

Error Analysis of TMI Rainfall Estimates over Ocean for Variational Data Assimilation

P. Bauer, J.-F. Mahfouf, W. S. Olson¹,
F. S. Marzano², S. Di Michele³,
A. Tassa³ and A. Mugnai³

Research Department

¹Joint Center for Earth Systems Technology, Greenbelt MD, USA

²Dipartimento di Ingegneria Elettrica, Università dell'Aquila,
L'Aquila, Italy

³Istituto di Fisica dell'Atmosfera, Consiglio Nazionale delle
Ricerche, Rome, Italy

Accepted for publication in Q. J. Roy. Met. Soc.

April 2002

This paper has not been published and should be regarded as an Internal Report from ECMWF.

Permission to quote from it should be obtained from the ECMWF.



European Centre for Medium-Range Weather Forecasts
Europäisches Zentrum für mittelfristige Wettervorhersage
Centre européen pour les prévisions météorologiques à moyen terme

For additional copies please contact

The Library
ECMWF
Shinfield Park
Reading
RG2 9AX
library@ecmwf.int

Series: ECMWF Technical Memoranda

A full list of ECMWF Publications can be found on our web site under:

<http://www.ecmwf.int/pressroom/publications/>

©Copyright 2002

European Centre for Medium Range Weather Forecasts
Shinfield Park, Reading, RG2 9AX, England

Literary and scientific copyrights belong to ECMWF and are reserved in all countries. This publication is not to be reprinted or translated in whole or in part without the written permission of the Director. Appropriate non-commercial use will normally be granted under the condition that reference is made to ECMWF.

The information within this publication is given in good faith and considered to be true, but ECMWF accepts no liability for error, omission and for loss or damage arising from its use.

Abstract

An intercomparison of retrieval errors from different TRMM passive microwave rainfall products was carried out to assess the definition of observation error for experiments of rainfall assimilation in a variational framework. Depending on algorithms and their spatial resolution and sampling, a large variety of error estimates occurred. The error propagation to the ECMWF model grid (here 45 km and 60 km) was investigated from error simulations and observed data with and without accounting for spatial error correlation.

All algorithms used in this study (TRMM standard product 2A12 V.5 and two alternative algorithms, namely PATER and BAMPR) employ a Bayesian retrieval framework. The Bayesian errors obtained from each algorithm from different case studies showed values well above 100% at low rain rates (0.1 mm/h) and around 50% at high rain rates (20–50 mm/h) at the original product resolution and sampling. These Bayesian errors corresponded very well with those from an independent evaluation which was carried out by comparing TMI estimates to PR retrievals at the same (here $\approx 27 \times 40 \text{ km}^2$) resolution.

The impact of spatial averaging on retrieval errors was simulated using fits to the Bayesian errors and realistic log-normal rainfall probability distributions. By neglecting spatial correlation the range of errors is reduced from 70–200% to 20–50% at low rain rates and from 25–70% to 5–20% at high rain rates. To account for spatial data correlation, TMI observations were averaged to the ECMWF model grid. First, the decorrelation of rain rates as a function of separation distance from all products was calculated. The introduction of spatial error correlation affected both error reduction and dispersion of errors per rain rate interval. The final error estimates ranged from 50–150% at low rain rates to 20–50% at high rain rates. The analysis suggests that once the spatial correlation pattern of a product is known, the probability density distribution of real observations inside the model grid does not produce larger scatter so that a simple scaling may suffice to calculate rainfall retrieval errors at the model resolution.

1 Introduction

Assimilation of observations into numerical weather prediction (NWP) models requires knowledge of both systematic and random errors in the observations. Systematic observation errors, once identified, are corrected through bias correction schemes assuming that the model output is unbiased. Random errors are included in both background and observation error covariance matrices and represent how much weight is given to model first-guess or observation in the assimilation procedure. Even though rainfall retrieval errors are known to be rather large, they may still provide useful information within an assimilation system since the model background errors will be of similar magnitude.

Following rainfall assimilation experiments in a three-dimensional assimilation system (Treadon 1997), the first attempt of exploiting the impact of the assimilation of rainfall data obtained from satellite measurements on analysis and forecast in an operational four-dimensional variational assimilation system (e.g. Rabier et al. 2000) was conducted for the Euro-TRMM project (e.g. Marécal and Mahfouf, 2000; Marécal and Mahfouf, 2002) using observations from the Tropical Rainfall Measuring Mission (TRMM; Kummerow et al. 1998). Part of this effort was the investigation of model sensitivity to data product (i.e. retrieval algorithm) and to associated errors (Marécal et al. 2002).

For this purpose, two new algorithms have been developed: (1) Precipitation Radar (PR) Adjusted TRMM Microwave Radiometer (TMI) Estimation of Rainfall (PATER; Bauer et al. 2001) and (2) Bayesian Algorithm for Microwave-based Precipitation Retrieval (BAMPR; Mugnai et al. 2001). Both methods allow the inclusion of PR data for constraining the primary TMI retrievals but only TMI retrievals will be analyzed here. These algorithms were compared to the TRMM standard product 2A12 that contains hydrometeor profiles retrieved from TMI data employing the algorithm proposed by Kummerow et al. (1996). Depending on the algorithm, the errors were defined differently as a product of both the employed inversion procedure and the training data sets. Algorithms and errors are compiled in Sections 2 and 3.

This paper summarizes the error definition for these algorithms and how they may be used within a variational assimilation system. Apart from their quantitative differences, these errors depend on (1) the data used for training the algorithm and (2) the geometrical layout; i.e., spatial resolution and sampling of the final product. The first issue includes errors introduced by radiative transfer simulations based on hydrometeor profiles which do not well represent the observed ones, errors in the assumed particle size distributions, their composition and shape as well as insufficient surface emissivity modeling, errors associated with the numerical solution to the radiative transfer problem in scattering atmospheres, and radiometer noise. The order in this list roughly corresponds to the magnitude of the contribution from each element to the total error. General information on all three algorithms is summarized in Table 1.

Table 1: Algorithm Specifications

Algorithm	Reference	Resolution	Sampling cross-track / along-track	Sensors
PATER	Bauer et al. (2001)	27 x 44 km ²	10 km / 14 km	TMI (PR)
BAMPR	Mugnai et al. (2000)	10 x 16 km ²	10 km / 14 km	TMI (PR)
2A12	Kummerow et al. (1996)	10 x 16 km ²	5 km / 14 km	TMI

The usage of retrieval products of this kind in global models requires a scaling to the model grid resolution. Our work presents a procedure to unify the errors at the spatial resolution of two versions of the ECMWF model, that is T_L319 and T_L511 which correspond to ≈ 60 km and ≈ 45 km grid box sizes. This scaling is carried out in a general way neglecting spatial error correlation and by inclusion of spatial error correlation using both idealized rainfall probability distributions and real data. These two approaches are described in Section 4.

Another motivation for our study was the stimulation of further investigations on retrieval error estimation in the framework of satellite rainfall data assimilation in preparation for its operational use at NWP centers. With the increasing computational efficiency of global models, the assimilation of new remote sensing data becomes feasible. Please note, that this paper does not address the accuracy of the forecast model. Secondly, all problems related to temporal sampling errors; i.e., the error resulting from limited satellite data coverage per grid box over the assimilation window is not addressed.

2 Bayesian Retrieval

The Bayesian retrieval framework (e.g. Lorenc 1986) is the basis for all algorithms used in this study so that it is only briefly summarized at this point. It makes use of the probabilities, P , of observations, \mathbf{y}^o , (here in terms of brightness temperature vectors, \mathbf{TB}) and atmospheric states (in terms of hydrometeor profiles or their reduction to surface rain rate, \mathbf{x}) due to the inherent estimate uncertainty that \mathbf{x} always produces \mathbf{y}^o . $P(\mathbf{x})$ is the a-priori probability of \mathbf{x} while $P(\mathbf{y}^o)$ is the probability that observation \mathbf{y}^o occurs. $P(\mathbf{y}^o|\mathbf{x})$ is the conditional probability that \mathbf{y}^o can be measured for atmospheric state \mathbf{x} whereas $P(\mathbf{x}|\mathbf{y}^o)$ is the conditional probability that atmospheric state \mathbf{x} is present once \mathbf{y}^o is observed.

$P(\mathbf{x}, \mathbf{y}^o)$ describes the common probability that atmospheric state \mathbf{x} is present and observation \mathbf{y}^o is made. The joint probability is:

$$P(\mathbf{x}, \mathbf{y}^o) = P(\mathbf{x}|\mathbf{y}^o)P(\mathbf{y}^o) = P(\mathbf{y}^o|\mathbf{x})P(\mathbf{x}) \quad (1)$$

which leads to the Bayes rule for the a–posteriori probability:

$$P(\mathbf{x}|\mathbf{y}^o) = \frac{P(\mathbf{y}^o|\mathbf{x})P(\mathbf{x})}{P(\mathbf{y}^o)} \quad (2)$$

In the case of rainfall retrievals, $P(\mathbf{x})$ originates from the a–priori knowledge on cloud variability and is usually taken from cloud resolving model simulations.

$P(\mathbf{y}^o|\mathbf{x})$ may be transformed to $P[\mathbf{y}^o - \mathbf{y}(\mathbf{x})]$ which is the probability of the deviations of the observations, \mathbf{y}^o , from synthetic observations, $\mathbf{y}(\mathbf{x})$, obtained from state \mathbf{x} using a radiative transfer model. This departure is sensitive to both observation and radiative transfer modeling errors. Assuming that $P[\mathbf{y}^o - \mathbf{y}(\mathbf{x})]$ is a multi–dimensional Gaussian function and that observation and radiative transfer errors are uncorrelated (Lorenz 1986):

$$P[\mathbf{y}^o - \mathbf{y}(\mathbf{x})] \propto \exp \left\{ -\frac{1}{2} [\mathbf{y}^o - \mathbf{y}(\mathbf{x})]^T [\mathbf{E} + \mathbf{F}]^{-1} [\mathbf{y}^o - \mathbf{y}(\mathbf{x})] \right\} \quad (3)$$

where $(\mathbf{E} + \mathbf{F})$ are the covariance matrices of observation (i.e. instrument noise) and modeling errors.

The best estimate of atmospheric state (i.e. the analysis, \mathbf{x}_a) may be produced from the expected value which is identical to the minimum variance solution in this context:

$$\mathbf{x}_a = \int \mathbf{x} P(\mathbf{x}|\mathbf{y}^o) d\mathbf{x} \propto \int P[\mathbf{y}^o - \mathbf{y}(\mathbf{x})] P(\mathbf{x}) \mathbf{x} d\mathbf{x} \quad (4)$$

Alternatively, a maximum a–posteriori probability approach can be chosen where $\mathbf{x}_a = \text{Max} P[\mathbf{y}^o - \mathbf{y}(\mathbf{x})] P(\mathbf{x})$ (e.g. Marzano et al. 1999).

Again, for well defined retrieval problems, the deviation from the background state has to be accounted for in (4). For this, iterative optimization schemes towards the smallest deviations in observable and state space are required. In case of precipitation retrievals, the optimization is likely to fail due to the large number of unknowns. Therefore, $P(\mathbf{x})$ enters the integration through the number of profiles contained in an off–line database. This database is usually obtained from mesoscale cloud model simulations because these provide consistent three–dimensional hydrometeor distributions (Olson et al. 1996). Of course, the lack of representativeness of these simulations may cause serious problems in the evaluation of (4) (Bauer 2001). Under the assumption of realistic $P(\mathbf{x})$, a way to numerically calculate (4) is given by (Olson et al. 1996):

$$\mathbf{x}_a = \frac{\sum_i J^o[\mathbf{y}^o, \mathbf{y}(\mathbf{x}_i)] \mathbf{x}_i}{\sum_i J^o[\mathbf{y}^o, \mathbf{y}(\mathbf{x}_i)]} \quad (5)$$

with cost–function J^o :

$$J^o[\mathbf{y}^o, \mathbf{y}(\mathbf{x}_i)] = \exp \left\{ -\frac{1}{2} [\mathbf{y}^o - \mathbf{y}(\mathbf{x}_i)]^T [\mathbf{E} + \mathbf{F}]^{-1} [\mathbf{y}^o - \mathbf{y}(\mathbf{x}_i)] \right\} \quad (6)$$

and where \mathbf{x}_i represents an individual state profile in the database. To be exact, the modeling errors, \mathbf{F} , should be quantified for each \mathbf{x}_i ; however, this would require the knowledge of the dependence of radiative transfer errors on local hydrometeor profiles. The inherent uncertainty is given by the integration of analysis values from those contained in the database:

$$\sigma_a^2 = \frac{\sum_i J^o[\mathbf{y}^o, \mathbf{y}(\mathbf{x}_i)] (\mathbf{x}_i - \mathbf{x}_a)^2}{\sum_i J^o[\mathbf{y}^o, \mathbf{y}(\mathbf{x}_i)]} \quad (7)$$

It has to be noted that this error estimate is subject to the representativeness of the database and the assumptions made for observation and modeling errors.

3 Algorithms

3.1 PATER

The PATER algorithm employs a comprehensive set of cloud resolving model simulations where the final retrieval database only contains those hydrometeor profiles whose simulated brightness temperatures were found in a large set of TMI observations. This was carried out to ensure an improved representativeness in terms of the assumptions made between (4) and (5). Also, (5) was defined only using two empirical orthogonal functions (EOF's) without any constraints on the retrieval, and without using the 85.5 GHz channels. Important to notice is that this algorithm derives rainwater content (w in g/m^3) rather than rain rate (RR in mm/h); for conversion, the relation $RR = 20.95 \cdot w^{1.12}$ has been used. The error of this fit is well below 1% for $w \in [0, 2 gm^{-3}]$ and a Gamma drop size distribution with a shape parameter of one.

3.2 BAMPR

BAMPR is based on the retrieval scheme implemented by Marzano et al. (1999) of which the radiometer-only version was used here (Mugnai et al. 2001). As for the other algorithms, selected mesoscale simulations were combined for database construction and the Bayesian framework represents the retrieval part. The radiative transfer was applied to slanted hydrometeor columns at cloud model resolution along the TMI viewing angle. The simulated TB's were then convolved using TMI footprint dimensions. An additional feature of BAMPR is the adjustment of raindrop size distributions in case of insufficient matches between simulations and observations since these were identified as the main errors source in the BAMPR simulation databases by Panegrossi et al. (1998). The database was sub-divided into less and more intense rainfall regimes. Thus, the inversion employs a scene identification before applying the inversion methodology itself. A minimum variance scheme was adopted without limiting assumptions on error and a-priori probability distributions.

3.3 2A12

The TRMM standard TMI algorithm 2A12 is based on the Goddard Profiling Algorithm (GPROF) which was developed for application to aircraft data and later to satellite data from the Special Sensor Microwave / Imager (Kummerow et al. 1996, Olson et al. 1996). The referenced articles also describe the cloud model simulations contained in the database; i.e., represented by all $P(\mathbf{x})$. It employs the solution given in the previous chapter where \mathbf{y}^o contains \mathbf{TB} as well as emission/scattering indices. It also uses constraints driven by a stratiform/convective classification and distance-from-convective measures (Kummerow et al. 2001). The latter have been introduced between versions 4 and 5.

4 Errors

From PATER, two independent uncertainty estimates may be obtained. The first is the 'Bayesian' error from (7). Since PATER employs EOF's, the sum of errors in (5) is replaced: $(\mathbf{E} + \mathbf{F})^{-1}$ becomes $[\mathbf{E}_s(\mathbf{E} + \mathbf{F})\mathbf{E}_s^T]^{-1}$ where \mathbf{E}_s denotes the eigenvector matrix associated with the simulations. Radiative transfer errors are assumed to be [2, 4, 4, 6 K] at [10.65, 19.35, 21.3, 37.0 GHz]. The second error estimate originates from the intercomparison of PR and TMI estimates of near-surface rain liquid water contents at the same resolution (Bauer et al. 2001). It can be shown that in the range of 1–15 mm/h this error is mostly driven by the database ambiguity, that is the standard deviation of multiple solutions having the same TB-signature (Bauer 2001). For lower rain

rates, the background noise originating from surface, non-precipitating clouds, and atmosphere supersedes the rain signal while for rain rates above 15 mm/h the signature from precipitating ice reduces the contribution of rainwater to the signal.

The BAMPR error covariances; i.e. $(\mathbf{E} + \mathbf{F})$ in (7), were derived from sensitivity tests. For this purpose, reference parameters such as the slope of particle size distributions, the presence of melting particles, sea surface wind speed, approximations to radiative transfer modeling, the temperature profile, and field-of-view dimensions were randomly varied within a given range. The resulting TB's were compared to those from the reference configuration, and the statistics of these errors were computed (Tassa et al. 2002).

Since the 2A12 version 5 (V.5) products contained no error information, a rough estimate of 25% on the 60 km model grid (Kummerow 1999, personal communication) was used in the previously mentioned assimilation studies. In preparation for version 6, however, an explicit error estimate through usage of (7) was implemented with an updated version of 2A12 V.5 (called V5.1 hereafter while the official release V5.1 may still contain further updates).

Table 2: Cases for data in shown in Figure 4.

Date	Orbit	Center Lat./Lon.	Raintype
PATER:			
1997/12/16	00295	12 N / 144 E	Cyclone 'Paka', Pacific Ocean
1998/02/10	01171	15 S / 60 E	Cyclone 'Anacelle', Indian Ocean
1998/02/16	01273	5 S / 145 W	Deep convection, Pacific ITCZ
1998/08/26	04283	33 N / 75 W	Hurricane 'Bonnie', West Atlantic
1998/08/29	04328	23 S / 142 W	Frontal system, East Pacific
1999/01/21	06620	3 N / 81 E	Deep convection, Indian Ocean
2000/08/02	15438	6 N / 16 W	Squall line, West African coast
2000/08/25	15795	18 N / 70 E	Monsoon, West Indian coast
2000/08/28	15835	13 N / 134 W	Convection, Central Pacific
2000/08/30	15876	34 N / 72 W	Frontal system, North American East coast
2000/09/11	16059	35 N / 33 W	Extratropical depression, East Atlantic
2000/09/17	16151	3 N / 70 E	Monsoon scattered convection, Indian Ocean
BAMPR:			
1998/08/25	04267	30 N / 74 W	Hurricane 'Bonnie', West Atlantic
1999/08/19	04176	11 S / 179 W	Convection, South Pacific Convergence Zone
1999/12/24	11938	13 S / 71 E	Hurricane 'Astrid', West Atlantic
2A12:			
1997/12/19	00336	16 N / 136 E	Supertyphoon 'Paka'

Figure 4 shows examples of these errors collected from several cases which are listed in Table 4. Comparing the results in Figure 4, it is noted that despite the different approaches used for the error calculation and the different magnitudes of modeling errors assumed for $\mathbf{E} + \mathbf{F}$, the gross behavior of all algorithms is fairly similar. For moderate rainfall rates; i.e. between 1 and 20 mm/h, errors decrease from $\approx 100\%$ to $\approx 50\%$. This represents the range where the signal-to-noise ratio is best because surface effects are suppressed by the rather large cloud opacity while extreme events with large cloud ice contents occur predominantly in more intense systems. The errors also suggest that the variability of cloud-resolving model profiles used by all algorithms is fairly similar. Since the non-uniqueness of the TB-rainfall relationship is the major error source in this range, the contribution of modeling errors to (7) is rather small. Therefore, the accurate error estimation requires the analysis of both cloud model profile distributions represented in the databases (i.e. $P(\mathbf{x})$) and non-uniqueness of the TB-rainfall

relationship. Another observation is that the lower limit of rain detection can be set to ≈ 0.1 mm/h because the retrieval errors are ≈ 100 – 200% . Below this threshold the assumption of Gaussian error distributions will not hold anymore.

As another means of validation, average differences between TMI and PR retrievals at a unified resolution (27×40 km²) are shown in Figure 4 as lines superimposed on the scatterplot. The cases from which the data were computed are identical to those from the PATER Bayesian error calculation (see Table 4). Please note that the comparison to PR data covers both systematic and random errors; i.e., if the TMI estimates would be bias-corrected by those of the PR, the resulting errors in Figure 4 would be lower. The average differences between PATER and PR retrievals are at the upper limit of the Bayesian errors which suggests that the retrievals are not completely unbiased; however, the magnitudes are very similar. For 2A12 V5.1 the average difference to PR retrievals is almost identical to that of PATER. Since the 2A12 V5.1 Bayesian errors are a little larger than those of PATER, the curve follows well the average distribution of the data.

5 Spatial averaging

Before the rainfall retrievals can be assimilated into the model, spatial averaging has to be carried out to match the model resolution at which the comparison with model rain rates enters the cost-function [see Equation 1 in Marécal and Mahfouf (2002)]. For this, the number of estimates inside the model grid, N , is needed given that this is fully included in the swath and that all these estimates are successful retrievals (including 'no rain'). This number depends on the sampling rather than the spatial resolution of the products; however, the resolution is important regarding the spatial correlation of the retrievals in each grid box as will be discussed below since resolution should be inversely related to spatial correlation.

BAMPR and PATER data are sampled like the lower frequency TMI channels: 104 fields of view along scan with a separation distance of ≈ 10 km, while the distance between scan lines is roughly 14 km resulting from the velocity of the sub-satellite point and the TMI antenna rotation frequency. 2A12 data are interpolated to the gridpoints of the 85.5 GHz-channels. Thus, the along-scan separation distance reduces to ≈ 5 km while the along-track distance is identical to that of the other products. Compared to the model grid size used by Marécal and Mahfouf (2002); i.e., ≈ 60 km, $N_{2A12} \approx 50$, $N_{PATER, BAMPR} \approx 25$, while for the current operational model version (≈ 45 km) $N_{2A12} \approx 30$ and $N_{PATER, BAMPR} \approx 15$. Figure 5 illustrates both sampling and resolution for each product near scan-center and scan-edge, respectively.

If RR^t is the true rain rate, we can define the error of each retrieval with respect to RR^t at the i -th pixel as $RR_i - RR_i^t$. In our case, the random component of this error is characterized by a standard deviation, σ_i , which is a function of rain rate provided by an algorithm at the i -th pixel. As a variance, σ_i^2 , the error may either originate from the Bayesian estimator (7) or from the comparison to PR data. The standard deviation of the mean rain rate, \overline{RR} , for spatially correlated errors is:

$$\sigma_{\overline{RR}} = \frac{1}{N} \left[\sum_{i=1}^N \sum_{j=1}^N C(i, j) \sigma_i \sigma_j \right]^{1/2} \quad (8)$$

which reduces to:

$$\sigma_{\overline{RR}} = \frac{1}{N} \left[\sum_{i=1}^N \sigma_i^2 \right]^{1/2} \quad (9)$$

in case of spatially uncorrelated errors because $C(i, j) = \delta_{ij}$.

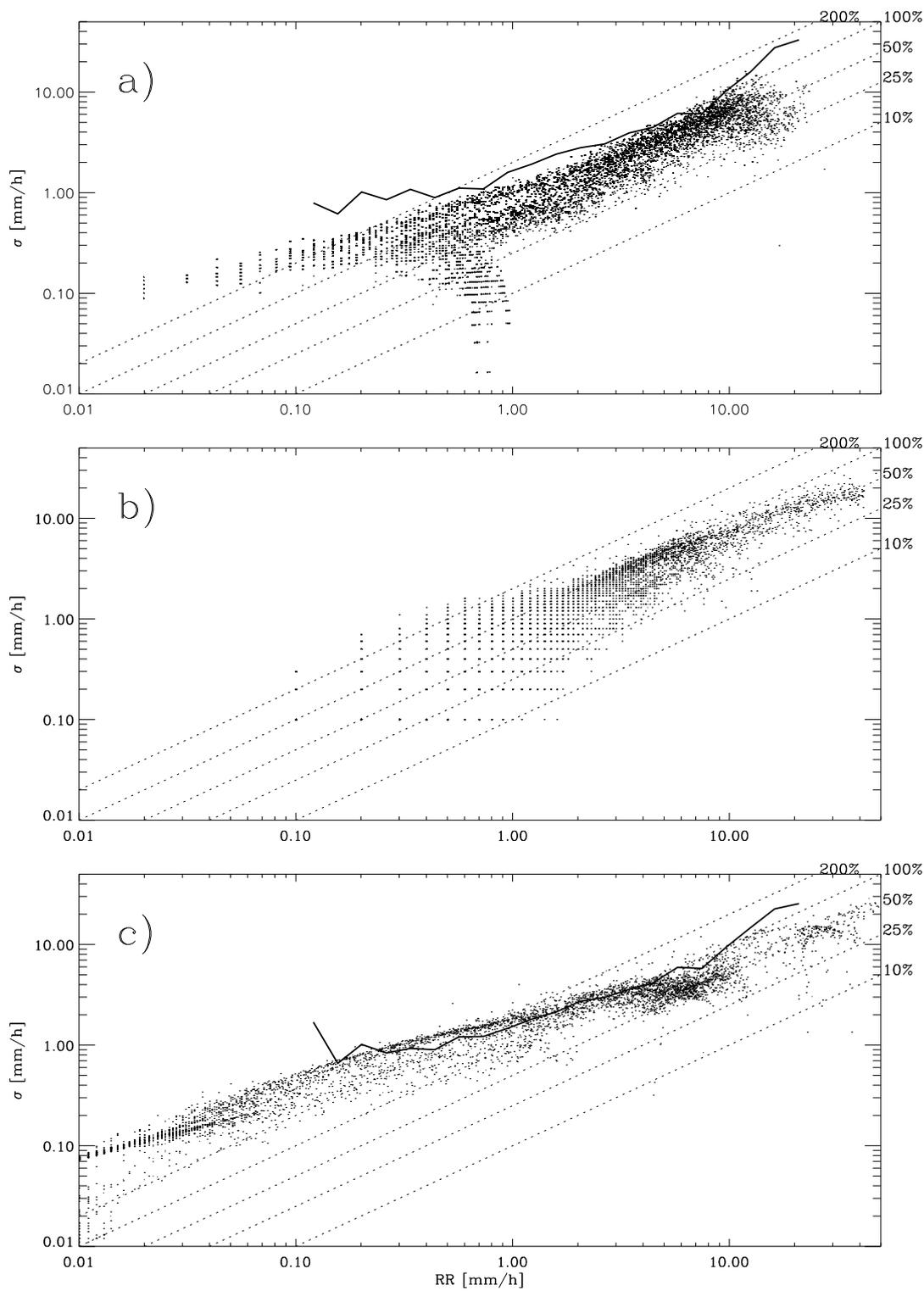


Figure 1: Bayesian retrieval errors from (7) as a function of rain rate at product resolution for PATER (a), BAMPR (b), and 2A12 V5.1 (c). Superimposed are lines denoting average differences between PATER and PR (a) and 2A12 V5.1 and PR (c) retrievals, respectively.

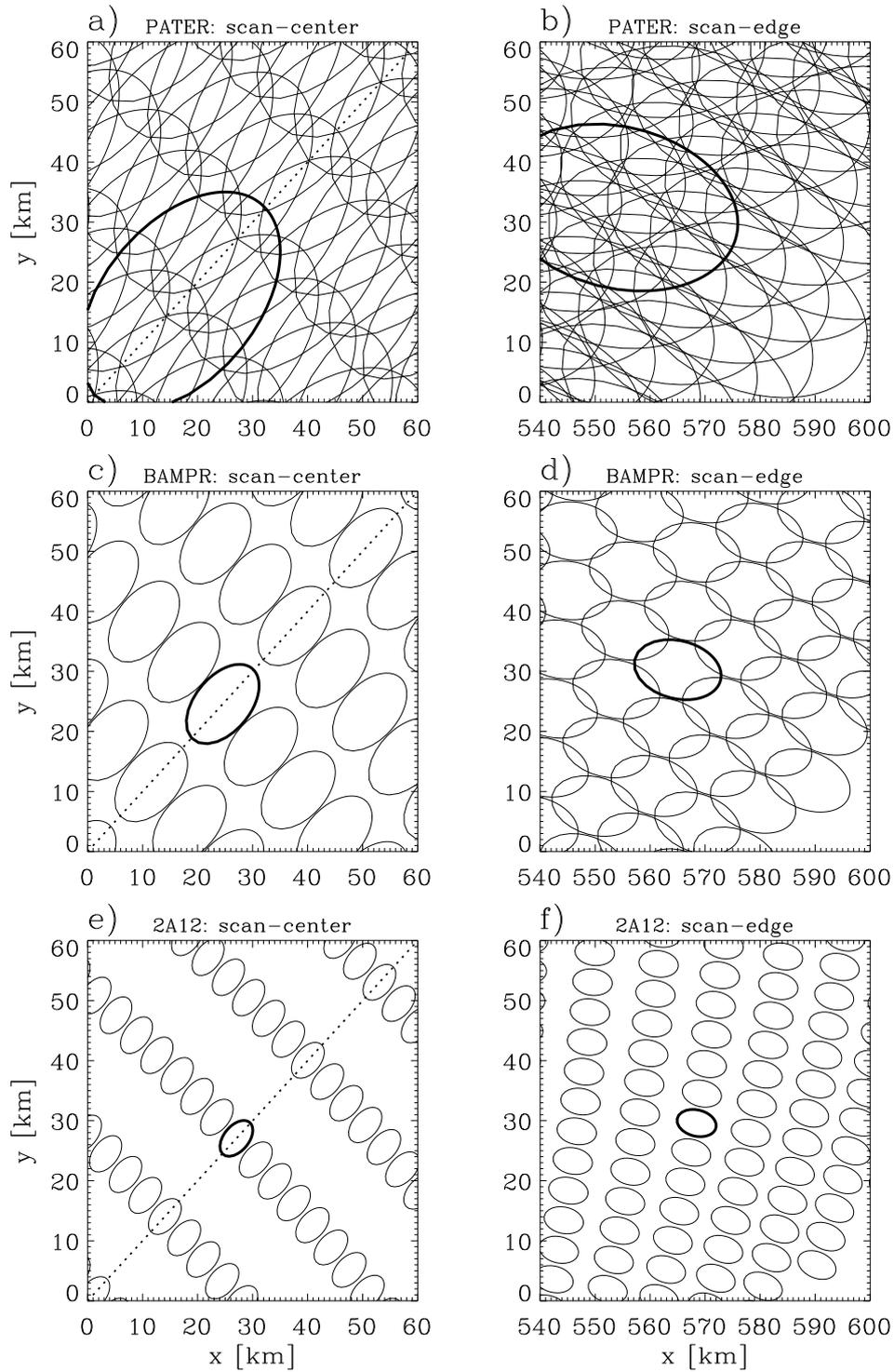


Figure 2: Example of sampling and spatial resolution of PATER near the TMI scan-center (a) and scan-edge (b). Same for BAMPR (c, d) and 2A12 (e, f).

5.1 Simulated effects

To investigate the gross effects of spatial averaging, a simulation of error scaling to two different model grids was carried out. It was assumed that the entire grid box is covered with rainfall observations. Inside the box a log-normal rainfall probability distribution (pdf) was assumed with:

$$P(x) = \frac{1}{\sigma_x \sqrt{2\pi}} \exp \left[-\frac{(x - \bar{x})^2}{2\sigma_x^2} \right], \quad \bar{x} = \int_{-\infty}^{\infty} xP(x)dx \quad (10)$$

where $x = \log(RR)$, and $\sigma_x^2 = \overline{[x - \bar{x}]^2}$ the associated variance. The standard deviation was assumed to be constant with $\sigma_x = \sqrt{1.21}$ following Kedem et al. (1990). The transfer to $\log(RR)$ was carried out to improve the integration with sparse discrete data. Thus, the scaling to model grid size is realized by an integration of rainfall dependent errors over the rainfall probability distribution with mean $\bar{x} = \sum_{i=1}^N x_i p(x_i) w_i$ where the w_i represent the weights of the Gaussian quadrature at discrete x_i with $\sum_{i=1}^N p(x_i) w_i \approx 1$. The number of x_i is obtained from spatial sampling vs. model grid box size and the summation limits are at $\bar{x} \pm \sigma_x$. As in (8), the variance of the mean rain rate is:

$$\sigma_{\bar{x}}^2 = \frac{\sum_{i=1}^N \sum_{j=1}^N P(x_i) P(x_j) w_i w_j \sigma_i \sigma_j}{\left(\sum_{i=1}^N P(x_i) w_i \right)^2} \quad (11)$$

which reduces to:

$$\sigma_{\bar{x}}^2 = \frac{\sum_{i=1}^N P(x_i)^2 w_i^2 \sigma_i^2}{\left(\sum_{i=1}^N P(x_i) w_i \right)^2} \quad (12)$$

for spatially uncorrelated errors. Here, σ_i is the error associated with local rain rate RR_i .

To facilitate the usage of errors, fits were derived to compute errors as a function of rain rate for all algorithms (see Figure 5.1a). For PATER; however, the standard difference between TMI and PR retrievals after calibrating the PATER estimates were taken (Bauer et al. 2001) because only calibrated retrievals were used in the assimilation experiments (Marécal et al. 2002):

$$\sigma_{PATER}[mm/h] = \max[23.464w^{0.12}\sigma_w, 0.2], \quad \sigma_w[g/m^3] = [0.1 + 0.45 \cdot \log_{10}(w)]w \quad (13)$$

Therefore these errors are smaller than those shown in Figure 4a. The fit to the BAMPR data produced:

$$\sigma_{BAMPR}[mm/h] = 0.7 RR \quad (14)$$

and for 2A12 V5.1:

$$\sigma_{2A12}[mm/h] = 1.357RR^{0.7} \quad (15)$$

was obtained. Figures 5.1b and 5.1c present the results from applying (12) to the error curves shown in Figure 5.1a. As expected, the errors are reduced for increasing model grid box size and the linearity of the BAMPR fit is conserved. The reduction of errors is significant and reaches 50% for 1 and 10 mm/h average rain rates on the 60 km grid. At low rain rates the errors remain at 50–100% which may point at rain detection error magnitudes in the vicinity of rainclouds.

5.2 Data

The evaluation of the error scaling with actual TMI retrievals was prepared by the calculation of the spatial correlations, C , of retrieved rain rates from each algorithm. It was assumed that these are identical to the spatial correlations of the errors. This assumption is justified in the situation where errors are proportional to rain rates.

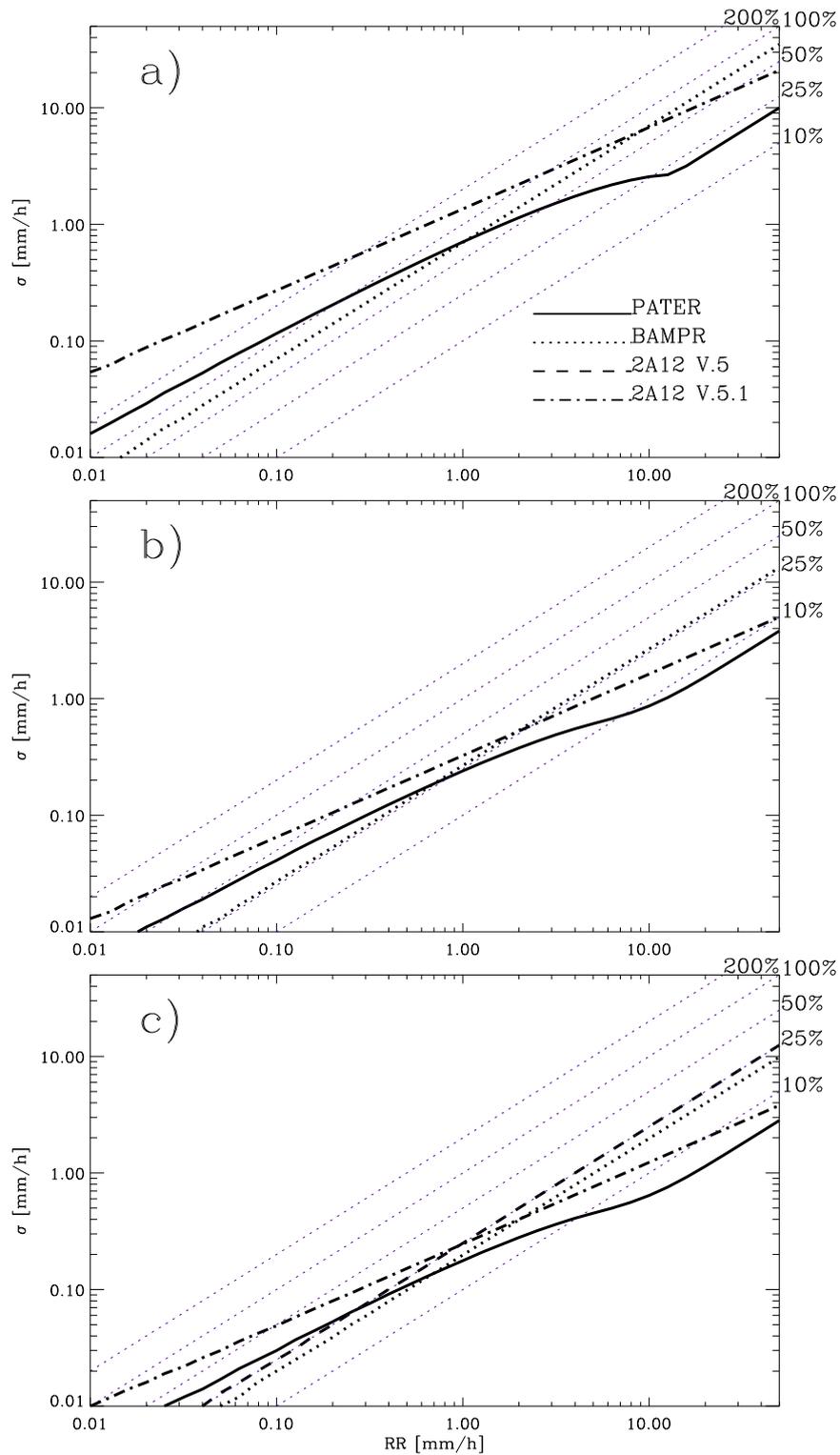


Figure 3: Error-fits at product resolution (a) and after spatial averaging to 45 km model grid (b) and 60 km model grid (c) applying (12). The error of product 2A12 V.5 is defined for the 60 km model grid only.

Figure 4 shows that this is reasonable between 2 and 20 mm/h for all three algorithms. TMI data covering a 6h-period (3:00–9:00 UTC) on August 25 1998 were taken and within 120 km model grid boxes the spatial correlations from each algorithm were calculated. The total number of data points was between $\approx 20,000$ –40,000 depending on the algorithm. This range indicates already that the rain detection skill is different between the algorithms.

Figure 5.2 shows the results as a function of separation distance, s , for PATER, BAMPR, and 2A12 from the data (Figure 5.2a) and from the following fits (Figure 5.2b):

$$C_{PATER} = a_0 + a_1 s + a_2 s^2, C_{BAMPR} = \frac{\exp(-b_2 s)}{b_0 + b_1 s}, C_{2A12} = \frac{\exp(-c_2 s)}{c_0 + c_1 s} \quad (16)$$

with $a_0 = 1.0416$, $a_1 = -0.016082$, $a_2 = 7.5697 \cdot 10^{-5}$, $b_0 = 0.945962$, $b_1 = 0.0681345$, $b_2 = 0.000735678$, and $c_0 = 0.999447$, $c_1 = 0.053814$, $c_2 = -0.00399736$. The functional form of C_{2A12} and C_{BAMPR} follows the one which was employed in the stochastic rainfall model of Bell et al. (1990).

Generally, with increasing distance the correlation decreases exponentially; however, the different spatial resolutions of the products are expressed by different gradients. Since PATER provides the lowest resolution product involving a strong overlap of neighboring pixels (see Figure 5a, b), a higher correlation for $s < 60\text{km}$ (where $C \approx 1/e$) is noticed. This is about twice the size of the field of view and without any overlap. 2A12 shows a stronger decrease of C with s ; i.e., a larger spatial variability for $s < 60\text{km}$ while beyond 2A12 and PATER behave similarly. Despite its lower spatial resolution with respect to 2A12, BAMPR retrievals show less correlation. A possible reason is the usage of the 85.5 GHz channels in the BAMPR database which are only weakly correlated with the surface rain rate as well as the use of different drop size distributions. As a result, neighboring retrievals may provide large differences in rainfall while TB's are still strongly correlated. Therefore, spatial resolution alone does not explain BAMPR's spatial correlation pattern. Also, the error correlation contained in the forward calculations may differ between the algorithms.

The above correlation patterns were applied by averaging TMI retrievals from 2A12, BAMPR, and PATER using the fits from Figure 5.1 over the same period; i.e., applying (8) for a model grid resolution of 60 km. Please note, that for 2A12 the fit given in (15) was used. As a reference, the computations were repeated without spatial correlation; i.e., using (9). Comparing the two clusters in each panel of Figure 5.2, it is noted that the inclusion of spatial correlation causes an increase of error magnitude but a decrease of scatter. Both effects are caused by the constraint which the correlation imposes on the error compensation. In any case, the error reduction is fairly continuous and almost linear with $\log(RR)$.

The clusters of dots in Figure 5.2 compare very well with the results from the error simulation, i.e., the solid, dotted, and dash-dotted lines in Figure 5.1c. Apparently, the assumption of log-normal pdf's with fixed variance holds well for averaged rain rates. In summary, including spatial correlation increases the total rain rate errors by $\approx 50\%$ because completely uncorrelated estimates are more efficient in random error reduction. Since PATER shows the highest spatial correlations, its increase of errors is most pronounced. In conclusion, the influence of spatial correlation is significant but reduces the scatter and may therefore be accounted for by a simple scaling. This however, depends on the algorithm performance that is the spatial correlation pattern of its rainfall estimates as a function of its nominal spatial resolution as well as the generated rain rate pdf's.

6 Conclusions

As an outcome of the EuroTRMM project, an intercomparison of retrieval errors from different TRMM passive microwave rainfall products and their propagation to average errors on global model gridscale was carried out. All algorithms used for this study employ a Bayesian retrieval framework. The differences originate from

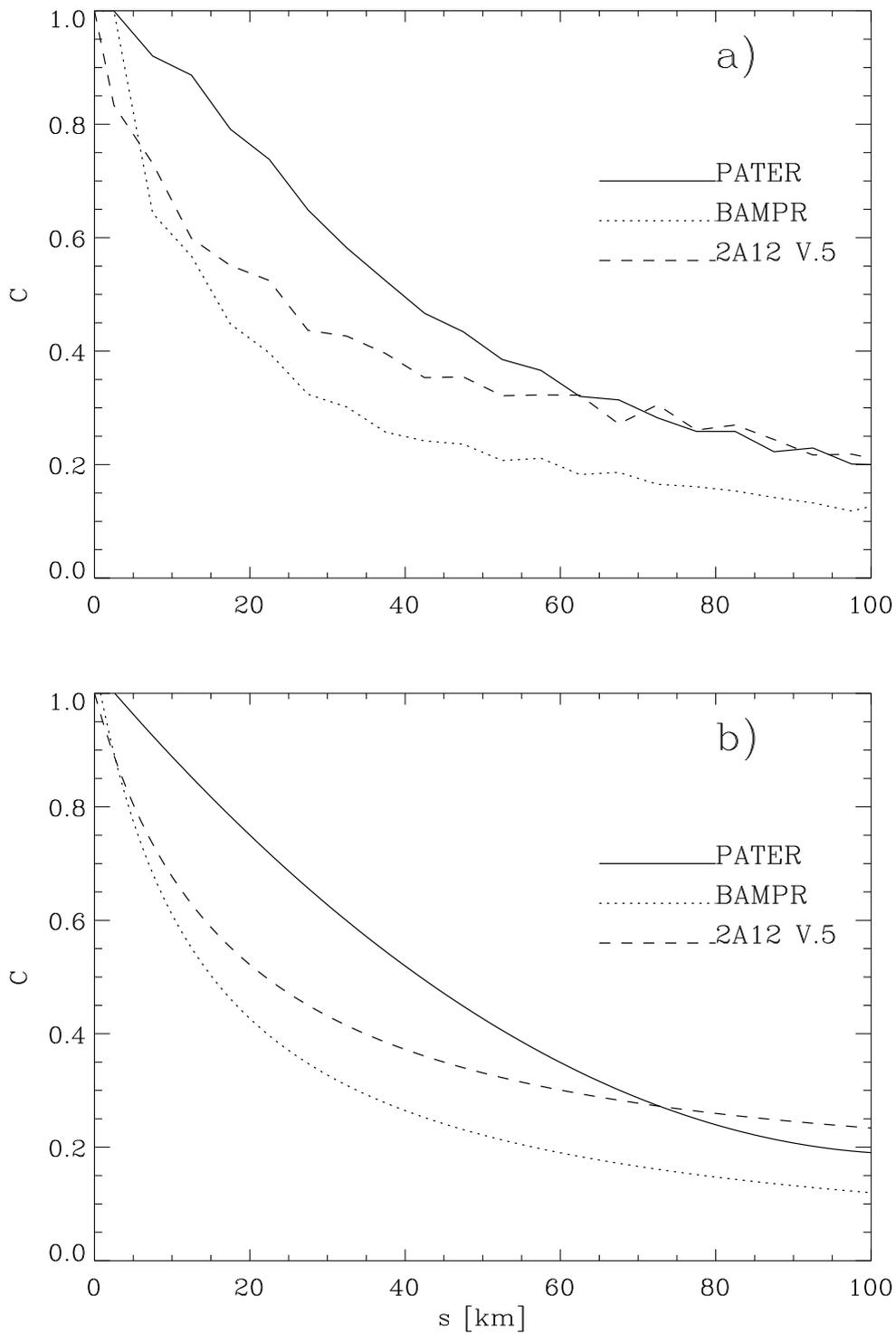


Figure 4: Spatial correlation of TMI products obtained from data (a) and represented by fits (b) in (16).

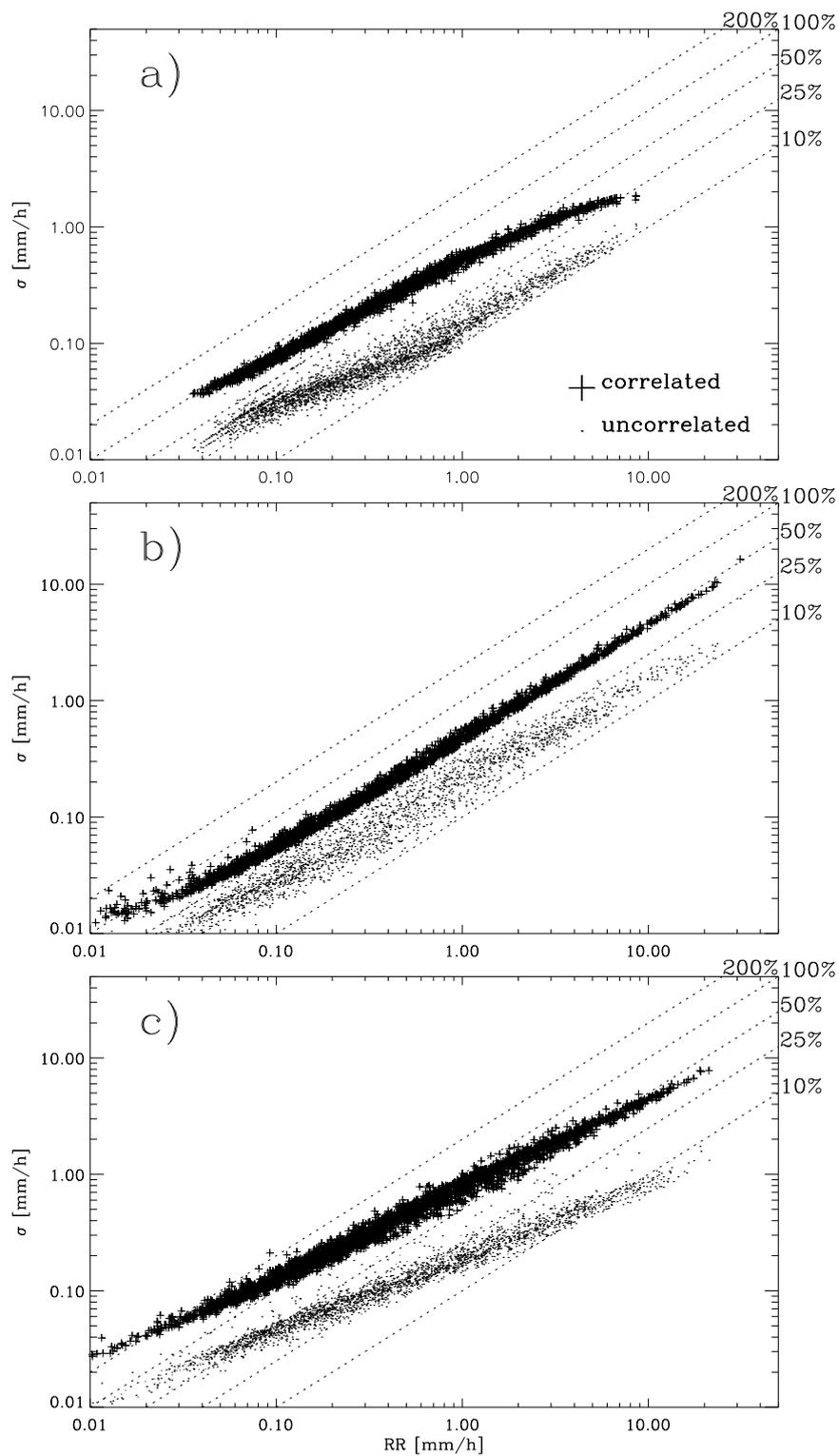


Figure 5: Errors after spatial averaging on 60 km model resolution with and without accounting for spatial correlation from PATER (a), BAMPR (b), and 2A12 V.5 (c) over 6h observation period on August 27, 1998.

different training databases and different probability distribution functions of retrieval variables and observables as well as individual algorithm constraints. If these products are used for variational data assimilation, error estimates are required assuming that the products are either unbiased or have been bias-corrected (with respect to the model fields). One possible output from the Bayesian algorithms is the retrieval error defined as the minimum variance solution for the errors associated with that for rain rates (or hydrometeor profiles).

Bayesian errors between 70–200% at low rain rates (0.1 mm/h) and between 20–70% at high rain rates (20–50 mm/h) were found at the original product resolution and sampling. The large errors at low rain rates originate from large contribution of information to the signal from effects which are not directly related to rainwater, that is surface effects and non-raining cloud emission. As an example, the large errors at low rainfall rates shown in Figure 4c represent surface effects since very high windspeeds were reported in the vicinity of typhoon Paka. In the middle range (1–20 mm/h), TB's are more sensitive to rainwater which reduces the direct retrieval error. However, the variability inside the database is large which means that the ambiguity of the TB-rain rate relation is high. Therefore there are many possible solutions having the same TB's. At high rain rates, this variability decreases again since extreme events are usually not well covered in the databases. On the other side, the strong contribution from precipitating ice increases the direct rain rate retrieval error so that the net effect is the same as for moderate rainfall. Generally, the errors appeared to depend linearly on $\log_{10}(RR)$. Therefore their characterization in terms of $\log_{10}(RR)$ may be useful which also facilitates their interpretation at low rain rates.

An independent evaluation was carried out by comparing TMI estimates to PR retrievals at the same resolution (27x40 km²). Interestingly, the results show very similar errors as those obtained from the Bayesian estimator. Important to notice is that TMI-PR errors contain both systematic and random contributions while the Bayesian error estimates are assumed to be random only. A possible interpretation of error decomposition would be that at low rain rates, the error consists mainly of a bias while at moderate rain rates the random component is dominant.

Two different evaluations of the impact of spatial averaging on global model grid scales were carried out to estimate the errors to be introduced into a variational data assimilation system, as for example outlined in Treadon (1997) and Marécal and Mahfouf (2002). The first evaluation used the error fits from a set of case studies (Figure 5.1). Assuming spatially uncorrelated errors and realistic probability distribution functions of rain, the errors reduced to 30–50% at low rain rates and to 5–30% at high rain rates. The second evaluation used TMI observations on the ECMWF model grid. First, the spatial correlation of each data product was calculated since the number of observations inside a model grid box only depends on the data sampling but not the spatial resolution. The spatial correlation patterns reflected the product resolution (e.g. overlap of neighboring pixels) very well. PATER produced the weakest decrease of correlation with separation distance. BAMPR retrievals showed the strongest decrease even though its spatial resolution is between that of PATER and 2A12. This is likely due to the usage of the 85 GHz channels in the retrieval algorithms and a less homogeneous database. This may cause larger local gradients in retrieved rain rates than obtained from the other algorithms.

Finally, the spatial correlations were used to calculate 60 km model grid box averaged errors as in the previous exercise. The previously modeled reduction of errors is partly compensated by the spatial correlation. The differences between algorithms become slightly spread since their correlation patterns are different. The final error estimates ranged from 50–150% at low rain rates to 20–50% at high rain rates. Neglecting spatial correlation further reduced the errors to 20–50% at low rain rates and to 5–20% at high rain rates. The modeled and observed behavior of spatially uncorrelated error propagation agreed well where observed and modeled rain rate pdf's were similar. An important conclusion is that once the spatial correlation pattern of a product is known, the probability density distribution of real observations inside the model grid does not produce larger scatter. This suggests that a simple scaling may suffice to calculate the error reduction to obtain an average error on the model grid.

Aiming at the direct assimilation of rain-affected radiances, the issue of grid-box averaged radiances has to be solved differently because the relationship between hydrometeor profiles and TB's is highly non-linear. Therefore, the TB of an averaged profile is not identical with the averaged TB's of sub-gridscale profiles. Chevallier and Bauer (2002) have shown that the grid-box averaged TB's are very sensitive to the model assumption on cloud overlap and its implementation in the radiative transfer model. Once this problem is solved the calculation of associated modeling errors is rather straight forward.

Acknowledgements

This project was funded by the European Space Agency (ESA), contract No. 12651/97/NL/NB. The authors gratefully acknowledge the usage of TRMM data which was made available through the TRMM program by the National Aeronautics and Space Administration (NASA). We are also very grateful to the three referees which provided very useful suggestions for the improvement of this article.

References

- Bauer, P. 2001: Over-ocean rainfall retrieval from multi-sensor data of the Tropical Rainfall Measuring Mission (TRMM). Part I: Design and evaluation of inversion databases. *J. Ocean. Atmos. Tech.*, **18**, 1315–1330.
- Bauer, P., Amayenc, P., Kummerow, C.D., and Smith, E.A. 2001: Over-ocean rainfall retrieval from multi-sensor data of the Tropical Rainfall Measuring Mission (TRMM). Part II: Algorithm implementation. *J. Ocean. Atmos. Tech.*, **18**, 1838–1855.
- Bell, T.L., Abdullah, A. Martin, R.L., and North, G.R. 1990: Sampling errors for satellite-derived tropical rainfall: Monte Carlo study using a space-time stochastic model. *J. Geophys. Res.*, **95**, 2195–2205.
- Chevallier, F. and Bauer, P. 2002: Model rain and clouds over oceans: Comparison with SSM/I observations *Month. Wea. Rev.*, submitted.
- Kedem, B., Chiu, L.S., and North, G.R. 1990: Estimation of mean rain rate: Application to satellite observations. *J. Geophys. Res.*, **95**, 1965–1972.
- Kummerow, C.D., Olson, W.S., and Giglio, L. 1996: A simplified scheme for obtaining precipitation and vertical hydrometeor profiles from passive microwave sensors. *IEEE Trans. Geosci. Remote Sens.*, **34**, 1213–1232.
- Kummerow, C.D., Barnes, W., Kozu, T., Shiue, J., and Simpson, J. 1998: The Tropical Rainfall Measuring Mission (TRMM) sensor package. *J. Atmos. Ocean. Tech.*, **15**, 809–817.
- Kummerow, C.D., Hong, Y., Olson, W.S., Yang, S., Adler, R. F., McCollum, J., Ferraro, R.R., Petty, G.W., Shin, D.-B., and Wilhelm, T.T. 2001: The evolution of the Goddard Profiling Algorithm (GPROF) for rainfall estimation from passive microwave sensors. *J. Appl. Meteor.*, **40**, 1801–1820.
- Lorenc, A.C. 1986: Analysis methods for numerical weather prediction. *Quart. J. R. Met. Soc.*, **112**, 1177–1194.

- Marzano, F.S., Mugnai, A., Panegrossi, G., Pierdicca, N., Smith, E.A., and Turk, J. 1999: Bayesian estimation of precipitating cloud parameters from combined measurements of spaceborne microwave radiometer and radar. *IEEE Trans. Geosci. Remote Sens.*, **37**, 596–613.
- Marécal, V. and Mahfouf, J.–F. 2000: Variational retrieval of temperature and humidity profiles from TRMM precipitation data. *Month. Wea. Rev.*, **128**, 3853–3866.
- Marécal, V. and Mahfouf, J.–F. 2002: Four–dimensional variational assimilation of total columnar water vapor in rainy areas. *Month. Wea. Rev.*, **130**, 43–58.
- Marécal, V., Mahfouf, J.–F., and Bauer, P. 2002: Comparison of TMI rainfall estimates and their impact on four–dimensional variational assimilation. *Quart. J. R. Met. Soc.*, accepted.
- Mugnai, A., di Michele, S., Marzano, F.S., and Tassa, A. 2001: Cloud–model based Bayesian techniques for precipitation profile retrieval from TRMM microwave sensors. Proc. EuroTRMM Workshop on Assimilation of Clouds and Precipitation in NWP Models, Reading, UK, November 2000, 323–346
- Olson, W.S., Kummerow, C.D., Heymsfield, G.M., and Giglio, L. 1996: A method for combined passive–active microwave retrievals of cloud and precipitation profiles. *J. Appl. Meteor.*, **35**, 1763–1789.
- Panegrossi, G., Dietrich, S., Marzano, F.S., Mugnai, A., Smith, E.A., Xiang, X., Tripoli, G.J., Wang, P.K., and Poiaras Baptista, J.P.V. 1998: Use of cloud model microphysics for passive microwave–based precipitation retrieval: Significance of consistency between model and measurement manifolds. *J. Atmos. Sci.*, **55**, 1644–1673.
- Rabier, F., Jaervinen, H., Klinker, E., Mahfouf, J.–F., and Simmons, A. 2000: The ECMWF operational implementation of four–dimensional variational assimilation. Part I: Experimental results with simplified physics. *Quart. J. R. Met. Soc.*, **126**, 1143–1170.
- Rodgers, E., Olson, W.S., Halverson, J., Simpson, J., and Pierce, H. 2000: Environmental forcing of Supertyphoon Paka’s (1997) latent heat structure. *J. Appl. Meteor.*, **39**, 1983–2006.
- Tassa, A., Di Michele, S., Mugnai, A., Marzano, F., and Poiaras Baptista, J.P.V. 2002: Cloud–model based Bayesian technique for precipitation profile retrieval from TRMM microwave imager. *Rad. Sci.*, submitted.
- Treadon, R.E. 1997: Assimilation of satellite derived precipitation estimates with the NCEP GDAS. *PhD Dissertation*, Dept. of Meteorology, Florida State University, Tallahassee, FL, USA, pp. 348.