. 3. Monthly scattergram of daily values, GPCP vs. MAP for June and July 1997.

A perfect estimate always produces a skill of 1, while an estimate that is not better than the reference produces a skill of 0, and estimates that are worse than the reference have negative skills.

Here we used the true skill statistics (TSS) also known as Hanssen and Kuipers score. The TSS is computed with

$$TSS = \frac{hz - fm}{(h+m)(f+z)} \tag{5}$$

which could also be expressed as the probability of detection (POD) minus the probability of false detection (POFD).

$$TSS = \frac{h}{h+m} - \frac{f}{f+z} = POD - POFD \tag{6}$$

The TSS reaches from minus one to one.

3 Results

Following the structure of the verification method described in the previous section, first the continuous statistics will be discussed. A first impression of the reliability of the GPCP-1DD products is given by the scattergrams of daily GPCP vs. MAP data for June and July 1997 (Fig. 3). The correlation coefficients for June (r=0.56) and for July (r=0.58) are in the range of what one could expect from correlations between satellite estimates and rain gauge measurements. Nevertheless, the correlation must be viewed to be low, only 31 % of the variations of daily surface precipitation is explained by the satellite algorithms. More detailed information is given in the time series of daily continuous statistics in Fig. 4. In the investigated period June to July 1997 the mean daily rain amount averaged over the MAP region is about 3 mm, for both GPCP and MAP data. At single days more than 10 mm have been observed. Fig. 4 shows that the GPCP precipitation time series are in good correspondence with the MAP time series. Neither a time-lag nor a systematic deviation could be detected. But both the mean error as well as the correlation coefficient show high daily variations. Days where the GPCP products give a good representation of the MAP observations are followed by poor GPCP estimates. Moreover, for about 10 % of all investigated days, the mean error of the GPCP products exceeds more than 2 mm/day or the correlations with the MAP observations are extremely poor with zero or negative correlation coefficients. This indicates, that the reliability of the GPCP-1DD products depends highly on specific weather situations or rain events.

The time series of the used measures, probability of detection (POD), false alarm ratio (FAR) and true skill statistics (TSS) are shown in (Fig. 5). A default threshold value of 0.1 mm/day was defined to distinguish between rain yes/no events. Thus precipitation values below 0.1 mm/day are treated as zero values corresponding to the measurement resolution of rain gauges. If one looks for example at the period 18 to 20 July 1997, the first two days are characterized by nearly perfect GPCP precipitation estimates. The differences between area averaged GPCP and MAP values are zero and the correlation is high (0.7 - 0.8). The accuracy of these two days is described by high values of POD (0.6 - 0.7) and low values of FAR (0.0 - 0.1) and a high skill by TSS (0.5 - 0.6). This is in opposite to the last day of this period. On 20 July not only the mean error but also the correlation with the MAP observations is zero. POD and FAR which are of same order (0.3) and TSS (0.1) confirm the low reliability of the GPCP estimate on this day.

The summary statistics given in Fig. 6 opens the possibility to compare the verification results with those of other authors. Therefore, additionally to the above described scores the commonly used verification measures hit rate (HR), critical success index (CI) also known as threat score (TS) and bias score (BS) have been calculated. Detailed descriptions of these scores are given by e.g. Wilks (1995). The results for June and July 1997 are very similar, with slightly better GPCP estimates in July. A further question of interest is, how higher precipitation thresholds effect the verification results. Fig. 7 shows the decreasing reliability of the GPCP-1DD products with increasing thresholds. The lower accuracy and skill of higher precipitation predictions is also known from model verifications (Cherubini et al., 1999).

4 Conclusions

In this paper we presented the first verifications of the experimental version of the new GPCP-1DD products. These verifications over the very dense gauged European Alps (Frei and Schär, 1998) show, that the accuracy of the GPCP-1DD precipitation estimates is comparable with those from other satellite estimates. For example in *The First WetNet Precipitation Intercomparison Project*, summarized in Barrett (1994), correlation coefficients between r=0.3 and r=0.8 have been calculated.

But it should be noted, that the MAP rain gauge measurements have not been corrected for systematic measurement errors. It could be expected, that corrected measurements are about 7 % higher than the raw measurements used in this study (Rudolf, 1995). Actual studies over the Alps concerning synoptic stations only confirm these results (Rubel et al., 2000). Concerning this, the mean error of the GPCP precipitation estimates increases from minus 0.6 mm/day (Fig. 6) to minus 0.8 mm/day. Therefore the under-estimation of the GPCP products is about 20 %. On the other hand the results presented have been obtained from a very short period; the final verification period will cover sev-

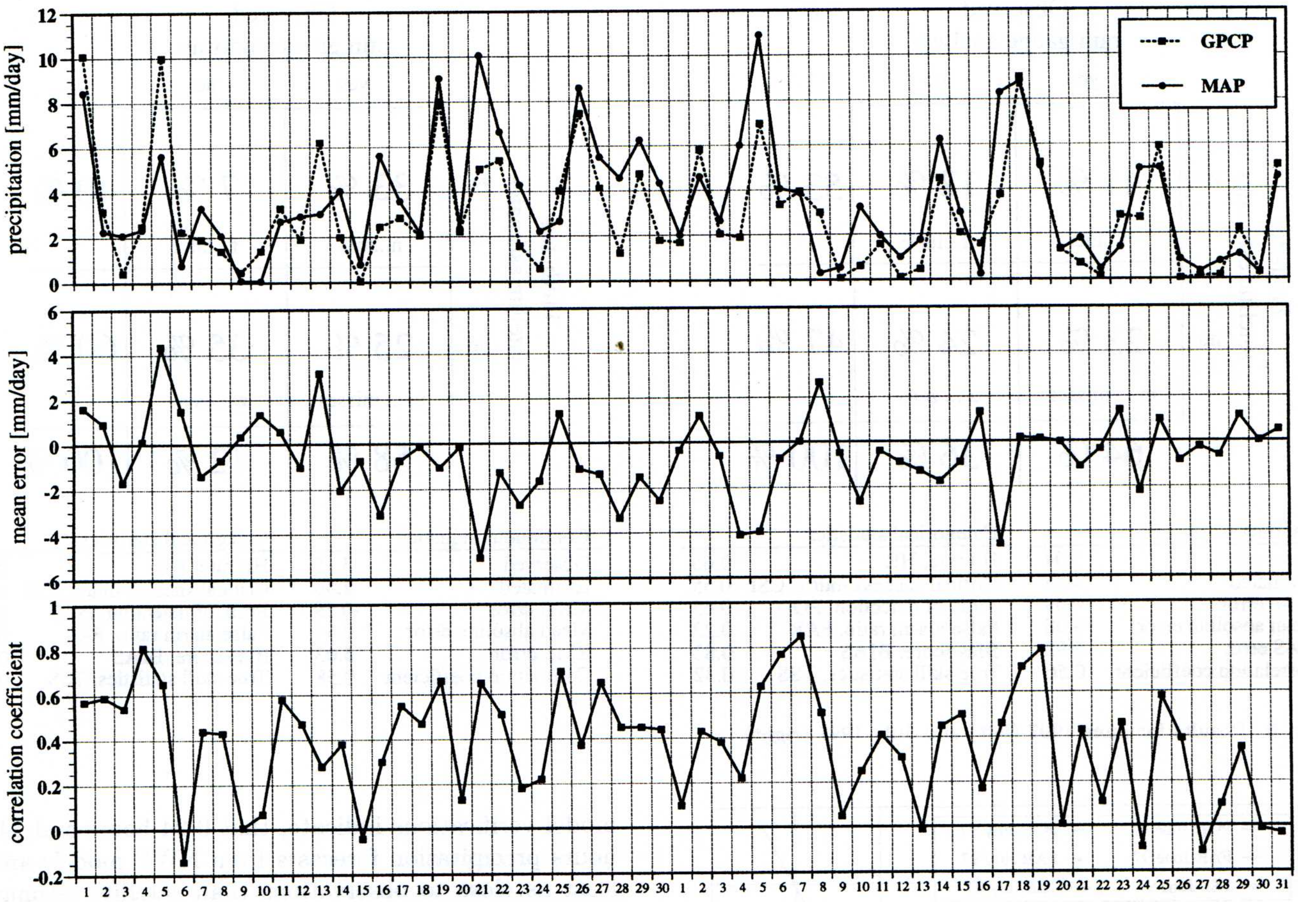

Fig. 4. Time series of continuous statistics: mean daily precipitation amounts estimated from satellite (GPCP), analyzed from rain gauges measurements (MAP), mean errors of GPCP estimates (difference GPCP minus MAP) as well as correlation coefficients for June to July 1997.

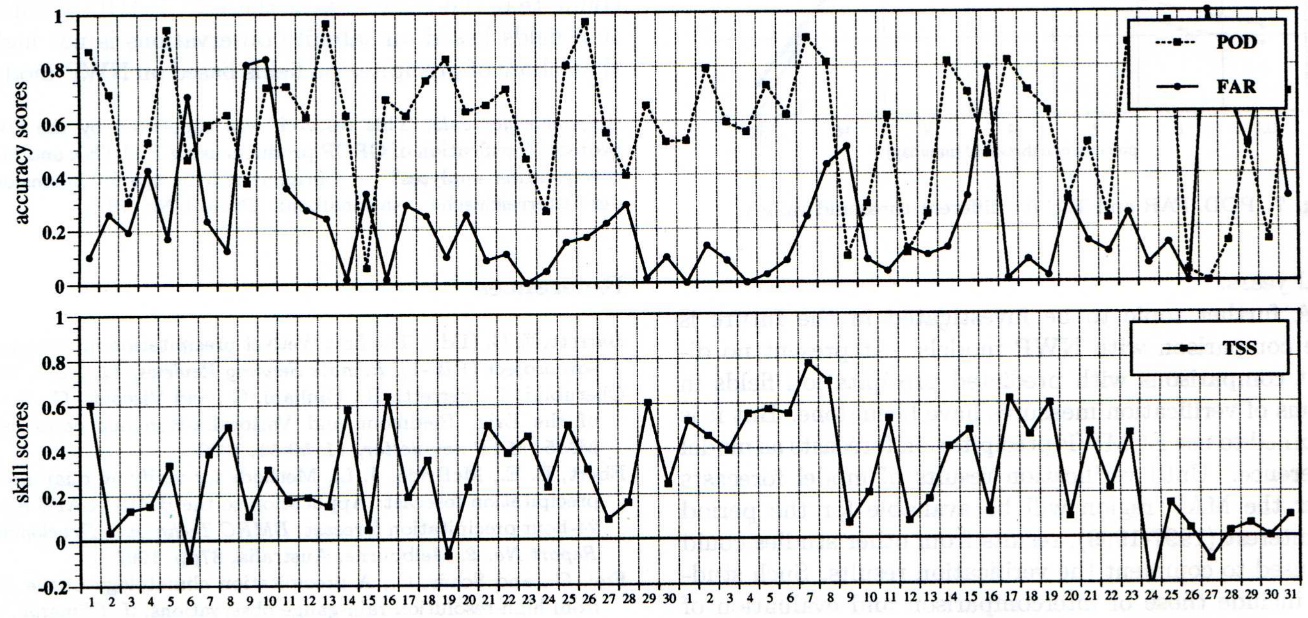

Fig. 5. Time series of categorical statistics: accuracy measures probability of detection (POD) and false alarm ration (FAR) as well as true skill statistics (TSS) for June to July 1997.

		Observed rain gauge analysis		
		yes	no	
Predicted satellite estimates	yes	41 % h...hits	12 % f...false	53 %
	no	23 % m ... misses	24 % z ... zero	47 %
		64 %	36 %	100 %

Continuous statistics	
Observed	3.94
Estimated	3.35
Mean error	-0.59
Mean absolute error	3.38
RMS error	5.92
Correlation coefficient	0.56

Categorical statistics	
Hit rate, HR	0.66
Critical success index, CSI	0.55
Prob. of detection, POD	0.65
False alarm ratio, FAR	0.22
Bias score, BIAS	0.83
True skill statistics, TSS	0.32

		Observed rain gauge analysis		
		yes	no	
Predicted satellite estimates	yes	33 % h...hits	7 % f...false	40 %
	no	25 % m ... misses	35 % z ... zero	60 %
		58 %	42 %	100 %

Continuous statistics	
Observed	3.15
Estimated	2.56
Mean error	-0.59
Mean absolute error	2.64
RMS error	5.03
Correlation coefficient	0.58

Categorical statistics	
Hit rate, HR	0.68
Critical success index, CSI	0.51
Prob. of detection, POD	0.57
False alarm ratio, FAR	0.18
Bias score, BIAS	0.69
True skill statistics, TSS	0.40

Fig. 6. Results for June 1997 (left) and July 1997 (right).

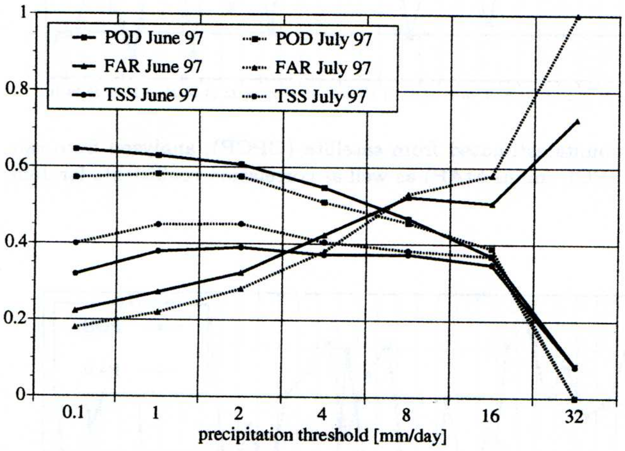

Fig. 7. POD, FAR and TSS for different threshold values.

eral years.

A further topic to be investigated in the future is the comparison with NWP models. At present no direct comparisons with predicted precipitation fields in terms of verification measures have been done. But it is planned to use ECMWF precipitation forecasts as model reference. Until verification results of model forecasts over the MAP region will be available for the period of interest (1997-1999), results from other studies could be used to comment the verification results. Such studies include those of intercomparison and evaluation of precipitation forecasts for MAP seasons 1995 and 1996 (Mladek et al., 2000) and the well documented verification of the Australian Limited Area Prediction System (LAPS) model (Ebert et al., 1997). The scores of these model verifications indicate, that 0-24 hours and 24-48 hours precipitation forecasts from NWP models are at least as close to observations than satellite estimates; both calculated on the same scale. But satellite estimates do not have the systematic model errors due to orography. Of course, they have other systematic ones. Therefore users should be very careful in interpreting results based on GPCP-1DD products, taking into account that the reliability of the GPCP-1DD precipitation fields based on satellite observations is not higher than those of precipitation fields based on NWP-models.

Acknowledgements. This research was supported by the DWD project "Verification of GPCP products using BALTEX and MAP precipitation analyses" (in German) and by the Jubiläumsfonds der Österreichischen Nationalbank (Project No. 7875).

References

Barrett, E. C. (Ed.), The first WetNet precipitation intercomparison project (PIP-1). *Remote Sensing Reviews, 11*, 1-373, 1994.

Cherubini, T., Ferretti, R., Giuliani, G., and Visconti, G., Scores of the Brig, Piedmont and Vaison-La-Romaine cases using MM5. *MAP newsletter, 11*, 50-51, 1999.

Ebert, E. E., McBride, J. L., Methods for verifying quantitative precipitation forecast: Application to the BMRC LAPS model 24-hour precipitation forecast, *BMRC Techniques Development Report No. 2*, Melbourne, Australia, 87pp, 1997.

Frei, C., and Schär, C., A precipitation climatology of the Alps from high-resolution rain-gauge observations, *J. Climatol., 18*, 873-900, 1998.

Huffman, G. J., Adler, R. F., Arkin, P., Chang, A., Ferraro, R., Gruber, A., Janowiak, J., McNab, A., Rudolf, B., and Schneider, U., The global precipitation climatology project (GPCP) combined precipitation dataset. *B. Amer. Meteor. Soc., 78*, 5-20, 1997.

Kreil, C., and Volkert, H., Precipitation in the northern Alpine region: Case-study-type validation of an operational forecast model, *Meteorol. Atmos. Phys., 72*, 161-173, 2000.

Mladek, R., Barckicke, J., Binder, P., Bougeault, P., Brzovic, N., Frei, C., Geleyn, J. F., Hoffmann, J., Ott, W., Paccagnella, T., Patruno, P., Pottier, P., and Rossa, A., Intercomparison and evaluation of precipitation forecasts for MAP seasons 1995 and 1996, *Meteorol. Atmos. Phys., 72*, 111-129, 2000.

Rubel, F., Scale dependent statistical precipitation analysis. Proceedings of the International Conference on Water Resources & Environmental Research: Towards the 21st Century, Kyoto, Japan, Volume I, 317-324, 1996.

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

Rubel, F., and Hantel, M., BALTEX precipitation analysis: Results from the BRIDGE preparation phase. *Phys. Chem. Earth*, This issue, 2000.

Rubel, F., Ungersböck, M., Skomorowski, P., Auer, I., Rudolf, B., Fuchs, T., and Rapp, J., On the correction of systematic rain gauge measurement errors, MAP Climate Workshop, Interlaken, Switzerland, 31 March to 1 April, 2000.

Rudolf, B., The temporal-spatial structure of global precipitation estimates (in German). *Berichte des Deutschen Wetterdienstes, 196*, 153pp, 1995.

Rudolf, B., Hauschild, H., Rüth, W., and Schneider, U., Comparison of raingauge measurements, precipitation estimates from satellite data and model results. *Adv. Space Res., 18*, 53-62, 1996.

Rudolf, B., Gruber, A., Adler, R., Huffman, G., Janowiak, J., and Xie, P. P., GPCP precipitation analyses based on observations as a basis of NWP and climate model verification. *WCRP-109, WMO/TD-No. 985*, Geneva, 196-200, 2000.

Stanski, H. R., Wilson, L. J., and Burrows, W. R., Survey of common verification methods in meteorology. *World Weather Watch Tech. Rep. No. 8, WMO/TD-No. 358*, Geneva, 114pp, 1989.

Susskind, J., P. Piraino, L. Iredell, and A. Mehta, Characteristics of the TOVS pathfinder path A dataset. *B. Amer. Meteor. Soc., 78*, 1449-1472, 1997.

WCRP, Global Precipitation Climatology Project - Implementation and Data Management Plan. *WMO/TD-No. 367*, Geneva, 47pp, 1990.

Wilks, D. S., Statistical Methods in the Atmospheric Sciences. An Introduction. Academic Press, San Diego, 467pp, 1995.