geoland

Integrated Project geoland

GMES products & services, integrating **EO** monitoring capacities
to support the implementation of European directives and policies
related to "land cover and vegetation"

CSP – Algorithm Theoretical Basis Document (ATBD)
WP 8318 – CSP Precipitation

CSP-0350-RP-0008-ATBD8318
Issue 2.00

Book Captain: **Markus Kottek (IMP)**
Contributing Auhtors: **Franz Rubel (IMP)**

Document Release Sheet

Book captain:	**Markus Kottek (IMP)**	Sign	Date	
Approval:	**Franz Rubel (IMP)**	Sign	Date	
Endorsement:	Co-ordinator (MEDIAS)	Sign	Date	
Distribution:	Medias France OLF ONC OFM			

Change Record

Issue/Rev	Date	Page(s)	Description of Change	Release
I1.00	November 8, 2005	all	Methodology update	I2.00

Contents

List of Figures

List of Tables

1 Background of the document

1.1 Executive summary

This document is the baseline for coding a processing line in order to generate the CSP (Core Service bio-geophysical Parameters) precipitation product. After a review of the users requirements a concise description of the methodology generating the precipitation product is given. This description consists of the whole retrieval algorithm including a declaration of the input data followed by remarks of the product quality, the validation procedure, the risk of failure and the used references.

1.2 Scope and objectives

The objective of the Biometeorological Group at the IMP[1] is to develop a global scale daily precipitation product based on existing multi-satellite estimates of precipitation and bias-corrected rain gauge measurements to provide the three CSP observatories, the ONC (Observatory of Natural Carbon fluxes), the OFM (Observatory of Foodsecurity & crop Monitoring) and the OLF (Observatory of global Land cover & Forest change). A major goal is to improve the multi-satellite estimates due to a calibration with the bias-corrected rain gauge data. The bias-correction of the ground-based precipitation measurements is needed because of the under-catch of operational rain gauges. This under-catch is of the order of 5 - 30 % on average. Currently, there does exist no daily operational global precipitation product which is based on bias-corrected rain gauge analyses. The global daily rain gauge data provided by the GPCC[2] will be collected for the period 1997 - 2003 and corrected for systematic measurement errors. This database will be used to calibrate the daily multi-satellite estimates of precipitation provided by the GPCP[3]. The accuracy of the existing operational satellite estimates of precipitation in terms of objective verification scores over Europe is well known. For an operational application it is necessary to improve these satellite estimates. Based on first results of the FP5 demonstration project ELDAS[4] a combined gauge-satellite data set for the period 1997 - 2003 will be compiled and verified over selected regions of the globe. The IMP will deliver daily $1°$ precipitation fields based on the GPCP-1DD (1 degree daily) global multi-satellite product calibrated with global bias-corrected rain gauge analyses based on about 6 000 synoptic stations.

1.3 Content of the document

This document is structured as follows: Chapter 2 on the facing page gives a brief but concise review of the users requirements in terms of product charecteristics and date of delivery. In Chapter 3 on the next page the methodology of retrieving the CSP precipitation product is described. This includes an overview consiting of the general framework and a review of existing methods as well as the whole retrieval algorithm followed by remarks about the validation procedure, the product quality and the risk of failure. The document is completed by some links to other geoland activities and a listing of the used references in Chapter 4 on page 20.

---

[1] Institute of Medical Physics, University of Veterinary Medicine Vienna, Austria
[2] Global Precipitation Climatology Centre, Offenbach, Germany
[3] Global Precipitation Climatology Project, NASA, USA
[4] Development of a European Land Data Assimilation System to predict floods and droughts

---

1.4 Related documents

1.4.1 Input

Overview of former deliverables acting as inputs to this document.

Document ID	Descriptor
CSP-0350-RP-0002	User Requirements
CSP-0350-RP-0005	Service portfolio
CSP-0350-PPR-0022	Quarterly report
CSP-0350-HA-0001	CSP handbook
CSP-0350-RP-0008	ATBD

Tab. 1: Related documents: input.

1.4.2 Output

Overview of other deliverables for which this document serves as input.

Document ID	Descriptor
CSP-0350-RP-0010	Test and benchmarks report
CSP-0350-DAT-0012	The input and output data sets used for tests and benchmarks
CSP-0350-PRD-0000	CSP precipitation product WP8330

Tab. 2: Related documents: output.

2 Review of users requirements

Overview of the users requirements in terms of products characteristics and dates of deliveries. The requirements presented in Tab. 3 are fully presented in the Service Portfolio Document (CSP-0350-RP-0005-ServicePortfolioWP8210).

3 Methodology description

3.1 Overview

The general framework of generating the CSP precipitation product consists first of collecting both input data the GPCP-1DD multi-satellite estimates of precipitation and the global daily rain gauge data for the period 1997 - 2003. Then the global daily rain gauge data are corrected for systematic measurement errors. This is done by a statistical correction model (Rubel and Hantel, 1999) with the main purpose to correct for wind-induced and evaporation and wetting losses. After this the corrected rain gauge data are interpolated to a regular $1°$ longitude/latitude grid by *ordinary block kriging* (Rubel, 1996a). This method considers both the inhomogenous distribution of the stations as well as the spatial structure of the precipitation process at the scale considered. Having the satellite estimates and the rain gauge measurements on the same grid both fields are merged due to *bivariate ordinary co-kriging* (Rubel, 1996b) to calculate the CSP precipitation product. Similar methods of this framework can be

Thematic content:	Precipitation fields
Input data sources:	Geostationary sensors IR data
	TOVS NOAA 12/14 data
	Synoptic rain gauge measurements
Methodology:	Bias-correction of the rain gauge measurements
	Kriging of rain gauges to regular grid
	Co-kriging of rain gauges and satellite product
Geographic resolution:	$1°$
Geographic projection:	LAT/LON plate carrée
Update frequency:	Daily
Delivery formats:	Raw binary, GRIB, ASCII
Data type:	Raster
Medium:	FTP
Dates of deliveries:	1st version for the year 2000 at T0+12
	2nd version for the year 2003 at T0+24
	3rd version for the period 1997 - 2003 at T0+30

Tab. 3: Review of user requirements.

found in Krajewski (1987) and Mitra *et al.* (2003). The improved quality of the satellite estimates is objectively measured in terms of verification scores (Rubel and Rudolf, 2001) over selected regions of the globe.

3.2 The retrieval algorithm

The statistical interpolation methods are linear or nonlinear, unbiased, least-squares spatial interpolation techniques. They are advanced applications of Gauss' Theory of Errors and therefore optimal in a statistical sense. Because of the assumptions of homogeneity and isotropy, which have to be made in the practical implementation, they do not necessarily yield optimal results. That is the reason why various assumptions have been proposed in the past, which lead to different, more or less expensive, interpolation methods. Here, for the objective analysis of the precipitation fields, the well known methods of ordinary block kriging (large scale analysis) and bivariate ordinary co-kriging (meso-scale analysis) are used.

3.2.1 Input data

Multi-satellite estimates: Fig. 1 shows an example of the GPCP-1DD fields (Huffman *et al.*, 2001) which are based on a combination of different satellite estimates. In the $40°$ N-S belt infrared (IR) estimates from geostationary satellites calibrated with Special Sensor Microwave/Imager (SSM/I) data are used. Outside this region the rain estimates are based on TIROS (Television Infrared Operational Satellite) Operational Vertical Sounder (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 globale 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. The GPCP-1DD precipitation estimates are calibrated with monthly precipitation data being near-realtime available SYNOP and CLIMAT reports. These SYNOP and
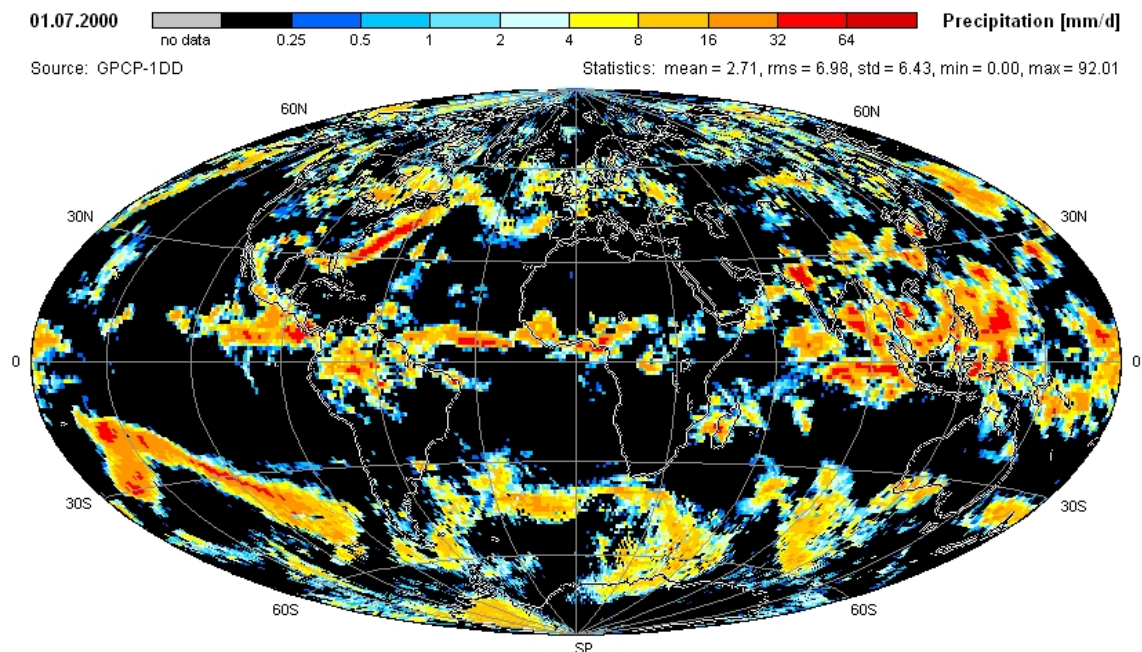
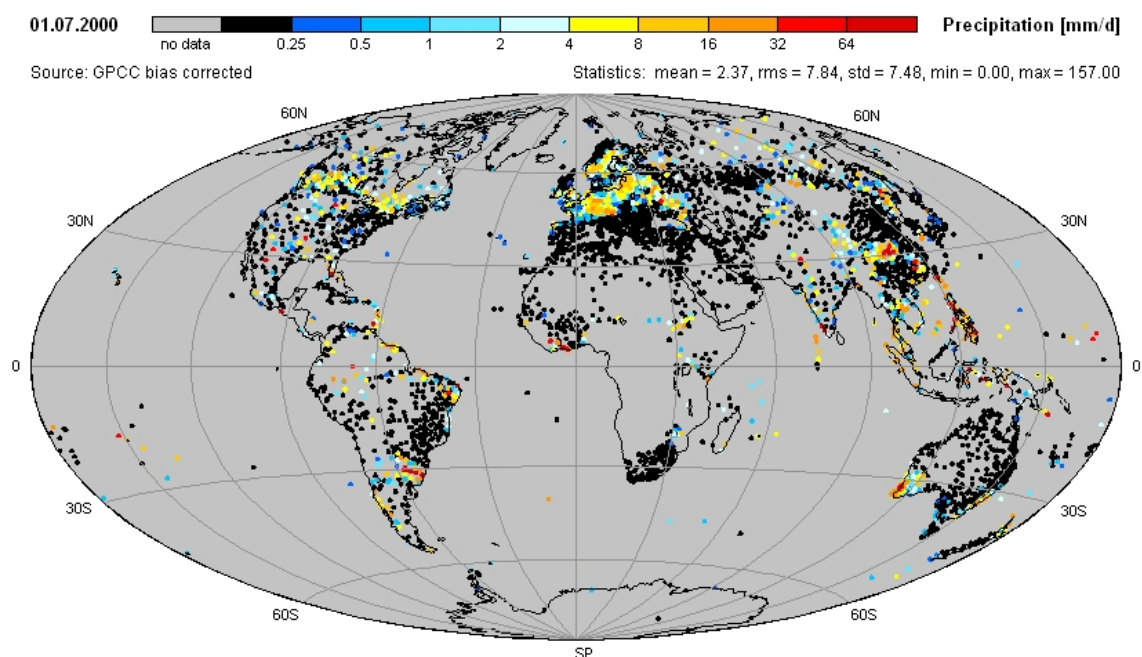**Fig. 1:** GPCP-1DD multi-satellite estimates of precipitation for July 01, 2000.

Fig. 2: Spatial distribution of the GPCC synoptic rain gauges with appropriate bias-corrected precipitation amounts for July 01, 2000.

CLIMAT reports are internationally disseminated via the Global Telecommunication System (GTS), processed by the Deutscher Wetterdienst (DWD) and evaluated by the GPCC (Rudolf *et al.*, 1992).

Rain gauge measurements: The second input data are those of the global synoptic stations available from the GTS network, provided by the GPCC. The spatial distribution of the synoptic stations is displayed in Fig. 2 as an example for July 01, 2000, and shows that the density is highest in Europe and South-Eastern Asia, while for large areas of Africa no observations are available. These observations are corrected for systematic measurement errors with a statistical correction model. Its main purpose is to correct for the wind-induced losses, which is the largest error. The correction formulas use observed wind speed and temperature as well as estimated rain intensity. For evaporation and wetting losses, which represent the second largest error of the precipitation measurements, climatological corrections are applied. Further the correction take instrument-specific properties into account; these comprise differentiation between unshielded (e.g. HELLMANN) and shielded (e.g. TRETJAKOV) gauges.

3.2.2 Kriging of rain gauge measurements

Assuming n contemporary gauge observations of the precipitation process $Z_g(\mathbf{u}_i)$, $\mathbf{u} \equiv (x, y)$, $i = 1, ..., n$, as suggested in Fig. 3, the areal averaged precipitation $\widehat{Z}(\mathbf{u}_0)$ can be estimated from the linear combination

$$\widehat{Z}(\mathbf{u}_0) = \sum_{i=1}^{n} \lambda_{gi} Z_g(\mathbf{u}_i), \tag{1}$$

combining the n neighbouring sample points of precipitation linearly with appropriate weights λ_g.
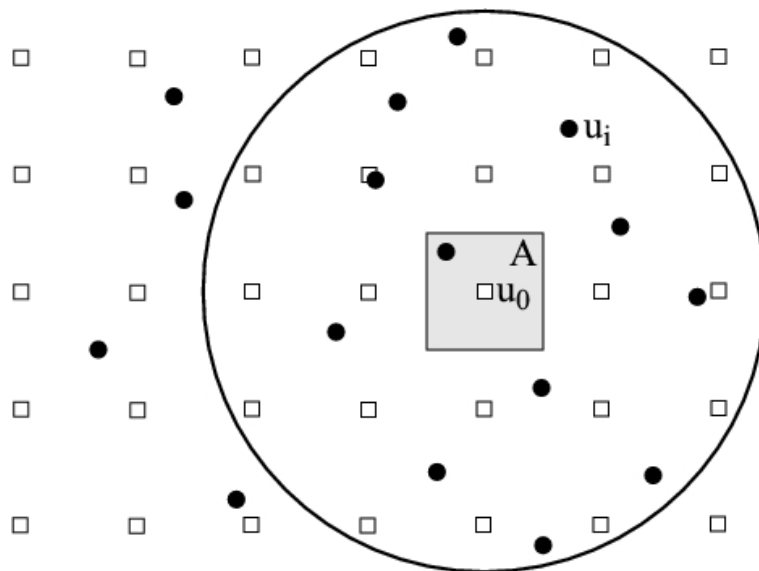

Fig. 3: For each area A, represented by a grid point $\mathbf{u}_0$, the n closest observations $Z_g(\mathbf{u}_i)$ within the decorrelation distance are used to estimate the value of the area averaged precipitation $\widehat{Z}(\mathbf{u}_0)$.

These weights λ_g can be obtained under the condition of minimising the estimation variance (mean square interpolation error)

$$\sigma_{\mathrm{E}}^2 = \mathrm{var}[Z(\mathbf{u}_0) - \widehat{Z}(\mathbf{u}_0)] = \mathrm{E}[(Z(\mathbf{u}_0) - \widehat{Z}(\mathbf{u}_0))^2] \tag{2}$$

from the following system of linear equations

$$\begin{vmatrix} R(\mathbf{u}_1,\mathbf{u}_1) & \cdots & R(\mathbf{u}_1,\mathbf{u}_n) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ R(\mathbf{u}_n,\mathbf{u}_1) & \cdots & R(\mathbf{u}_n,\mathbf{u}_n) & 1 \\ 1 & \cdots & 1 & 0 \end{vmatrix} \cdot \begin{vmatrix} \lambda_{g1} \\ \vdots \\ \lambda_{gn} \\ \mu_g \end{vmatrix} = \begin{vmatrix} R(\mathbf{u}_1,\mathbf{u}_0) \\ \vdots \\ R(\mathbf{u}_n,\mathbf{u}_0) \\ 1 \end{vmatrix}. \tag{3}$$

$R(\mathbf{u}_i,\mathbf{u}_j)$ are the correlations between rain gauge observations at the locations $\mathbf{u}_i$ and $\mathbf{u}_j$, $R(\mathbf{u}_i,\mathbf{u}_0)$ are the correlations between the rain gauge observations at the locations $\mathbf{u}_i$ and the grid area integrated true precipitation process $Z(\mathbf{u}_0)$ and μ_g is the scalar Lagrangian multiplier. The correlation functions $R(\mathbf{u}_i,\mathbf{u}_j)$ are defined as covariance functions normalized with the variance $\sigma^2 = \mathrm{cov}[\mathbf{u}_i,\mathbf{u}_i]$,

$$R(\mathbf{u}_i,\mathbf{u}_j) = \frac{\mathrm{cov}[\mathbf{u}_i,\mathbf{u}_j]}{\sigma^2} = \frac{\mathrm{E}[Z(\mathbf{u}_i) \cdot Z(\mathbf{u}_j)] - m^2}{\sigma^2}. \tag{4}$$

For further considerations the system (3) can be truncated by using the gauge-gauge correlation matrix $\mathbf{R}_{gg}$, the gauge-area correlation vector $\mathbf{R}_{g0}$ and the vector of weights to be estimated $\boldsymbol{\lambda}_g$,

$$\begin{vmatrix} \mathbf{R}_{gg} & \mathbf{1} \\ \mathbf{1}^\top & 0 \end{vmatrix} \cdot \begin{vmatrix} \boldsymbol{\lambda}_g \\ \mu_g \end{vmatrix} = \begin{vmatrix} \mathbf{R}_{g0} \\ 1 \end{vmatrix}. \tag{5}$$

Thus, if the correlations are known, the vector $\boldsymbol{\lambda}_g$ can be calculated from (5) and the linear combination (1) for the area averaged precipitation can be evaluated.

The corresponding estimation variance (kriging variance) can be written as

$$\sigma_{\mathrm{E}}^2 = R(\mathbf{u}_0,\mathbf{u}_0) - \sum_{i=1}^{n} \lambda_{gi} R(\mathbf{u}_i,\mathbf{u}_0) + \mu_g. \tag{6}$$

The estimation variance is given seperately for each gridpoint and is a measure of the quality of the analysis. It depends on the size of the analysed area as well as on the station density and the statistical structure of the precipitation field. Therefore, the estimation variance defined for the grid area A - that is the normalized mean square interpolation error - decreases with increasing station density and increasing decorrelation distance.

3.2.3 Co-kriging of rain gauge and satellite data

The linear combination for cokriging of satellite estimates $Z_s(\mathbf{u})$ and rain gauge data $Z_g(\mathbf{u})$ as represented in Fig. 4 can be written as

$$\widehat{Z}(\mathbf{u}_0) = \sum_{j=1}^{n_s} \lambda_{sj} Z_s(\mathbf{u}_j) + \sum_{i=1}^{n_g} \lambda_{gi} Z_g(\mathbf{u}_i), \tag{7}$$

combining the n_s neighbouring satellite estimates linearly with appropriate weights $\boldsymbol{\lambda}_s$ and doing the same for the n_g neighbouring gauge analyses. The corresponding weight vectors $\boldsymbol{\lambda}_s$ and $\boldsymbol{\lambda}_g$, that need
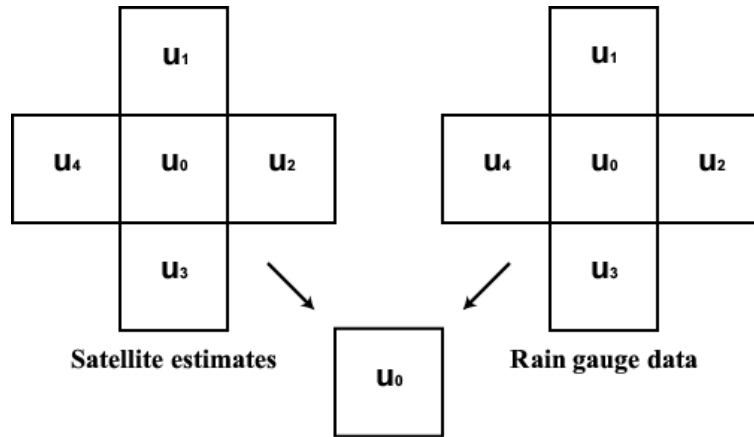
Fig. 4: Schematic representation of the configuration of the data included in the cokriging system.

to be estimated, can be computed from the following system of linear equations

$$
\begin{vmatrix}
\mathbf{R}_{ss} & \mathbf{R}_{sg} & \mathbf{1} & \mathbf{0} \\
\mathbf{R}_{gs} & \mathbf{R}_{gg} & \mathbf{0} & \mathbf{1} \\
\mathbf{1}^\top & \mathbf{0}^\top & 0 & 0 \\
\mathbf{0}^\top & \mathbf{1}^\top & 0 & 0
\end{vmatrix}
\cdot
\begin{vmatrix}
\boldsymbol{\lambda}_s \\
\boldsymbol{\lambda}_g \\
\mu_s \\
\mu_g
\end{vmatrix}
=
\begin{vmatrix}
\mathbf{R}_{s0} \\
\mathbf{R}_{g0} \\
0 \\
1
\end{vmatrix}
\tag{8}
$$

with the satellite-satellite correlation matrix $\mathbf{R}_{ss}$, the gauge-gauge correlation matrix $\mathbf{R}_{gg}$ and the satellite-gauge crosscorrelation matrix $\mathbf{R}_{sg} = \mathbf{R}_{gs}^\top$. Since the true area integrated precipitation is unknown, the satellite-area correlation vector $\mathbf{R}_{s0}$ and the gauge-area correlation vector $\mathbf{R}_{g0}$ are approximated from $\mathbf{R}_{ss}$ and $\mathbf{R}_{gg}$.

Analogous to the ordinary block kriging equation (5) the weight vectors $\boldsymbol{\lambda}_s$ and $\boldsymbol{\lambda}_g$ can be obtained from the cokriging equation (8) if the correlations $\mathbf{R}_{ss}$, $\mathbf{R}_{gg}$ and $\mathbf{R}_{sg}$ are known.

Once the coefficients λ_s and λ_g are determined, one can calculate the estimation variance (cokriging variance) as

$$
\sigma_{\mathrm{E}}^2 \;=\; R_{gg}(\mathbf{u}_0, \mathbf{u}_0) - \sum_{j=1}^{n_s} \lambda_{sj}\, R_{s0}(\mathbf{u}_j, \mathbf{u}_0) - \sum_{i=1}^{n_g} \lambda_{gi}\, R_{g0}(\mathbf{u}_i, \mathbf{u}_0) + \mu_g.
\tag{9}
$$

In this expression R_{gg} was substituted for R_0.

3.2.4 Empirical estimation of correlation functions

The statistical structure of precipitation is usually described by space and time correlation functions. Because precipitation is a complex physical phenomenon, it is not possible to fully describe precipitation processes by just one statistical model. Ideally, precipitation events should be classified by type, e.g. stratiform or convective type precipitation. Each type of precipitation can then be assumed to be a part of an ensemble of homogeneous realizations with defined statistical properties. The large- and mesoscale precipitation, e.g. over Europe, contains stratiform and convective components of different intensities and compositions at the same time. Moreover, it is hardly possible to separate them. For this reason the analysis of the precipitation fields is performed by the use of only a single correlation function for the whole model domain.

We assume that there exist n contemporary observations of precipitation process $Z(\mathbf{u}_i)$, $\mathbf{u} = (x, y)$ and $i = 1, ..., n$. Further, if there exist multiple realizations $t = 1, ..., T$ of this process $Z(\mathbf{u}_i)$, then the correlation can be defined as

$$
R(\mathbf{u}_i, \mathbf{u}_j) = \frac{\sum_{t=1}^{T}[Z_t(\mathbf{u}_i) - m(\mathbf{u}_i)][Z_t(\mathbf{u}_j) - m(\mathbf{u}_j)]}{\sqrt{\sum_{t=1}^{T}[Z_t(\mathbf{u}_i) - m(\mathbf{u}_i)]^2 \sum_{t=1}^{T}[Z_t(\mathbf{u}_j) - m(\mathbf{u}_j)]^2}}, \tag{10}
$$

where $m(\mathbf{u}_i)$ is the mean of the precipitation process at the location $\mathbf{u}_i$.

Assuming that the precipitation process is homogenous and isotropic within the model domain, the correlations depend only on the interstation distance ρ_{ij}; they are independent of the geographical locations of the stations. The result of the empirical analysis of the correlation coefficients is a scatter plot of correlations. To identify the structure of precipitation a hypothesis about the theoretical model has to be made. For the precipitation over the global land covers the nonlinear model

$$
R(\rho_{ij}) = c_1 \exp(-c_2 \, \rho_{ij}^{c_3}) \tag{11}
$$

with 3 coefficients c_1, c_2 and c_3 is chosen. For the first version of the CSP precipitation product this model is applied constantly on the whole globe. For the second version a global climate classification (Fig. 6) will be used to regionalise this constant correlation model. Note that the exponential correlation function has the required property to be positive definite.

3.2.5 Scale dependence of correlation functions

For a simple characterization of the correlation function a typical distance, the so-called decorrelation distance, will be used. It is defined as the distance for which the correlation decreases to 1/e. From the curve of the correlation function one can infer the type of precipitation. A strong decrease of the correlation function corresponding to a short decorrelation distance is characteristic for convective type precipitation. A weak decrease corresponding to a wide decorrelation distance is typical for stratiform precipitation.

The dependence of the correlation functions on the scale is summarized in Fig 5 on the next page. If one considers the accumulation time of precipitation, which corresponds to a typical extension of the model domain, then typical decorrelation distances are 2 km for 1-minute, 10 km for 15-minute, 50 km for hourly, 200 km for 12-hourly and daily, 300 km for monthly and 500 km for annual accumulated values of precipitation. Creating a precipitation product on a daily scale typical decorrelation distances of about 200 km have to be considered.

3.2.6 The Köppen Geiger climate classification

A global gridded climate classification calculated by the use of Wladimir Köppen's and Rudolf Geiger's (Essenwanger, 2001) climate formula is applied to regionalise the used correlation model for the second version of the CSP precipitation product. The underlying data for this climate classification are recently available gridded data sets of monthly mean temperature and precipitation (Mitchell and Jones, 2005). This CRU TS 2.1 data set covers all global land areas excluding Antarctica and provides long-term time-series for the period 1901 to 2002 for a set of meterological parameters on a
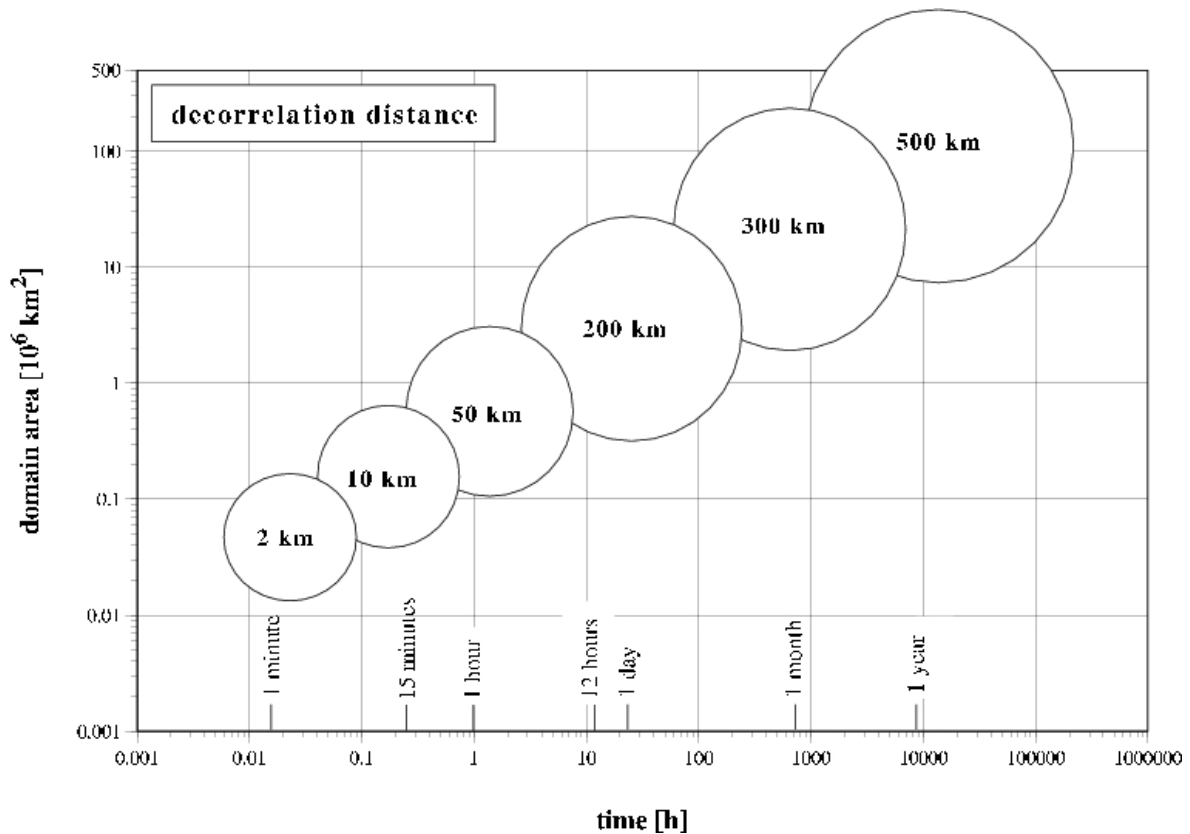
Fig. 5: Scale dependence of correlation function used for statistical analysis of precipitation, represented here by decorrelation distance for increasing spatial and temporal resolution.

global 0.5° lon/lat grid. This data-set is provided by the Climate Research Unit (CRU) and is freely available on the internet[5].

By the climate formula of Köppen and Geiger every climate is specified in principle with three letters. The first one identifies the main climate region; only uppercase letters are used:

```
A   ...   tropical humid climate without frost,
B   ...   dry climate,
C   ...   temperate mid-latitude climate,
D   ...   cold mid-latitude climate,
E   ...   polar climate.
```

The second one identifies the precipitation in consideration of annual amount and seasonal distribution; upper- and lowercase letters are used:

---

[5]http://www.cru.uea.ac.uk/

```
S  ...  steppe,
W  ...  desert,
w  ...  winterdry semihumid,
s  ...  summerdry semhumid,
m  ...  mixed; dry period compensated by high annual
          precipitation amount; monsoonal,
f  ...  fullhumid.
```

The third one identifies the temperature in a standard height of 2 m; only lowercase letters are used:

```
a  hot summer,
b  warm summer,
c  cool summer,
d  cool summer and extreme continental,
h  hot dry climate,
k  cold dry climate.
```
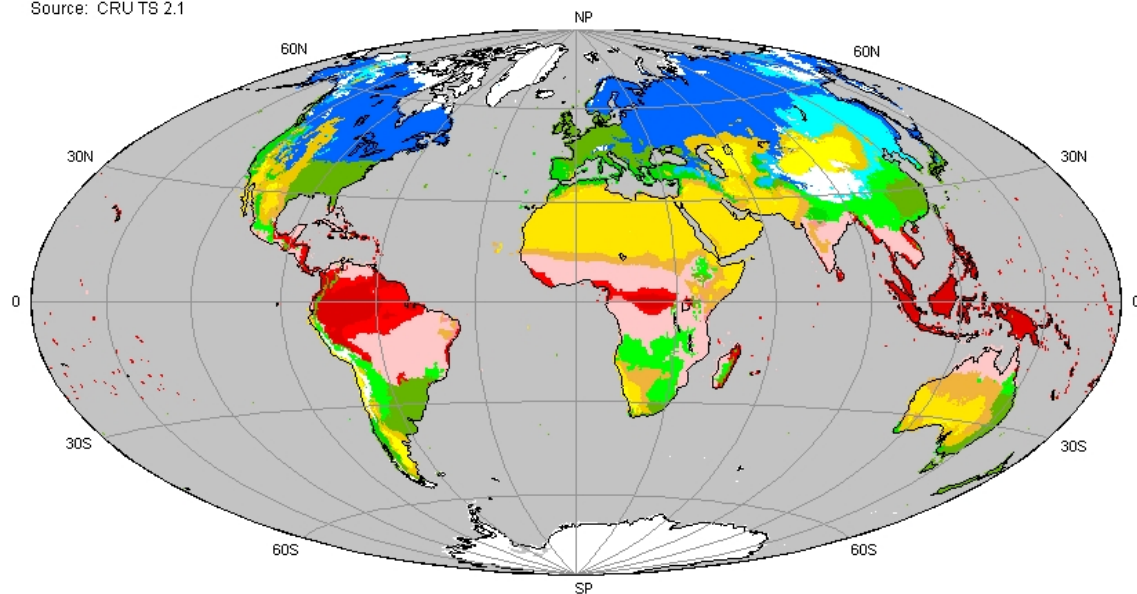

Fig. 6: Global Köppen climate classification as calculated by IMP using CRU TS 2.1 temperature and precipitation for the period 1951 to 2000. In opposite to historical climate classifications not available in digital form this product is well documented (publication in preparation).

The dry climates B are further subdivided into dessert W and steppe S and into cold k and hot h. The polar climate E is subdivided into EF and ET where F stands for frost (temperature of the warmest month $< 0°$ C) and T for tundra. The Antarctic is manually set to polar frost climate EF without having any temperature data for this region. For all other main climates the second letter marks the precipitation conditions, where f means wet all year round, s and w means summer or winter dry, respectively. For the tropical climate A a medium state m is defined where a dry season occurs but it is compensated by the precipitation of the following months. For the temporal and cold

climates C and D Köppen and Geiger introduced the third letter in order to further specify summer or winter temperature conditions with hot summer a, warm summer b, cool summer c and strong winter d. Note that a third letter d can not occur by definition in a temperate climate C.
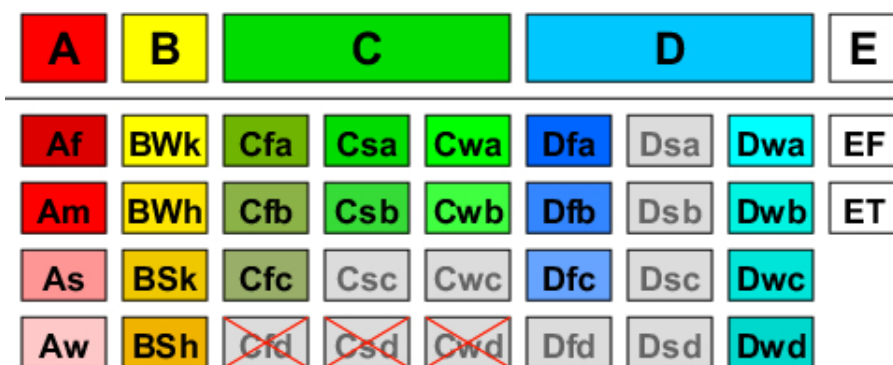

Fig. 7: Schematic illustration of all possible 31 different climate classes defined by Köppen and Geiger. Red cancelled climates cannot occur by definition, grey climates are rare climates and will not occur in a large areal amount.

As depicted in Fig. 7 this leads all together to 31 different climate classes, whereas Fig. 6 on the preceding page shows only 15 of them. In this global climate map the two subtypes of the polar climate E are groupd together as well as the third letter temperature conditions for the temporal and cold climates C and D.

3.2.7 Regionalisation of the correlation model

The second version of the CSP precipitation product will be calculated by the use of the correlation model (11) regionalised according to the 5 main climatic zones of the here presented climate classification. For the polar climate E too few information is available to achieve reliable results, so we get the following correlation functions within the 4 main climates A to D (Fig. 8 on the next page), calculated from the GPCC synoptic stations over the period 1997 to 2003 on a monthly base:

3.3 The validation procedure

About 21 000 stations with rain gauge measurements over Europe, collected for the ELDAS project by the Biometeorology Group at IMP, will be used to verify the global precipitation fields. The preliminary distribution of the stations is documented by Rubel (2004); additional stations have been collected during GEOLAND. First of all areal precipitation amounts using the grid of the GPCP-1DD products will be calculated from these rain gauge observations. For that an ordinary block-kriging method will be applied. This method considers both the inhomogeneous distribution of the stations as well as the spatial structure of the precipitation process at the scale considered. Additionally to each areal precipitation estimate the normalized kriging variance is known. Note that this kind of interpolation error depends on the station density, the spatial auto-correlation function and the size of the grid boxes. The verification of the global precipitation products for GEOLAND will be performed by using only grid boxes with normalized kriging variances below 0.15. Having precipitation fields from gauges and satellites comparable on the same grid, a first visual comparison will be done by viewing the fields. To quantify the verification results continuous and categorical statistics will be
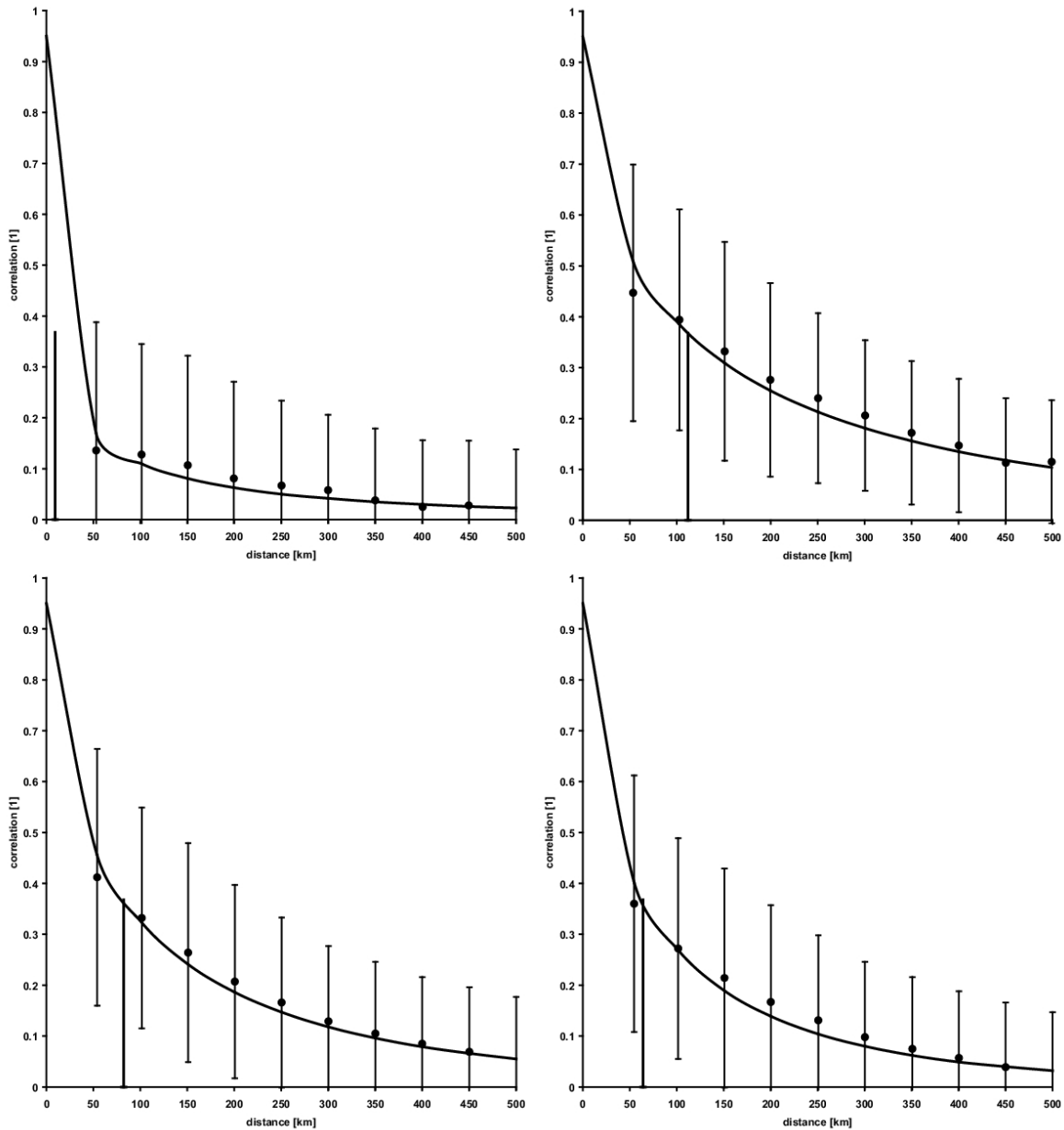
Fig. 8: Mean correlations and standard deviations with fitted autocorrelation function of type $R(\rho_{ij}) = c_1 \exp(-c_2 \, \rho_{ij}^{c3})$ and corresponding decorrelation distance (vertical line) for tropical climate A (upper left), dry climate B (upper right), temperate climate C (lower left) and cold climate D (lower right). Correlations calculated from GPCC synoptic stations over the period 1997 to 2003 on monthly base.

used. These measures comprises mean error (ME), mean absolute error (MAE), root mean square error (RMSE) and correlation coefficient (R). Here, because of the non-Gaussian probability density function of daily precipitation values, a nonparametric correlation coefficient, the Spearman rank-order correlation coefficient (R_s), is applied. Additionally a t-statistic that tests the significance of a non-zero R_s will be implemented. Categorical scores will supplement the continuous measures. The categorical statistics are based on two dimensional contingency tables (Fig. 9); one distinguishes

between accuracacy and skill scores.

Observed
Ground Truth

Fig. 9: Contingency table for the calculation of accuracy and skill scores.

3.3.1 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. 9) as

$$\text{ACC} \quad = \quad \frac{\text{correct estimates}}{\text{total estimates}} = \frac{z+h}{n}. \tag{12}$$

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,

$$\text{POD} \quad = \quad \frac{\text{correct rain estimates}}{\text{rain observations}} = \frac{h}{h+m}, \tag{13}$$

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

$$\text{FAR} \quad = \quad \frac{\text{false alarms}}{\text{rain estimates}} = \frac{f}{f+h}. \tag{14}$$

The best FAR value is zero and the worst is one.

3.3.2 Skill scores

With skill scores the improvements of the GPCP-1DD estimates over some reference estimates such as random chance, persistence or climatology can be measured. In general a skill score (SS) is defined as

$$\text{SS} \quad = \quad \frac{\text{ACC}_\text{estimate} - \text{ACC}_\text{reference}}{\text{ACC}_\text{perfect} - \text{ACC}_\text{reference}}. \tag{15}$$

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, which references for correct forecasts that would be made due to random chance. The TSS is computed from the contingency table (Fig. 9) as

$$\text{TSS} \quad = \quad \frac{\text{hz} - \text{fm}}{(\text{h} + \text{m})(\text{f} + \text{z})}, \tag{16}$$

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

$$\text{TSS} \quad = \quad \frac{\text{h}}{\text{h} + \text{m}} - \frac{\text{f}}{\text{f} + \text{z}} = \text{POD} - \text{POFD}. \tag{17}$$

The TSS reaches from minus one to plus one and is the perfect verification measure because it does not depend on the fraction of rain/no rain events as other scores do. Some of these other frequently used skill scores are the critical success index (CSI), the equitable threat score (ETS) or the Heidke skill score (HSS).

3.4 The product quality

The first version of the combined precipitation product (CSP v1.0, Fig. 10) has been validated for the year 2000. For verification purposes non-synoptic dense precipitation measurements based on about 21 000 station over the ELDAS domain were used.

As the accuracy of the existing multi-satellite estimates of precipitation is well known (in terms of objective verification scores), a quality improvement of about 10 % can be shown (Tab. 4).

3.5 Risk of failure

Up to the present there are identified no causes to fail the product or the methodology within the development cycle of the CSP precipitation product, except the input data will be no more available. In this case there does not exist any alternative option to generate the CSP precipitation product.

3.6 Links to other geoland activities

The here described methodology is similar to that one of EARS[6] applied to METEOSAT data. The product provided by IMP is on a global scale using 10 times more input data. An inter-comparison of the product will be realized to improve the reliability of each product. One methodology could be to compare rainfall products with a validation dataset generated by means of Jackknifing.

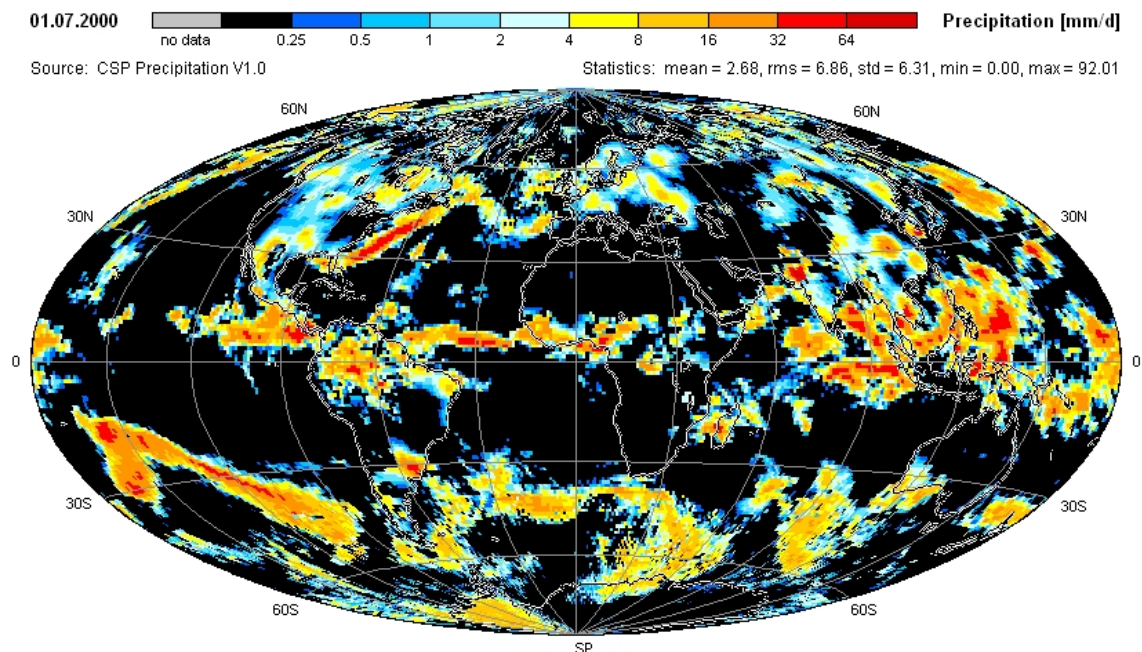[6]Environmental Analysis and Remote Sensing, Delft, Netherlands

Fig. 10: CSP v1.0 precipitation product as a combination of multi-satellite estimates and bias-corrected rain gauge measurements for July 01, 2000.

Continuous statistics		Continuous statistics	
Observed	1.78	Observed	1.78
Estimated	1.94	Estimated	1.73
Mean error	0.16	Mean error	-0.05
Mean absolute error	1.91	Mean absolute error	1.43
RMS error	3.97	RMS error	3.13
Rank-order correlation	0.50	Rank-order correlation	0.64
Categorical statistics		Categorical statistics	
Hit rate, HR	0.66	Hit rate, HR	0.72
Critical success index, CSI	0.50	Critical success index, CSI	0.59
Probability of detection, POB	0.55	Probability of detection, POB	0.68
False alarm ratio, FAR	0.17	False alarm ratio, FAR	0.18
BIAS score, BIAS	0.66	BIAS score, BIAS	0.83
True skill statistics, TSS	0.38	True skill statistics, TSS	0.46

Tab. 4: Verification statistics of GPCP-1DD (left) and CSP v1.0 (right) for summer (JJA) 2000.

4 References

Essenwanger, O. M., 2001: *World Survey of Climatology*, Ch. General Climatology, 1C, p. 102. Elsevier, Amsterdam.

Huffman, G. J., R. F. Adler, M. M. Morrissey, D. T. Bolvine, S. Curtis, R. Royce, B. McGavock, and J. Susskind, 2001: Global precipitation at one-degree daily resolution from multi-satellite observations. *J. Hydrometeorol.*, **2**, 36–50.

Kottek, M., and M. Hantel, 2005: *Observed Global Climate. New Series on Landolt-Börnstein, Numerical Data and Functional Relationships*, Ch. Global Maps 1991 - 1995, p. 190. Springer, Berlin.

Krajewski, W. F., 1987: Cokriging Radar-Rainfall and Rain Gage Data. *J. Geophys. Res.*, **92**, 9571–9580.

Mitchell, T., and P. Jones, 2005: An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *Int. J. Climatol.*, **25**, 693–712.

Mitra, A. K., M. Das Gupta, S. V. Singh, and T. N. Krishnamurti, 2003: Daily Rainfall for the Indian Monsoon Region from Merged Satellite and Rain Gauge Values: Large-Scale Analysis from Real-Time Data. *J. Hydrometeorol.*, **4**, 769–781.

Rubel, F., 1996a: PIDCAP Quick Look Precipitation Atlas. *Österreichische Beiträge zu Meteorologie und Geophysik*, **15**, 98pp.

Rubel, F., 1996b: Scale dependent statistical precipitation analysis. In *Proceedings of the International Conference on Water Resources & Environmental Research: Towards the 21st Century, 29-31 Oct. 1996, Kyoto, Japan*, Vol. 1, pp. 317–324.

Rubel, F., 2004: A new European precipitation dataset for NWP model verification and data assimilation studies. In Cote, J., Ed., *Research Activities in Atmospheric and Oceanic Modelling*, pp. 11–12. WMO/TD No. 1220, Report No. 34, Section 2.

Rubel, F., and M. Hantel, 1999: Correction of daily rain gauge measurements in the Baltic Sea drainage basin. *Nord. Hydrol.*, **30**, 191–208.

Rubel, F., and M. Kottek, 2005: *Observed Global Climate. New Series on Landolt-Börnstein, Numerical Data and Functional Relationships*, Ch. Data Management, p. 19. Springer, Berlin.

Rubel, F., and B. Rudolf, 2001: Global daily precipitation estimates proved over the European Alps. *Meteorol. Z.*, **10**, 403–414.

Rudolf, B., H. Hauschild, M. Reiss, and U. Schneider, 1992: Contributions to the Global Precipitation Climatology Centre. *Meteorol. Z.*, **1**, 7–84. In German.