TECHNICAL MEMORANDUM

429

The assimilation of SSM/I and TMI rainfall rates in the ECMWF 4D-Var system

Jean-François Mahfouf[1], Peter Bauer and Virginie Marécal[2]

Research Department

[1] Meteorological Service of Canada, MRB/RPN, 2121 Trans-canada Highway, Dorval, QC, H9P 1J3, Canada
[2] LPCE/CNRS, 3A, Avenue de la Recherche Scientifique, 45071 Orléans cedex 2, France

Submitted to Quarterly Journal of the Royal Meteorological Society

November 2003

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

Series: ECMWF Technical Memoranda

A full list of ECMWF Publications can be found on our web site under:
http://www.ecmwf.int/publications/

Contact: library@ecmwf.int

Abstract

A recent version of the European Centre for Medium Range Weather Forecasts (ECMWF) four-dimensional variational (4D-Var) assimilation system (40 km horizontal resolution with a 12-h window) is used to examine the comparative impact of including satellite derived rainfall rates from SSM/I and TMI radiometers within the tropics. The methodology is similar to the one proposed by Marécal and Mahfouf (2002) where total column water vapour retrievals in rainy areas from a simplified 1D-Var assimilation are introduced in the 4D-Var system. An improved methodology for the estimation of rain rate retrieval errors proposed by Bauer *et al.* (2002) is used. Three one-month experiments are undertaken: a Control run (no rain rate assimilation), a TMI run (assimilation of TMI derived rain rates) and a SSM/I run (assimilation of SSM/I derived rain rates). The impact of TMI and SSM/I assimilations is positive on forecast scores both in the extra-tropics and in the tropics. Results from SSM/I run show a larger positive impact which tends to demonstrate that the increased number of data from SSM/I with respect to TMI is more beneficial to the assimilation than more accurate data.

1 Introduction

Most global operational data assimilation systems use three-dimensional or four-dimensional variational (3D/4D-Var) techniques to provide the initial conditions of Numerical Weather Prediction (NWP) models. Such systems are flexible enough to extract efficiently the information contained in asynchroneous satellite data. As a consequence, there has been significant improvements in forecast skill scores during the recent years particularly in the Southern Hemisphere (Simmons and Hollingsworth 2001). A limitation of present day systems is to restrict the assimilation of satellite data to clear sky regions for infra-red radiances and to non-precipitating areas for microwave radiances. Recently, various studies have been initiated to evaluate the potential of assimilating satellite data in cloudy and rainy regions within 3D/4D-Var systems (Chevallier *et al.* 2002, Janisková *et al.* 2002, Moreau *et al.* 2003). A major difficulty associated with this exercise is the complexity of the observation operator defining the model equivalent of a satellite radiance (or of a geophysical product). This operator is usually strongly non-linear with thresholds associated with the triggering of microphysical processes that describe cloud and rain physics (condensation/evaporation, auto-conversion, sedimentation). Additional uncertainties are associated with specifying the shape, the density and the distribution of hydrometeors (these aspects being particularly relevant when modeling cloudy and rainy radiances). Recently, Marécal and Mahfouf (2003) have shown that these non-linearities can prevent a direct assimilation of rainfall rates in the European Centre for Medium Range Weather Forecasts (ECMWF) 4D-Var (using the incremental formulation proposed by Courtier *et al.* (1994)) from being successful. Therefore, alternative methodologies have to be defined for the time being. Marécal and Mahfouf (2000, 2002) have proposed an approach (defined as “1D-Var+4D-Var”) for the assimilation of rain rates that is compatible with an operational implementation at ECMWF. They use a 1D-Var to adjust moisture and temperature profiles from the ECMWF model in order to produce a rainfall rate from the moist physics (mass-flux convection scheme and large-scale condensation) closer to an observed value. Then, 1D-Var Total Column Water Vapour (TCWV) retrievals are considered as new observations and are assimilated in 4D-Var. Such methodology has produced improved tropical analyses and reduced some biases on the hydrological cycle. Rain rate products were derived from the microwave imager TMI on-board the Tropical Rainfall Measuring Mission (TRMM). Despite the longer than expected life time of TRMM, rain rate products derived from SSM/I on-board operational Defense Meteorological Satellite Program (DMSP) satellites offer a longer temporal continuity. One advantage of SSM/I vs. TMI is a global coverage (instead of only the tropics) and an improved spatial coverage (two satellites with a wider swath). On the other hand, the lower orbit of the TRMM satellite together with the 10 GHz additional frequency on TMI makes the rain retrievals more accurate than with SSM/I (reduction of beam filling effects and of brightness temperature saturation at high rain intensities).

The objective of this paper is to evaluate within a recent version of the ECMWF 4D-Var system the trade-off between the accuracy and the number of rainfall rate observations, in terms of quality of analyses and forecasts. Some considerations about the chosen methodology are given in Section 2. A central aspect of this study concerns the definition of rain rate retrieval errors. It is described in Section 3 based on the work of Bauer *et al.* (2002). A slightly modified version of the 1D-Var is presented in Section 4 to accomodate more efficiently the large number of SSM/I rainfall rate observations. Results from three 4D-Var assimilations are discussed in Section 5. A final discussion on the major results from this study and their relevance to the future Global Precipitation Mission (GPM) is given in Section 6.

2 Methodology suitable for an incremental 4D-Var

2.1 General considerations

Over the last twenty years, many methods have been developed to improve global atmospheric analyses (particularly in the tropics) using information from satellite-derived rain rates. They were mostly based on empirical inversions of simple convection schemes to retrieve moisture profiles and diabatic normal mode initialization to adjust the divergent circulation to satellite-derived heating rate profiles (Krishnamurti *et al.* 1984, Heckley *et al.* 1990, Puri and Miller 1990). With the advent of variational data assimilation systems it becomes possible to include rainfall rates like any other kind of observation. Various studies have been published in the literature but never in a context close to operational applications (Zupanski and Mesinger 1995, Zou and Kuo 1996, Fillion 2002). So far, the only methodology suitable for a global operational incremental 4D-Var has been proposed by Marécal and Mahfouf (2000, 2002) using the “1D-Var+4D-Var” approach described in the introduction. It is important to recognize that the variational assimilation of rainfall rates requires linearised versions of moist physical processes, but that their usefulness can be questionable in that context (Errico and Reader 1999, Fillion and Bélair 2003). The existence of pathologies where switches in the convection/large-scale precipitation schemes are activated can have a detrimental effect on the convergence of the minimization of the variational problem. The “1D-Var+4D-Var” approach represents a compromise where potential problems are first identified through a series of quality controls in 1D-Var and where only the information from non-problematic profiles is introduced in 4D-Var. This methodology could also be suitable for the assimilation of radiances in rainy areas even though the observation operator would be more complex (including both moist physical processes and a radiative transfer scheme). We would like to emphasize that given the state of the art of current variational data assimilation systems, satellite-derived rain rates and moist physical parameterizations, it is already possible to extract some useful information for improving the initial state of NWP models and their resulting forecasts.

2.2 Specific features of the ECMWF system

Experiments reported hereafter have been performed using a recent version (Cycle 23R4) of the ECMWF incremental 4D-Var (Rabier *et al.* 2000, Mahfouf and Rabier 2000, Klinker *et al.* 2000) where the observation departures are computed with a high resolution non-linear model having a T_L511 spectral resolution (corresponding to a physical grid of about 40 km) and the minimization is done with a low resolution linearised model having a T_L159 spectral resolution (corresponding to a physical grid of about 120 km). Both models have 60 vertical levels. The assimilation window is 12-h (Bouttier 2001) and the major data assimilated are both conventional (surface stations, upper air radiosondes) and derived from various satellites (cloud wind motions from geostationary satellites, water vapour and surface wind speed from SSM/I radiances, raw radiances from AMSU-A on board NOAA polar orbiting satellites). Data assimilation experiments reported in Section 5 include all the above data plus satellite-derived rain rates from SSM/I or TMI.

3 Satellite-derived rain rate products

3.1 Retrieval technique

The method used for the retrieval of near–surface precipitation from SSM/I and TMI brightness temperatures applies the Bayesian retrieval methodology (Evans *et al.* 1995) because the inversion is largely underconstrained in the case of rainfall. A–priori databases are constructed from combined cloud resolving model–radiative transfer simulations that were evaluated against observed brightness temperatures using a simple EOF–analysis (Bauer 2001). This quality control reduces the database to those situations whose statistical distribution matches best an independent set of observations.

Instead of rainfall rates, the retrieval of water contents is preferred because of the uncertain dependence of drop terminal fall velocities on drop size. The algorithm generates estimates of rain liquid water contents at 2 km altitude (averaged over a 1 km altitude range), its error, and both integrated rain–water and ice paths. In contrast to the original concept (Bauer *et al.* 2001), a fixed retrieval altitude (2 km) was chosen and the resolution of the products was changed to that of the 37.0 GHz channels (see Table 1) to avoid overlap of adjacent samples. The latter presents advantages when spatially averaged retrievals and retrieval errors are generated for climatologies or data assimilation purposes (Bauer *et al.* 2002).

The databases are adjusted to either TMI or SSM/I technical specifications applying the spatial resolution of each instruments' microwave channels, respectively. Since the uncertainties of cloud and radiative transfer modelling are most expressed at higher microwave frequencies, the algorithms were defined for each sensor without using the 85 GHz channels.

The algorithm for TMI refered as PATER (Precipitation Radar (PR) adjusted TMI estimation of rainfall) works as described in Bauer *et al.* (2001). The rain water contents from TMI have been calibrated from PR estimates (TRMM product 2A25) averaged at the TMI product spatial resolution employing idealized antenna gain function over the common swath of the two instruments. A similar principle is applied to simulate SSM/I–type observations using TMI Brightness Temperatures (TB's). This approach has two major advantages: (1) the mismatch between DMSP and TRMM overpasses in time and space is avoided and therefore much larger datasets can be collected for the comparison; (2) the calibration of SSM/I retrievals with PR estimates can be carried out as for the TMI. If, after calibration, differences between globally averaged estimates from TMI and SSM/I remain, these have to be due to sampling differences rather than algorithm differences.

As evident from Table 1, the TMI sensor has better spatial resolution and sampling than the SSM/I. This allows the spatial averaging of TMI TB's with SSM/I antenna gain functions and TMI sampling. Figure 1 illustrates this procedure for 19.35, 37.0, and 85.5 GHz, respectively, and Figure 2 shows an example of original TMI and simulated SSM/I TB's for a tropical storm that was observed in the Indian Ocean on 25 January 2002. The reduction of spatial resolution by a factor of ≈ 4 is expressed by much smoother features in the SSM/I TB–distributions. Consequently, the dynamic range of the retrieved rain rates will be reduced and small–scale features will be missing with this radiometer. The coverage of the inner swath by PR observations with about 5 km resolution and 5 km sampling that is contiguous coverage. Given the Effective Field Of View (EFOV) dimensions from Table 1, about 10 PR samples are averaged for comparison with a TMI retrieval while about 50 PR observations are compared to an SSM/I retrieval.

The calibration of both PATER–TMI and PATER–SSM/I rain rate retrievals is carried out as follows:

1. 64 cases were selected which represent a large variety of synoptic situations over all tropical latitudes during a 8-month period.
2. PATER-TMI and PATER-SSM/I rain rates are retrieved and compared to PR–retrievals that were aver-

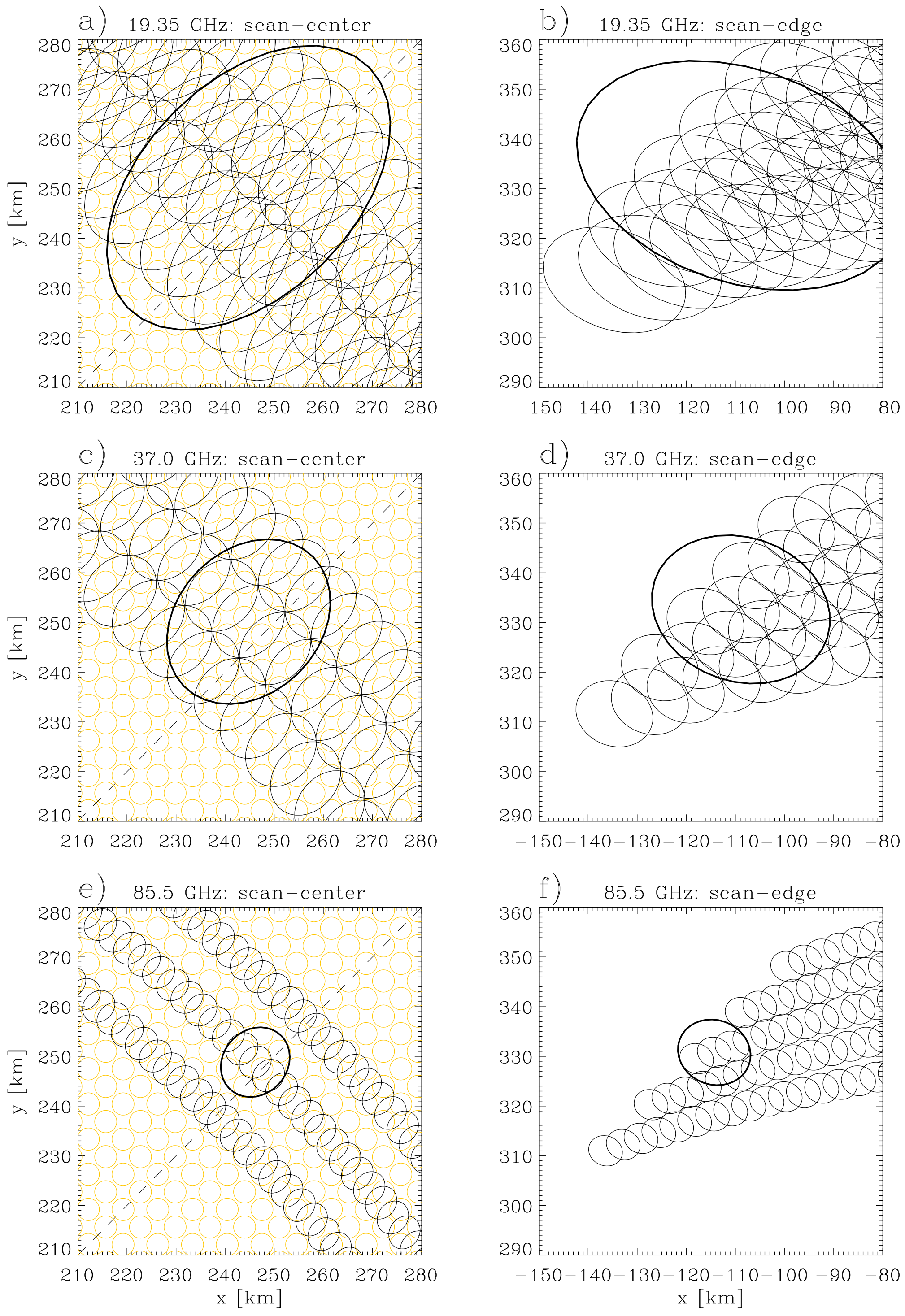

Figure 1: Demonstration of synthetic SSM/I antenna patterns (thick black, solid) overlayed on TMI antenna patterns (thin black, solid) at 19.35 GHz (a,b), 37.0 GHz (c,d), and 85.5 GHz (e,f). PR field–of–views are indicated as light grey circles. Left/right pannels show scan geometry at scan–center/scan–edge.

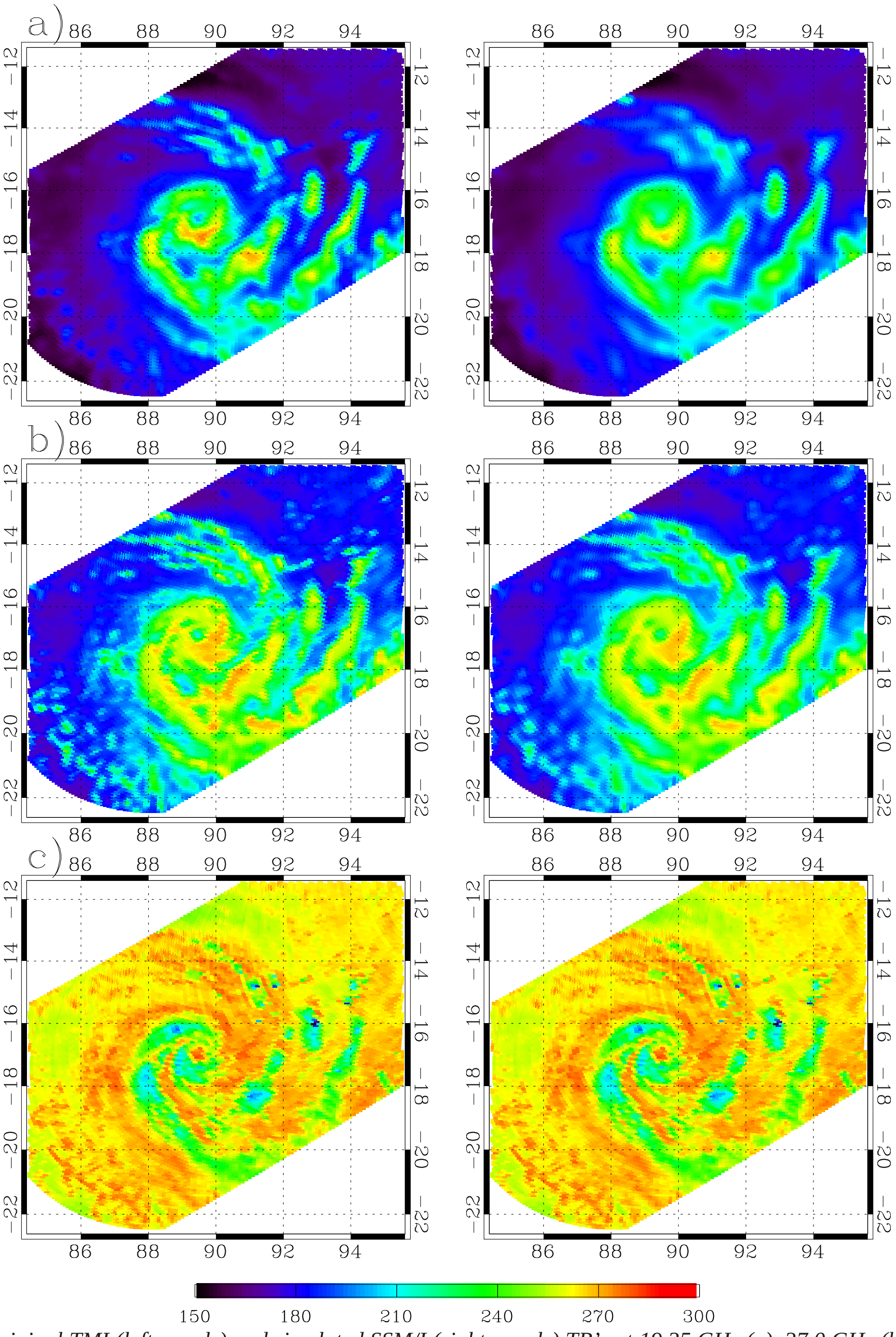

Figure 2: Original TMI (left panels) and simulated SSM/I (right panels) TB's at 19.35 GHz (a), 37.0 GHz (b), and 85.5 GHz (c). All TB's are at horizontal polarization.

Table 1: Specifications of SSM/I, TMI, and PR sensors. Note that figures represent effective field–of–view (EFOV) dimensions.

Sensor	Channel [GHz]	Spatial resolution (cross–track)x(along–track) [km]	Spatial sampling [km]
SSM/I	19.35	43 x 69	25
	22.235	40 x 50	25
	37.0	29 x 37	25
	85.5	13 x 15	12.5
TMI	10.65	39 x 63	10
	19.35	20 x 30	10
	21.3	18 x 27	10
	37.0	13 x 16	10
	85.5	6 x 7	5
PR	13.8	4 x 4	5

aged to the resolution of each radiometer product.

3. For discrete bins in rain liquid water content, as obtained from the radiometer algorithms, the probability distribution functions (pdf) of ratios between radiometer and PR liquid water contents were generated.

The proposed technique works successfully in removing a large part of the bias between PATER and PR estimates. Both retrievals from the passive instruments overestimate the rain water content deduced from the radar except for very small values around 0.01 gm^{-3} where the agreement between the passive and active retrievals is reasonable (Figure 3). The improved accuracy of the TMI retrievals clearly shows up within the range 0.01 to 0.1 gm^{-3} where the bias correction factor is half the value applied to SSM/I retrievals. The calibration curve of TMI vs. PR presented by Bauer *et al.* (2001) indicated mostly an underestimation of TMI retrievals. The difference with the present curves comes from the fact that samples with very extreme differences between PR and TMI-SSM/I retrievals have been excluded therefore producing somewhat smoother curves. In very heavy rain, the PR sometimes produces unrealistic features that have been filtered out. However, due to the non-linear nature of rain liquid water distributions from any retrieval source, any calibration cannot not completely remove biases.

3.2 Estimation of retrieval errors

Rain water contents w (g m^{-3}) produced by the PATER algorithms are converted in rain rate R (mm h^{-1}) from a relation proposed by Bauer *et al.* (2002):

$$R = 20.95 \times w^{1.12} \quad (1)$$

assuming a Gamma dropsize distribution with a shape parameter of one. Each rain rate R_i has a corresponding accuracy σ_i obtained from the Bayesian retrieval error (Bauer *et al.* 2002). Averaged rain rates have been obtained as in Marécal and Mahfouf (2000) by binning each observation within model grid boxes (on the reduced Gaussian grid of the non-linear model in order to avoid spatial interpolations of temperature T and specific humidity q profiles). The estimation of averaged rain rate error must include the spatial correlation of this quantity within each model grid box. As shown by Bauer *et al.* (2002) the standard deviation of a mean

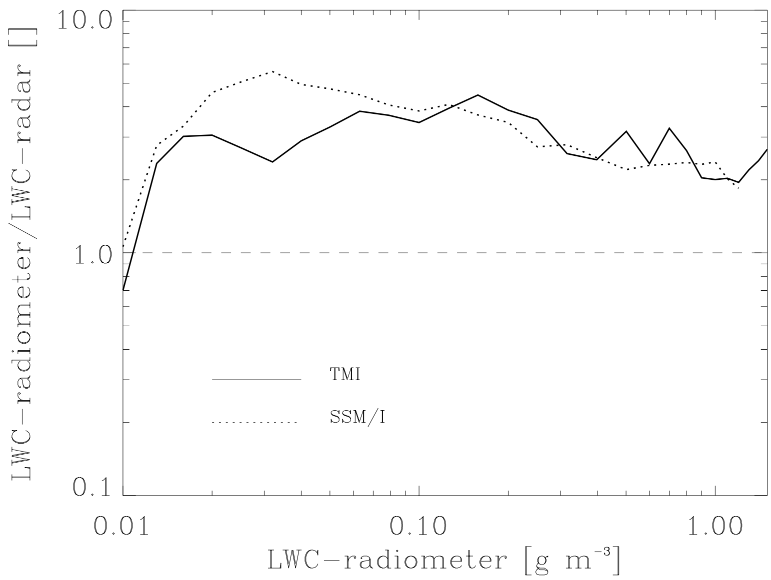

Figure 3: Average calibration factor as a function of rain water content for the TMI (solid curve) and SSM/I (dotted line) retrievals

rain rate $\overline{R}$ defined by :

$$\overline{R} = \frac{1}{N}\sum_{i=1}^{N} R_i \tag{2}$$

writes:

$$\sigma_{\overline{R}} = \frac{1}{N}\left[\sum_{i=1}^{N} C(i,j)\sigma_i\sigma_j\right]^{1/2} \tag{3}$$

where $C(i,j)$ is the spatial correlation of errors between two points i and j separated from a distance d_{ij}.

Therefore, before computing the averaged rain rates and their associated errors, it is first necessary to use a sample of satellite retrievals in order to compute $C(i,j)$. On a Gaussian grid associated with a T_L159 truncation all pairs of rainy points within each model grid box were gathered with their associated errors. The use of a coarser grid than the actual interpolation grid (T_L511) allows the computation of correlations over distances up to 120 km. With TMI, even though the pixel resolution is about 35 km, the sampling is about 12 km which gives a very good spatial sampling of d_{ij}. Distances are binned in 5 km categories for TMI to compute spatial correlations. For SSM/I this distance is increased to 25 km because the satellite has a much coarser sampling. The polynomial fit used in Bauer *et al.* (2002) is kept (with different coefficients):

$$C(r) = 1 + a_0 \times r + a_1 \times r^2 + a_2 \times r^3 \tag{4}$$

Values obtained from the regression fit are given in Table 2 and the curves $C(r)$ are plotted for TMI and SSM/I in Figure 4. The spatial correlations are very consistent between the two retrievals, since both TMI and SSM/I retrievals are calibrated from the PR. As mentioned by Bauer *et al.* (2002) the more rapid spatial decorrelation noticed in their study with other algorithms comes from the 85.5 GHz channel that is not considered in PATER type algorithms. It is interesting to note that the correlation at 100 km is about 0.3 whereas it was only 0.2 in results presented by Bauer *et al.* (2002). This difference comes from the use of a too small sample in their study (four times smaller than here) making the correlation at large distances less reliable. However, this difference is not important for the present study since the maximum distance between two pixels within a grid box at the chosen model resolution is only 60 km.

Table 2: Regression coefficients for the polynomial fit of the horizontal correlations of rain rate errors (Equation (4))

	a_0	a_1	a_2
TMI	-0.015715	1.39636E-4	-5.05216E-7
SSM/I	-0.0153113	1.21264E-4	-3.91943E-7

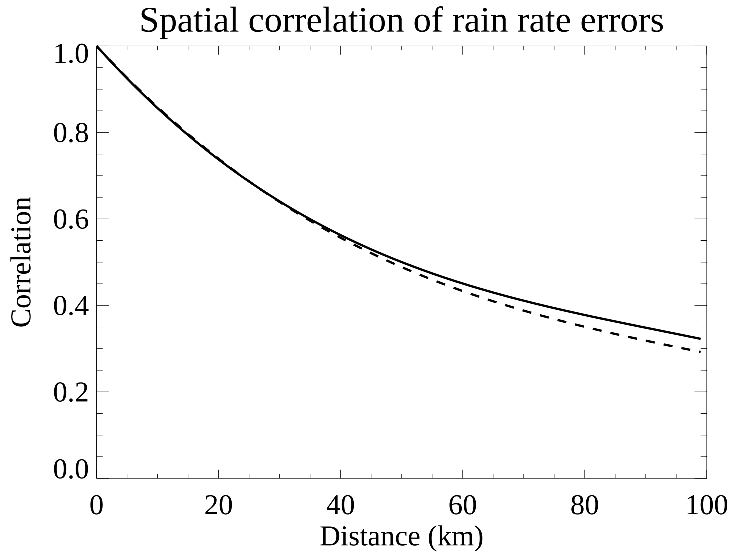

Figure 4: Spatial correlation of TMI (solid line) and SSM/I (dashed line) rainfall rate retrieval errors used to compute mean grid box errors

Two days of data have been used to compute the triplets $(d_{ij}, \sigma_i, \sigma_j)$. This is a compromise between computing time and stability of the statistics. Filters are applied to the data before retaining a mean grid box rain rate for further assimilation : when the number of rainy points within a grid box is smaller than 10 % of the total number of points, the rain rate is set to zero (the scale of the precipitating system is assumed to be too small to be adequately described by the model). We also assume a minimum number of 8 points (resp. 2) for TMI (resp. SSM/I) data to describe the averaged rain rate within each grid box (otherwise the corresponding point is discarded). This threshold depends both on the model resolution and on the spatial sampling of the data. The distribution of the number of points per grid box could be used to define more objectively these thresholds.

For a given 6-hour period (26 May 2001 : 0300-0900 UTC) the TMI and SSM/I rain rates are plotted with their associated errors both at pixel resolution (30 km for TMI and 50 km for SSM/I) and averaged on the model T_L511 resolution (Figure 5). At pixel resolution, important differences are noticed between the products. First, the lack of the 10 GHz channel and also larger pixels reduce SSM/I rain rate intensities to maximum values around 15 mmh^{-1}. On the contrary, TMI PATER produces significant rain rates above 10 mmh^{-1} with an accuracy at pixel resolution around 2 mmh^{-1}. Such already small errors are reduced to 1 mmh^{-1} after averaging at model scale. Low TMI rain rates ($<$ 1mmh^{-1}) are associated with errors larger than 100 % at pixel resolution but are significantly reduced below this threshold when averaged at model scale. Ambiguities in the cloud data base are reflected by the fact that some rain rates can be retrieved with very different accuracies. For SSM/I rain rates, the larger footprint reduces the differences between scatter plots at pixel and model resolutions. Rain rate errors are usually larger for SSM/I than for TMI and there is a non-negligible amount of very low rain rates ($<$ 0.01 mmh^{-1}), that may come from the different rejection criteria between the two products. However, it is unlikely that such small amounts could have a significant impact when assimilated due to the very large corresponding errors at model scale ($>$ 1000 %). By comparison, the distribution of model instantaneous rain

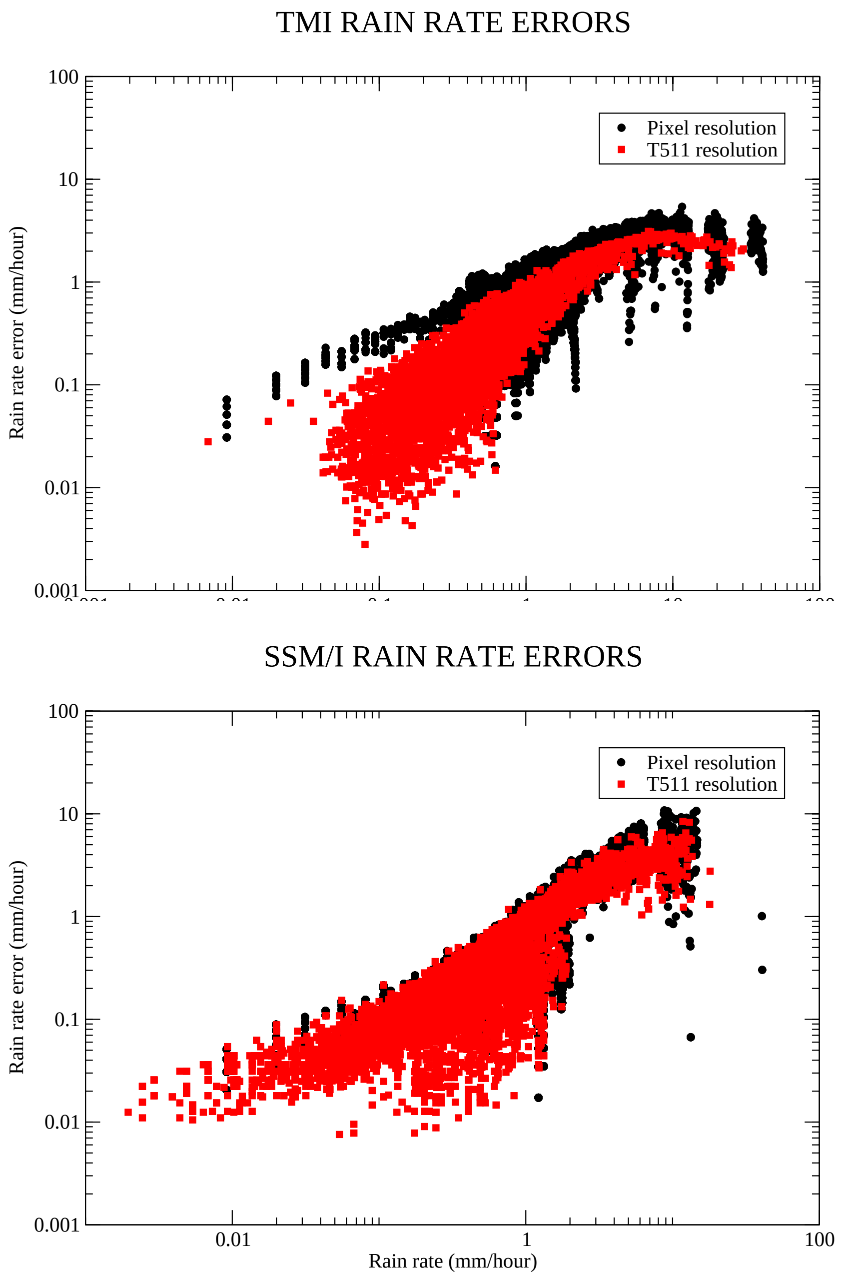

Figure 5: Errors in rain rate retrievals as a function of rain intensity at pixel resolution and model resolution (T_L511) for TMI products (upper panel) and SSM/I products (lower panel)

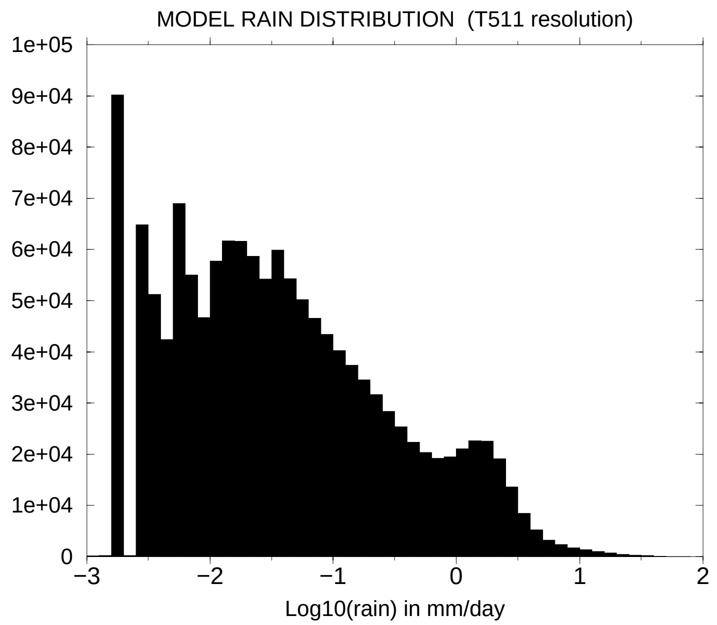

Figure 6: Instantaneous model rain rates at T_L511 resolution over tropical oceans sampled over a two-day period (25-16 May 2001)

rates over tropical oceans plotted in Figure 6 exhibits two modes. The first one peaks at a small value of 0.03 mmh^{-1}. Light model rain rates often correspond to no-rain observations where the accuracy is arbitrarily set to 0.01 mmh^{-1} for both products. The second mode is around a value of 1.5 mmh^{-1}, where it is interesting to note that the accuracy of the two products is quite similar. From these remarks one can anticipate that the important differences noticed between SSM/I and TMI errors at high and low rain rates could have a small impact on the behaviour of the 1D-Var assimilation.

4 A simplified 1D-Var approach for TCWV retrievals

Following an initial proposal of Fillion and Errico (1997), Marécal and Mahfouf (2002) defined a 1D-Var problem where temperature and specific humidity profiles x are modified to improve the model fit to an observed rain rate R_o. This problem is solved by searching for the minimum of the following cost function :

$$J(x) = \frac{1}{2}(x - x_b)^T \mathbf{B}^{-1}(x - x_b) + \left[\frac{R(x) - R_o}{\sigma_o}\right]^2 \tag{5}$$

where x_b is a background profile (short-range forecast), $\mathbf{B}$ the associated error covariance matrix and R the non-linear observation operator.
The gradient of J writes :

$$\nabla J(x) = \mathbf{B}^{-1}(x - x_b) + \mathbf{R}^{\mathrm{T}}\left[\frac{R(x) - R_o}{\sigma_o^2}\right] \tag{6}$$

The operator $\mathbf{R}^{\mathrm{T}}$ can be evaluated in finite differences as in Marécal and Mahfouf (2000) which is expensive or can be obtained from the adjoint coding of the moist physical processes which is rather tedious. The method proposed herafter allows to find an approximate solution of the 1D-Var problem at a reduced cost without requiring to code the adjoint of the convection and large-scale precipitation schemes.

If the observation operator R that computes the surface rainfall rate from atmospheric profiles x can be approximated as:

$$R(x) \simeq R(x_b) + \mathbf{R}(x - x_b) \qquad (7)$$

then the analysis x_a (i.e. profiles of T and q) defined by $\nabla J(x_a) = 0$ can be written as:

$$x_a = x_b + \mathbf{BR}^{\mathrm{T}}(\mathbf{RBR}^{\mathrm{T}} + \sigma_o^2)^{-1}[R_o - R(x_b)] \qquad (8)$$

It is worth noting that the matrix inversion reduces to a scalar since there is only one observation. The various quality controls proposed by Marécal and Mahfouf (2000) in order to reject unsuitable profiles both a priori and a posteriori can still be applied.

This technique has been first compared to the 1D-Var over a number of profiles. A vertical profile from a 31-level version of the ECMWF model chosen in the study of Fillion and Mahfouf (2000) is used for a first comparison of the two methods. Profiles of T and q are modified in order to produce a model rain rate closer to a simulated observation taken as 1.5 times the background value with an associated error of 25 % the background rain rate. The operational ECMWF mass-flux scheme (Tiedtke 1989, Gregory *et al.* 2000) defines the observation operator $R(x)$ of the 1D-Var problem. The convergence with the 1D-Var is fast (4 iterations), whereas by definition the simplified approach (named “OI” since it has some similarities with an optimum interpolation) provides the solution in one iteration. Specific humidity increments are shown in Figure 7. Some differences between the 1D-Var and the OI can be noticed, even though the vertical structure is very similar (with a first maximum where the variance of the forecast errors is largest and another maximum near the surface associated with the maximum of the rainfall Jacobian of the convection scheme). The TCWV increment (which is the information given to the 4D-Var) is slightly larger with the OI than with the 1D-Var (1.14 kgm^{-2} vs. 0.99 kgm^{-2}). In theory, if the convection scheme was linear, the solution provided by the two methods should be the same (and the 1D-Var should converge in one iteration). When the 1D-Var converges, values obtained after 2 iterations are usually very close to the final ones. The main reason is that during the first iteration, the initial conditions given to the minimizer M1QN3 (Quasi-Newton algorithm developed by Gilbert and Lemaréchal (1989)) are such that the step in the direction of the gradient is rather small in order to avoid convergence problems at the beginning of the minimization.

The assimilation techniques have been compared on a larger sample using a T_L159L60 version of the ECMWF model and TMI 2A12 rain rates averaged at that resolution over a 6-hour period (02 Apr 2001 from 0300 UTC to 0900 UTC). The main results are summarized in Table 3. The number of observations to be introduced in the 4D-Var is about 10 % smaller with the OI than with the 1D-Var. Since the a-priori quality control (QC) defined by Eq. (5) and (6) from Marécal and Mahfouf (2000) is unchanged using the OI instead of the 1D-Var, differences come from the a-posteriori QC where the analyzed rain rate R_a is assumed to be bounded by $|R_a - R_o| < \sigma_o$. Most problematic profiles correspond to situations where the assimilation should decrease the model rain rate but the increments produced by the OI are too large thereby switching off completely model precipitation. After assimilation, the statistics of observation departures (bias and standard deviation) are very similar between the two techniques. The most important analyzed quantity to examine is TCWV since these retrievals are to be included in 4D-Var. As already mentioned by Marécal and Mahfouf (2002), the mean TCWV increments are negative which reflects atmospheric situations where the model produces surface rainfall when none is observed. In such situations, the OI tends to dry more the atmosphere than the 1D-Var. However the “no-rain” information is ambiguous for retrieving an atmospheric water vapour profile whatever the method. The standard deviation of TCWV increments is around 1.5 kgm^{-2} with both techniques (such values are also consistent with results from Marécal and Mahfouf (2002)). Finally the most striking result between the two techniques is the gain in computing time by almost a factor of 10 brought by the OI with respect to the 1D-Var.

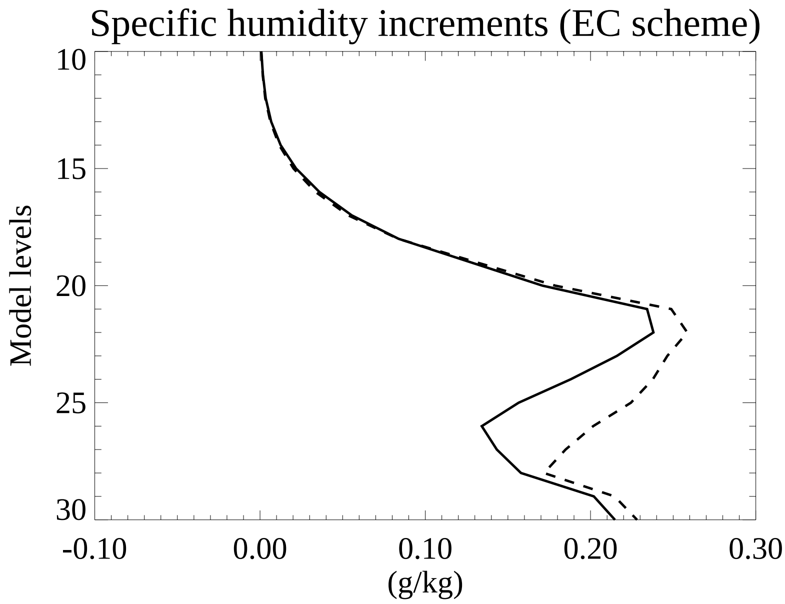

Figure 7: Increments of specific humidity produced by an assimilation of a simulated rainfall rate observation (=1.5 x background value) using the ECMWF mass-flux convection scheme. The solid lines correspond to an assimilation using the 1D-Var technique and the dashed lines correspond to an assimilation using the OI technique. The atmospheric profile from the ECMWF model is the same as in Fillion and Mahfouf (2000)

Table 3: Comparison between 1D-Var and OI assimilation techniques for a T_L159L60 version of the ECMWF model over a 6-h time window (02 Apr 2001 from 0300 UTC to 0900 UTC). Statistics are presented in terms of differences between observations (O) and both the background (B) and the analysis (A).

Method	CPU (sec)	Number of obs.	Mean rain rate (O - B) [mmh^{-1}]	Std rain rate (O - B) [mmh^{-1}]	Mean rain rate (O - A) [mmh^{-1}]	Std rain rate (O - A) [mmh^{-1}]	Mean TCWV increments [kg/m^2]	Std TCWV increments [kg/m^2]
OI	190	747	-0.268	0.733	0.0104	0.111	-0.719	1.575
1D-Var	1415	871	-0.278	0.757	0.0138	0.099	-0.423	1.542

Results presented in this section showed that a simplified 1D-Var approach (defined as “OI”) for retrieving specific humidity and temperature profiles in rainy areas provides very similar results compared to a comprehensive 1D-Var and brings a gain in computing time by about a factor ten. The a-posteriori QC increases slightly the number of rejected data (by about 10 %). However, since this method is proposed to make affordable the increased number of observations considered in this study, this data reduction should not affect significantly results presented hereafter.

5 Results from 4D-Var assimilations

A series of three 4D-Var assimilation experiments starting on 1200 UTC 01 May 2001 and lasting one month where performed with version CY23R4 of the ECMWF forecasting system. Simplified 1D-Var assimilations of rain rates are performed every 6 hours using observations 3 hours before and after the analysis time (and assumed to be valid at that time). Then these two batches of TCWV retrievals in rainy areas are introduced in

4D-Var as new observations. The associated errors are specified as in Marécal and Mahfouf (2002) for both SSM/I and TMI. The first assimilation is a “Control” using the operational configuration of the forecasting system at the time, the second assimilation includes TMI PATER rain rates in addition to all data from the Control, while the third experiment is similar to TMI PATER but assimilates rain rates from SSM/I instead (two satellites are used : DMSP 13 and 14). The usage of satellite derived rain rates is limited to the oceans of the tropical belt (40N / 40S) where microwave retrieval techniques are the most reliable.

5.1 Impact on analyses

First the data usage is very different than in previous 4D-Var experiments (Marécal and Mahfouf 2002, Marécal *et al.* 2002). The use of a higher model resolution increases the number of TMI data by about a factor two (2500 every 6-hours with respect to about 1200 with a T_L319 model resolution. The availability of two satellites for SSM/I retrievals increases further the number of TCWV observations up to about 6000 per 6-hour period. In comparison, the number of clear-sky TCWV SSM/I retrievals (Gérard and Saunders 1999, Marécal *et al.* 2001) is about 900 due to a sampling of the pixels along and across the satellite scans for the assimilation.

Figures 8 and 9 display the statistics of TCWV departures in rainy areas (simplified 1D-Var using rain rate observations) and in clear sky areas (1D-Var using observed brightness temperatures) over the one-month assimilation period. The mean background departures (O-B) are negative, indicating that observations in rainy areas will have a tendency to dry the atmosphere (to switch off model rain at locations where no precipitation is observed). This bias around -0.7 kgm^{-2} is similar to what was obtained with a model at lower resolution for a single cycle (Section 4) but is larger than values obtained by Marécal and Mahfouf (2002) as a consequence of the use of the OI technique instead of the 1D-Var (coming from the ambiguity of the “no rain” information). The 4D-Var assimilations lead to a reduction of the biases indicating that the model trajectory at the end of the minimisation is closer to the observations. The small increase in standard deviation already noticed in Marécal and Mahfouf (2002) is a consequence of the assimilation of TCWV as further explained below. No significant differences between SSM/I and TMI are noticed on these statistics. When examining the statistics associated with the assimilation of clear-sky TCWV noticeable differences are found. First, the mean background departures (O-B) are positive whereas they are negative in rainy areas. SSM/I clear sky observations indicate that the model is too dry and therefore tend to add humidity as shown by a reduction of the bias in the analysis departures (O-A). This somewhat contradicting information between clear-sky and rainy areas translates into larger positive values of (O-B) when TMI PATER is assimilated and even larger with SSM/I PATER. This conflict in the assimilation system (even though observations are not at the same location, at least for SSM/I) comes from the horizontal correlations imposed by the **B** matrix. The drying in rainy areas also reduces humidity in surrounding clear-sky areas (particularly because the horizontal scale is probably too large in rainy conditions). Despite an increase in the positive bias for both the background and the analysis, the assimilation has brought the model closer to observed clear sky TCWV observations for each experiment.

5.2 Global forecast scores

Forecast scores measured in Figure 10 by the anomaly correlation of the geopotential at 500 hPa in mid-latitudes over the whole month (29 cases) show a slight improvement over both hemispheres with TMI PATER, and a larger improvement in the Northern Hemisphere (NH) with SSM/I PATER but associated with a slight degradation in the Southern Hemisphere (SH). This NH improvement is detailed in the next section.

Tropical scores are presented in terms of Root Mean Square (RMS) errors for the temperature at 850 hPa and 200 hPa up to day-4 (Figure 11). Each set of forecasts is compared with respect to its own analysis. The use of TCWV in rainy areas is slightly positive to the 850 hPa scores with SSM/I PATER but slightly negative with

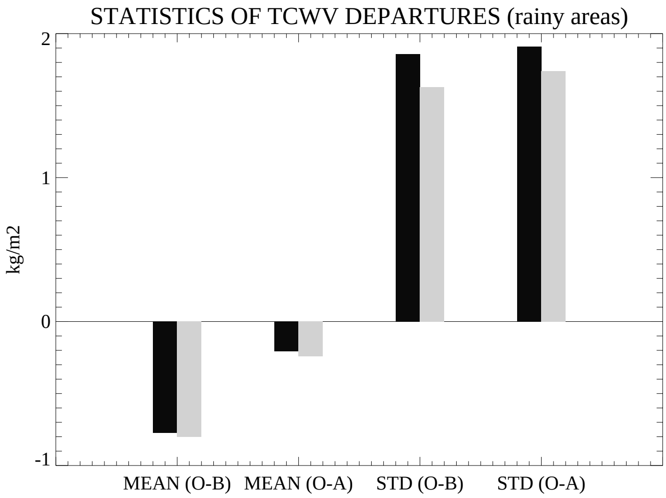

Figure 8: Global mean and standard deviation (STD) of TCWV model departures from a simplified 1D-Var retrievals in rainy areas for both the background (O-B) and the analysis (O-A). Black bars correspond to TMI rain rate assimilation and grey bars to SSM/I rain rate assimilation

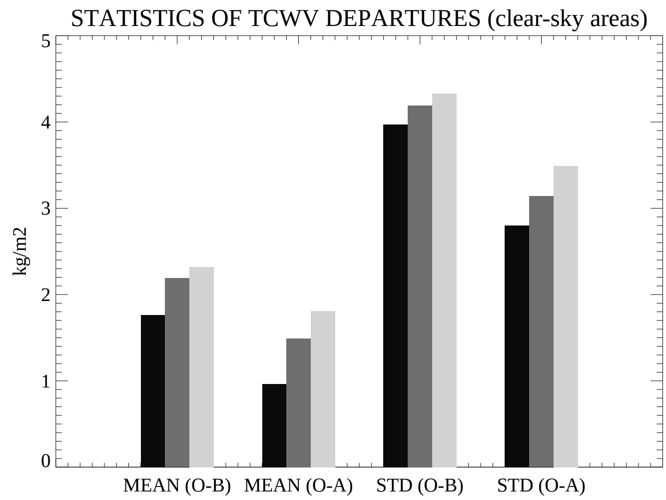

Figure 9: Global mean and standard deviation (STD) of TCWV model departures from 1D-Var retrievals in clear-sky areas for both the background (O-B) and the analysis (O-A). Black bars correspond to the "Control" assimilation, dark grey bars to TMI rain rate assimilation and light grey bars to SSM/I rain rate assimilation

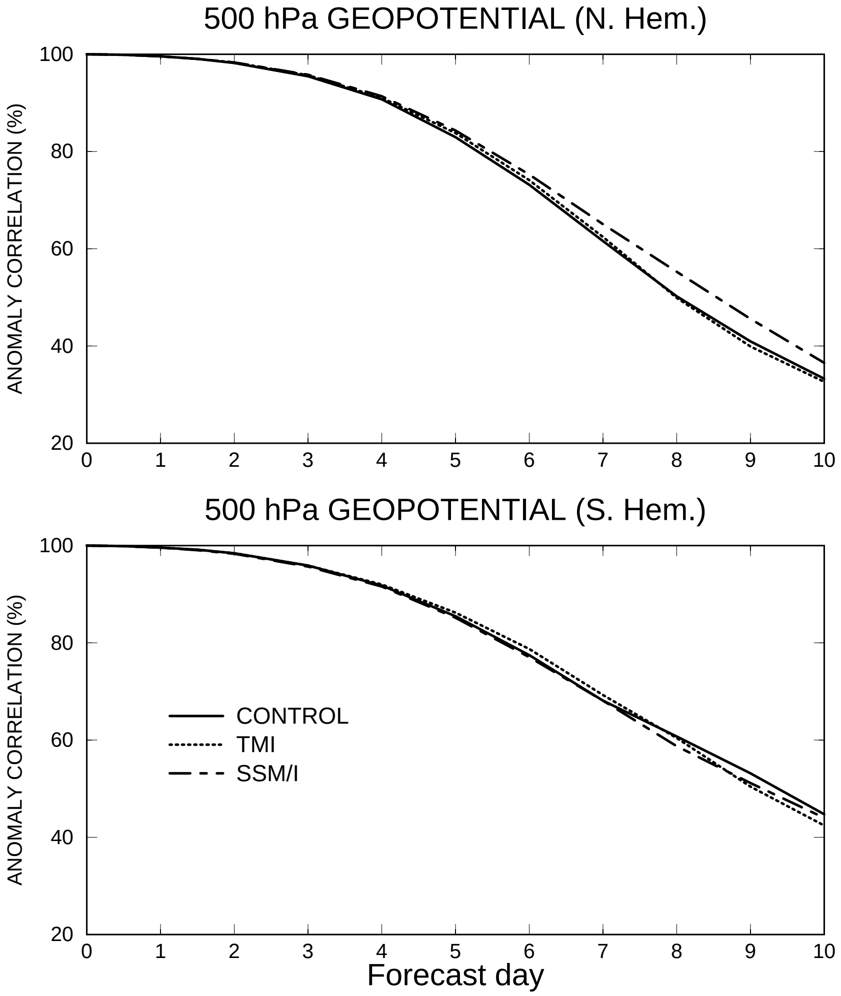

Figure 10: Anomaly correlation of the geopotential at 500 hPa averaged over 29 forecasts (1-29 May 2001) issued from 4D-Var for the Northern Hemisphere (upper panel) and the Southern Hemisphere (lower panel). The Control experiment corresponds to the solid line, the TMI experiment to the dashed line and the SSM/I experiment to the dotted line.

TMI PATER. At 200 hPa there is a noticeable reduction of the RMS error with both rain rate assimilations. This signal mostly reflects a reduction of the model bias (not shown). The intensity of moist convection has been significantly reduced during the assimilation period when considering TCWV in rainy areas. The reduction of the upper tropospheric bias shows that such information can be maintained during the whole range of the forecasts in the tropics. These results are consistent with those previously obtained by Marécal and Mahfouf (2002) at 200 hPa even though they also found a positive impact on tropical winds at 850 hPa that is not recovered in the present study.

5.3 Case study

Marécal and Mahfouf (2002) showed that the impact of assimilating TMI derived rain rates was mostly neutral for extra-tropical scores. However, in the present study, a positive impact has been noticed over the Northern Hemisphere when assimilating either TMI or SSM/I rain information. This impact mostly comes from a significant improvement of several particularly bad medium range forecasts over the US during May 2001 with the

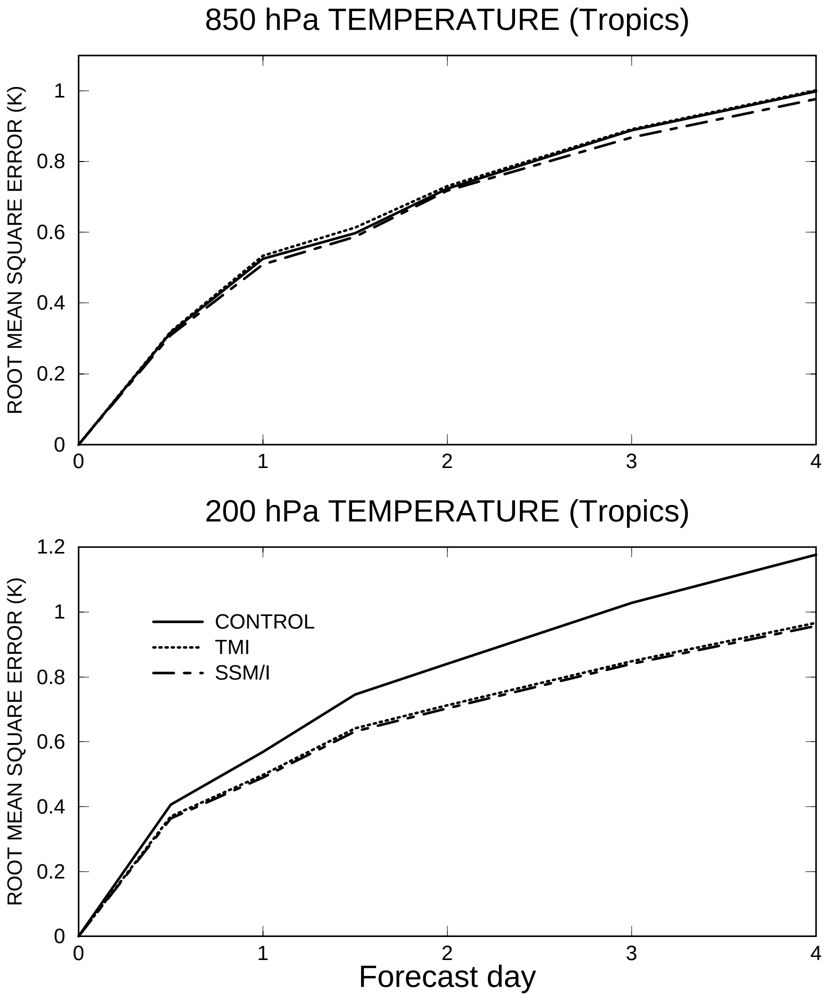

Figure 11: Root-mean square error of the temperature at 850 hPa (upper panel) and at 200 hPa (lower panel) averaged over 24 forecasts (1-24 May 2001) issued from 4D-Var in the tropics. The Control experiment corresponds to the solid line, the TMI experiment to the dotted line and the SSM/I experiment to the dash-dotted line

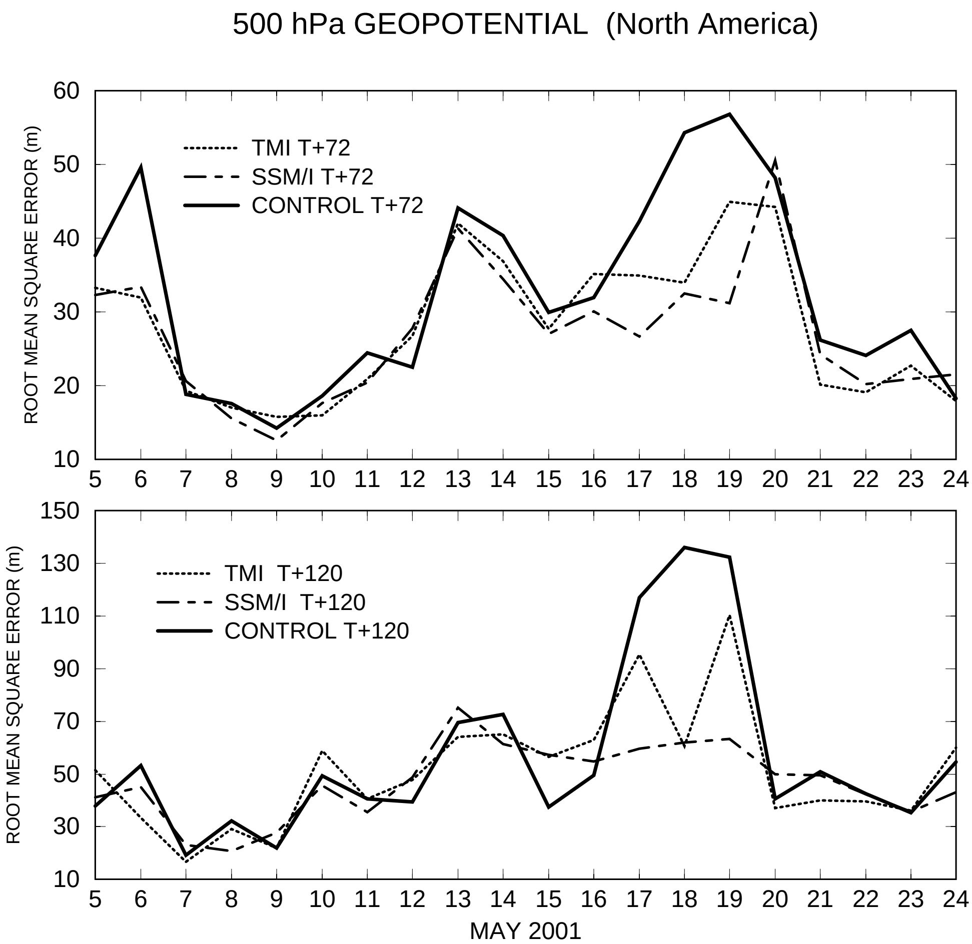

Figure 12: Time series of RMS errors of the geopotential at 500 hPa over North America for day-3 (upper panel) and day-5 (lower panel) forecasts issued from a “Control” assimilation, rom an assimilation with TMI derived rain rates and from an assimilation with SSM/I derived rain rates (5-24 May 2001)

ECMWF model. Indeed, the assimilation period was chosen for this reason. Through a number of diagnostics (A. Persson, personnal communication) the origin of the forecast errors was traced back in the initial conditions over the western tropical Pacific Ocean. It appeared that the interactions between a decaying tropical cyclone *Cimaron* and the mid-latitude westerlies played a significant role on the development of a strong cut-off low over Central US in the medium range. From the conclusions of these studies and also on impact studies performed at ECMWF on the use of QuikScat sea surface wind data (H. Herbasch, personnal communication), it was likely that if these rainy systems were sampled by TMI and/or SSM/I a significant impact could be expected on forecast scores over US.

Figure 12 shows time series of day-3 and day-5 forecasts during a 3-week period over the Northern American continent for the RMS error of the geopotential at 500 hPa. For three forecasts starting on the 17, 18 and 19 May the RMS error at day-5 is larger than 100 m. The impact of TMI data is to improve significantly the forecast starting on 1200 UTC 18 May 2001 (RMS around 60 m) and to have also a positive impact (but less pronounced) on the two other poor forecasts. The impact in the medium range is also present at shorter ranges (over this period there is an almost systematic improvement at 72 h). The impact of SSM/I rain rates in even more dramatic since the three poor forecasts are all improved with RMS errors reduced by more than a factor of two. The improvement in the short range is also larger than with TMI. The synoptic situation at 1200 UTC 23 May 2002 is displayed in Figure 13. The analysis shows a deep cut-off centered over the Great Lakes with a minimum of 5300 m. In the operational ECMWF day-5 forecast the minimum is located too far south and is not deep enough (5500 m). Moreover the axis of the trough has a north-east/south-west orientation whereas it is tilted along an axis north-west/south-east in the analysis. The forecast starting from TMI analysis produces a 500 hPa geopotential field in better agreement with the analysis (the cyclone is deeper and the location of the minimum is closer to the analysis). The forecast starting from SSM/I produces similar large scale patterns. The too strong low off the East coast of the continent produced with TMI is weaker when using SSM/I which improves the agreement with the analysis. On the other hand, the high simulated over Quebec is much weaker in the operational analysis.

6 Discussion and conclusions

Studies of Marécal and Mahfouf (2002) and Marécal *et al.* (2002) on the assimilation of satellite derived rain rates in the ECMWF 4D-Var system using a “1D-Var+4D-Var” approach have been extended to a more recent version of the system (higher horizontal resolution, longer time window) and on the use of SSM/I derived rain rates. The conclusions of Marécal and Mahfouf (2003) regarding the direct assimilation of rain rates in the ECMWF incremental 4D-Var were that, for the time being, the “1D-Var+4D-Var” approach is a more robust technique (problematic profiles are identified in the 1D-Var through a number a quality control checks). The rain rate observations being interpolated on the model grid, going from a T_L319 resolution (as in Marécal and Mahfouf (2002)) to a T_L511 resolution (as in the present study) has significantly increased the number of observations to be presented to the 4D-Var. The use of a 12-h assimilation window also requires two batches of 6-h 1D-Var to be performed before starting the 4D-Var. These computing constraints have led to the definition of a simplified 1D-Var (named “OI”) where the analysis is obtained without an iterative procedure by assuming that the linearisation of the observation operator (moist physics) around the background state is a reasonable approximate of the non-linear behaviour. Comparative studies have shown that the OI and the 1D-Var produce very similar results. When the tangent-linear approximation for the moist physics is valid, the convergence of the 1D-Var is fast and the OI provides a comparable solution. When the tangent-linear approximation breaks down, the 1D-Var fails to converge (solution close to the background) and the OI solution is unrealistic. In such situations the a-posteriori QC defined by Marécal and Mahfouf (2000) rejects the corresponding profiles before introduction in the 4D-Var. The most significant difference comes from situations where the model rain rate

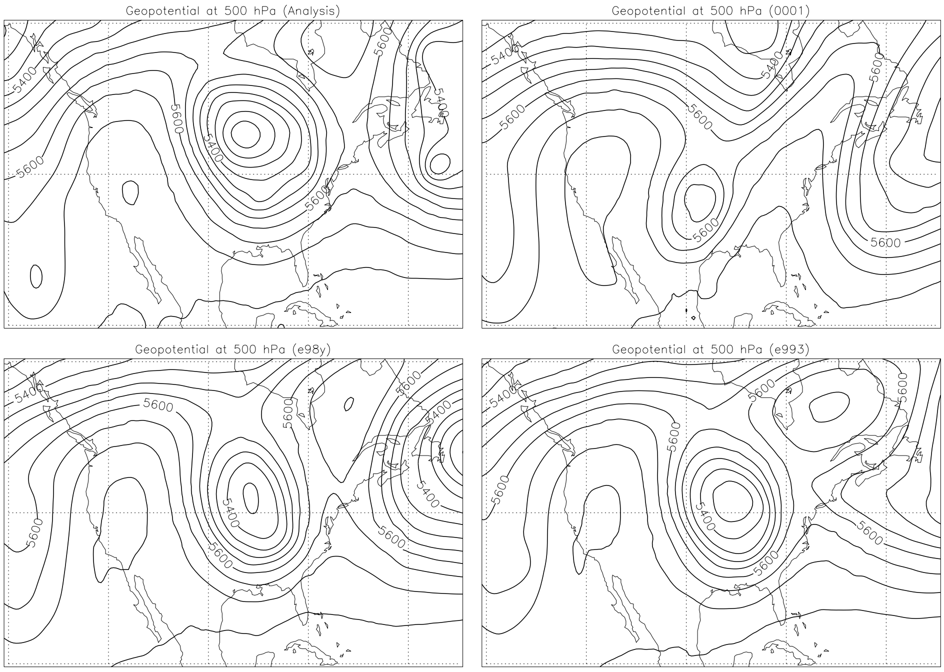

Figure 13: Geopotential at 500 hPa on 23 May 2001 1200 UTC: ECMWF operational analysis (upper left), 5-day operational ECMWF forecast (upper right), 5-day forecast from 4D-Var assimilation using TMI rain rates (lower left), 5-day forecast from 4D-Var assimilation using SSM/I rain rates (lower right)

has to be switched off (these are pathological profiles but nevertheless bring positive impacts on the analyses as demonstrated in previous studies), the OI approach provides a dryer atmospheric state than the 1D-Var. Regarding the data usage, two important aspects were modified. Following the methodology of Bauer *et al.* (2002) the specification of rain rate retrieval errors at the model grid scale has been improved by considering spatial correlations of errors within grid boxes. An improved SSM/I rain rate retrieval algorithm has been developed and calibrated from TRMM PR estimates as in Bauer *et al.* (2001). As a consequence, it has been possible to examine the comparative impact of SSM/I and TMI rain rate products on the ECMWF 4D-Var. The usage of SSM/I has been limited to the latitude band : 40 N/40 S (same coverage as TMI). The availability of two DMSP satellites has increased the number of rain rates retrievals with SSM/I by a factor of two with respect to TMI. TMI retrievals are more accurate than SSM/I particularly at high intensities ($>$ 10 mmh^{-1}). Moreover, the dynamical range of TMI rain rates is larger than with SSM/I. In terms of global impact on the analyses and forecasts the two rain rate products give similar results, but the use of SSM/I is slightly more beneficial (larger reduction of model biases and more significant improvements on case studies). To the question : is it better to have a small number of accurate rain rate observations (i.e. from TMI) or to have more observations but with a reduced accuracy (i.e. from SSM/I) ? the answer from the current study is that it is more beneficial to increase the number of observations. There are various explanations to this result. The accuracy of SSM/I and TMI are similar within a range of 0.1 mmh^{-1} to 5 mmh^{-1} which is around the second mode of the model rainfall distribution. Rather accurate TMI high intensity rain rates (coming from the usage of the 10 GHz channel) are not assimilated efficiently through the 1D-Var because the model physics at 40 km resolution can hardly produce instantaneous rain rates above 10 mmh^{-1} as shown by the model rainfall distribution (therefore quality controls tend to reject such high observed values). It appears that the global behaviour of the 4D-Var system is mostly sensitive to the “no-rain” information, that is probably as valuable with SSM/I than with TMI since the same observation error is set. Marécal *et al.* (2002) have shown that different rain rate accuracies can lead to different local impact on the associated dynamics of rainy meteorological systems but with a negligible impact on objective forecast scores. Indeed, when projecting the model background errors in the observation space, the large sensitivity of the moist processes implies that background errors are most of the time much larger than the observation errors (i.e. in the 1D-Var procedure, a strong weight is given to the observation term ; one can see that the resulting mean TCWV increments introduced in 4D-Var are very similar using either TMI or SSM/I). The impact of SSM/I is also larger probably because there is more consistency (spatial and temporal) with the assimilation of clear-sky water vapour retrievals (i.e. the rain observations can counter act the effect of "clear-sky" corrections taking place in rainy areas).

From the conclusions of this study, increasing the number of (less accurate) rain rate observations in the ECMWF 4D-Var had a positive impact on the quality of medium range forecasts. The constellation design of the future Global Precipitation Mission (GPM) is therefore highly desirable for data assimilation applications. The current limitations of the 4D-Var system both in terms of spatial resolution and description of physical processes, prevent from taking advantage of the more accurate high rainfall rates available with TMI. Therefore an additional 10 GHz channel does not appear to be a strong requirement for global data assimilation systems so far. Results from direct assimilation of microwave radiances currently undertaken at ECMWF will provide complementary conclusions to this study. Finally, further studies should focus on the impact of using rain rate retrievals over mid-latitudes and over continents.

Acknowledgements

This study was partly funded by an ESA contract in preparation to the European contribution to the Global Precipitation Mission.

7 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 data bases. *J. Ocean. Atmos. Tech.*, **18**, 1315-1330.

Bauer, P., P. Amayenc, C.D. Kummerow and E.A. Smith, 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.

Bauer P., J.-F. Mahfouf, W.S. Olson, F.S. Marzano, S. Di Michele, A. Tassa, and A. Mugnai, 2002: Error analysis of TMI rainfall estimates over ocean for variational data assimilation. *Quart. J. Roy. Meteor. Soc.*, **128**, 2129-2144.

Betts, A.K., and M.J. Miller, 1993: The Betts-Miller scheme. *The Representation of Cumulus Convection in Numerical Models of the Atmosphere, Meteor. Monogr.*, No 46, Amer. Meteor. Soc., 107-121.

Bouttier, F., 2001: The development of 12-hourly 4D-Var. ECMWF Technical Memorandum No 348.

Chevallier, F., P. Bauer, J.-F. Mahfouf and J.-J. Morcrette, 2002: Variational retrieval of cloud profile from ATOVS observations. *Quart. J. Roy. Met. Soc.*, **128**, 2511-2526.

Courtier, P., J.-N. Thépaut, and A. Hollingsworth, 1994: A strategy for operational implementation of 4D-Var using an incremental approach. *Quart. J. Roy. Met. Soc.*, **120**, 1367-1388.

Errico, R.M. and K.D. Reader, 1999: An examination of the accuracy of the linearization of a mesoscale model with moist physics. *Quart. J. Roy. Met. Soc.*, **125**, 169-195.

Evans, F., J. Turk, T. Wong, and G. Stephens, 1995: A Bayesian approach to microwave precipitation retrieval. *J. Appl. Meteor.*, **34**, 260–279.

Fillion, L., 2002: Variational assimilation of precipitation data and gravity wave excitation. *Mon. Wea. Rev.*, **130**, 357-371.

Fillion, L. and R.M. Errico, 1997: Variational assimilation of precipitation data using moist convective parameterization schemes: a 1D-Var study. *Mon. Wea. Rev.*, **125**, 2917-2942.

Fillion, L. and J.-F. Mahfouf, 2000: Coupling of moist-convective and stratiform precipitation for variational data assimilation. *Mon. Wea. Rev.*, **128**, 109-124.

Fillion, L. and S. Bélair, 2003: Tangent-linear aspects of the Kain-Fritsch moist convective parameterization scheme. *Mon. Wea. Rev.* (submitted).

Gérard, E. and R.W. Saunders, 1999: 4D-Var assimilation of SSM/I total column water vapour in the ECMWF model. *Quart. J. Roy. Met. Soc.*, **125**, 3077-3101.

Gilbert, J.-C. and C. Lemaréchal, 1989: Some numerical experiments with variables-storage quasi-Newton algorithms. *Math. Programming*, **45**, 407-435.

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

Heckley, W.A., G. Kelly, and M. Tiedtke, 1990: On the use of satellite-derived heating rates for data assimilation within the tropics. *Mon. Wea. Rev.*, **118**, 1743-1757.

Janisková, M., J.-F. Mahfouf, and J.-J. Morcrette, 2002: Preliminary studies on the variational assimilation of cloud-radiation observations. *Quart. J. Roy. Met. Soc.*, **128**, 2713-2736.

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

Krishnamurti,T.N., K. Ingles, S. Cooke, T. Kitade, and R. Pasch, 1984: Details of low-latitude, medium-range numerical weather prediction using a global spectral model. Part 2 : Effects of orography and physical initialization. *J. Meteor. Soc. Japan*, **62**, 613-648.

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

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

Marécal V., E. Gérard, J.-F. Mahfouf, and P. Bauer, 2001 The comparative impact of the assimilation of SSM/I and TMI brightness temperatures in the ECMWF 4D-Var system. *Quart. J. Roy. Met. Soc.*, **127**, 1123-1142.

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

Marécal V. and J.-F. Mahfouf, 2003: Experiments on 4D-Var assimilation of rainfall data using an incremental formulation. *Quart. J. Roy. Meteor. Soc.*, **129**, 3137-3160.

Marécal V., J.-F. Mahfouf and P. Bauer, 2002 Comparison of TMI rainfall estimates and their impact on four-dimensional variational assimilation *Quart. J. Roy. Met. Soc.*, **128**, 2129-2144.

Moreau, E., P. Bauer and F.Chevallier, 2003: Variational retrieval of rain profiles from spaceborne passive microwave observations. *J. Geophys. Res.* **doi:10.1029/2002JD003315**, ACL 11 1–18.

Puri, K. and M.J. Miller, 1990: The use of satellite data in the specification of convective heating for diabatic initialization and moisture adjustment in numerical weather prediction models. *Mon. Wea. Rev.*, **118**, 67-93.

Rabier, F., H. Järvinen, E. Klinker, J.-F. Mahfouf, and A. Simmons, 2000: The ECMWF operational implementation of four-dimensional variational assimilation. Part I: Experimental results with simplified physics. *Quart. J. Roy. Meteor. Soc.*, **126**, 1143-1170.

Simmons, A. and A. Hollingsworth, 2001 Some aspects on the improvement in skill of numerical weather prediction. ECMWF Technical Memorandum No 342.

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

Zou, X. and Y.-H. Kuo, 1996: Rainfall assimilation through an optimal control of initial and boundary conditions in a limited-area mesoscale model. *Mon. Wea. Rev.*, **124**, 2859-2882.

Zupanski, D. and F. Mesinger, 1995: Four-dimensional variational assimilation of precipitation data. *Mon. Wea. Rev.*, **123**, 1112-1127.