# The Goddard Profiling Algorithm (GPROF): Description and Current Applications

William S. Olson, Song Yang, John E. Stout, and Wei-Kuo Tao

Atmospheric scientists use different methods for interpreting satellite data. In the early days of satellite meteorology, the analysis of cloud pictures from satellites was primarily subjective. As computer technology improved, satellite pictures could be processed digitally, and mathematical algorithms were developed and applied to the digital images in different wavelength bands to extract information about the atmosphere in an objective way. The kind of mathematical algorithm one applies to satellite data may depend on the complexity of the physical processes that lead to the observed image, and how much information is contained in the satellite images both spatially and at different wavelengths. Imagery from satellite-borne passive microwave radiometers has limited horizontal resolution, and the observed microwave radiances are the result of complex physical processes that are not easily modeled. For this reason, a type of algorithm called a Bayesian estimation method is utilized to interpret passive microwave imagery in an objective, yet computationally efficient manner.

Scientists participating in NASA's Tropical Rainfall Measuring Mission program have developed a Bayesian algorithm for determining surface rainfall rate and precipitation vertical structure from satellite microwave radiometer imagery. Called GPROF (for Goddard Profiling Algorithm), it has been applied to both tropical and midlatitude radiometer data to yield maps of precipitation and precipitation vertical structure. In addition, GPROF can be used to estimate the latent heat released by the condensation of atmospheric moisture through statistical correlations between latent heating and precipitation structure. A knowledge of the distributions of atmospheric latent heating is important because latent heating produces warmer, more buoyant air, and so atmospheric vertical circulations are driven by latent heating to a large extent. The tropical Hadley Circulation is one example.

GPROF rain rate estimates agree well with independent rain estimates from both ground-based and satellite-borne radar. Over the globe, the greatest monthly rainfall is generally found in the equatorial belt of precipitation called the Intertropical Convergence Zone, where persistent, organized weather systems produce a large percentage of the rainfall and latent heating. Midlatitude weather systems are characterized by greater horizontal extent but shallower vertical precipitation and latent heating structures.

Note to NASA reviewers: we were invited to write this short paper in the form of a book chapter at the request of Eurainsat steering committee members Vincenzo Levizzani, Joseph Turk, and Peter Bauer. Eurainsat is a European Commission activity for "European satellite rainfall analysis and monitoring at the geostationary scale".

# The Goddard Profiling Algorithm (GPROF): Description and Current Applications

William S. Olson
Joint Center for Earth Systems Technology
and NASA/Goddard Space Flight Center

Song Yang
George Mason University
and NASA/Goddard Space Flight Center

John E. Stout
George Mason University
and NASA/Goddard Space Flight Center

Mircea Grecu
Goddard Earth Sciences and Technology Center
and NASA/Goddard Space Flight Center

## INTRODUCTION

The use of Bayesian estimation methods in passive microwave radiometry follows from a recognition that the total information content of radiometer observations is insufficient to determine a "unique" estimate of surface rain rate or precipitation vertical profile. In other words, for a given set of multi-frequency microwave observations at a given location, there exist several precipitation profiles that are radiatively consistent with the observations, and so iterative methods for seeking a unique solution (ref. Smith et al. 1994) would not necessarily return a better estimate. Also, being non-iterative, Bayesian methods are relatively computationally efficient, since iterative forward radiance calculations are not required.

Here, the Goddard Profiling Algorithm (GPROF) is described, and applications of the most recent implementation (Version 6) of the algorithm are presented and critiqued.

## ALGORITHM DESCRIPTION

GPROF is based upon a Bayesian technique originally described in Kummerow et al. (1996) with an extension to latent heating estimation by Olson et al. (1999). A summary of more recent developments in the algorithm can be found in Kummerow et al. (2001). In the algorithm, cloud-resolving model simulations, coupled to a radiative transfer code, are used to generate a large supporting database of simulated precipitation/latent heating vertical profiles and corresponding upwelling microwave radiances. Given a set of observed multichannel microwave radiances from a particular sensor, the entire database of simulated radiances is scanned; the "retrieved" profile is composited from those profiles in the database that correspond to simulated radiances

consistent with the observed radiances. Formally, a GPROF estimate of profile parameters, $\hat{E}[x]$, is given by

$$\hat{E}[x] = \sum_k x_k \frac{\exp\left\{-0.5\left(I_s(x_k) - I_0\right)^T\left(S_I + O_I\right)^{-1}\left(I_s(x_k) - I_0\right) + C\right\}}{\hat{M}}, \quad (1)$$

where the model profile vector $x_k$ contains all parameters, including the surface rain rate, convective rain rate, liquid/ice-phase precipitation and latent heating profiles, corresponding to the simulated radiance indices, $I_s(x_k)$. The radiance indices, constructed from radiances at the different radiometer channel frequencies/polarizations, are the normalized polarization and scattering indices defined by Petty (1994). $I_0$ is a vector of sensor observed radiance indices, similarly defined. $S_I$ and $O_I$ are error covariance matrices of the simulated and observed microwave radiance indices, respectively. Additional information regarding the observed profile, such as estimates of the area fractions of convective and stratiform rain within the nominal satellite footprint (14 km x 14 km for TMI) and the freezing level, is included in the constraint term, $C$. The summation in (1) is over all simulated profiles/radiance indices in the supporting cloud-radiative model database. $\hat{M}$ is a normalization factor.



**GPROF Algorithm**

$I_o$ observed

~10 km

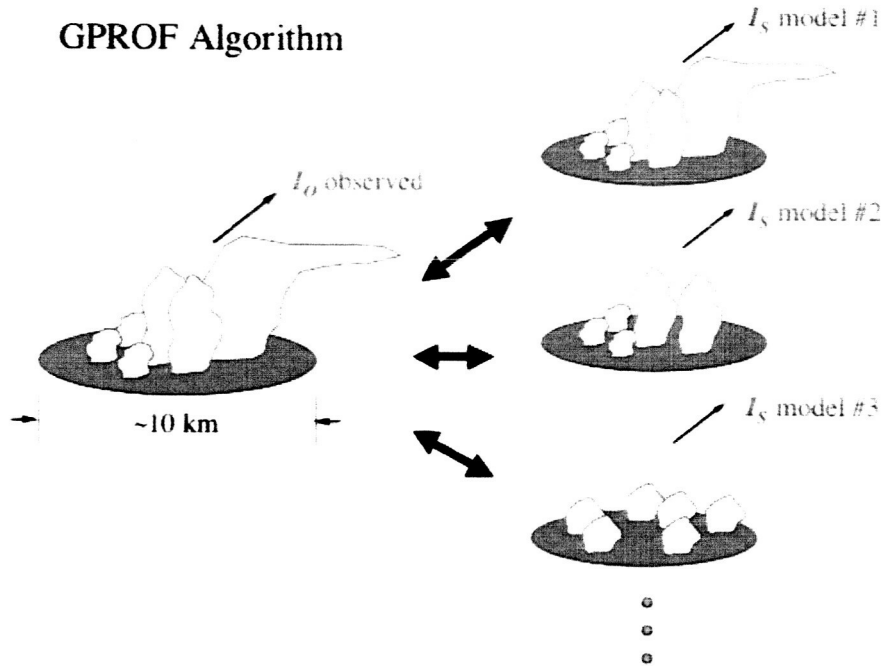$I_s$ model #1

$I_s$ model #2

$I_s$ model #3

Fig. 1. Schematic of the GPROF algorithm.

The GPROF algorithm is shown schematically in Fig. 1. Simulated radiometer footprints and calculated radiance indices of the algorithm's database are shown at right, while the observed footprint and radiance indices are at left. Observed radiance indices are compared to each set of simulated radiance indices in the database- the precipitation/latent heating parameters of those simulations that are more radiatively consistent with the observations (e.g. model #1 in the figure) contribute strongly to the GPROF estimate of the parameters, while those simulations that are less consistent radiatively contribute much less. Therefore, the algorithm performs a kind of radiative filtering of the database.

Since multichannel passive microwave observations contain limited information regarding precipitation and related parameters, there are, in fact, a distribution of these parameters that are consistent with any set of observations at a given footprint location. Equation (1) gives the mean of this distribution, but it is also possible to calculate the variance of the distribution for a single estimated parameter using

$$\hat{\sigma}^2[x] = \hat{E}\left\{\left(x - \hat{E}[x]\right)^2\right\}, \tag{2}$$

which yields a measure of the uncertainty in the estimate of $x$ due to the limited information content of the observations. The uncertainty represented by (2) would exist even if the cloud-radiative model simulations in the GPROF supporting database and the radiometer observations were perfect, and so additional uncertainties in GPROF estimates due to modeling or observational errors may occur. However, since true validation of precipitation-related quantities using independent observations is difficult, (2) at least provides a lower bound on the error of GPROF estimates- a basic "building block" for estimates of the random error in derived products. Algorithm-derived estimates of random error for two case studies will be presented in GPROF Applications, below.
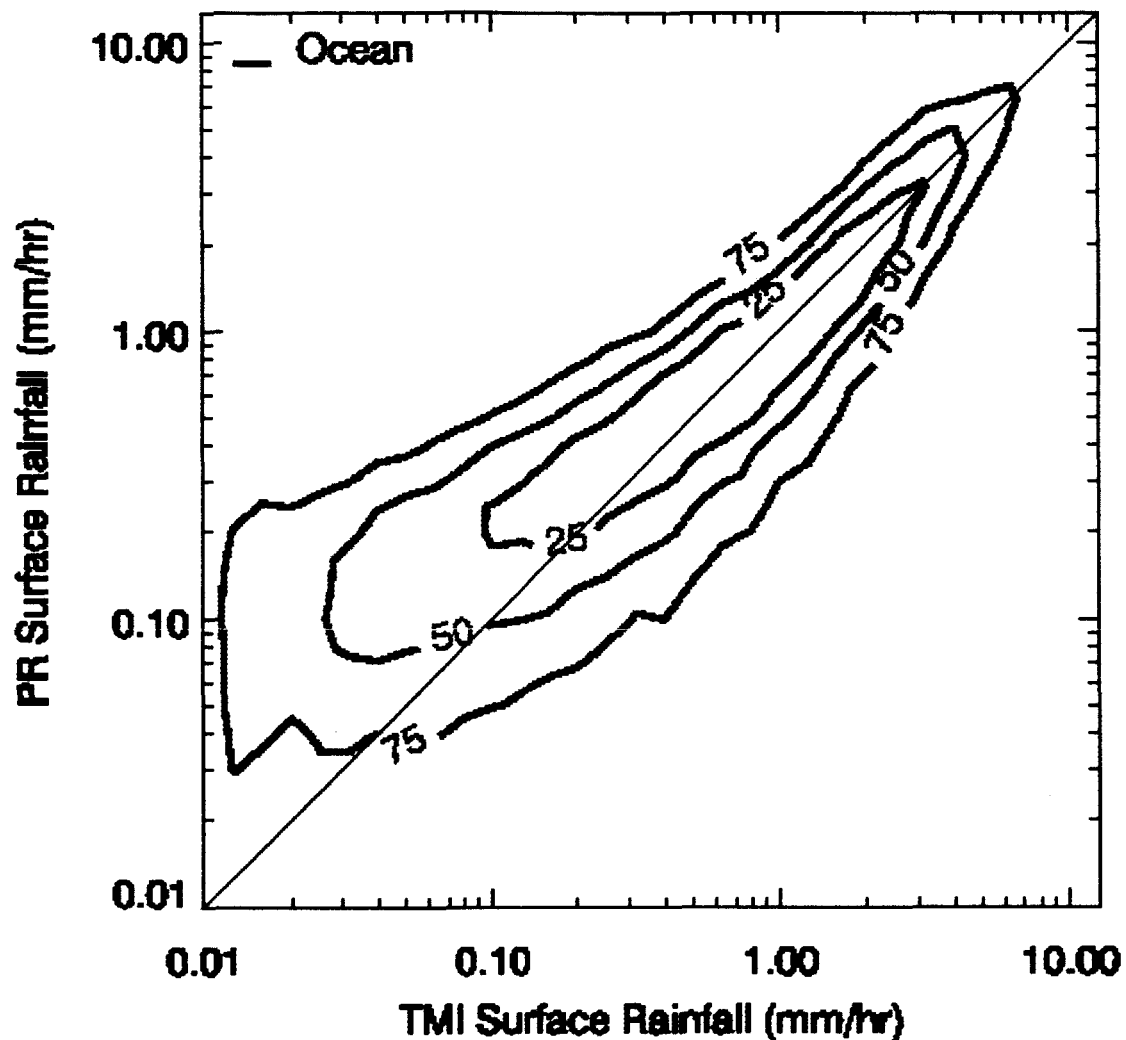
EVALUATION OF RAIN RATE ESTIMATES

Fig. 2. Scatterplot histogram of Version 6 PR and TMI instantaneous, 60 km resolution rain rate estimates over ocean for the month of July, 2001.

In collaboration with several TRMM investigators, independent estimates of rain rate, convective proportion, and latent heating were collected and compared to GPROF algorithm estimates based upon TRMM Microwave Imager (TMI) radiance observations. Although independent evaluation of surface rain rate estimates is important in its own right, surface rain rate can also be viewed as a proxy for vertically integrated latent heating, and the convective proportion is an indicator of heating profile shape. Ground validation radar estimates of rain rate and convective proportion coincident with TRMM overpasses of significant rain events were collected from the Melbourne, FL and Kwajalein (Marshall Islands) sites over an entire year (1998). Surface rainfall rate estimates from the Version 5 TMI, PR, and TMI-PR algorithms and ground-based radar measurements from Melbourne, FL and Kwajalein, Marshall Islands showed varying degrees of agreement. The Version 6 TMI estimates generally showed greater consistency with ground validation radar rain observations. The analysis yielded a

correlation of 0.83 between Version 6 TMI estimates and Kwajalein ground radar rain rates at 50 km resolution. The correlation between Version 6 TMI and PR estimates at Kwajalein at the same resolution was 0.84.

Since comparisons to ground validation radars can bias error statistics toward local conditions, TMI rain rate and convective proportion estimates were also compared to PR estimates over the entire TRMM observing domain. Shown in Fig. 2 is a comparison of all coincident Version 6 PR and TMI rain estimates over ocean for the month of July 2001. Estimated rain rates greater than a few tenths of a mm h$^{-1}$ are strongly correlated; the low bias of TMI estimates at very low rain rates does not contribute appreciably to the total rain of the distribution. Error modeling of TMI rain estimates suggests that 70-90% of the random difference between TMI and PR instantaneous rain estimates at half-degree resolution can be explained by random errors in the TMI estimates; the remainder is due to errors in the PR estimates and differences in the spatial sampling of rain by the two instruments within half-degree boxes.

GPROF APPLICATIONS

Instantaneous Precipitation Estimates

Shown in Figs. 3 and 4 are applications of GPROF Version 6 to TMI observations of two precipitation systems. Hurricane Bret developed in the Gulf of Mexico in mid-August, 1999, and attained hurricane strength on 21 August; see Lawrence et al. (2001). The GPROF estimates of precipitation in Fig. 3 are derived from TMI observations of Bret on 21 August at 2245 UTC, as the storm moved northward towards the coast of Texas, USA. Estimated surface rain rates are most intense (> 10 mm h$^{-1}$) in the nearly circular eyewall of Bret, with somewhat less intense rains in bands to the northeast of the eye. Note that the resolution of instantaneous rain estimates from GPROF is 14 km, and so rains are possibly more intense at sub-footprint resolution. The eyewall and innermost rain band are surrounded by less intense, stratiform rains (< 5 mm h$^{-1}$).

Errors in estimates of instantaneous rain rates, given by (2), are shown in the top-right panel of Fig. 3. Note that the errors in estimates of lighter rains can be equal to or greater than 100%; while for the most intense rains, errors are ~60%. These errors are characteristic of satellite passive microwave estimates of rain rate- although errors in rain rate estimates generally increase with increasing rain rate, percentage errors tend to decrease with rain rate. These errors are substantially reduced by space- and time-averaging; see Bauer et al. (2002).

Total precipitation water content (the combined water content of all precipitating hydrometeors) and $Q_I$-$Q_R$ estimates along the transect A-B are shown in the lower panel of Fig. 3. Note that the deepest precipitation structures and greatest heating rates along the transect are associated with the eyewall of Bret. The heating vertical structure is "convective" in the eyewall, in the sense that heating rates are positive through most of the troposphere with a maximum at midlevels. Rain rates and heating rates are greater in the left branch of the eyewall, relative to the right branch in the transect, which is
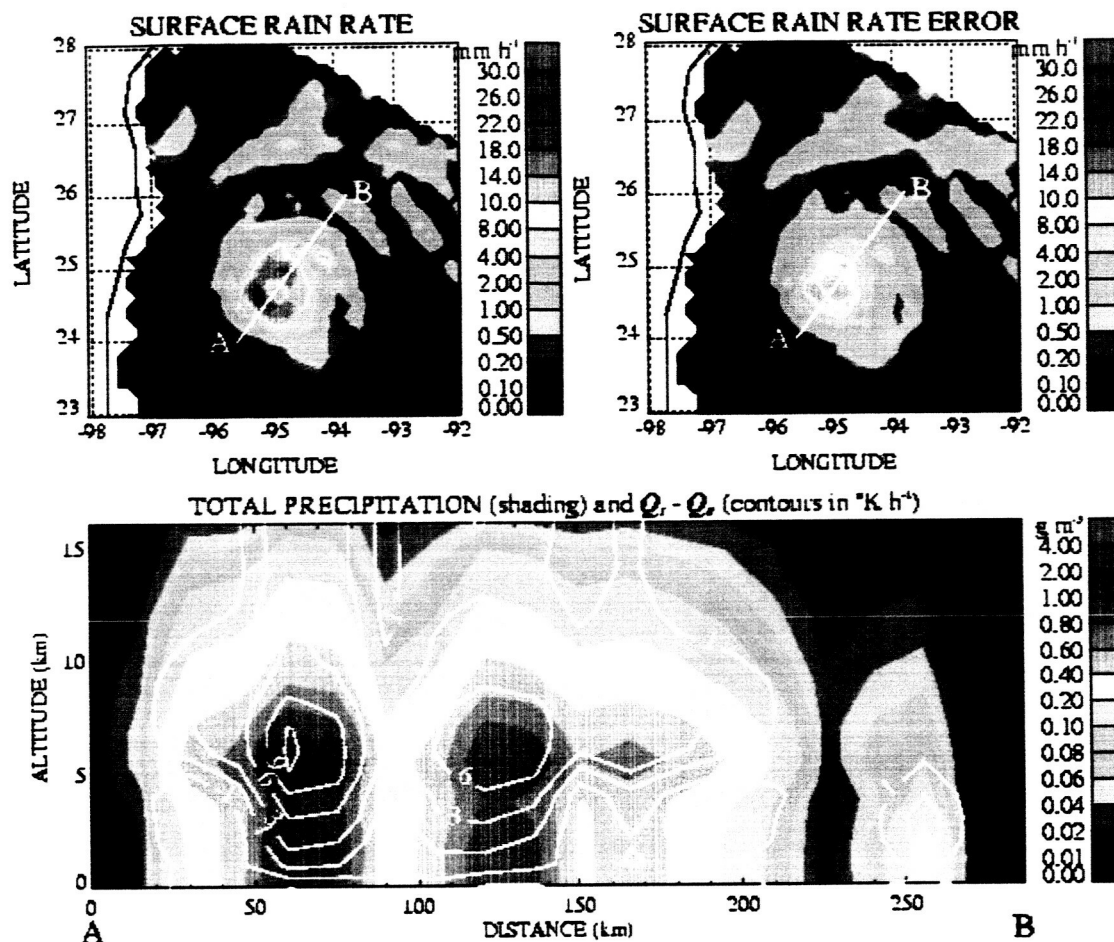
Fig. 3. GPROF estimates of surface rain rate (upper left panel), surface rain rate error (upper right panel), and cross sections of total precipitation and $Q_1 - Q_R$ along the A-B transect (lower panel) from TMI observations of Hurricane Bret on 21 August 1999. In the lower panel, total precipitation (rain, graupel, and snow) water contents are shaded and heating rates are contoured in white at levels of $-3, -1, 1, 3, 6, 9,$ and $12$ K h$^{-1}$.

consistent with the higher-resolution PR-derived distribution of rain intensity (not shown). Although the rain rates are a relative minimum in the eye, the TMI does not detect a "clear" eye due to the limited horizontal resolution of the instrument and its oblique viewing angle (52.8° from vertical). The quasi-stratiform rainband to the northeast of the eye exhibits maximum water contents just above the freezing level (estimated to be ~4.7 km), with relatively weak heating aloft and evaporative cooling below the freezing level. The outer rainband near "B" in the transect has convective structure but is much shallower in the vertical than the eyewall.

Although GPROF applications to observations of tropical precipitation systems have been a primary focus of research in the past, the recent expansion of the algorithm's cloud resolving model database to include midlatitude simulations has made extra-
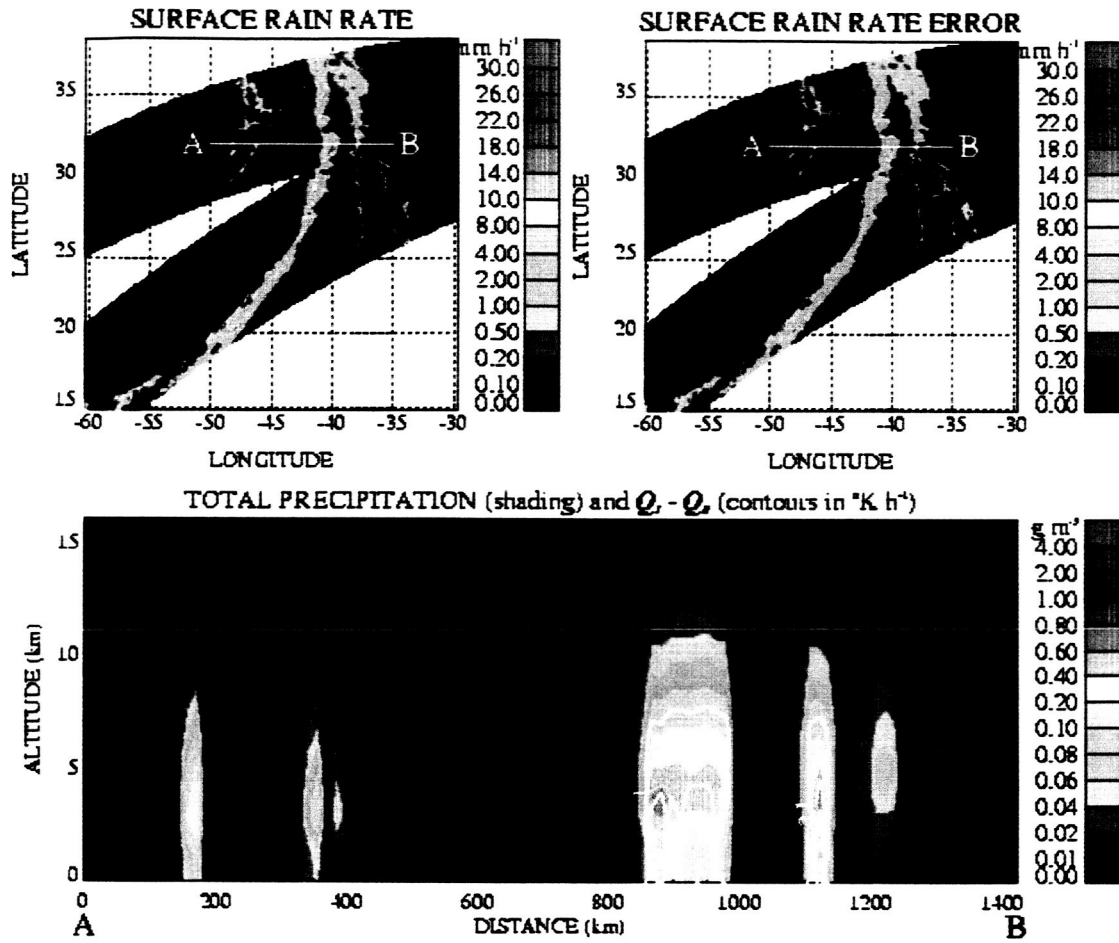
Fig. 4. Same as Fig. 3, but for TMI observations of an extra-tropical cyclone over the North Atlantic on 17 January 2000. The two swath sections depicted represent observations that were separated by ~90 minutes of time, causing a slight discontinuity in the GPROF estimates near the inter-swath boundary.

tropical applications possible. The GPROF estimates shown in Fig. 4 are derived from two TRMM overpasses of a baroclinic system on 17 January 2000 near 09 UTC. The cold-frontal rainband is primarily contained in the lower swath, with a frontal occlusion and post-frontal precipitation contained in the upper swath. Maximum rain rates are seen in the upper portion of the cold frontal band. Note that the vertical precipitation and heating structures along the A-B transect are much shallower than those retrieved from the Bret data. Maximum water contents (0.6 g m$^{-3}$) and heating rates (3 K h$^{-1}$) are consistent with weaker updrafts in the extra-tropical system.

Finally, note that the instantaneous heating estimates shown in Figs. 3 and 4 are sometimes subject to large random errors, given the indirect inference of heating from TMI precipitation signatures. These figures are presented to indicate the qualitative plausibility of GPROF-derived heating structures, even though the individual heating rate
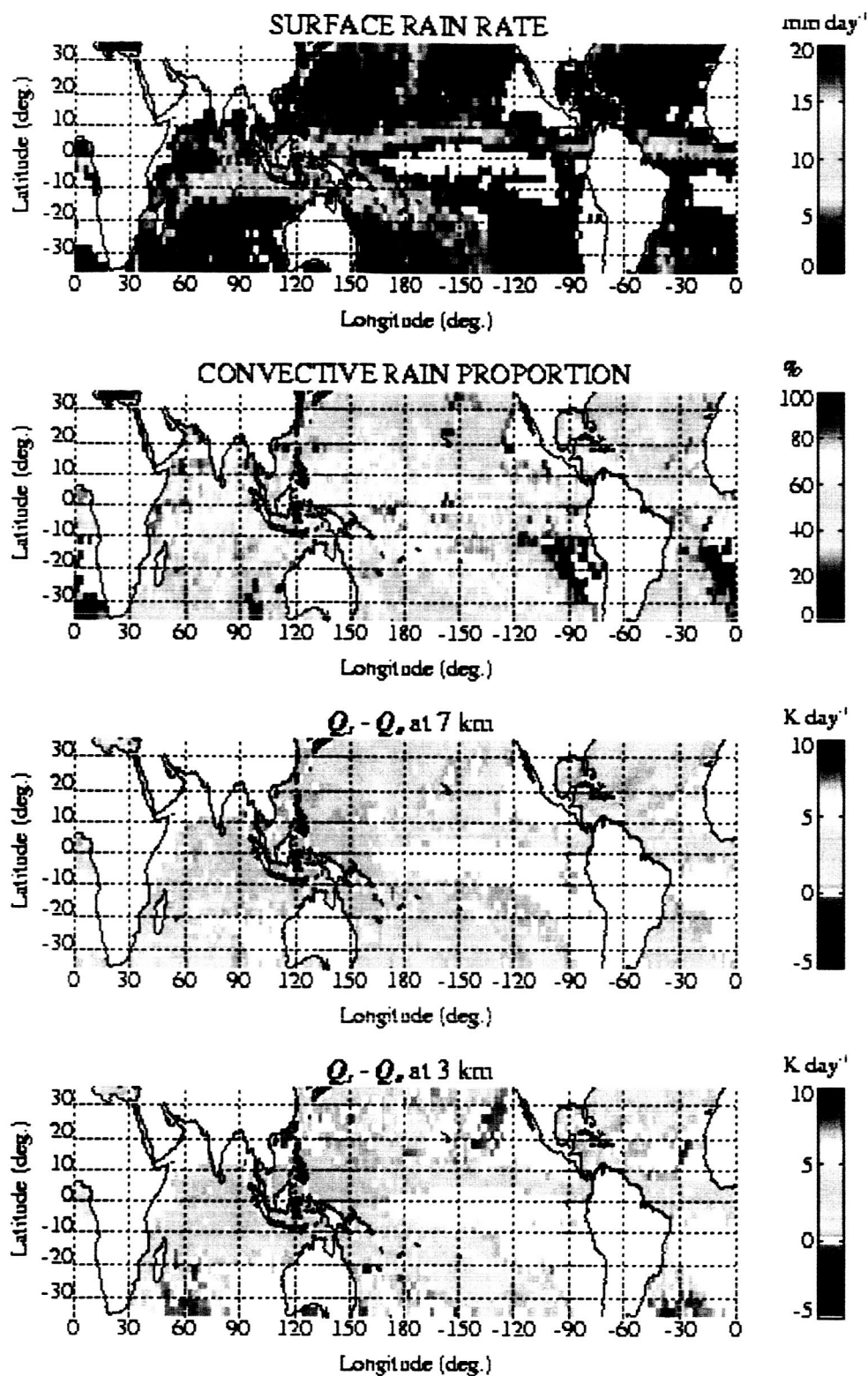
Fig. 5. Mean surface rain rates, convective rain proportions, and latent heating rates at 7 and 3 km altitude, derived from TMI observations from January 2000, using the GPROF algorithm.

estimates may contain large quantitative errors. Random errors in GPROF heating estimates are reduced significantly with space- and time-averaging.

Large-scale Precipitation and Latent Heating

GPROF estimates of precipitation and latent heating based upon TMI data can be aggregated to produce estimates of their large-scale distributions in the Tropics and Sub-tropics. Shown in Fig. 5 are GPROF estimates aggregated in $2.5°$ x $2.5°$ latitude/longitude boxes over the month of January 2000. Note that the distributions in Fig. 5 are not "smooth" in part because the temporal sampling by TMI of a given $2.5°$ box is limited (about 1 day$^{-1}$ near the equator). Temporal sampling by microwave radiometers is expected to improve in the future, as additional satellite radiometers will be launched as part of the follow-on mission to TRMM. Nevertheless, the main features of global rain distributions are captured by TMI observations, including the Inter-Tropical Convergence Zone (ITCZ), the South Pacific Convergence Zone, and rains along the southern portion of the mid-latitude storm tracks in the Northern Hemisphere. Rains in the Tropics are predominantly convective; however, regions of peak rainfall along the ITCZ exhibit a relative minimum of convection, indicating the contribution of organized mesoscale convective systems to the total rainfall. Latent heating basically follows the pattern of surface rain rate. Heating is generally a maximum near 7 km altitude where the rainfall is most intense- another signature of organized precipitation systems. In regions where less organized convection dominates the spectrum of precipitation systems, the heating maximum occurs at lower altitudes.

CONCLUDING REMARKS

This brief survey is intended to illustrate some of the strengths and weaknesses of precipitation and latent heating estimation using a Bayesian method (GPROF). It should be noted that any remote sensing method is limited by the information content of the input observations. GPROF was designed to exploit not only the multifrequency sensing capability of microwave radiometers, but also information drawn from the horizontal distributions of observed radiances. In the future, the identification of precipitation system type (isolated convection, short-lived convective lines, squall lines, tropical cyclones, extra-tropical cyclones) from radiometer observations may lead to more specific estimates from the algorithm. Latent heating estimates require contextual information and should benefit from the identification of system type.

Another important area of study is the construction of the algorithm's supporting database. First, there are potential biases in the cloud-radiative model calculations; in particular the simulation of precipitation-sized ice aloft has been studied (Tao et al. 2003). To the extent that biases in the cloud-radiative model simulations can be quantified and the physical mechanisms understood, they should be corrected. In addition, the general representativeness of the database must be examined. That is, the population of simulated precipitation structures in the database should reflect the natural population of these structures. High-resolution Precipitation Radar (PR) observations can be used as a guide to determine the natural population of precipitation structures.

Missing structures should be included in the algorithm's database and artificial structures eliminated. Since the *a priori* distribution of precipitation structures in the database has a significant impact on precipitation and heating estimates (Shin and Kummerow, 2004), due to the limited information content of the radiometer observations, proper database construction is critical.

The authors are currently conducting the "validation phase" of the Version 6 GPROF algorithm. Rain rate estimates from the PR and ground-based radar provide reference estimates. Latent heating vertical profiles from the algorithm are being compared to rawinsonde budget estimates, such as those from SCSMEX (Johnson and Ciesielski, 2002), while heating profiles inferred from dual-Doppler radar observations in combination with radiosonde-derived thermal structure provide another reference.

References

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. Q. J. R. Meteorol. Soc., 128, 2129-2144.

Johnson, R. H., and P. E. Ciesielski, 2002: Characteristics of the 1998 summer monsoon onset over the northern South China Sea. *J. Met. Soc. Japan*, **80**, 561-578.

Kummerow, C., W. S. Olson, and L. Giglio, 1996: A simplified scheme for obtaining precipitation and vertical hydrometeor profiles from passive microwave sensors. *IEEE Trans. Geosci. Remote Sensing*, **34**, 1213-1232.

Kummerow, C., Y. Hong, W. S. Olson, S. Yang, R. F. Adler, J. McCollum, R. Ferraro, G. Petty, D.-B. Shin, and T. T. Wilheit, 2001: The evolution of the Goddard Profiling Algorithm (GPROF) for rainfall estimation from passive microwave sensors. *J. Appl. Meteor.*, **40**, 1801-1820.

Lawrence, M. B., L. A. Avila, J. L. Beven, J. L. Franklin, J. L. Guiney, and R. J. Pasch, 2001: Atlantic hurricane season of 1999. *Mon. Wea. Rev.*, **129**, 3057-3084.

Olson, W. S., C. D. Kummerow, Y. Hong, and W.-K. Tao, 1999: Atmospheric latent heating distributions in the Tropics derived from passive microwave radiometer measurements. *J. Appl. Meteor.*, **38**, 633-664.

Petty, G. W., 1994: Physical retrievals of over-ocean rain rate from multichannel microwave imagers. Part I: Theoretical characteristics of normalized polarization and scattering indices. *Meteor. Atmos. Phys.*, **54**, 79-99.

Shin, D.-B. and C. Kummerow, 2004: Parametric rainfall retrieval algorithms for passive microwave radiometers. *J. Appl. Meteor.*, in press.

Smith, E. A., C. Kummerow, and A. Mugnai, 1994: The emergence of inversion-type profile algorithms for estimation of precipitation from satellite passive microwave measurements. *Remote Sensing Reviews*, **11**, 211-242.

Tao, W.-K., C.-L. Shie, J. Simpson, S. Braun, R. H. Johnson, and P. E. Ciesielski, 2003: Convective systems over the South China Sea: Cloud-resolving model simulations. *J. Atmos. Sci.*, **60**, 2929-2956.