TECHNICAL MEMORANDUM

Comparison of TMI rainfall estimates and their impact on 4D–Var assimilation

V. Marécal[1], J.–F. Mahfouf and P. Bauer

Research Department

[1]Laboratoire de Physique et Chimie de l'Environnement, Orléans, France

Abstract

The objectives of this paper are to perform a comparison between three rainfall rate estimates from TMI (Tropical Rainfall Measuring Mission Microwave Imager) data and to study their impact in the ECMWF four-dimensional variational (4D-Var) assimilation system. The three algorithms for rainfall estimation considered are: PATER (Precipitation Radar Adjusted TMI Estimation of Rainfall), BAMPR–P (Bayesian Algorithm for Microwave–based Precipitation Retrieval), and 2A12 level 5 (NASA operational algorithm). Results from the comparison show that BAMPR–P and 2A12 retrievals provide on average higher rain rates by about a factor 1.5 with respect to PATER mainly due to the different spatial resolutions and rain detection methods used in the three algorithms. The three TMI products are then compared to the rainfall rates from the ECMWF model. Globally, the rain occurence simulated by the ECMWF model is higher than in TMI estimates. The model also produces lower rainfall rates in precipitating systems. PATER retrievals are generally closer to the model than those from other algorithms.

Three 15–day assimilation experiments ("Rain–PATER", "Rain–BAMPR-P" and "Rain–2A12") were run using the three TMI rainfall rate estimates in the ECMWF 4D-Var assimilation system. Fairly small differences were found between the global analyses and forecasts from all experiments and a control experiment without rainfall data assimilation. This is explained by the small number of TMI rain observations which are used per assimilation cycle compared to the other types of observational data. The local impact on analyses was studied on two cases, namely a tropical cyclone ("Bonnie") and a midlatitude front propagating into the subtropics. For these two cases, the differences between the three TMI products and their associated errors lead to significant changes in the humidity and wind analyses.

1 Introduction

The assimilation of satellite derived rain rates in numerical weather prediction (NWP) models has made important progress during the last decade due to advances in data assimilation techniques and spaceborne instruments (Treadon 1997; Krishnamurti *et al.* 2001; Hou *et al.* 2001; Marécal and Mahfouf 2002). Data assimilation systems based on the variational technique have been recently implemented in a number of meteorological operational centres. The design of variational assimilation systems allows an increasing usage of non-conventional observations such as satellite radiances and/or geophysical products. New satellite instruments that can provide accurate estimates of precipitation from space such as the Precipitation Radar (PR) on board the Tropical Rainfall Measuring Mission (TRMM) and the Advanced Microwave Scanning Radiometer (AMSR) on board EOS-Aqua and ADEOS-II are becoming available. However, given the large differences in rain estimates provided by current retrieval algorithms (Smith *et al.* 1998) it seems important to assess their significance in terms of data assimilation for NWP.

The European Community (EC) and the European Space Agency (ESA) jointly funded a three year project (EuroTRMM, between 02/1998 and 02/2001) based on the use of TRMM data. The objective of EuroTRMM was twofold: (1) to process TRMM data and generate rainfall products in order to provide assimilation or verification data for numerical weather prediction models; and (2) to attempt the assimilation of precipitation data in the European Centre for Medium–Range Weather Forecasts (ECMWF) forecasting system, and to use the precipitation data for testing and tuning convection schemes. Objective (1) included the development of two algorithms for rainfall rate estimation from TMI data: PATER (PR–Adjusted TMI Estimation of Rainfall, Bauer 2001, Bauer *et al.* 2001) and BAMPR–P (Bayesian Algorithm for Microwave-based Precipitation Retrieval, Mugnai *et al.* 2001).

Within the EuroTRMM project, a method allowing the assimilation of surface rain rates in the operational ECMWF 4D–Var analysis system was developed (Marécal and Mahfouf 2002). The method was tested using NASA operational surface rain rates (2A12 product version 5). Results showed a positive impact on both analysis and forecast performances in the tropics, in particular better trajectories of tropical cyclones were

produced in the analyses and the forecast errors for wind and upper tropospheric temperature were reduced.

The objective of this paper is to conduct a comparative study between PATER, BAMPR–P, 2A12 products and the model surface rain rates and to study their individual impact on the assimilation system. Results are presented from a 2–week period in August 1998 (18/08–02/09) and two case studies from that period. A short description of PATER, BAMPR–P and 2A12 algorithms is given in section 2. In Section 3 the algorithm performance is analyzed with focus on sampling and resolution characteristics. In section 4 the comparison between the model surface rain rates from the ECMWF analysis and the three TMI rain products averaged at model resolution is discussed. Section 5 provides a brief description of the method used to perform rain rate assimilation and an outline of the assimilation experiments. Results from these analysis experiments are discussed in section 6. Concluding remarks and a discussion on issues related to rain assimilation from the experiments are given in section 7.

2 Description of TMI algorithms

2.1 Bayesian framework

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 the atmospheric state $\mathbf{x}$ whereas $P(\mathbf{x}|\mathbf{y}^o)$ is the conditional probability that the atmospheric state $\mathbf{x}$ is present once $\mathbf{y}^o$ is observed.

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

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

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

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

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 (Lorenc 1986):

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

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:

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

Alternatively, a maximum a–posteriori probability approach can be chosen where $\mathbf{x}_a = \max P[\mathbf{y}^o - \mathbf{y}(\mathbf{x})]P(\mathbf{x})$ (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):

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

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

with cost–function J^o 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:

$$\sigma_a{}^2 = \frac{\sum_i J^o[\mathbf{y}^o, \mathbf{y}(\mathbf{x}_i)](\mathbf{x}_i - \mathbf{x}_a)^2}{\sum_i J^o[\mathbf{y}^o, \mathbf{y}(\mathbf{x}_i)]} \tag{6}$$

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.

2.2 Retrieval Errors

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 section where $\mathbf{y}$ contains **TB** as well as emission/scattering indices. It also uses constraints driven by a stratiform/convective classification and distance–from–convection measures (Olson *et al.* 2001). The latter have been introduced between versions 4 and 5. Since the product itself contains no error information, a rough estimate of 25% was given (Kummerow 1999, personal communication) to be valid on the 60 km model grid used in the previously mentioned assimilation studies.

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). Due to the strong correlation between radiometer channels, an empirical orthogonal function (EOF) analysis was carried out. It was found that 2 EOF's explained about 98% of the variance contained in the TMI channels between 10.7 and 37.0 GHz. Both 85.5 GHz channels were excluded because the uncertainties in the combined cloud–radiative transfer modeling increase strongly with frequency. Therefore, the Bayesian framework explained in Section 2 was reformulated for 2 EOF's so that the final PATER algorithm only consists of a two–dimensional look–up table. Important to notice is that this algorithm derives rainwater content (in g/m^3) rather than rain rate (in mm/h); for conversion, the relation

$RR = 20.95 \cdot w^{1.12}$ has been used where $[RR] = mm/h$ and $[w] = g/m^3$. The error of this fit is well below 0.5% for $w \in [0, 2\,g/m^3]$ assuming a Gamma drop size distribution with a shape parameter of one.

The PATER algorithm allows an intercomparison of PR and TMI estimates of near–surface rain liquid water contents at the same resolution (Bauer *et al.* 2001). Systematic differences between the two independent retrievals are collected in a dynamically updated database as a function of rainwater content and are subsequently used for the correction of TMI–biases. This assumes that PR retrievals are unbiased. Within the same procedure, the remaining standard deviation is calculated which serves as an estimate of the absolute TMI retrieval error. 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 signature (Bauer 2001). Below, background noise originating from surface, non–precipitating clouds, and atmosphere supersedes the rain signal while above the signature from precipitating ice reduces the information on rainwater.

For the assimilation experiments described here, a preliminary fit to these errors was used:

$$\sigma_{PATER}[mm/h] = \max[23.33w^{0.12}\sigma_w, 0.2RR] \tag{7}$$
$$\text{with} \qquad \sigma_w[g/m^3] = [0.1 + 0.45 \cdot log_{10}(w)]w$$

BAMPR is based on the retrieval scheme implemented by Marzano *et al.* (1999) of which the radiometer–only version (BAMPR–P) 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–P 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–P simulation databases by Panegrossi *et al.* (1998).

The database was sub–divided into weaker 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. Error covariances were derived from sensitivity tests. All TMI channels are used in the retrieval and products are provided at the resolution of the 37.0 GHz channel footprint. The BAMPR–P errors are represented by:

$$\sigma_{BAMPR-P}[mm/h] = \max[0.1, 0.7RR] \tag{8}$$

The retrieval errors associated with the TMI rain products from the three algorithms are compared in Fig. 1.

3   Comparison of TMI rainfall estimates

All TMI rain rate products have been defined with different spatial resolutions, namely 27 km x 44 km for PATER and about 10 km x 16 km for both BAMPR–P and 2A12 (see Table 1). The latter, however, is interpolated to all 85 GHz footprint locations so that twice as many data points are provided per scan line than for the other products. Averaging these products to match the model grid size will lead to different averaged rain rates because of these specifications. Principally, the lower the spatial resolution of a product is, the lower the retrieved rain rates will be. This effect has to be separated from differences in the retrievals themselves which are caused by, for example, different retrieval databases.

To analyze the resolution/sampling part of the problem, a sensitivity test was carried out using rainfall rate retrievals from the Precipitation Radar onboard TRMM. This eliminates retrieval algorithm differences and
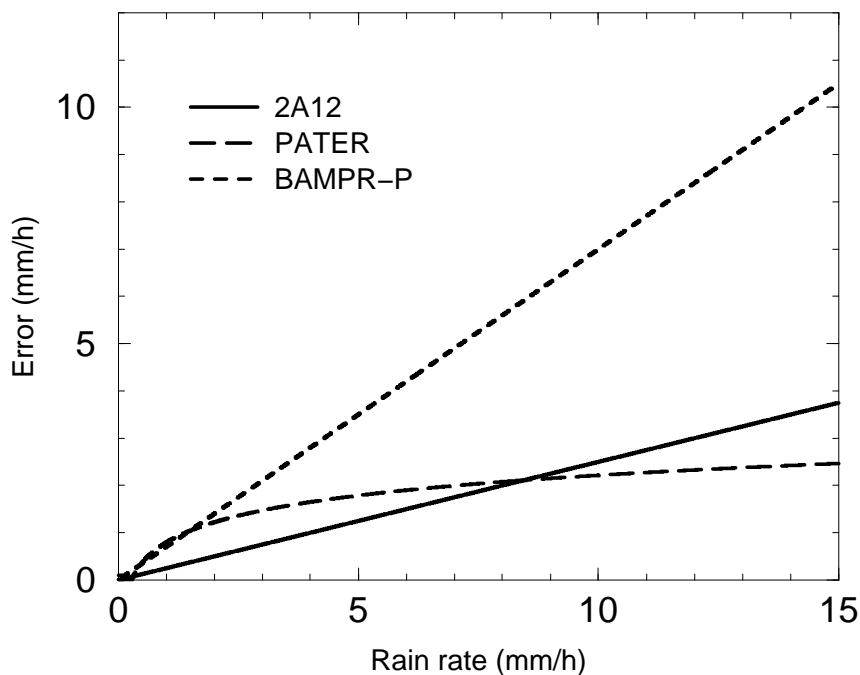
Figure 1: TMI rain rate errors used in 1D–Var as a function of the rain rate in mm/h for PATER, BAMPR–P and 2A12 products.

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–P	Mugnai *et al.* (2001)	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

allows the isolation of how sampling and spatial resolution affect the model grid spatial averages use for the assimilation. The PR data are available over a 220 km swath centered on the TMI scan with a spatial resolution of about 4 km along 49 beams. The conversion from effective reflectivities (including attenuation) to rain rates was carried out as described in Bauer *et al.* (2001) as were the 12 cases selected for evaluating the PATER algorithm.

PR data were averaged to match TMI effective field of view (EFOV) dimensions as given by Bauer and Bennartz (1998) for channels 1, 4, and 5 (10.7, 37.0, 85.5 GHz). Note that the 10.7 GHz resolution is enhanced by a deconvolution technique. The averaging was carried out using synthetic TMI antenna pattern response functions centered at the TMI footprint locations of the lower frequency channels. Thus the sampling and imaging of the TMI is reproduced as is the rainfall statistics from the selected cases. In a second step, the TMI–equivalent rain rates were again averaged to produce 60 km x 60 km rain rates as used for the assimilation described in the following section. The probability distribution functions (pdf) of these rain rates are displayed in Figure 2.

Figure 2a shows the pdf's at product resolution and sampling, respectively. Rain rates are given in dB(RR) = $10 \cdot log_{10}RR$ to account for the quasi–lognormal rainfall pdf–shape. Since all products were generated with the same sampling, the differences originate only from the spatial resolution. At PATER resolution, a shift of
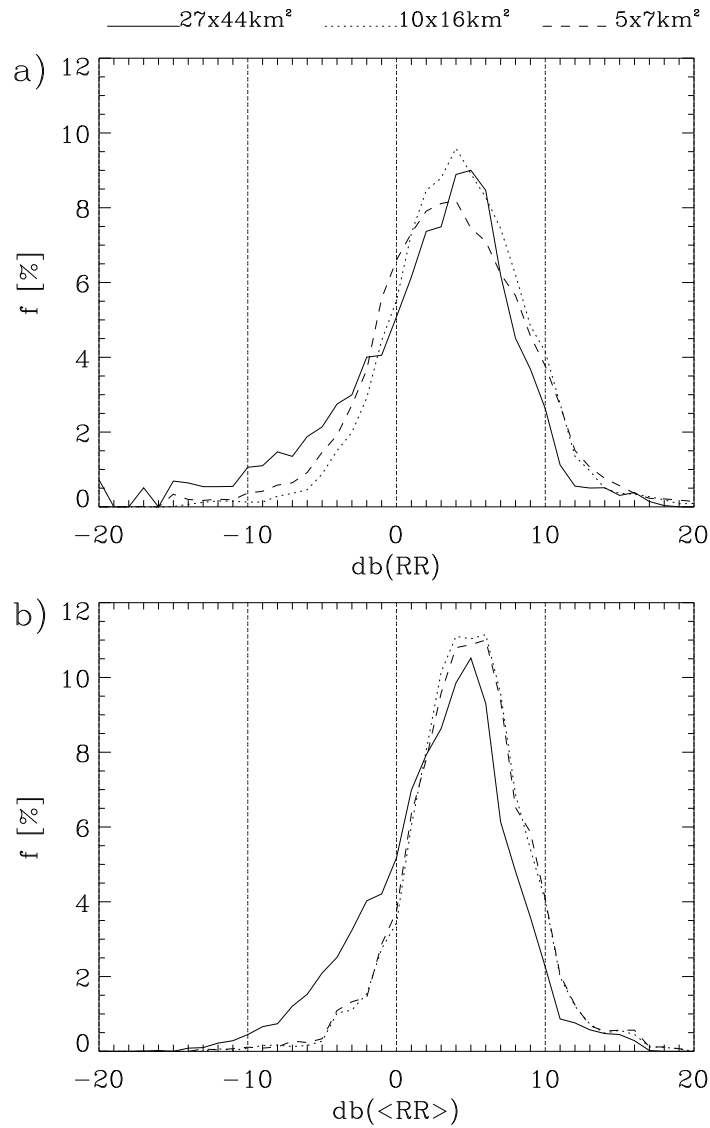
Figure 2: Probability distribution of rain rates (in dB) at product resolution (a) and averaged to 60 km grid boxes (b) for PATER (solid), BAMPR–P (dotted), and 2A12 (dashed) from 12 representative TRMM PR overpasses including fronts, tropical cyclones, as well as scattered and deep convection.

the pdf towards smaller rain rates by about 1–2 dB(RR) with respect to the others is noted. At 0.1, 1, and 10 mm/h a 1 dB shift roughly corresponds to 0.02 mm/h, 0.25 mm/h, and 2 mm/h i.e. 20–25%. After the spatial averaging to 60 km model grid cells, the pdf's become more narrow and averaged rain rates are smaller than at the original resolutions (see Figure 2b). The previously observed shift is increased up to 4 dB at low rain rates. This corresponds to a systematic reduction of average rain rates by about 50–100% as a function of original product resolution. The increased shift can be explained by the strong overlap of neighbouring pixels at PATER resolution and TMI sampling vs. no overlap of BAMPR–P/2A12 pixels with the same sampling.

Comparing rainfall occurrence for the entire 2–week period shows that at product resolution PATER obtained 3,120,000 rain contaminated pixels while BAMPR–P and 2A12 retrieved 2,210,000 and 5,230,000 rain events, respectively. The latter has to be divided by two since the 2A12 sampling is twice as high as for the other algorithms. Thus PATER retrieves 20% more rain events than 2A12 while BAMPR–P finds 20% less events. After the averaging procedure, 1,140,000, 970,000, and 1,200,000 samples remain for PATER, BAMPR–P, and 2A12. This suggests that apart from algorithm differences also rain detection uncertainties of 10–20% will determine the global distributions to be assimilated.

Figure 3 presents the 2–week zonally averaged rain rates and their standard deviations from all products at product resolution (Figure 3a, c) and averaged to model resolution (Figure 3b, d). Zonal averages are obtained using a one degree regular grid and by accounting for rainy points only. As predicted by the previous analysis, PATER estimates are systematically lower than those from BAMPR–P and 2A12) by a factor of 1.5–2. As shown above, a significant portion of this originates from the lower spatial resolution of PATER. The same holds for the standard deviations. PATER rainfall distributions are fairly smooth with little variation over latitude. Both BAMPR–P and 2A12 reveal a large variability above -10 degrees latitude caused by very large local rainfall gradients in the ITCZ, the Eastern Arabian Sea, the Caribbean, and East of Japan (not shown here). This points at a different response to mainly convective rainfall by each algorithm.

In summary, when assimilating satellite rainfall products, a large contribution to the uncertainty of spatially averaged distributions originates from rain detection skill. The spatial averaging procedure has to account for both spatial resolution and sampling of the products which also influences the horizontal correlation of retrievals and errors. The latter plays a significant role when calculating the spatially averaged errors (Bauer *et al.* 2002).

4 ECWMF model analyses vs. TMI retrievals

The ECWMF model is a global spectral model. Its prognostic variables are divergence, vorticity, temperature, specific humidity, cloud content, and cloud fraction. Rainfall rate is a diagnostic quantity produced by the parameterizations of convection (Tiedtke 1989, Gregory *et al.* 2000) and large scale condensation (Tiedtke 1993). The model version used in this study has 50 vertical levels (up to the 0.1 hPa level) and a T_L319 truncation corresponding to a horizontal grid–mesh of about 60 km.

To perform the comparison with the model, the TMI rain rate estimates from PATER, BAMPR–P and 2A12 were averaged to the model resolution (as in Marécal and Mahfouf 2000) to be consistent with the model rain rate which represents the mean of precipitation and/or no–precipitation areas in the model grid box. A simple binning method was used to average the satellite–derived rain rate fields. TMI rain products are averaged spatially over 6–hour periods and compared to the model instantaneous fields from the analysis at the middle time of each 6–hour period. This implies a maximum difference of three hours between the model and the observations. Model instantaneous rain rates are preferred for the comparison to accumulated rain rates since the TMI swath generally overpasses a particular area once within 6 hours.

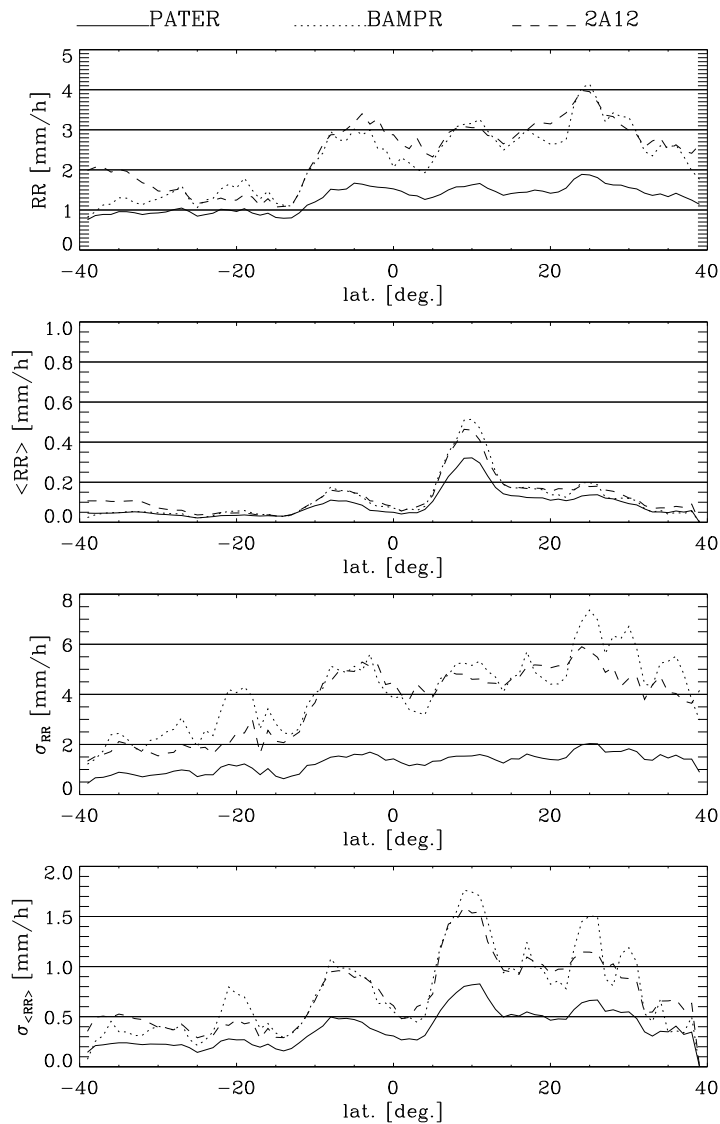Figures 4 and 5 display the model surface rain rates together with the TMI estimates at model resolution for

Figure 3: 2–week zonally averaged rain rates and standard deviations at product (a,c) and model resolution (b,d) from PATER (solid), BAMPR–P (dotted), and 2A12 (dashed).

two cases studies : hurricane "Bonnie" and a frontal system in the Southern hemisphere. For "Bonnie", three regions of interest have been labeled: regions A, B and C (see Figure 4b). For the front case, two overpasses are available within the time window considered (between 30/08/1998 at 21 UTC and 31/08/1998 at 03 UTC). In the following, the North and South overpasses are noted overpass 1 and overpass 2, respectively (see Figure 5b).

In both Figures 4 and 5, the rain rate field from the model analysis has a larger spatial extent than any of the three TMI products. For the "Bonnie" case (Figure 4) model results are closer to the PATER product in terms of intensity in the central part of the cyclone than either BAMPR–P or 2A12 retrievals. Both provide very large rain rates. For the front case (Figure 5), 2A12 is very similar to the model for overpass 2 while BAMPR–P and PATER provide weaker rain rates. For overpass 1, the model rain field provides slightly less rain than all three TMI estimates.

To give a more global picture of the model behaviour compared to TMI observations, statistics (observation
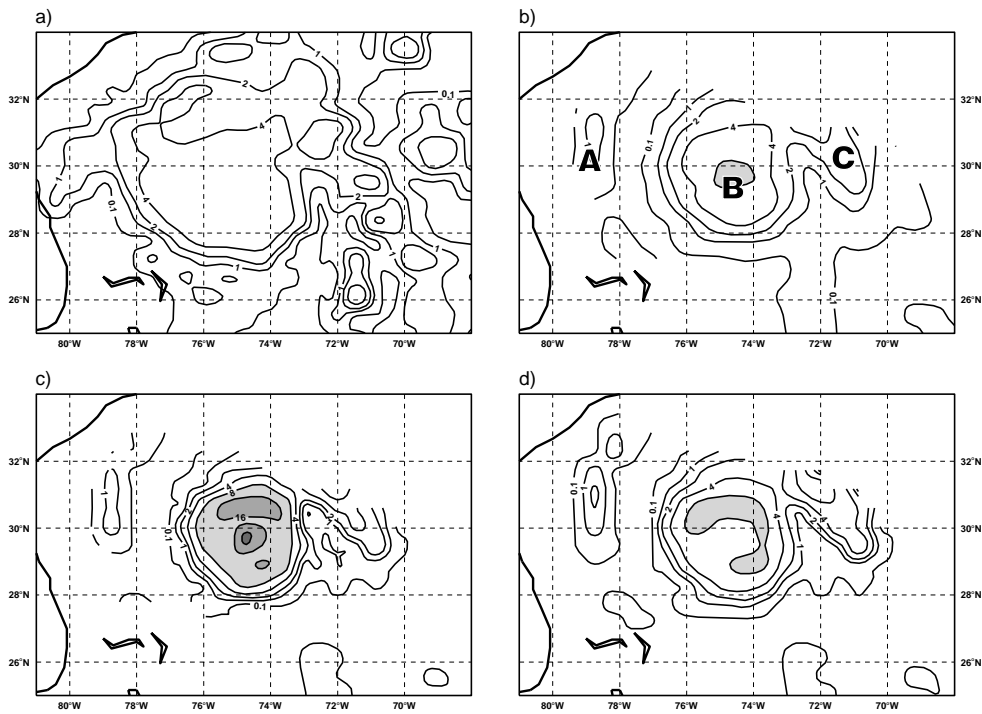
Figure 4: Instantaneous surface rain rate in mm/h for the "Bonnie case" on 25/08/1998 at 1800 UTC (a) from the model, (b) from PATER, (c) from BAMPR–P, and (d) from 2A12 algorithms applied to TMI observations.

- model) over a 15–day period starting on 18/08/1998 at 09 UTC and ending on 02/09/1998 at 15 UTC were computed. Results are shown in Figure 6. Two sets of observations are considered in these statistics: all TMI rain rate observations including zero rain rates (denoted "ALL OBS") or positive TMI rain rates only (denoted "OBS > 0"). Biases for the "ALL OBS" case are all negative showing that the model produces too much precipitation on average compared to the observations. This is because the model tends to generate precipitation more often than observed as also illustrated in Figures 4 and 5. One of the main reasons is that the scale which is actually resolved by the model is larger than one grid length leading to smoother fields than in the observations. For the "OBS > 0" case, larger root–mean square differences (RMS's) and positive biases are found for both BAMPR–P and 2A12. This is because BAMPR–P and 2A12 algorithms produce rather intense rain rates compared to PATER algorithm (previously discussed in section 3) and compared to the model. The PATER product is consistently closer to the model than BAMPR–P and 2A12 estimates with larger correlations and much smaller RMS's. The PATER algorithm produces on average smoother and less intense rain fields than the other two.

Correlation coefficients for "ALL OBS" and "OBS > 0" are fairly low (between 0.26 and 0.38). The comparison is performed on a fine grid-mesh (60 km) meaning that a shift of only few grid boxes of the model rainy areas compared to observations leads to a large decrease of the correlation. Moreover, the model can only resolve explicitly rainy systems having horizontal scales of several grid boxes. This representativeness problem has been confirmed by comparing the model rain rates with TMI products filtered spatially using a Cressman technique with a radius of 120 km. With smoother observed precipitation fields, the correlations of the "OBS > 0" and "ALL OBS" are increased by about 0.1 and the biases and RMSs are decreased by a factor 2–3 (not shown).
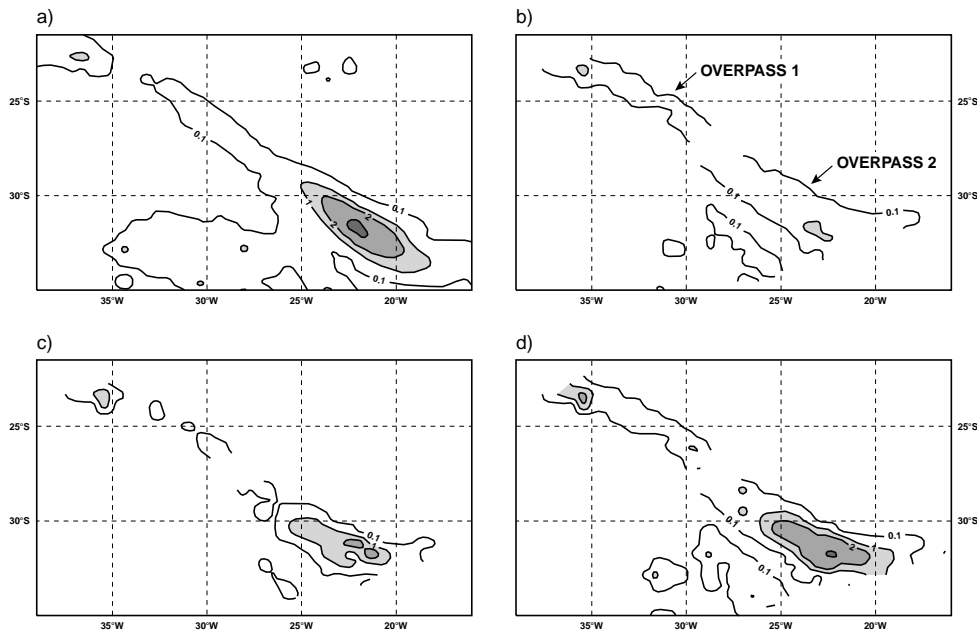
Figure 5: Same as Figure 4 but for the "Front case" on 31/08/1998 at 00 UTC.

5   Assimilation experiments

The operational ECMWF assimilation system is based on an incremental four–dimensional variational (4D–Var) method (Rabier *et al.* 2000; Mahfouf and Rabier 2000; Klinker *et al.* 2000). 4D–Var aims at an optimal balance between observations and the dynamics of the atmosphere by finding a model solution which is as close as possible, in a least–squares sense, to the background information (short–term forecast) and to the observations available over a given time period (six hours in this study).

To assimilate TMI surface rain rate observations, a "1D–Var + 4D–Var" approach was developed. In a first step, 1D–Var provides adjusted model profiles of temperature and humidity that minimize the difference between model and observed surface rain rates within both model and observation errors (Marécal and Mahfouf 2000). The total column water vapour (TCWV), which is the vertical integral of the specific humidity, is then computed from the 1D–Var humidity profiles. In a second step, 1D–Var TCWV estimates are used as a new type of observations in 4D–Var (Marécal and Mahfouf 2002). Thus in the "1D–Var + 4D–Var" approach, the rain rate information from TMI is converted to a humidity information that can be easily assimilated in 4D–Var. Figure 7 shows a schematic flowchart of the "1D–Var + 4D–Var" method. Details on the 1D–Var and on the "1D–Var + 4D–Var" can be found in Marécal and Mahfouf (2000, 2002). The 1D–Var approach was chosen because it allows the influence of this new type of observations to be evaluated without requiring too much technical developments and computing resources. Note that since 4D–Var takes into account the temporal evolution of any variable within the assimilation window, a modification of the analyzed humidity will induce a consistent change of the global solution and consequently of the thermodynamic and dynamic fields.

The errors of the observed rain rates at model resolution needed in the 1D–Var retrieval were already introduced in section 2 and Figure 1. However, rain detection errors are difficult to estimate since they depend on how close the observation is with respect to the cloud. In the vicinity of cloud sytems this error would be larger than in clear–sky situations. Such error estimate dependency has not been included in this study for non–raining observations. Instead, an arbitrarily small value has been chosen. The fairly large discrepancy between rain estimation errors shown in Figure 1 already indicates a large source of uncertainty for the assimilation of these
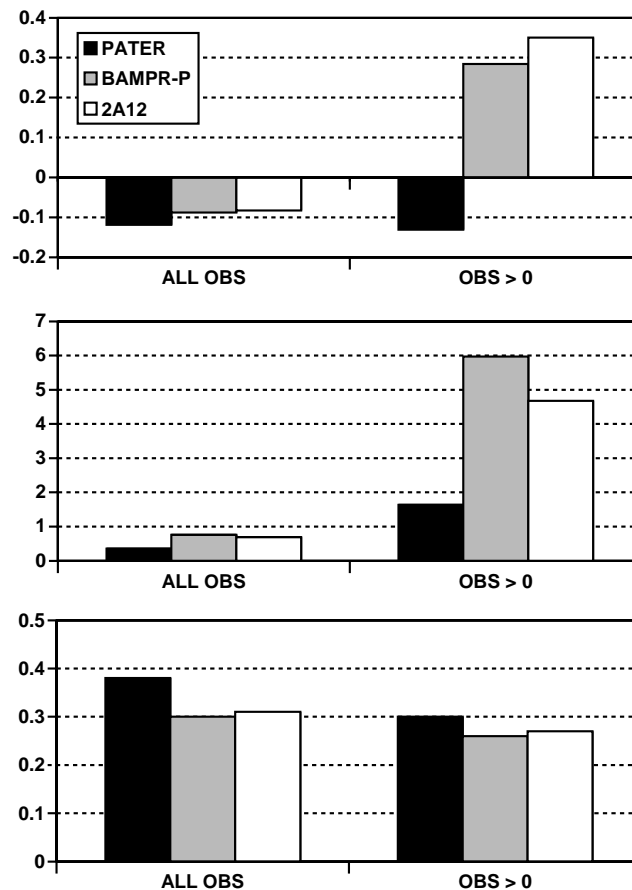
Figure 6: Comparison between ECMWF model surface rain rates and PATER, BAMPR–P and 2A12 products over ocean. Global mean statistics computed for the period between 18/08/1998 at 0900 UTC and 02/09/1998 at 1500 UTC. (a) Bias (observations-model) in mm/h, (b) RMS in mm/h and (c) correlation.

products. While large errors pass more data into the system, their local impact is smaller because they have less weight in the objective function defining the variational problem. With small errors, the 1D-Var quality control defined by Marécal and Mahfouf (2000) will exclude more observations but they will have a stronger impact on the analysis. Since these errors have a rain rate dependency, the relative effect of "number of observations" vs. "impact on analysis" will depend on both model and observation rain probability distributions. A more detailed analysis of rainfall observation errors averaged to global model grid scales is given in Bauer *et al.* (2002).

The period selected for the 4D–Var assimilation tests is from 18/08/1998 at 1200 UTC to 02/09/1998 at 1200 UTC. Four experiments were set up to assess the impact of assimilating 1D–Var TCWV in rainy areas in the ECMWF 4D–Var system. The first experiment is the "Control" where only the operational data set of the considered model version is assimilated. Note that the "Control" experiment rainfall rate fields were already used in section 5 where they were referred to as the model rainfall rate fields. The other three experiments called "Rain–PATER", "Rain–BAMPR–P" and "Rain–2A12" are identical to "Control" except that they also assimilate in 4D–Var all quality–controlled 1D–Var TCWV from PATER, BAMPR–P and 2A12 rain rates, respectively. The 1D–Var quality control consists of retaining TCWV retrievals only if the 1D–Var adjustment provides a rain rate close to the observation within the observation error (Marécal and Mahfouf 2000). Consequently, the behavior of the 1D–Var and then of the 4D–Var experiments will not only be changed by the use of different rain estimates but also by different rain rate errors.
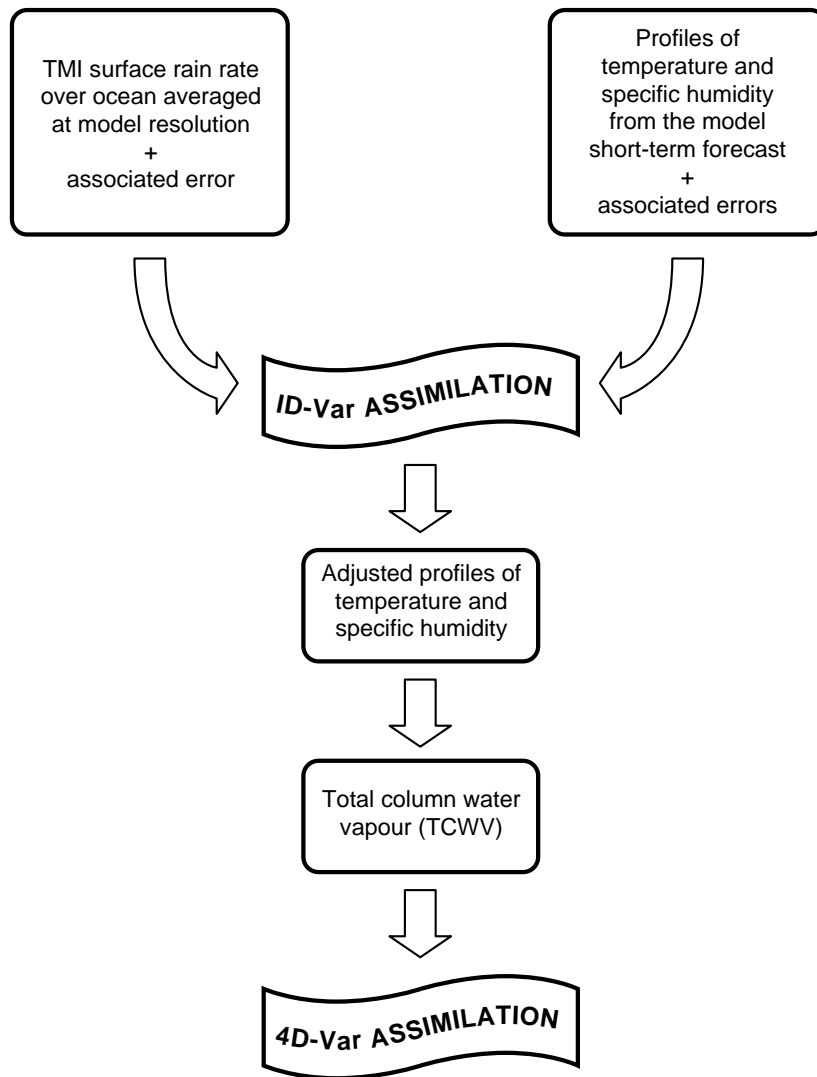
Figure 7: Schematic flowchart of the "1D–Var + 4D–Var" method for the assimilation of TMI rain rates.

6 Results

6.1 Global impact

Figure 8 shows global background departure statistics from the 15–day period of the 4D–Var assimilation of 1D–Var TCWV for the three "Rain" experiments ("Rain–PATER", "Rain–BAMPR–P" and "Rain–2A12"). The observations considered here are the 1D–Var TCWV retrieved from TMI rain rate estimates.

On average 1D–Var TCWV observations from all three "Rain" experiments tend to decrease the model water vapour to reduce the production of precipitation by the model in observed non–rainy areas. Standard deviations for both "Rain–PATER" and "Rain–BAMPR–P" are smaller than for "Rain–2A12". For "Rain–PATER", this is because PATER rain rates are relatively close to the model rain rates and thus require on average small modifications of the humidity field in 1D–Var to adjust the observed rain rates. For "Rain–BAMPR–P", the reason is the large error associated with the BAMPR–P product leading to a weak constraint on the observation in 1D–Var and thus to small humidity adjustments. The number of TCWV observations per assimilation cycle
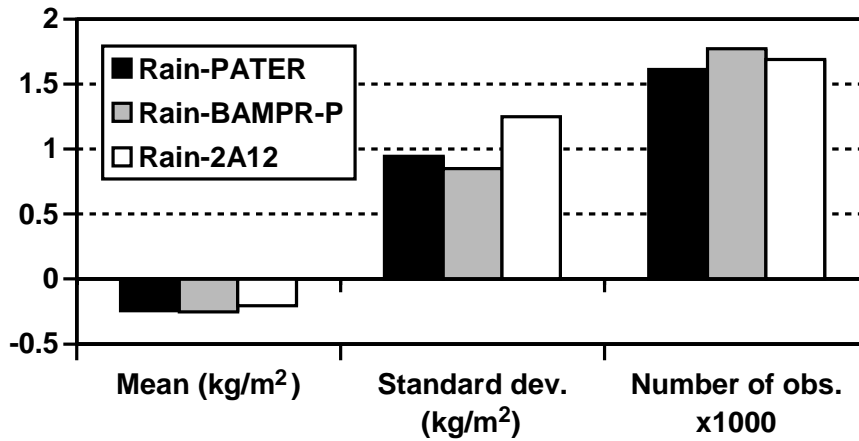
Figure 8: Global mean background departure statistics of 1D–VAR TCWV (in kg m^{-2}) from "Rain–PATER", "Rain–BAMPR–P" and "Rain–2A12" experiments for the 15–day period.

is larger for "Rain–BAMPR–P" than for "Rain-PATER". This shows that 1D–Var is more successful (in the 1D–Var quality control sense) in adjusting temperature and humidity within the observation error of BAMPR–P rain rates because of its larger rain rate error. The comparison of the number of observations "Rain–2A12" with the others is difficult since 2A12 retrievals are interpolated from a higher spatial resolution (see Table 1) than the other two. This leads to more observations close to the coasts and islands and thus to an increase of data in 1D–Var and then in 4D–Var.

The impact on the global mean analyzed TCWV and wind velocity fields is fairly small. Differences between the four experiments are below 1% (not shown). This is mainly related to the low frequency of occurrence of rainy areas within TMI coverage (about 5 % of the data is rain flagged) leading to a small number of 1D–Var TCWV (1,000 to 2,000 per assimilation cycle) to be assimilated with respect to other types of data ($\sim$150,000 per assimilation cycle). Nevertheless, the local impact can be important since rain–related processes lead to latent heat exchanges that modify the dynamics. This will be investigated in the following subsection by focusing on the analysis results obtained on the two case studies presented in section 5.

6.2 Case studies

6.2.1 Bonnie

The TCWV fields from the four analyses for the "Bonnie case" are shown in Figure 9. The three regions (A, B and C) noted in this figure are those where the analysis fields were most strongly modified. In region A, the three "Rain" experiments provide consistently drier atmospheres with respect to the "Control" with maximum differences of $\sim$4 kg m^{-2}, $\sim$3 kg m^{-2} and $\sim$2 kg m^{-2} for "Rain-PATER", "Rain–BAMPR–P" and "Rain–2A12", respectively. This is related to the model behaviour (i.e. "Control" analysis) that is producing too much precipitation compared to any of the three TMI estimates (see Figure 4). The magnitude of the modifications in the "Rain" experiments is well correlated to the difference between the model and the three TMI rain rates. For instance, 2A12 produces larger rain rates than BAMPR–P and PATER and consequently "Rain–2A12" exhibits the moistest atmosphere of the three "Rain" experiments.

In regions B and C, the three "Rain" experiments produce TCWV fields that are different with respect to each other by up to 4 kg m^{-2}. All three provide a moister atmosphere than the "Control". In these two regions, there is no obvious correlation between the intensity of the TMI rain rate fields (Figure 4) and the TCWV analyzed
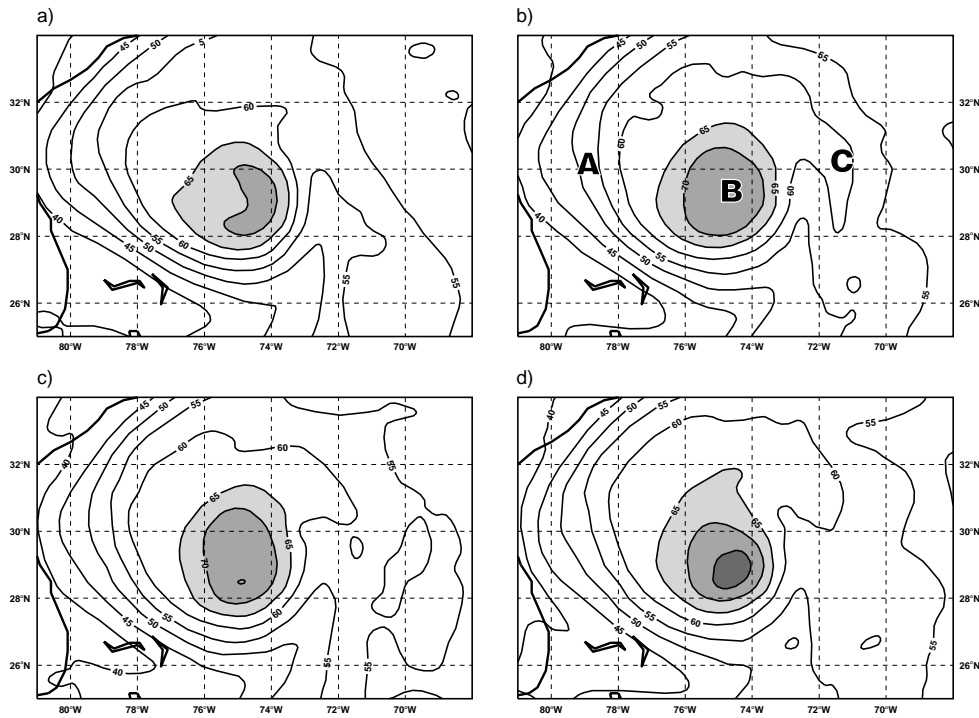
Figure 9: TCWV field (in kg m^{-2})from 4D–Var analysis on 25/08/1998 at 1800 UTC ("Bonnie case") (a) from "Control" experiment, (b) from "Rain–PATER", (c) from "Rain–BAMPR–P", and (d) from "Rain–2A12" experiments.

field (Figure 9): the largest (weakest) rain rates are not related to the moistest (driest) TCWV field. There are two main reasons for this. Firstly, the assimilation experiments are run in a cycling mode meaning that each analysis benefits from all the previous analyses and thus from the TMI data that were already assimilated. This is the reason why the "Rain-PATER" experiment is producing an increase of TCWV in regions B and C (≈ 6 kg m^{-2} at maximum) with respect to the "Control" although a reduction of precipitation was required to fit PATER rain rate observations on the particular overpass chosen (25/08/1998 around 18 UTC). Secondly, 4D–Var analyses depend on the availability of 1D–Var TCWV observations. This is subject to the success of the 1D–Var minimization that is related to both the rain rate estimate used and its associated error. For instance, the BAMPR–P algorithm produces large rain rates with large errors in the central part of the cyclone. This leads to a weak constraint in 1D–Var and thus to 1D–Var TCWV estimates close to the model background state. For 2A12 the large rain rates correspond to a smaller error meaning that in this case 1D–Var is less often successful. However, once it is successful, the impact is more pronounced because of the large TCWV estimates provided by 1D–Var.

Figure 10 displays the impact on the analyzed wind fields at 700 hPa from the four experiments. All "Rain" experiments provide a more intense hurricane compared to the "Control" experiment with stronger ascents and increased horizontal winds in the central part of the hurricane (regions B and C). Nevertheless, the three "Rain" experiments produce analyzed horizontal and vertical winds that are significantly different. This is the direct consequence of the changes in the analyzed TCWV. The three "Rain" experiments give wind fields that are very consistent with the corresponding TCWV fields (Figure 9). For instance, the maximum values of TCWV are found in "Rain–2A12" and correspond to the most intense ascending motions. In region A the impact on the wind field for the three "Rain" experiments is smaller compared to "Control" than in regions B and C because of the smaller TCWV modifications (mean around 2 kg m^{-2} for region A and 4 kg m^{-2} for regions B and C). "Rain-PATER" provides a larger reduction of the convergence in region A compared to "Rain–BAMPR–P" and "Rain–2A12" which is well correlated with a weaker TCWV field. Near the top of the hurricane, at 200 hPa,
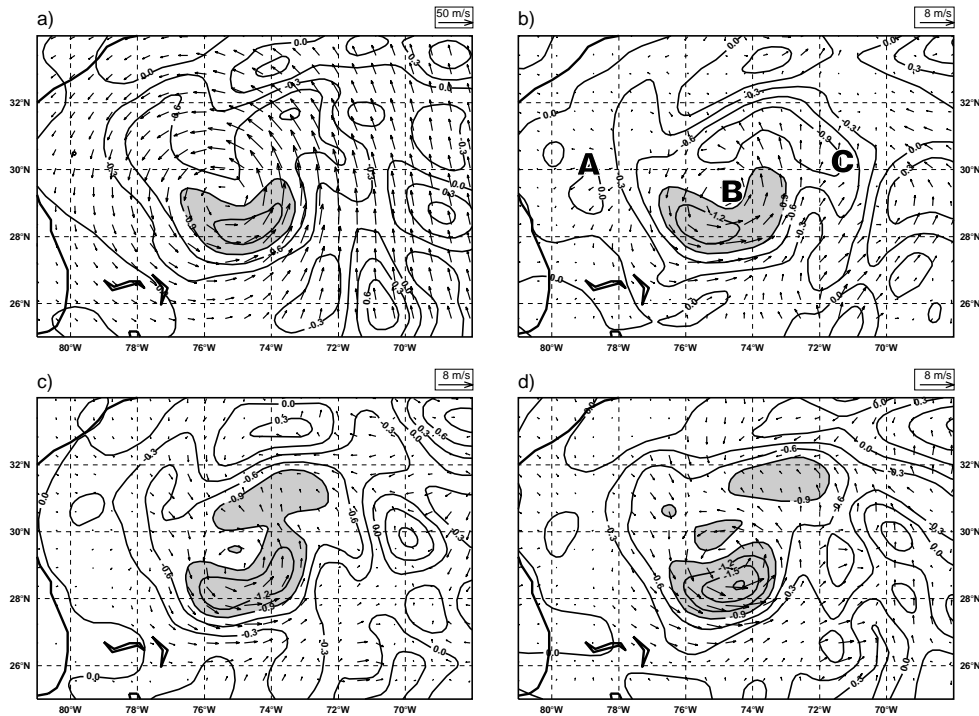
Figure 10: Wind field from 4D–Var analysis on 25/08/1998 at 1800 UTC ("Bonnie case") (a) from "Control" experiment, (b) from "Rain–PATER", (c) from "Rain–BAMPR–P", and (d) from "Rain–2A12" experiments. Contours are for the vertical velocity in Pa/s. Contours are every 0.3 Pa/s wih grey shading starting below -0.9 Pa/s. In (a) arrows display the horizontal wind. In (b) (respectively in (c) and (d)) arrows represent the difference between "Rain–PATER" analysis (respectively "Rain–BAMPR–P" analysis and "Rain–2A12" analysis) and the "Control" analysis.

the changes of the analyzed TCWV field due to the assimilation of TMI rain rates lead to a reinforcement of the divergent flow with relative differences of several m/s between the three "Rain" experiments (not shown).

6.2.2 Front case

Figure 11 displays the analyzed TCWV fields from the four analysis experiments for the front case. For overpass 1, all three "Rain" experiments generally provide a moister atmosphere compared to "Control". This result is consistent with the rain rate fields (see Figure 5) which shows that the model analysis (i.e. "Control") produces slightly weaker rain rates compared to the three TMI estimates. The three "Rain" experiments exhibit differences of up to 3 kg m^{-2} for overpass 1. For overpass 2, a consistent reduction of the TCWV amounts is found for the three "Rain" experiments with respect to "Control". "Rain–2A12" TCWV's are the highest with maximum values greater than 40 kg m^{-2} while "Rain-PATER" and "Rain–BAMPR–P" maximum values are $\sim$37 kg m^{-2} and $\sim$39 kg m^{-2}. The wind field analysis at 700 hPa from the four 4D–Var assimilation experiments is displayed in Figure 12. One noticeable feature is that the largest wind increments ("Rain" experiment minus "Control") are found in the areas where TMI rain rates are assimilated. This illustrates clearly that rain–derived observations (i.e. 1D–Var TCWV) have a direct and significant impact on the analyzed dynamics. For overpass 2, the decrease of precipitation needed to fit TMI observations is converted into a decrease of TCWV leading to a consistent weakening of the front. Again, for both overpasses 1 and 2, the three "Rain" experiment wind fields are noticeably different in terms of intensity and structure. As for the hurricane case, this is related to the TCWV field differences. For this case study, "Rain–2A12" is generally providing smaller increments (in absolute terms) because 2A12 product is closest to the model rain rates and thus requires less adjustment in
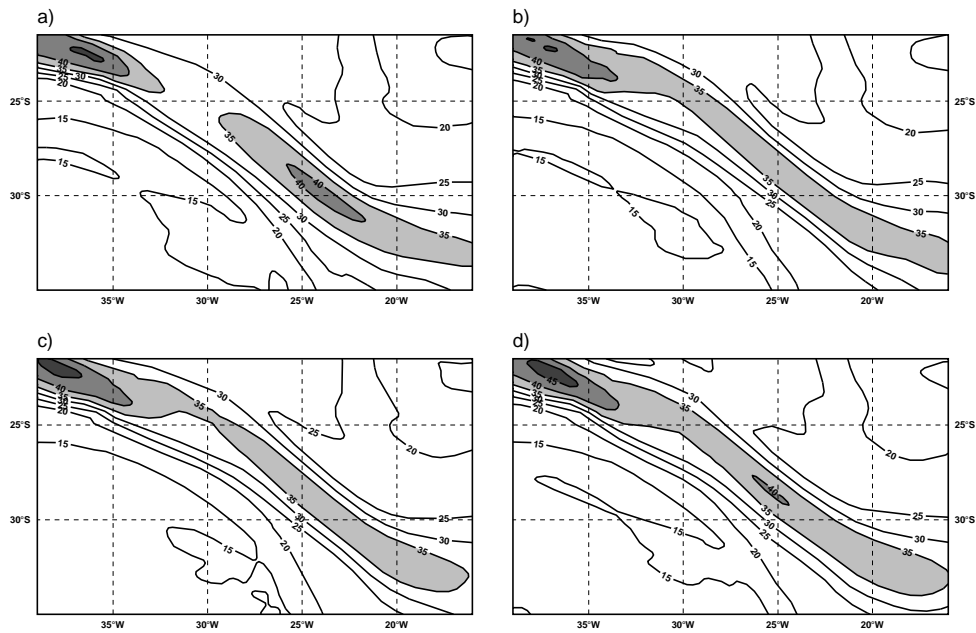
Figure 11: Same as Figure 9 but for the analysis on 31/08/1998 at 0000 UTC ("Front case").

1D–Var retrieval.

7   Conclusion and discussion

In the framework of the EuroTRMM project, two TMI algorithms for rain rate retrieval algorithms were developed and tested: PATER (Bauer 2001; Bauer *et al.* 2001) and BAMPR–P (Mugnai *et al.* 2001). These two algorithms were compared to TRMM standard algorithm 2A12 version 5 (Kummerow *et al.* 1996). PATER rainfall rates exhibit significant differences when compared to BAMPR-P and 2A12. PATER probability distribution function is shifted towards low rainfall rate values. This is mainly explained by the coarser spatial resolution used in PATER algorithm that smoothes the large rainfall rates and by differences in the rain detection methods that modify significantly the contribution of the low rainfall rates.

The three TMI rain rate products (PATER, BAMPR–P and 2A12) were compared to the ECMWF model surface rain rates over oceans (from the model analyses) for two case studies, and globally over a 15-day period. To allow this comparison, PATER, BAMPR–P and 2A12 rain rates were averaged to the model resolution, i.e., 60 km. The rain rate comparison exhibited large discrepancies between the model and the three sets of observations. The model tends to produce precipitation areas which are too extended compared to the observations. This is because the model produces light precipitation very often and can only resolve scales larger than the nominal model grid size. In areas where rain is observed, the model generally produced weaker rain rates compared to BAMPR–P and 2A12 products. Global mean results showed that PATER rain rates are significantly closer to the model rain rates than the others.

The usefulness of assimilating TMI derived rain rates in the ECMWF analysis system was demonstrated recently by Marécal and Mahfouf (2002). Because of the large discrepancies found between the three TMI products which were considered in EuroTRMM, their individual impact on the ECMWF analyses has been evaluated. The approach chosen for the assimilation of surface rain rates is based on the "1D–Var + 4D–Var" method described in Marécal and Mahfouf (2000, 2002). Four assimilation experiments were run over a period
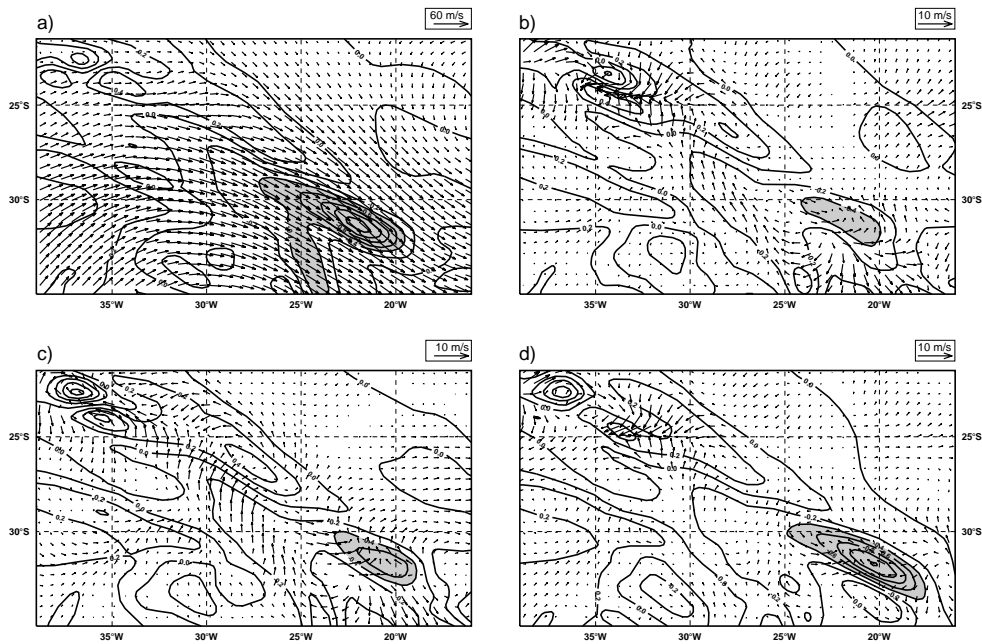
Figure 12: Same as Figure 10 but for the analysis on 31/08/1998 at 0000 UTC ("Front case").

of 15 days: a "Control" and three experiments assimilating 1D–Var TCWV's retrieved from the three TMI estimates of rain rates (PATER, BAMPR–P and 2A12). The errors associated with each of the rain products were significantly different also affecting the 1D–Var retrieval performance. For instance, BAMPR–P rain rates are associated with large errors providing a weak constraint on the 1D–Var minimization and consequently less minimization failures (i.e. 1D–Var quality control failures) compared to PATER or 2A12.

Despite the large differences between PATER, BAMPR–P and 2A12 products, the four experiments were found very similar for the global mean analysis fields of humidity and wind. This is related to the low occurrence of rainy areas which are sampled by TMI within 6 hours (length of the assimilation window). This leads to a small number of TMI–derived TCWV observations to be assimilated in 4D–Var ($\sim$1% of total) compared to the total number of data already assimilated operationally. Even though the global forecast performances for the "Rain" experiments show an improvement of the winds and upper tropospheric temperature in the tropics when compared to "Control", differences found between the three "Rain" experiments were not found significant. In this context, one has to focus on the local impact on 4D–Var analyses.

Two case studies were used to illustrate the local impact: one overpass over hurricane "Bonnie" and two overpasses over a frontal system. They were chosen because the processes involved in the rain generation for hurricanes and fronts are somewhat different. Our results demonstrate that for both cases, the differences between the three rain rate estimates and their associated errors lead to noticeable modifications of the humidity analyses (several kg m^{-2}) in areas where TMI data were assimilated. The consequence is a significant change of the analyzed wind fields (several m s^{-1}) between the "Control" and the "Rain" experiments and also among the three "Rain" experiments. These changes were very consistent with the TCWV fields. In summary, it was shown that the 4D–Var analysis is fairly sensitive to the rain rate product and its associated error used in 1D–Var.

The "1D–Var + 4D–Var" approach used in this paper should be regarded as a first step that demonstrated the usefulness of rain data in the ECMWF assimilation system In this approach, the rain rate information is converted into a humidity information to be assimilated in 4D–Var. By doing so, the rain rate information is filtered and thus its impact on the dynamics is weakened. A more advanced assimilation approach, in which

rain rates are directly assimilated in the 4D–Var, is currently tested. Many issues linked to this approach have to be investigated before being able to perform an optimal operational assimilation. Some of them concern the assimilation methodology itself (configurations of the incremental 4D–Var and of the model physics) while others rather concern the observations.

As for the "1D–Var + 4D–Var" approach, an error on the satellite–derived rain rates has to be specified in the direct 4D–Var assimilation of rain rates. This is even more crucial for the latter since the 4D–Var minimization performance could be easily affected by an erroneous specification of the rain rate observation errors. Thus, it is worth emphasizing that, for variational assimilation purposes, it is absolutely necessary to have a reliable estimate of the observed rain rate error together with any rain rate product. A more detailed study on satellite observation errors and their averaging at model scale is given by Bauer *et al.* (2002). In our study, TCWV observations were assumed to be spatially uncorrelated. A usual way to overcome this problem in operational systems is to perform a thinning by discarding observations which are too close to each other. This implies an undesirable reduction of the number of observations to be assimilated in rainy situations.

Another related issue is the statistical distribution of the observation errors. So far, a Gaussian distribution was used in the "1D–Var + 4D–Var" approach for the sake of simplicity. Actually, these errors are more likely non–Gaussian. As shown by Errico *et al.* (2000) the results of the minimization in a 1D–Var context are very sensitive to the assumed distribution form. Therefore, both errors have to be available and their probability distribution has to be accounted for.

Since many data types are already assimilated in the ECMWF 4D–Var system, the weight given to a new type of observation should be carefully chosen. This weight depends on observation errors and background departures (difference between the observations and the model equivalent in the observation space). Rain rates can vary by several orders of magnitude on small spatial scales so that the background departures can be very large. On one hand, a strict quality control is needed to avoid assimilating observations associated with large departures because of their possible dominant weight in the assimilation. On the other hand, since rainy areas are scarce, it is important to retain as many rain rate observations as possible in the assimilation process. This means that a microwave algorithm providing rain rates close to the model values should be preferred in a data assimilation system where model errors are not included.

Acknowledgements

References

Bauer, P. and Bennartz, R. 1998: TMI imaging capabilities for the observation of rainclouds. *Rad. Sci.*, **33**, 335–349.

Bauer, P. 2001: Over–ocean rainfall retrieval from multi–sensor data of the Tropical Rainfall Measuring Mission. Part I: Design and evaluation of inversion databases. *J. Atmos. Ocean. 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. Part II: Algorithm implementation. *J. Atmos. Ocean. Tech.*, **18**, 1838–1855.

Bauer, P., Mahfouf, J.–F., Olson, W.S., Marzano, F.S., Di Michele, S., Tassa A., and Mugnai, A. 2002: Error analysis of TMI rainfall estimates over ocean for variational data assimilation. *Q. J. Roy. Meteor. Soc.*, in revision.

Errico, R. M., Fillion, L., Nychka, D., and Lu. Z.–Q. 2000: Some statistical considerations associated with the data assimilation of precipitation observations. *Q. J. Roy. Meteor. Soc.*, **126**, 339–359.

Gregory, D., Morcrette, J.–J., Jakob, C., Beljaars, A.C.M., and Stockdale, T. 2000: Revision of convection, radiation, and cloud schemes in the ECMWF Integrated Forecasting System. *Q. J. Roy. Meteor. Soc.*, **126**, 1685–1710.

Hou A.Y., Zhang, S., da Silva, A., Olson, W., Kummerow, C., and Simpson, J. 2001: Improving global analysis and short-range forecast using rainfall and moisture observations derived from TRMM and SSM/I passive microwave sensors. *Bull. Amer. Meteor. Soc.*, **81**, 659–679.

Klinker, E., Rabier, F., Kelly G., and Mahfouf, J.–F. 2000: The ECMWF operational implementation of four–dimensional variational assimilation. Part III: Experimental results and diagnostics with operational configuration. *Q. J. Roy. Meteor. Soc.*, **126**, 1191–1215.

Krishnamurti, T.N., Surendran S., Shin, D.W., Correa-Torres, R.J., Vijaya Kumar, T.S.V., Williford, E., Kummerow, C., Alder, R.F., and Simpson, J. 2001: Real-time multianalysis-multimodel superensemble forecasts of precipitation using TMI and SSM/I products. *Mon. Wea. Rev.*, **129**, 2861–2883

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 Sensing*, **34**, 1213–1232.

Lorenc, A.C. 1986: Analysis methods for numerical weather prediction. *Q. J. Roy. Meteor. Soc.*, **112**, 1177-1194.

Mahfouf, J.–F., and Rabier, F. 2000: The ECMWF operational implementation of four–dimensional variational assimilation. Part II: Experimental results with improved physics. *Q. J. Roy. Meteor. Soc.*, **126**, 1171–1190.

Marécal, V. and Mahfouf, J.–F. 2000: Variational retrieval of temperature and humidity profiles from TRMM precipitation data. *Mon. Wea. Rev.*, **128**, 3853–3866.

Marécal, V. and Mahfouf, J.–F. 2002: Four dimensional variational assimilation of total column water vapour in rainy areas. *Mon. Wea. Rev.*, **130**, 43-58.

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. Geosc. Remote Sens.*, **37**, 596–613.

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. Proceedings of the ECMWF/EuroTRMM Workshop on assimilation of clouds and precipitation. Reading, UK, 6–9 November 2000, 323-345

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.

Olson, W.S., Hong, Y., Kummerow, C.D., and J. Turk 2001: A texture–polarization method for estimating convective–stratiform precipitation area coverage from passive microwave radiometer data.*J. Appl. Meteor.*, **40**, 1577–1591.

Panegrossi, G., Dietrich, S., Marzano, F.S., Mugnai, A., Smith, E.A., Xiang, X., Tripoli, G.J., Wang, P.K., and Poiares Baptista, J.P.V. 1998: Use of cloud model microphysics for passive microwave-based precipitation retrieval: Significance of consistency between model and measurements manifolds. *J. Atmos. Sci.*, **55**, 1644-1673.

Rabier, F., Järvinen, 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. *Q. J. Roy. Meteor. Soc.*, **126**, 1143–1170.

Smith, E.A., Lamm, J.E., Adler, R., Alishouse, J., Aonashi, K., Barrett, E., Bauer, P., Berg, W., Chang, A., Ferraro, R., Ferriday, J., Goodman, S., Grody, N., Kidd, C., Kniveton, D., Kummerow, C., Liu, G., Marzano, F., Mugnai, A., Olson, W., Petty, G., Shibata, A., Spencer, R., Wentz, F., Wilheit, T., and Zipser, E. 1998: Results from WetNet PIP-2 Project. *J. Atmos. Sci.*, **55**, 1483–1536.

Tiedtke, M. 1989: A comprehensive mass flux scheme for cumulus parameterization in large scale models. *Mon. Wea. Rev.*, **117**, 1779–1800.

Tiedtke, M. 1993: Representation of clouds in large–scale models. *Mon. Wea. Rev.*, **121**, 3040–3061.

Treadon, R.E. 1997: Assimilation of satellite derived precipitation with the NCAP GDAS. Ph.D. dissertation, The Florida State University, 348 pp.