

**NOAA NESDIS
CENTER for SATELLITE APPLICATIONS and
RESEARCH**

**GOES-R Advanced Baseline Imager (ABI)
Algorithm Theoretical Basis Document
For
Rainfall Rate (QPE)**

Robert J. Kuligowski, NOAA/NESDIS/STAR

Version 2.6
October 24, 2013

TABLE OF CONTENTS

LIST OF FIGURES	4
LIST OF TABLES	5
LIST OF ACRONYMS	6
ABSTRACT	7
1 INTRODUCTION	8
1.1 Purpose of This Document	8
1.2 Who Should Use This Document	8
1.3 Inside Each Section	8
1.4 Related Documents	8
1.5 Revision History	9
2 OBSERVING SYSTEM OVERVIEW	10
2.1 Products Generated	10
2.2 Instrument Characteristics	11
3 ALGORITHM DESCRIPTION	12
3.1 Algorithm Overview	12
3.2 Processing Outline	12
3.3 Algorithm Input	14
3.3.1 Primary Sensor Data	14
3.3.2 Ancillary Data	14
3.4 Theoretical Description	16
3.4.1 Physics of the Problem	17
3.4.1.1 Training Data: Matched ABI Predictors and Microwave Rain Rates	17
3.4.1.2 Rainfall Detection	20
3.4.1.3 Rainfall Rate Estimation	21
3.4.1.4 Independent rainfall rates	22
3.5 Mathematical description	22
3.5.1 ABI-MW footprint matching	22
3.5.2 Calibration: Rain / no rain discrimination	23
3.5.3 Calibration: Rainfall rate	24
3.5.4 Application to independent data	26
3.6 Algorithm Output	27
4 TEST DATA SETS AND OUTPUTS	30
4.1 Simulated/Proxy Input Data Sets	30
4.1.1 SEVIRI Data	30
4.1.2 Microwave-Derived Rainfall Rates	31
4.2 Output from Proxy Input Data Sets	32
4.2.1 Precision and Accuracy Estimates	33
4.2.1.1 Validation against TRMM PR	34
4.2.1.2 Validation against Nimrod data	36
4.2.2 Error Budget	39
5 PRACTICAL CONSIDERATIONS	40
5.1 Numerical Computation Considerations	40
5.2 Programming and Procedural Considerations	40
5.3 Quality Assessment and Diagnostics	40

5.4	Exception Handling	41
5.5	Algorithm Validation	42
6	ASSUMPTIONS AND LIMITATIONS	43
6.1	Performance	43
6.2	Assumed Sensor Performance	44
6.3	Pre-Planned Product Improvements	45
6.3.1	Smoothing along Region Seams	45
6.3.2	Incorporation of Numerical Model Moisture Fields to Correct for Subcloud Evaporation of Rainfall.....	45
6.3.3	Correction for Orographic Modulation of Rainfall.....	45
6.3.4	Incorporation of Time Change Information.....	45
6.3.5	Incorporation of Retrieved Cloud Microphysics Information	45
7	REFERENCES	46

LIST OF FIGURES

Figure 1. High Level Flowchart of the rain rate algorithm, illustrating the main processing sections.....	13
Figure 2. Full disk 10.8- μm color-enhanced image from SEVIRI for 1200 UTC on January 7, 2005.	31
Figure 3. Blended MW rainfall rates covering 1230-1300 UTC 7 January 2005.....	32
Figure 4. Sample rain rate algorithm output from 1245 UTC January 7, 2005.....	33
Figure 5. Coverage of TRMM PR data during October 6, 2005.	34
Figure 6. Scatterplot of Rainfall Rate algorithm vs. collocated TRMM rain rates; colors are related to pixel density with red highest and purple lowest. Solid line is the 1:1 line and the dashed line is the best-fit line.....	35
Figure 7. CDF of errors of Rainfall Rate product with rates of 9.5-10.5 mm/h vs. TRMM.	36
Figure 8. Coverage of Nimrod mosaic radar data.....	37
Figure 9. Same as Fig. 6, but vs. Nimrod.	38
Figure 10. Same as Fig. 7, but vs. Nimrod.	39

LIST OF TABLES

Table 1. F&PS Requirements for the Rainfall Rate / QPE algorithm.	11
Table 2. Channel numbers, wavelengths, and footprint sizes of the ABI bands.	11
Table 3. Contents of each data record of the IR-MW matched data file.	15
Table 4. Contents of the retrieval coefficient table file.	16
Table 5. Predictors computed from the data in the matched MW-ABI data file.	18
Table 6. Description of rainfall classes.	20
Table 7. Quality flags for the Rainfall Rate product.	27
Table 8. Diagnostic information for the Rainfall Rate product.	27
Table 9. Gridded quality information for the Rainfall Rate product.	28
Table 10. Metadata for the Rainfall Rate product.	29
Table 11. Comparison of Rainfall Rate algorithm validation with F&PS.	39
Table 12. Minimum acceptable values for each algorithm predictor.	41
Table 13. Channel mapping associated with ABI proxy data from SEVIRI during algorithm development and validation.	44

LIST OF ACRONYMS

ABI – Advanced Baseline Imager
AIT – Algorithm Integration Team
AMSR-E – Advanced Microwave Scanning Radiometer-EOS [Earth Observing System]
AMSU – Advanced Microwave Sounding Unit
ATBD – Algorithm Theoretical Basis Document
BADC – British Atmospheric Data Centre
BTD – Brightness Temperature Difference
CDF – Cumulative Distribution Function
CLASS – Comprehensive Large Array-data Stewardship System
COPS – Convective and Orographically-induced Precipitation Study
EDR – Environmental Data Record
F&PS – Functional and Performance Specification
GAS – GOES-R Archive System
GOES – Geostationary Operational Environmental Satellite
GPM – Global Precipitation Measurement
GPO – GOES-R Program Office
HSS – Heidke Skill Score
IR – Infrared
LEO – Low-Earth Orbit
L-U – Lower-Upper matrix decomposition
LZA – Local Zenith Angle
MRD – Mission Requirements Document
MW – Microwave
NAMMA – NASA African Monsoon Multidisciplinary Analysis
NASA – National Aeronautics and Space Administration
NESDIS – National Environmental Satellite, Data, and Information Service
NOAA – National Oceanic and Atmospheric Administration
OSDPD – Office of Satellite Data Processing and Distribution
PDF – Probability Distribution Function
PR – Precipitation Radar
SCaMPR – Self-Calibrating Multivariate Precipitation Retrieval
SEVIRI – Spinning Enhanced Visible Infrared Imager
SSM/I – Special Sensor Microwave / Imager
STAR – Center for Satellite Applications and Research
TMI – TRMM Microwave Imager
TRMM – Tropical Rainfall Measuring Mission
QPE – Quantitative Precipitation Estimate

ABSTRACT

This Rainfall Rate Algorithm Theoretical Basis Document (ATBD) contains a high-level description (including the physical basis) of an algorithm for estimating pixel-scale rainfall rate from images taken by the Advanced Baseline Imager (ABI) flown on the Geostationary Operational Environmental Satellite-Series R (GOES-R) series of National Oceanic and Atmospheric Administration (NOAA) geostationary meteorological satellites. A brief overview of the GOES-R observing system is followed by a more specific description of the Rainfall Rate algorithm, validation efforts, and planned improvements.

1 INTRODUCTION

1.1 Purpose of This Document

The Rainfall Rate Algorithm Theoretical Basis Document (ATBD) provides a high-level description of and the physical basis for the estimation of pixel-scale rainfall rate from images taken by the Advanced Baseline Imager (ABI) flown on the Geostationary Operational environmental Satellite-Series R (GOES-R) series of National Oceanic and Atmospheric Administration (NOAA) geostationary meteorological satellites. The rainfall rate is produced as an Environmental Data Record (EDR) and the algorithm output is used by the rainfall potential algorithm to create nowcasts of rainfall rates and accumulations.

1.2 Who Should Use This Document

The intended users of this document are those interested in understanding the physical basis of the algorithms and how to use the output of this algorithm in a manner that is consistent with its underlying assumptions. This document also provides information useful to anyone maintaining or modifying the original algorithm.

1.3 Inside Each Section

This document is broken down into the following main sections.

- **System Overview:** Provides relevant details of the Rainfall Rate Algorithm and provides a brief description of the products generated by the algorithm.
- **Algorithm Description:** Provides all the detailed description of the algorithm including its physical basis, its input and its output.
- **Test Data Sets and Output:** Provides a description of the test data set used to characterize the performance of the algorithm and quality of the data products. It also describes the results from algorithm processing using simulated input data.
- **Practical Considerations:** Provides an overview of the issues involving numerical computation, programming and procedures, quality assessment and diagnostics and exception handling.
- **Assumptions and Limitations:** Provides an overview of the current limitations of the approach and gives the plan for overcoming these limitations with further algorithm development.

1.4 Related Documents

This document currently does not relate to any other document outside of the Parallax Algorithm Theoretical Basis Document and to the specifications of the GOES-R Ground

Segment Mission Requirements Document (MRD) and Functional and Performance Specification (F&PS) and to the references given throughout.

1.5 Revision History

Version (0.1) of this document was created by Dr. Robert J. Kuligowski of NOAA/NESDIS [National Environmental Satellite, Data, and Information Service]/STAR [Center for Satellite Applications and Research] and its intent was to serve as a draft submission to the GOES-R Program Office (GPO) for initial comments.

Version (1.0) of this document was created by Dr. Robert J. Kuligowski of NOAA/NESDIS and its intent was to accompany the delivery of the 80% algorithm to the GOES-R AWG Algorithm Integration Team (AIT).

Version (2.0) of this document was created by Dr. Robert J. Kuligowski of NOAA/NESDIS and its intent was to accompany the delivery of the 100% algorithm to the GOES-R AWG Algorithm Integration Team (AIT).

2 OBSERVING SYSTEM OVERVIEW

This section will describe the products generated by the ABI Rainfall Rate Algorithm and the requirements it places on the sensor.

2.1 Products Generated

The Rainfall Rate Algorithm produces a field of instantaneous rainfall rates associated with the most recently available GOES imagery. In terms of the F&PS, it is responsible directly for the Rainfall Rate / QPE [Quantitative Precipitation Estimate] product within the Hydrology product sub-type, and meets the requirements listed in Table 1. The Rainfall Rate Algorithm design calls for a quantitative rainfall rate in millimeters per hour on the same grid as the 2-km ABI IR bands. These products are intended for use by operational meteorologists and hydrologists for flood forecasting. There are no diagnostic products for external use aside from the official Rainfall Rate product and accompanying quality flags, but the calibration coefficient tables and rainfall class grids will be available internally for diagnostic purposes.

Requirement Description	Requirement Value
Name	Rainfall Rate / QPE
User	GOES-R
Geographic Coverage	Full Disk
Temporal Coverage Qualifiers	Day and night
Product Extent Qualifier	Quantitative out to at least 70 degrees LZA or 60 degrees latitude—whichever is less—and qualitative beyond
Cloud Cover Conditions Qualifier	N/A
Product Statistics Qualifier	Over rain cases and mesoscale-sized surrounding regions
Vertical Resolution	N/A
Horizontal Resolution	2.0 km
Mapping Accuracy	2.0 km
Measurement Range	0 – 100 mm/hr
Measurement Accuracy	6 mm/hr at a rate of 10 mm/hr with higher values at higher rates (<i>pending addition: “Quantitative for convective rainfall and qualitative for stratiform rainfall.”</i>)
Product Refresh Rate / Coverage Time (Mode 3)	15 min
Refreshment Rate / Coverage Time (Mode 4)	15 min
Vendor Allocated Ground Latency	266 sec
Product Measurement Precision	9 mm/hr at a rate of 10 mm/hr with higher values at higher rates (<i>pending addition: “Quantitative for convective rainfall and qualitative for stratiform rainfall.”</i>)

Table 1. F&PS Requirements for the Rainfall Rate / QPE algorithm.

Note that for pixels outside the local zenith angle and latitude cutoffs, rainfall rates will still be generated by the algorithm, but their use will be discouraged and they will not be validated for comparison against spec.

2.2 Instrument Characteristics

The rainfall rate will be produced for each pixel observed by the ABI between 60°S and 60°N latitude that has a local zenith angle (LZA) of less than 70 degrees. The final channel set is still being determined as the algorithms are developed and validated. Table 2 summarizes the current channel use by the Rainfall Rate Algorithm. Note that these particular bands had METEOSAT Spinning Enhanced Visible InfraRed Imager (SEVIRI) equivalents and therefore are the only ones used in the current version of the algorithm. However, the operational version of the code will be modified to include inputs from the available ABI bands without SEVIRI equivalents if they are shown to have a positive impact on algorithm performance.

<i>Channel Number</i>	<i>Wavelength (μm)</i>	<i>Resolution (km)</i>	<i>Used in Rain Rate</i>
1	0.47	1.0	
2	0.64	0.5	
3	0.865	1.0	
4	1.378	2.0	
5	1.61	1.0	
6	2.25	2.0	
7	3.9	2.0	
8	6.19	2.0	✓
9	6.95	2.0	
10	7.34	2.0	✓
11	8.5	2.0	✓
12	9.61	2.0	
13	10.35	2.0	
14	11.2	2.0	✓
15	12.3	2.0	✓
16	13.3	2.0	

Table 2. Channel numbers, wavelengths, and footprint sizes of the ABI bands.

In addition to the data from the individual bands, the algorithm also uses brightness temperature differences (BTD's) between pairs of selected bands, and also uses some spatial gradient information from the infrared (IR) window band (14); see Section 3.4.1.1 for details. Therefore, the performance of the Rainfall Rate Algorithm is sensitive to any imagery artifacts or instrument noise. The channel specifications are given in the Mission Requirements Document (MRD) section 3.4.2.1.4.0. The performance outlined therein was assumed during development efforts.

3 ALGORITHM DESCRIPTION

This section contains a complete description of the algorithm at the current level of maturity (which will improve with each revision).

3.1 Algorithm Overview

The rain rate algorithm identifies raining pixels and derives rain rates on a pixel level in ABI imagery. Its calibration is based on matches of ABI data with microwave (MW)-derived rainfall rates, which are considered to be the most accurate estimates of instantaneous rainfall rate available from satellite data. The ABI rain rate algorithm is based on the Self-Calibrating Multivariate Precipitation Retrieval (SCaMPR) algorithm first described in Kuligowski (2002).

The algorithm derives rainfall rate fields in two steps:

1. Identify pixels that are experiencing rainfall. The predictors and predictor coefficients for detecting rainfall are derived using discriminant analysis in a calibration against MW-retrieved rainfall areas.
2. Retrieve rainfall rates for pixels where rainfall has been detected. The predictors and predictor coefficients for retrieving rainfall rate are derived using stepwise forward linear regression in a calibration against MW-retrieved rainfall rates.

The rain rate algorithm provides estimates of instantaneous rainfall rate at the same pixel resolution as the ABI. In addition to its use in estimating rainfall rates from current ABI data, the estimates are also extrapolated forward in time in the GOES-R Rainfall Potential Algorithm, and these nowcasts are in turn used as input for the Probability of Rainfall Algorithm (see corresponding ATBD's for additional details).

3.2 Processing Outline

The processing outline of the rain rate algorithm is summarized in Fig. 1. The rain rate is designed to run on individual pixels, with some information required from pixels in the 5x5 neighboring region.

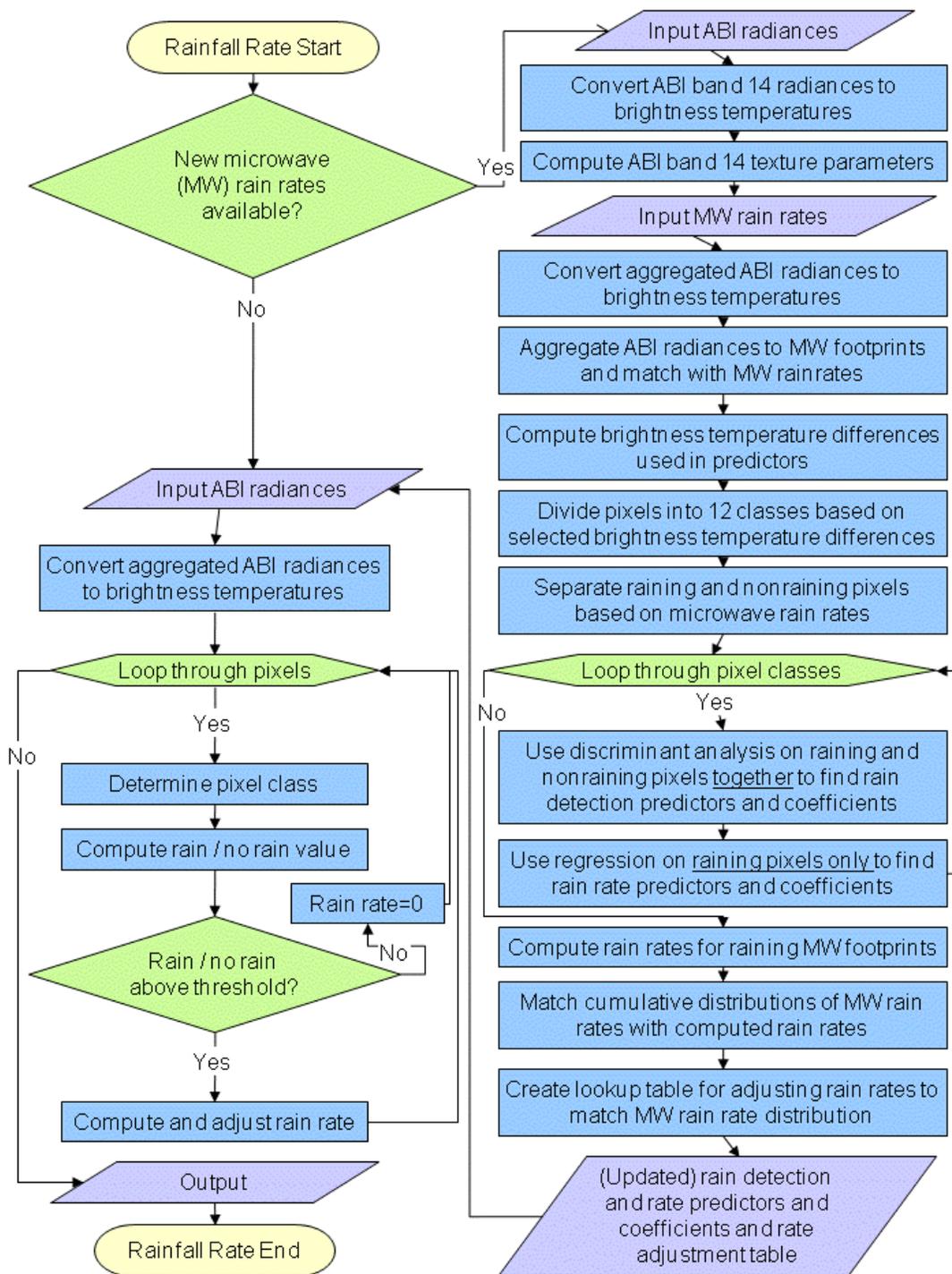


Figure 1. High Level Flowchart of the rain rate algorithm, illustrating the main processing sections.

3.3 Algorithm Input

This section describes the input needed to process the rainfall rates. While the rainfall rate is derived for each pixel, it does require limited knowledge of the surrounding pixels (5x5). The Rainfall Rate Algorithm can run with information from only one pixel, but only if certain predictors are removed from the predictor set.

3.3.1 Primary Sensor Data

The list below contains the primary sensor data used by the Rainfall Rate Algorithm. Primary sensor data refers to information that is derived solely from the ABI observations and geolocation information.

- Parallax-corrected, calibrated brightness temperatures (or radiances that will then be converted to brightness temperatures) for channels 8, 10, 11, 14, and 15 (see separate ATBD for description of parallax correction algorithm)
- Pixel latitude, longitude, and local zenith angle (LZA)
- Minimum channel 14 brightness temperature over a 5x5 array centered on the pixel of interest—note that if any of the pixels in the 5x5 array are missing, this value is still computed based on the available pixels.
- Average channel 14 brightness temperature of the nearest 4 pixels along the scan line (2 on each side) and nearest pixel in the two adjacent scan lines—unless at least 3 of these 6 values are missing, in which case this value is not computed.
- Any relevant ABI quality control flags

3.3.2 Ancillary Data

The following list briefly describes the ancillary data required to run the Rainfall Rate Algorithm. Ancillary data is defined as data that requires information not included in the ABI observations or geolocation data. All three of these ancillary data sets would be considered to be non-ABI dynamic data (i.e., they are not other ABI-derived products); no static ancillary data (i.e., time-constant ancillary data such as topography or a land/sea mask) are required.

- **MW-derived rainfall rates**

Rainfall rates, presumably from MW data but also permissible from active radar, are required as a calibration target for the algorithm. These rainfall rates do not need to be available in real time, though the accuracy of the rain rate estimates tends to degrade slightly as the difference between the time period covered by the training data and the time of the retrieval from the ABI becomes longer. The MW rainfall rates will be obtained from an operational NESDIS Blended Microwave Rainfall Rate product that will combine rainfall rates from multiple platforms

(e.g., SSMIS, AMSU-B/MHS) and match their statistical distributions in order to resolve inconsistencies between the two. These rainfall rates are on a 1437 row x 2500 element Mercator grid with an equator resolution of 16 km centered on the equator and the Greenwich meridian.

- **Matched MW rain rates and ABI predictors**

These MW-derived rainfall rates are matched with ABI-derived predictors that have been aggregated to the spatial resolution of the MW rain rates (nominally 15 km). Each data point is on a separate data record (the data are not necessarily on a regularly-spaced grid, though they can be) and the contents of each record of this matched file are given in Table 3:

Bytes	Variable	Type	Value
0-3	MW pixel latitude	Real*4	-60.0 to 60.0
4-7	MW pixel longitude	Real*4	-180.0 to 180.0
8-11	MW rainfall rate (mm/h)	Real*4	0 to 50
12-15	ABI band 8 brightness temperature (K)	Real*4	174 to 325
16-19	ABI band 10 brightness temperature (K)	Real*4	174 to 325
20-23	ABI band 11 brightness temperature (K)	Real*4	174 to 325
24-27	ABI band 14 brightness temperature (K)	Real*4	174 to 325
28-31	ABI band 15 brightness temperature (K)	Real*4	174 to 325
32-35	$S = 0.568*(T_{\min,11.2} - 217 \text{ K})$	Real *4	-107.432 to 43.568
36-39	$G_t = T_{\text{avg},11.2} - T_{\min,11.2}$	Real*4	0 to 151
40-41	MW satellite ID	Integer*2	

Table 3. Contents of each data record of the IR-MW matched data file.

- **Retrieval coefficient table**

This table contains the ID's (from the matched file) of the selected predictors along with their calibration coefficients for both rain / no rain discrimination and rain rate calibration. A list of the contents of this table is provided in Table 4:

4-Byte Word	Variable	Type	Value
0	Digital day	Real*4	0 to 366.99
1-12	Number of previous time periods required for training	Integer*4	1 and higher
13-24	Heidke Skill Score from 2-predictor rain / no rain calibration (1 value for each of 12 classes)	Real*4	-1.0 to 1.0
25-36	Correlation coefficient from 2-predictor rain rate calibration (1 value for each of 12 classes)	Real*4	-1.0 to 1.0

37-48	First rain rate predictor ID for classes 1-12	Integer*4	1 to 16
49-60	Second rain rate predictor ID for classes 1-12	Integer*4	1 to 16
61-72	First rain / no rain predictor ID for classes 1-12	Integer*4	1 to 8
73-84	Second rain / no rain predictor ID for class 1-12	Integer*4	1 to 8
85-108	Rain rate intercept for classes 1-12	Real*8	
109-132	Class 1-12 rain rate multiplier for predictor 1	Real*8	
133-156	Class 1-12 rain rate multiplier for predictor 2	Real*8	
157-180	Class 1-12 rain / no rain multiplier for predictor 1	Real*8	
181-204	Class 1-12 rain / no rain multiplier for predictor 2	Real*8	
205-216	Class 1-12 rain / no rain threshold values	Real*4	
217-240	Class 1-12 nonlinear transformation multiplier for predictor 1 (0 if predictor is not transformed)	Real*8	
241-264	Class 1-12 nonlinear transformation multiplier for predictor 2 (0 if predictor is not transformed)	Real*8	
265-288	Class 1-12 nonlinear transformation exponent for predictor 1 (0 if predictor is not transformed)	Real*8	
289-312	Class 1-12 nonlinear transformation exponent for predictor 2 (0 if predictor is not transformed)	Real*8	
313-324	Class 1-12 nonlinear transformation intercept for predictor 1 (0 if predictor is not transformed)	Real*4	
325-336	Class 1-12 nonlinear transformation intercept for predictor 2 (0 if predictor is not transformed)	Real*4	
337-348	Look-up table values for adjusting rain rates for classes 1-12 at 0.1 mm/h	Real*4	
349-360	Look-up table values for adjusting rain rates for classes 1-2 at 0.2 mm/h	Real*4	
361-12336	Look-up table values for adjusting rain rates for classes 1-12 at 0.3...50.0 mm/h	Real*4	

Table 4. Contents of the retrieval coefficient table file.

3.4 Theoretical Description

As stated previously, retrieval of rainfall rate requires two steps: determining which pixels in satellite imagery will be associated with rainfall, and then deriving rainfall rates for those pixels. In the case of visible / IR instruments such as the ABI, the basic approach is to use the information about cloud-top properties that are inherent in the brightness temperature information (e.g., height, thickness, phase, particle size) to make inferences about the occurrence and rate of rain. This algorithm develops statistical relationships between the brightness temperature values and the occurrence of rainfall and its intensity. In the algorithm, discriminant analysis is used to determine the best predictors and predictor coefficients for occurrence, and stepwise forward linear regression is used for intensity.

3.4.1 Physics of the Problem

The difficulty in using visible- and IR-wavelength information for retrieving rainfall rates is that raining clouds are generally optically thick, meaning that the information at these wavelengths comes from the top portion of the cloud, and typically above precipitating hydrometeors that are actually of interest. Microwave-frequency information is relatively more valuable because raining clouds are generally not optically thick in that portion of the spectrum, meaning that MW signals are sensitive to the total water or ice path in the cloud rather than just the properties of the cloud top. However, since MW sensors are for the near future restricted to low-Earth orbit (LEO), rainfall information from such instruments will not be available on a continuous basis without a much more substantial LEO satellite constellation than is currently planned. The compromise has been to use the intermittently-available but relatively more accurate MW-based rainfall rates as a calibration target for visible and IR data from geostationary platforms, and to use the resulting calibration to retrieve rainfall rates at the full spatial and temporal resolution of the geostationary data.

The following subsections describe how this calibration process is performed in the Rainfall Rate Algorithm. The first subsection describes the MW data set that is used as a calibration target and how it is matched against the ABI predictors, and the subsequent four subsections describe respectively how the training data are assembled, how the rainfall detection algorithm is calibrated, how the Rainfall Rate Algorithm is calibrated, and then how the resulting calibration is applied to independent data to product the rain rate product.

3.4.1.1 Training Data: Matched ABI Predictors and Microwave Rain Rates

The MW rain rates serve as the calibration target for the rain rate algorithm, both in terms of identifying raining areas and in retrieving the intensity of rainfall. In practice, any reliable rainfall rate field could be used for calibration, including radar data, and it is not necessary that these fields be continuous in space or time—just that they represent instantaneous rates of rainfall rather than accumulations over time. Since rainfall rates from multiple microwave instruments are used, the rates should be bias-adjusted since differences among the input data sets will effectively act as noise in the training data set. This blending and bias adjustment is performed by a separate algorithm that was developed by S. Kidder of the Cooperative Institute for Research in the Atmosphere (CIRA) and is currently moving toward operational implementation at NESDIS.

To properly match the ABI predictors with the MW rain rates in space, differences in footprint size between the ABI and the MW sensors must be addressed. The procedure in this algorithm is to aggregate the ABI data onto the MW footprint, assuming perfectly round footprints with a diameter of 2 km for the ABI and 15 km for the MW. For each available MW footprint, those ABI footprints that at least partially overlap the MW footprint are identified, and the fraction of the coverage of the MW footprint by the ABI footprint is computed based on the location of the footprint centers and the above assumptions about footprint shape and size. The weight of each GOES footprint is

proportional to the total computed overlapping area with the MW footprint such that all of the weights add up to unity. For the aggregation process, the GOES brightness temperatures are converted to radiances and aggregated, and then the resulting radiance is converted back to brightness temperature. This prevents errors from the nonlinear relationship between radiance and brightness temperature.

Time matching of the ABI predictors with the MW rain rates should match the ABI and MW fields that are closest in time. For the current requirement of 15-min update frequency, the ABI and MW pixels should be within 7.5 minutes in time; if in the future the requirement is changed to a 5-min update cycle, the ABI and MW pixels should be within 2.5 minutes in time. Due to the latency of microwave data, previous ABI images will need to be available for matching with microwave data that may be up to 3 hours old at the time of availability.

Note that the predictors in this algorithm are not necessarily restricted to ABI data; predictors from other GOES-R instruments (e.g., lightning) can also be used, in addition to any other ancillary data that might prove to be relevant (e.g., stability profiles from numerical weather models). This point will be raised again in the subsections on training. The current list of predictors is presented in Table 5, where the subscript refers to the wavelength of the brightness temperature T; e.g., $T_{7.34}$ is the brightness temperature at 7.34 μm . Note that the matched MW-ABI data file contains the component SEVIRI brightness temperature values and also the derived values in Predictors 2 and 3; the additional predictors in Table 5 (i.e., brightness temperature differences) are computed internally by the calibration program to reduce the required size of the matched data files. The constant adjustments are performed in order to optimize the nonlinear predictor transformation described in Sections 3.1.4.3 and 3.4.2.2. In the case of brightness temperature differences, a constant is added in order to avoid negative values which have an undefined logarithm; in the case of the brightness temperature values, a constant is subtracted because lower (but positive) values are most sensitive to the nonlinear transformation.

Input ID	Description
1	$T_{6.2} - 174 \text{ K}$
2	$S = 0.568 \cdot (T_{\min,11.2} - 217 \text{ K}) + 25 \text{ K}$
3	$T_{\text{avg},11.2} - T_{\min,11.2} - (S - 25 \text{ K}) + 85 \text{ K}$
4	$T_{7.34} - T_{6.19} + 30 \text{ K}$
5	$T_{8.5} - T_{7.34} + 30 \text{ K}$
6	$T_{11.2} - T_{7.34} + 20 \text{ K}$
7	$T_{8.5} - T_{11.2} + 30 \text{ K}$
8	$T_{11.2} - T_{12.3} + 20 \text{ K}$
9-16	Nonlinear transformations of predictors 1-8

Table 5. Predictors computed from the data in the matched MW-ABI data file.

Note that if a particular pixel is present but there are missing pixels in the 5x5 neighboring region, T_{\min} (and hence S) will still be computed, but the missing pixels will

be ignored in calculating the predictors that require information from the neighboring region. However, if at least 3 of the 6 neighboring pixels (the closest 4 in the same scan line and the 2 corresponding pixels in the adjacent scan lines) used to compute G_t are missing G_t is not computed and is assigned a missing value of -999.0.

Finally, if any of these predictors has a value less than or equal to zero, the nonlinear transformation (see Section 3.4.1.3) cannot be performed and so no retrieval is performed for such pixels; their rainfall rate value is set to a missing value of -999.0.

The performance of the rain rate algorithm has been shown to improve when the data are divided into classes that can be determined a priori from available data. Specifically, the data have been divided into 4 latitude regions (60-30°S, 30°S-EQ, EQ-30°N, and 30-60°N) and further into 3 rainfall types according to selected ABI brightness temperature values from the matched MW-ABI data set:

- Type 1 (water cloud): $T_{7.34} < T_{11.2}$ and $T_{8.5} - T_{11.2} < -0.3$
- Type 2 (ice cloud): $T_{7.34} < T_{11.2}$ and $T_{8.5} - T_{11.2} \geq -0.3$
- Type 3 (cold-top convective cloud): $T_{7.34} \geq T_{11.2}$

These types were determined by experimenting with the changes in the relationship between $T_{11.2}$ and rainfall rate ($T_{11.2}$ is the band typically used for rain rate retrievals because of its sensitivity to cloud-top temperature with relatively small water vapor effects) for various brightness temperature threshold difference regimes. The thresholds are the points at which this relationship changes significantly from one side of the threshold to the other, which implies that they represent significantly different regimes for rainfall rate retrieval purposes. This results in a total of 12 rainfall classes in the algorithm which are summarized in Table 6. Separate files of matched MW rain rates and ABI data are maintained for each class.

Class	Latitude	Cloud Type
1	60-30°S	1 (water)
2	60-30°S	2 (ice)
3	60-30°S	3 (convective)
4	30°S-EQ	1 (water)
5	30°S-EQ	2 (ice)
6	30°S-EQ	3 (convective)
7	EQ-30°N	1 (water)
8	EQ-30°N	2 (ice)
9	EQ-30°N	3 (convective)
10	30-60°N	1 (water)
11	30-60°N	2 (ice)
12	30-60°N	3 (convective)

Table 6. Description of rainfall classes.

Note that if a pixel falls exactly on the boundary between two latitude regions, it is assigned to the higher of the two regions. For instance, a pixel at exactly 30.0°N will be assigned to one of classes 10-12, not to classes 7-9. Note also that if any of the three ABI brightness temperature values used to determine cloud type ($T_{7.34}$, $T_{8.5}$, or $T_{11.2}$) is missing for a given pixel, the rainfall class will be set to zero and no retrieval will be performed for that particular pixel (i.e., its value will be set to -999.0).

Separate matched data sets are maintained for each rainfall class, with the data points in reverse time order (i.e., the most recent data in the first record and the oldest data in the last record). These matched data sets are rolling-value data sets; i.e., older data are cycled out of the end of the file as newer data are brought in to the beginning of the file keep the data set up-to-date. Initial work with training data sets covering a fixed period of time (e.g., 24 hours) proved to be unsuitable because time variations in the intensity distribution of rainfall would affect the robustness of the calibration. For example, if an extended period of light rain or no rain were followed by heavy rain, the algorithm might be preferentially calibrated for light rain and thus perform poorly when the heavy rain began. To ensure a training data set that contains enough raining pixels for reliable results but is still short enough to reflect recent conditions, the number of raining (>1 mm/h) data points in the training data files are kept fixed. Specifically, as newer data become available and are added to the training data file, the oldest data points are removed until the number of raining pixels returns to the same value as before the newer data were added. Sensitivity studies showed that the best results were obtained when 5,000 raining pixels were required for Type 1 and Type 2 clouds and 1,000 raining pixels were required for Type 3 clouds. It should be noted that the number of training pixels read is controlled by the training code rather than by the file size; i.e., the training code will read through the training file (i.e., backward in time) until the required number of raining pixels is read. Therefore, it is not critical for the training file to be trimmed strictly to the required number of training pixels as extra pixels will simply be ignored.

It is important to note that the training of the coefficients and the retrieval of the actual rain rates can be done in parallel to save time: the coefficients would be updated whenever new target data become available, and then these updated coefficients would be used in retrieving the rain rates from the next available set of ABI imagery.

3.4.1.2 Rainfall Detection

The objective of the rainfall detection portion of the algorithm is to separate clear and cloudy but non-raining pixels from raining pixels. The training of the separation portion of the algorithm is applied to each cloud class separately using discriminant analysis (similar to linear regression but with a binary predictand—the value is 1 if the MW rain rate exceeds 1.0 mm/h and 0 otherwise. Note that the 1.0 mm/h threshold is used because of significant differences among MW instruments in sensitivity to drizzle and very light rain). For each of the 12 algorithm classes, discriminant analysis is performed using each of the available predictors, and the predictor that produces the best Heidke Skill Score (HSS; see Section 3.4.2.1 for definition) for rain / no rain discrimination (compared to

observations) is selected. This first predictor is then combined with all of the remaining predictors for a second set of tests to determine the best 2-predictor combination. The result is an equation for linearly combining one or two predictors, plus a threshold value above which the pixel is considered to be raining.

This threshold value is then optimized to produce an unbiased result, since experience has shown that the best HSS value is not necessarily associated with minimum bias. Specifically, the maximum and minimum values for the rain / no rain discriminator are computed, and for 1,000 intervals of equal size the bias is computed for each selected rain / no rain discriminator value (such that no pixels would be classified as raining when the minimum threshold value is used and all of them would be classified as raining when the maximum threshold value is used). A binary search is then used to identify the threshold value with a bias closest to unity (i.e., the number of pixels in the training data that are classified as raining by the scheme is as close as possible to the number of raining microwave pixels in the training data), and this is the threshold value that is used.

Once the predictors and coefficients have been selected and the threshold value has been determined, the resulting predictor ID's and coefficients are then written to a file for use by the prediction program. A more detailed description of this process is contained in Section 3.4.2.1.

3.4.1.3 Rainfall Rate Estimation

The objective of the rainfall rate portion of the algorithm is to determine rainfall rates for those pixels that were classified as raining by the discriminant analysis scheme, with separate equations for each algorithm class. Consequently, only those pixels that have non-zero target rainfall rates are used in developing the equations for retrieving rainfall rates.

Since the relationship between many of the predictors (e.g., IR window brightness temperature) and rainfall rates is known to be nonlinear, the first step is to supplement the predictor set with a second set of predictors that represent optimal nonlinear transformations of the original set: for each algorithm class, each original predictor is re-scaled to eliminate negative values, and then the predictors and target rain rates are regressed against each other in \log_{10} - \log_{10} space. The resulting slope and intercept become an exponent and multiplier in linear space, and this slope and intercept are used to create a nonlinear transformation of each predictor.

After creating the set of transformed predictors for each algorithm class (which are included with the original predictors in the predictor pool), all of the predictors are evaluated via linear regression against the target rain rates, with separate regressions performed for each algorithm class. The predictor that has the best correlation with the target rain rates is then combined with each of the remaining predictors, and the 2-predictor combination that produces the best correlation with the target data is selected. After this is done, a preliminary set of rain rates is retrieved and compared to the microwave training data to derive a set of coefficients for adjusting the retrieved rain fall rate distribution to match the distribution of the microwave rainfall rates. All of the

required coefficients are then written to a file for use by the prediction routine. A more detailed description of this process is contained in Section 3.4.2.2.

3.4.1.4 Independent rainfall rates

The equations produced by the calibration of the rainfall detection and rainfall rate estimation are used to derive the rainfall rates from current ABI imagery that comprise the algorithm output at the full ABI resolution. Note that the rainfall rates are produced using different equations for each of the 12 classes.

3.5 Mathematical description

3.5.1 ABI-MW footprint matching

The starting point for matching the ABI and MW footprints is to determine the distance between the centers of the two. The assumption of a spherical Earth (since the distances involved are less than 20 km) leads to the equation

$$Distance = \{[(lon_{ABI} - lon_{MW}) * \cos(lat_{MW})]^2 + [(lat_{ABI} - lat_{MW})]^2\} * 111 \text{ km} \quad (1)$$

where lat_{ABI} and lat_{MW} are the latitudes of the centers of the ABI and MW footprints, and lon_{ABI} and lon_{MW} are the corresponding longitudes. Presuming the distance is less than the sum of the two radii ($r_{ABI}=1$ km, $r_{MW}=7.5$ km; both footprints assumed to be perfect circles), so that the two circles have a common area. Two possible scenarios exist in that case: if the distance is less than the radius of the smaller (ABI) radius, then the ABI footprint is contained entirely within the MW footprint and so the common area is simply the ABI footprint area (πr_{ABI}^2). If the distance is greater than the ABI radius but smaller than the sum, then the common area of the two circles is expressed as

$$Area = r_{MW}^2 * (\alpha - \cos(\alpha) \sin(\alpha)) + r_{ABI}^2 * (\beta - \cos(\beta) \sin(\beta)) \quad (2)$$

where

$$\alpha = \arccos \left[\frac{-r_{ABI}^2 + distance + r_{MW}^2}{2 * r_{MW}^2 * distance} \right] \quad (3)$$

and

$$\beta = \arcsin \left[\frac{r_{MW} * \sin(\alpha)}{r_{ABI}} \right] \quad (4)$$

and

$$\beta = \pi - \beta \text{ if } distance < r_{MW} \cos(\alpha) \quad (5)$$

The weight of each ABI pixel in computing the MW value is simply the common area of the ABI and MW pixel divided by the total area of ABI pixels within the MW pixel; thus, the weights for each MW pixel always add up to unity. In other words, for each ABI

pixel i among a total of n ABI pixels (or partial pixels) within the MW pixel of interest, the weight w_i is computed as

$$w_i = \frac{Area_i}{\sum_{i=1}^n Area_i} \quad (6)$$

3.5.2 Calibration: Rain / no rain discrimination

Mathematically, a special case of multiple linear regression called discriminant analysis (in which the target values are 0 and 1 instead of continuous values) is used to separately calibrate the rain /no rain discrimination for each algorithm class. A two-predictor additive multiple regression model is used for each algorithm class c :

$$y_c = b_{c,0} + b_{c,1}x_{c,1} + b_{c,2}x_{c,2} + \varepsilon_c \quad (7)$$

where y is the target MW rain rate or rain / no rain value; the x 's are the two selected ABI-derived predictors; the b 's are the calibration coefficients; and ε is the residual error, which is to be minimized by solving the following system of normal equations for the coefficients b_0 , b_1 , and b_2 :

$$\begin{aligned} \sum_{j=1}^{n_c} y_{c,j} &= b_{c,0}n_c + b_{c,1} \sum_{j=1}^{n_c} x_{c,1,j} + b_{c,2} \sum_{j=1}^{n_c} x_{c,2,j} \\ \sum_{j=1}^{n_c} x_{c,1,j}y_{c,j} &= b_{c,0} \sum_{j=1}^{n_c} x_{c,1,j} + b_{c,1} \sum_{j=1}^{n_c} x_{c,1,j}^2 + b_{c,2} \sum_{j=1}^{n_c} x_{c,1,j}x_{c,2,j} \\ \sum_{j=1}^{n_c} x_{c,2,j}y_{c,j} &= b_{c,0} \sum_{j=1}^{n_c} x_{c,2,j} + b_{c,1} \sum_{j=1}^{n_c} x_{c,1,j}x_{c,2,j} + b_{c,2} \sum_{j=1}^{n_c} x_{c,2,j}^2 \end{aligned} \quad (8)$$

where the first part of the subscript is the algorithm class, the second is the predictor number, and the third part is the data point number (all sums are over the total number of data points in the class n_c .)

The calibration procedure begins by solving the 1-predictor version of these equations (i.e., no terms containing $b_{c,2,j}$ or $x_{c,2,j}$) for each of the first 8 (untransformed) predictors in the matched data set (see Table 5). Since the outputs of these equations will be continuous (i.e., non-binary), a threshold value must be determined for converting the output to binary values: outputs above the threshold are assigned a value of 1 (rain) and outputs below are assigned a value of 0 (no rain). This threshold is selected to produce the minimum amount of bias; i.e., the number of pixels in the training data set that are classified as raining should match as closely as possible the actual number of raining microwave pixels in the training data set. In order to do this, the equation outputs are computed for each training pixel and the highest and lowest values are preserved. This range of values is divided into 1000 equally spaced intervals, and for each of these 1000 threshold values the number of pixels classified as raining is computed. Since this number decreases monotonically as the threshold value increases (the threshold value must be exceeded), a simple binary search can then be used to identify the threshold

value that produces the best match to the microwave data in terms of the number of raining pixels.

Once the bias has been optimized, the HSS is computed for that particular predictor, and the predictor the highest Heidke Skill Score (HSS) is selected. The HSS is computed as follows:

$$HSS = \frac{2(c_1c_4 - c_2c_3)}{(c_1 + c_2)(c_2 + c_4) + (c_3 + c_4)(c_1 + c_3)} \quad (9)$$

where c_1 is the number of correct no-rain estimates, c_2 is the number of false alarms (i.e., the estimate has rain but the observation has no rain); c_3 is the number of failed detections (i.e., the estimate has no rain but the observation has rain); and c_4 is the number of correct rain estimates for the class of interest. Higher HSS values indicate greater skill, with 1 as a perfect value (i.e., $c_2=c_3=0$).

After the first predictor is selected, the procedure is repeated for each two-predictor pair containing the first selected predictor to obtain two predictors and the associated coefficients from Eq. (7) plus the threshold value for converting the continuous equation output into binary form.

It should be noted that the selection of only 2 predictors was the result of experimentation—additional predictors were shown not to have a positive impact on the performance of the algorithm. This may be due to the high degree of correlation among the visible and IR bands when depicting optically thick clouds.

3.5.3 Calibration: Rainfall rate

For each class, a separate pair of predictors is selected for the rain rate retrieval, using Eqs. (1)-(2) as the basis for selection but with continuous output. Prior to selection, the set of 8 predictors is supplemented by a set of non-linear transformations (see Table 5). These nonlinear transformations for each predictor p and class c $x_{p,c}^T$ use the power function; i.e.,

$$x_{p,c}^T = \alpha_{p,c} x_{p,c}^{\beta_{p,c}} \quad (10)$$

where the coefficients $\alpha_{p,c}$ and $\beta_{p,c}$ are found by solving the equation

$$\log_{10} y = \log_{10} \alpha_{p,c} + \beta_{p,c} \log_{10} x^T. \quad (11)$$

Solving this equation separately for each predictor and class yields the following least-squares solutions:

$$\beta_{p,c} = \frac{n_{p,c} \sum_{i=1}^{n_{p,c}} (\log_{10} x_{p,c,i}) (\log_{10} y_{c,i}) - \sum_{i=1}^{n_{p,c}} (\log_{10} x_{p,c,i}) \sum_{i=1}^{n_{p,c}} (\log_{10} y_{c,i})}{n_{p,c} \sum_{i=1}^{n_{p,c}} (\log_{10} x_{p,c,i})^2 - \left(\sum_{i=1}^{n_{p,c}} \log_{10} x_{p,c,i} \right)^2} \quad (12)$$

$$\log_{10} \alpha_{p,c} = \frac{\sum_{i=1}^{n_{p,c}} (\log_{10} y_{c,i} - \log_{10} x_{p,c,i})}{n_{p,c}} \quad (13)$$

For each predictor p in each class c , the coefficients $\alpha_{p,c}$ and $\beta_{p,c}$ are solved for by applying Eqs. (12) and (13) using the predictor values $x_{p,c,i}$ and the corresponding target microwave rainfall rates $y_{c,i}$. However, since the equation form in Eq. (10) has no intercept, it is constrained to pass through the origin (0,0), so a modified version was developed.

$$x_{p,c}^T = 10^{\alpha_{p,c}} \left(x_{p,c} + 1 + \gamma_{p,c} \right)^{\beta_{p,c}} - 1 \quad (14)$$

The third unknown ($\gamma_{p,c}$) cannot be solved for with only two equations, so the equation is optimized using a “brute force” approach. First, the value of $\gamma_{p,c}$ is initially set to 0 and the equation is solved using Eq. (12) and (13). The value of $\gamma_{p,c}$ is then incremented by 25 and Eq. (12) and (13) are solved again; i.e., the value of $\gamma_{p,c}$ is added to the each predictor value $x_{p,c}$ when solving the equation. The Pearson correlation coefficient is then computed for the transformed data:

$$\text{Correlation} = \frac{\text{cov}(x, y)}{\sigma_x \sigma_y} \quad (15)$$

where $\text{cov}(x,y)$ is the covariance of the predictor and target data, and σ_x and σ_y represent the standard deviations of the predictor and target data, respectively. The predictor data in this case consists of the values of the transformed predictors (i.e., $x_{p,c}^T$) and the target data consist of the microwave rainfall rates against which they have been matched (i.e., y_c).

If the equation fit (as measured by correlation coefficient) is improved, the value of $\gamma_{p,c}$ is incremented by 25 and the process is repeated (but is stopped if it reaches 2500 to avoid non-convergence). If the equation fit is degraded (i.e., lower correlation coefficient), then the process stops and the previous value of γ is used. This process of determining the coefficients $\alpha_{p,c}$, $\beta_{p,c}$, and $\gamma_{p,c}$ is repeated for each predictor and each class and applied to create the supplemental set of 8 transformed predictors for each class.

The total set of 16 predictors is then used for calibrating the rainfall rate retrieval. For each class, each predictor is first regressed against the target rain rates using Eq. (8), and the predictor with the highest Pearson correlation coefficient against the target microwave rainfall rates is selected.

The process is then repeated with the 15 possible combinations of the selected first predictor and each of the remaining predictors, and the pair with the highest correlation against the target data is selected as the predictor pair (and associated coefficient set) for that class.

As with the rain / no rain discrimination, experiments showed no positive impact from using more than two predictors, again perhaps because of the high degree of correlation among the visible and IR bands for optically thick clouds.

Previous work has shown that the rainfall rates retrieved using this approach generally exhibit a strong systematic dry bias—too wet for low rainfall rates and much too dry for higher rainfall rates. This is believed to be the result of significant scatter in the training data caused by spatial displacements between the coldest cloud tops and the heaviest rainfall rates. To address this problem, an adjustment for the retrieved rainfall rate is derived that adjusts its distribution to match the training microwave rainfall rates.

Specifically, for each rainfall class, the rainfall rates are retrieved using the coefficients derived above, and then are sorted from lowest to highest and matched against the training rainfall rates which have also been independently sorted lowest to highest. The result of this match is a lookup table (LUT) whereby the value of the retrieved rainfall rate is converted to the value of the corresponding microwave rainfall rate so that the distribution of the retrieved rainfall rate will match that of the microwave rainfall rates.

To create a useful LUT, linear interpolation is used to create a table with evenly spaced increments of 0.1 mm/h for the training rainfall rates. In addition, since the MW rainfall rates have a lower dynamic range (in part due to their coarser spatial resolution), but since extrapolation of the data could produce non-physical results, for all values between 50 mm/h (the maximum rainfall rate from the TRMM Microwave Imager) and 100 mm/h the input and output values are set equal to one another (i.e., a retrieved rainfall rate of 75 mm/h will be mapped to a final rainfall rate of 75 mm/h). Linear interpolation is then performed between the data point with the highest rainfall rates and the (50 mm/h, 50 mm/h) data point. This LUT is then written to the end of the retrieval coefficient table in Table 4, Section 3.2.2.

3.5.4 Application to independent data

The predictors and coefficients obtained during the calibration outlined in the previous two subsections are then applied to the current ABI imagery using Equation (1) with the appropriate coefficients and predictor values. To avoid the potential data quality issues with any pixel that is missing one but not all bands, if any of the input bands has a value less than 120 K, the value of the 10.8-micron (band 14) brightness temperature is set to the missing value of -999.0. The rain /no rain discriminator is then computed. For values below the threshold, a rain rate of zero is assigned; for values above the threshold, the rainfall rate is computed using Eq. (1) with the rain rate coefficients (and predictor transformations from Eq. (4) as needed), followed by the distribution adjustment in Eq. (10).

3.6 Algorithm Output

The final output of this algorithm is the Rainfall Rate product—a field of instantaneous rainfall rates in mm/h (rounded to the nearest integer) at the same resolution as the ABI IR data—2 km at nadir. Note that for output purposes these values are converted to short integers and multiplied by 10, so the effective values in the output files are tenths of mm per hour. This product will also be accompanied by a grid of corresponding quality flags, with values of 0 for good data and non-zero for data that are of questionable quality due to deficiencies in the input data, as described in Table 7:

Byte	Bit	Flag	Source	Value
0	0	Rainfall Rate output	RR	1=bad data; 0=OK
	1	Local zenith angle block-out zone	SDR	1=local zenith angle>70° or lat>60°; 0=OK
	2	Bad input data for 1 st rain / no rain predictor	SDR and RR	1=bad data; 0=OK
	3	Bad input data for 2 nd rain / no rain predictor	SDR and RR	1=bad data; 0=OK
	4	Bad input data for 1 st rain rate predictor	SDR and RR	1=bad data; 0=OK
	5	Bad input data for 2 nd rain rate predictor	SDR and RR	1=bad data; 0=OK
	6	Retrieval coefficients missing	RR	1=no retrieval coefficients; 0=OK
	7	Not used		

Table 7. Quality flags for the Rainfall Rate product.

Note that if any of bits 2-6 are set to 1 that all of them should be set to 1.

In addition, two quality information fields will be output: a gridded file containing flags indicating if the rainfall rate values were truncated at 0 mm/h or at 100 mm/h (Table 8) and a gridded file containing the rainfall class (1-12) of a particular pixel (Table 9):

Byte	Bit	Flag	Source	Value
0	0	Rain rate > 100 mm/h	RR	1=rain rate >100 mm/h but truncated at 100 mm/h; 0=rain rate <100 mm/h
	1	Rain rate < 0 mm/h	RR	1=rain rate <0 mm/h but truncated at 0 mm/h; 0=rain rate >0 mm/h

Table 8. Diagnostic information for the Rainfall Rate product.

Grid	Field	Source	Value
1	Precipitation class identifier	RR	Value of rain class, ranging from 1 to 12

Table 9. Gridded quality information for the Rainfall Rate product.

Finally, the metadata file will contain the information listed below in Table 10:

Type	Variable
Float	Total rain area (<i>number of pixels in image with rain rates > 1 mm/h</i>)
Float	Total rain volume (<i>total rain in rain area, mm/h</i>)
Long	Total number of pixels with a successful retrieval (i.e., bit 0 in
Long	Number of QA flag values: 8
Long	Number of retrievals with QA flag value 0 (<i>all bits set to 0</i>)
String	Definition of QA flag value 0: <i>Good rain rate retrieval</i>
Long	Number of retrievals with QA flag bit 0 set to 1
String	Definition of QA flag with bit 0 set to 1: <i>Bad rain rate retrieval</i>
Long	Number of retrievals with QA flag bit 1 set to 1
String	Definition of QA flag with bit 1 set to 1: <i>Local zenith angle block-out zone</i>
Long	Number of retrievals with QA flag bit 2 set to 1
String	Definition of QA flag with bit 2 set to 1: <i>Bad input data for 1st rain / no rain predictor</i>
Long	Number of retrievals with QA flag bit 3 set to 1
String	Definition of QA flag with bit 3 set to 1: <i>Bad input data for 2nd rain / no rain predictor</i>
Long	Number of retrievals with QA flag bit 4 set to 1
String	Definition of QA flag with bit 4 set to 1: <i>Bad input data for 1st rain rate predictor</i>
Long	Number of retrievals with QA flag bit 5 set to 1
String	Definition of QA flag with bit 5 set to 1: <i>Bad input data for 2nd rain rate predictor</i>
Long	Number of retrievals with QA flag bit 6 set to 1
String	Definition of QA flag with bit 6 set to 1: <i>Retrieval coefficients missing</i>
String	Definition of Rain Class 1: <i>Water cloud; 60-30 S latitude</i>
String	Definition of Rain Class 2: <i>Ice cloud; 60-30 S latitude</i>
String	Definition of Rain Class 3: <i>Overshooting top cloud; 60-30 S latitude</i>
String	Definition of Rain Class 4: <i>Water cloud; 30 S latitude-EQ</i>
String	Definition of Rain Class 5: <i>Ice cloud; 30 S latitude-EQ</i>
String	Definition of Rain Class 6: <i>Overshooting top cloud; 30 S latitude-EQ</i>

String	Definition of Rain Class 7: <i>Water top cloud; EQ-30 N latitude</i>
String	Definition of Rain Class 8: <i>Ice top cloud; EQ-30 N latitude</i>
String	Definition of Rain Class 9: <i>Overshooting top cloud; EQ-30 N latitude</i>
String	Definition of Rain Class 10: <i>Water cloud; 30-60 N latitude</i>
String	Definition of Rain Class 11: <i>Ice top cloud; 30-60 N latitude</i>
String	Definition of Rain Class 12: <i>Overshooting top cloud; 30-60 N latitude</i>

Table 10. Metadata for the Rainfall Rate product.

Additional diagnostic information will be provided by the corresponding retrieval coefficient table (Table 4, Section 3.3.2). All of these fields will be delivered to the GOES-R Archive System (GAS) and the Comprehensive Large Array-data Stewardship System (CLASS) as well as being retained in 2-day local storage for diagnostic purposes.

4 TEST DATA SETS AND OUTPUTS

4.1 Simulated/Proxy Input Data Sets

As described below, the data used to test the Rainfall Rate Algorithm included SEVIRI observations and blended MW rainfall rates. The test period chosen was the 6th through the 9th of January, April, July, and October 2005. While SEVIRI is obviously not operating over the GOES domain and does not have the exact same spectral coverage and spatial resolution, for rainfall rate applications, it is still preferred over simulated ABI data for objective validation, given the errors exhibited by the latter in depicting the correct intensity and location of precipitation features. The rest of this section describes the proxy and validation data sets used in assessing the performance of the Rainfall Rate Algorithm.

4.1.1 SEVIRI Data

SEVIRI provides 11 spectral channels with a spatial resolution of 3 km at nadir (coarser than the 2-km resolution of the ABI) and a temporal resolution of 15 minutes, and thus represents the best source of data currently available for testing and developing the Rainfall Rate Algorithm. Figure 2 is a full-disk SEVIRI image from 1200 UTC on January 7, 2005. The SEVIRI data was provided by the GOES-R Proxy Data Team.

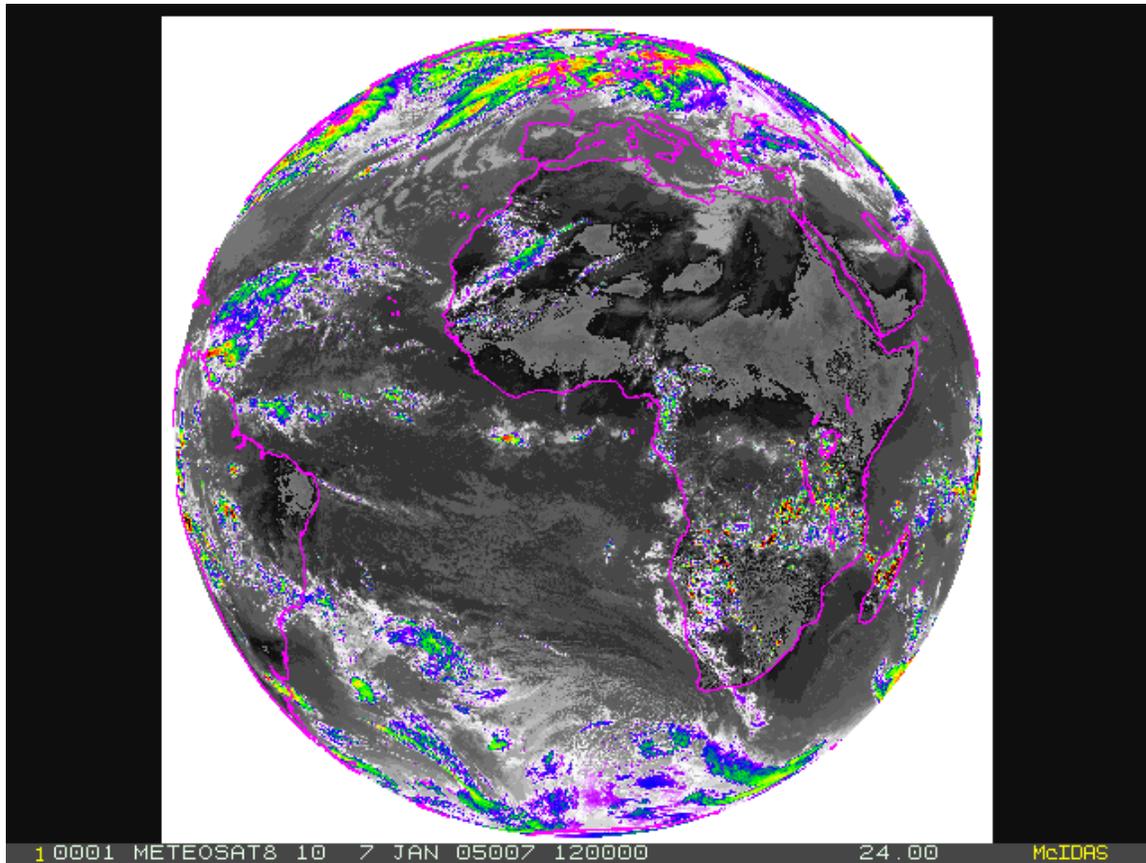
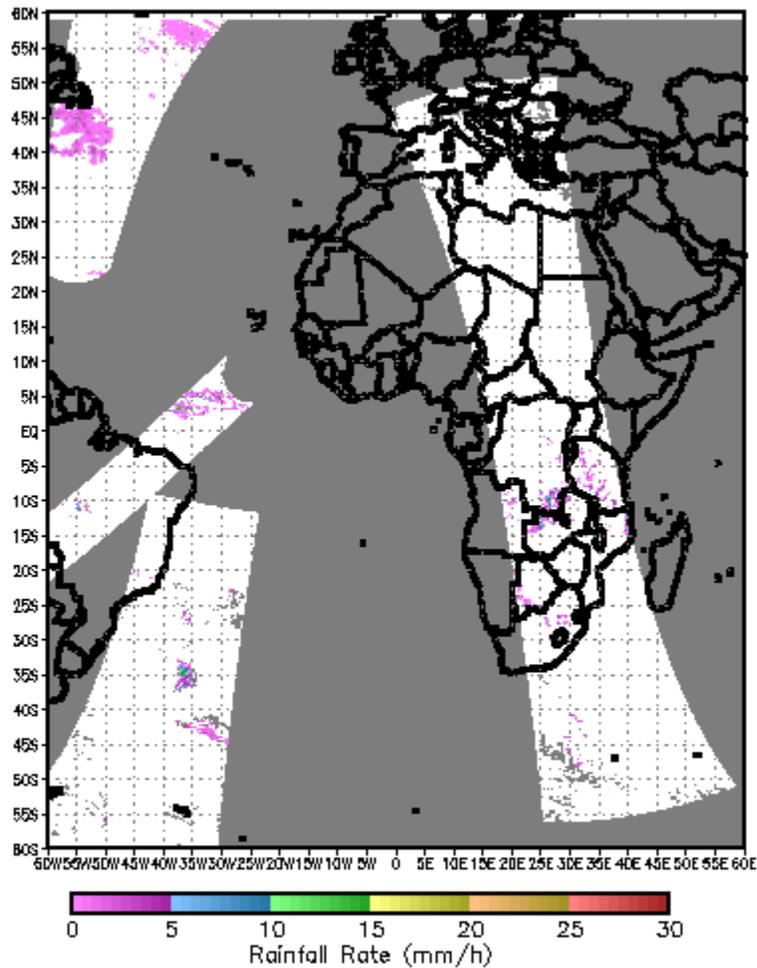


Figure 2. Full disk 10.8- μm color-enhanced image from SEVIRI for 1200 UTC on January 7, 2005.

4.1.2 Microwave-Derived Rainfall Rates

The target data for calibration is a blend of MW rain rates from multiple Special Sensor Microwave/Imager (SSM/I) and Advanced Microwave Sounding Unit (AMSU) sensors, plus the Advanced Microwave Scanning Radiometer-EOS (AMSR-E) and the Tropical Rainfall Measuring Mission (TRMM) Microwave Imager (TMI), all of which were reprojected onto a common lat/lon grid. These fields are available every half hour, and include data from all of the MW overpasses during that time period. An example is shown in Fig. 3. It should be noted that the testing and validation results in this ATBD were obtained using a somewhat different matched MW rain rate data set than the one planned for operations and described in Section 3.3.2; this is because the latter data set was not yet in operational production at the time of this writing and thus was not available for use in the initial testing and validation.



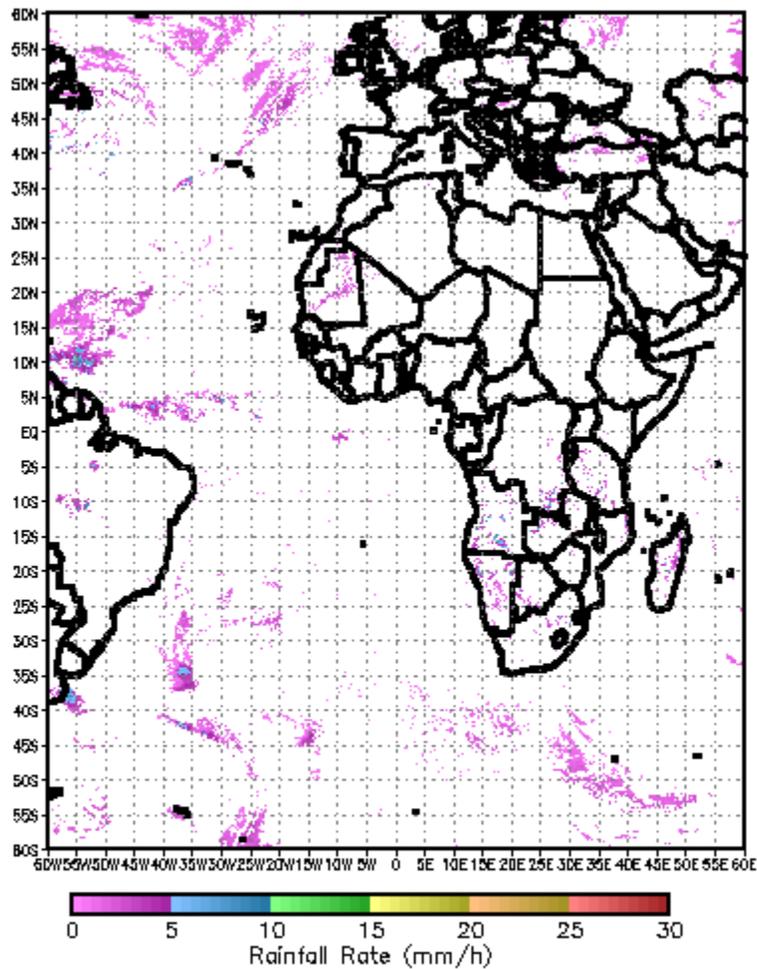
GRADS: COLA/IBES

2008-07-17-15:57

Figure 3. Blended MW rainfall rates covering 1230-1300 UTC 7 January 2005.

4.2 Output from Proxy Input Data Sets

As part of an intercomparison exercise to select which algorithm would be used for GOES-R, the rain rate algorithm was applied to SEVIRI data from the 6th through the 9th of January, April, July, and October 2005. The same time periods were used for the evaluation of algorithm precision and accuracy, along with all of January 2008. Figure 4 is an example of output from the Rainfall Rate Algorithm described in this document.



GRADS: COLA/IBES

2008-07-17-18:02

Figure 4. Sample rain rate algorithm output from 1245 UTC January 7, 2005.

4.2.1 Precision and Accuracy Estimates

The F&PS specifications for the Rainfall Rate algorithm (see Table 1) refer to instantaneous rainfall rates, so radar data (both space-based and ground-based) must be heavily relied on since gauges generally do not provide reliable information on instantaneous rates. However, such data are very difficult to obtain over Europe and Africa. Comparisons will be made against Tropical Rainfall Measuring Mission (TRMM) Precipitation Radar (PR) data and Nimrod radar data over Western Europe obtained from the British Atmospheric Data Centre (BADC).

Because slight errors in the spatial location of rainfall can significantly degrade statistics that are computed on a pixel-by-pixel basis (e.g., Ebert 2008), a “fuzzy” approach for fine-scale rainfall validation has been selected whereby the Rainfall Rate pixel is compared with the pixel within a 10-km radius that has the most similar value rather than with the directly corresponding pixel. This is a variant of the “single observation – neighborhood forecast” strategy described in Ebert (2008).

4.2.1.1 Validation against TRMM PR

Validation against the 5-km resolution TRMM PR 2A25 product was performed for the 5th-9th of January, April, July, and October 2005 plus all of January 2008. For illustration purposes, Fig. 5 shows the coverage of the TRMM PR during a typical 24-hour period.

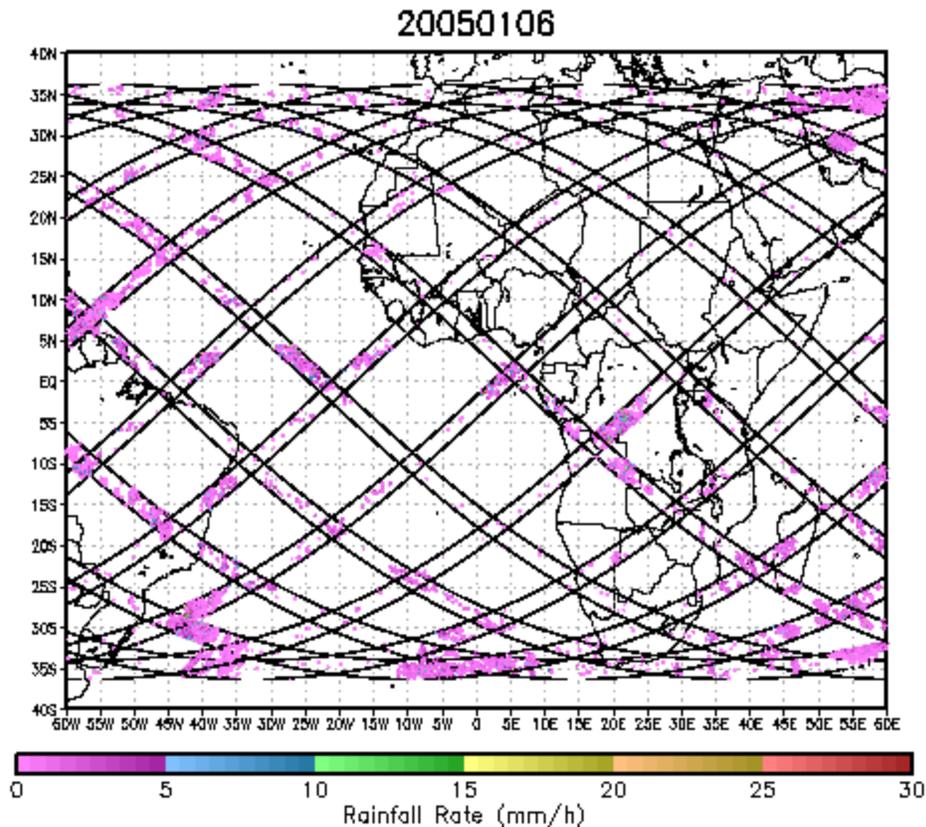
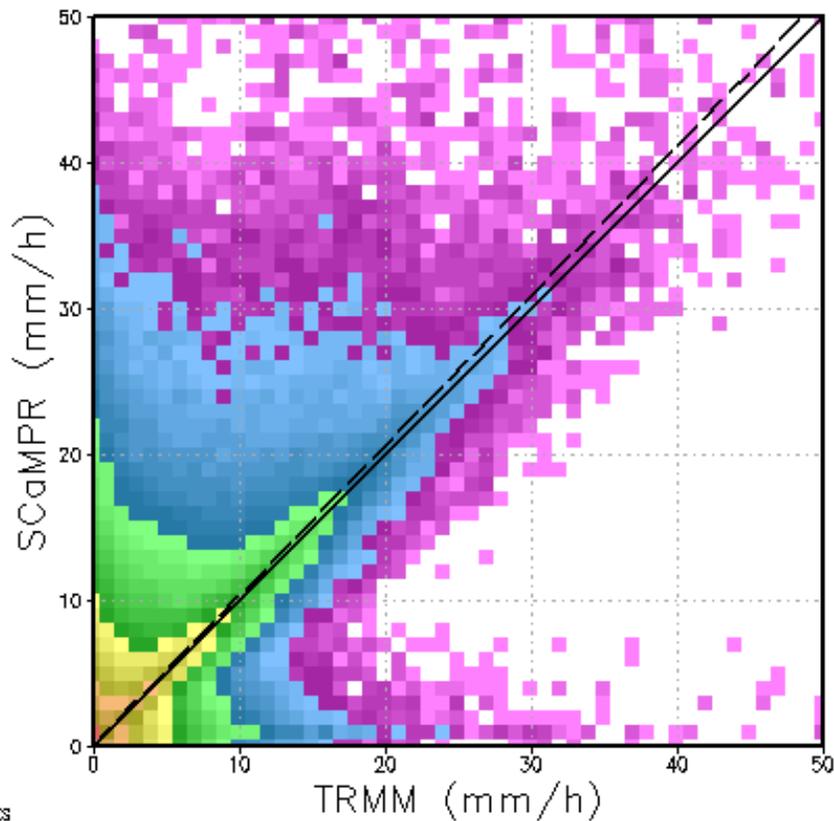


Figure 5. Coverage of TRMM PR data during October 6, 2005.

Figure 6 shows a scatterplot of the rainfall rates that was created using the “fuzzy” verification method described in the previous section, with the density of points indicated by color (red=more dense; purple=less dense) to eliminate the visually misleading effect of multiple overlapping points. The Rainfall Rate product displays a significant wet bias, but the best-fit line still corresponds quite well with the 1:1 line (i.e., a high occurrence of matching values between the estimates and observations).



GRADS: COLA/IBES

Figure 6. Scatterplot of Rainfall Rate algorithm vs. collocated TRMM rain rates; colors are related to pixel density with red highest and purple lowest. Solid line is the 1:1 line and the dashed line is the best-fit line.

Because rainfall rates are highly skewed toward low rates that are of much less hydrometeorological interest than higher rain rates, the F&PS precision spec focuses on the performance of the algorithm for rain rates of 10 mm/h (see Table 1). Specifically, when the rain rate at a given pixel is 10 mm/h, the corresponding observed value should be within 9 mm/h (i.e., within the range of 1 – 19 mm/h) 68% of the time. This appears to be a very loose requirement compared to most products; however, instantaneous rainfall rate is far more variable in both space and time than most other geophysical parameters, and even a “fuzzy” validation approach will not resolve this without using a significantly larger radius than the 10-km radius used for this algorithm.

The performance of the algorithm against the F&PS precision spec is illustrated in Fig. 7 by the cumulative distribution function (CDF) of absolute error in the Rainfall Rate product (the values along the abscissa) with respect to the TRMM PR for only those pixels with algorithm values between 9.5 and 10.5 mm/h (there are too few pixels with rain rates of exactly 10.0 mm/h to enable a statistically significant analysis). These errors were computed using the “fuzzy” verification strategy described in the previous section.

The dashed line indicates that 68% of the errors are below 8.9 mm/h, which is within the spec value of 9.0 mm/h.

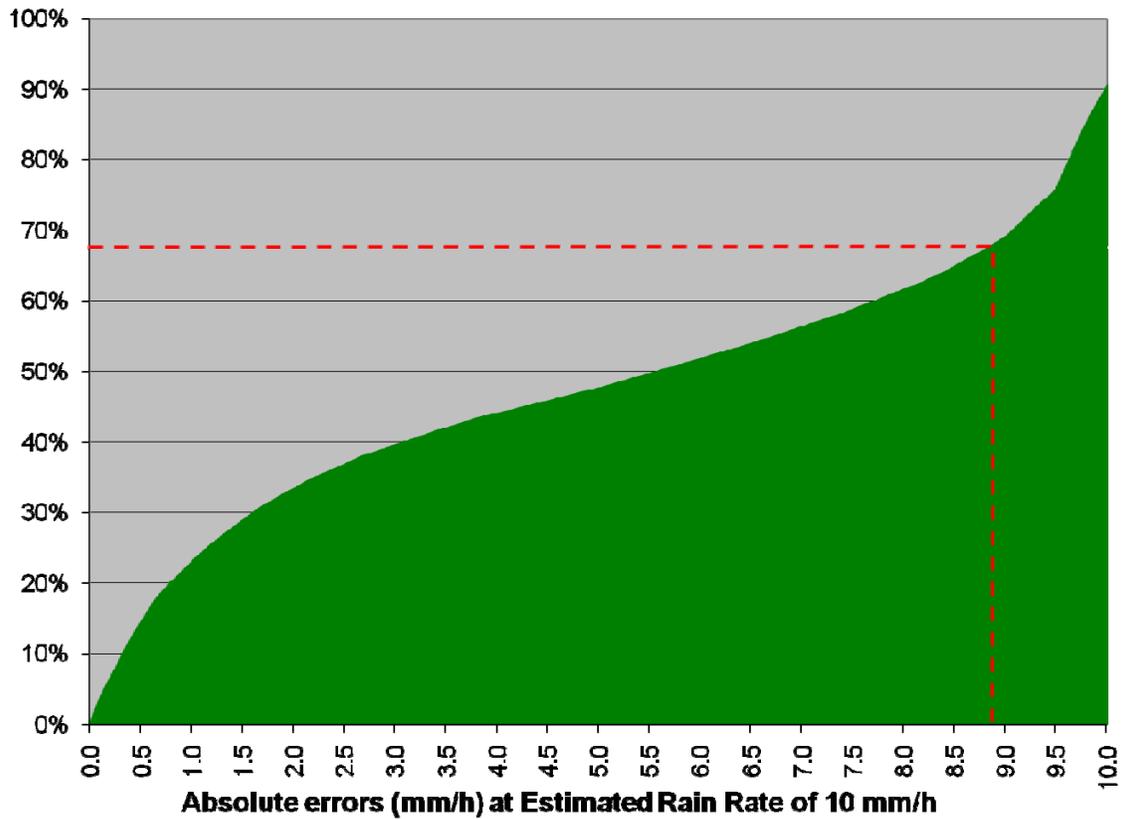


Figure 7. CDF of errors of Rainfall Rate product with rates of 9.5-10.5 mm/h vs. TRMM.

4.2.1.2 Validation against Nimrod data

Validation against the 5-km Nimrod composite radar product was performed for the 5th-9th of April, July, and October 2005 (January 5-9 was not available from the BADC archive) and all of January 2008. The coverage of these radars is illustrated in Fig. 8.

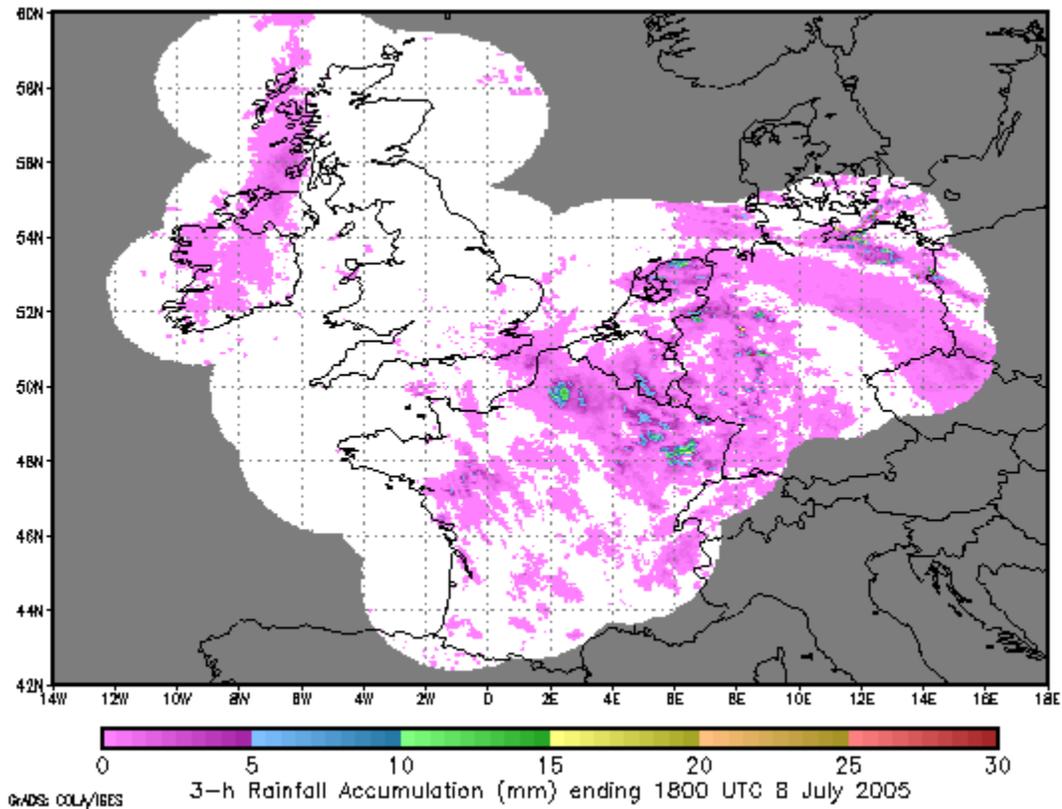
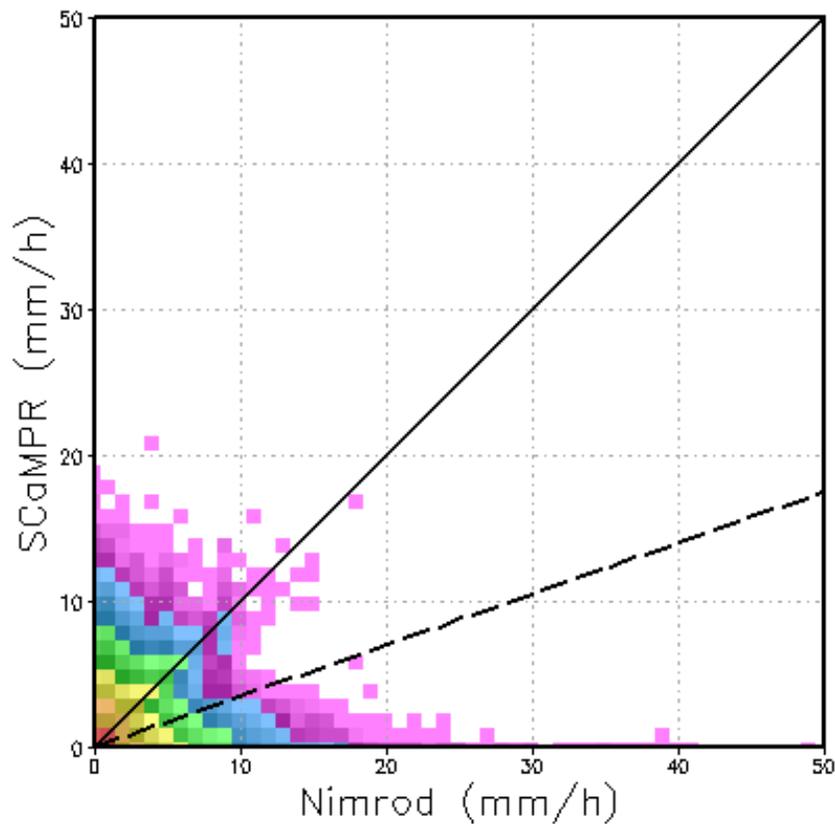


Figure 8. Coverage of Nimrod mosaic radar data.

Figure 9 shows a scatterplot analogous to Fig. 6 (and using the same “fuzzy” validation as against TRMM PR), but here the Rainfall Rate product displays a very strong systematic dry bias (indicated by the small slope of the dashed best-fit line), which in this case translates into a volume bias of approximately 14% (i.e., the retrieved rainfall volume is 14% lower than the Nimrod volume). This is not at all unexpected given that IR-based algorithms can exhibit significant detection problems for stratiform precipitation at higher latitudes.



GRADS: COLA/IBES

Figure 9. Same as Fig. 6, but vs. Nimrod.

Figure 10 is analogous to Fig. 7, except that it is for Nimrod. As the red dashed line indicates, approximately 68% of the errors are below 9.8 mm/h, which is outside the spec value of 9.0 mm/h; however, the requirement is for quantitative validation only for convective precipitation, and Western Europe is a region that is largely dominated by stratiform rainfall. A comparison of Fig. 10 with Fig. 7 illustrates the difficulty of retrieving precipitation from the midlatitudes, where stratiform precipitation is much more prevalent and for which cloud-top imagery does not contain sufficient information to estimate rainfall to the same degree as in the tropics.

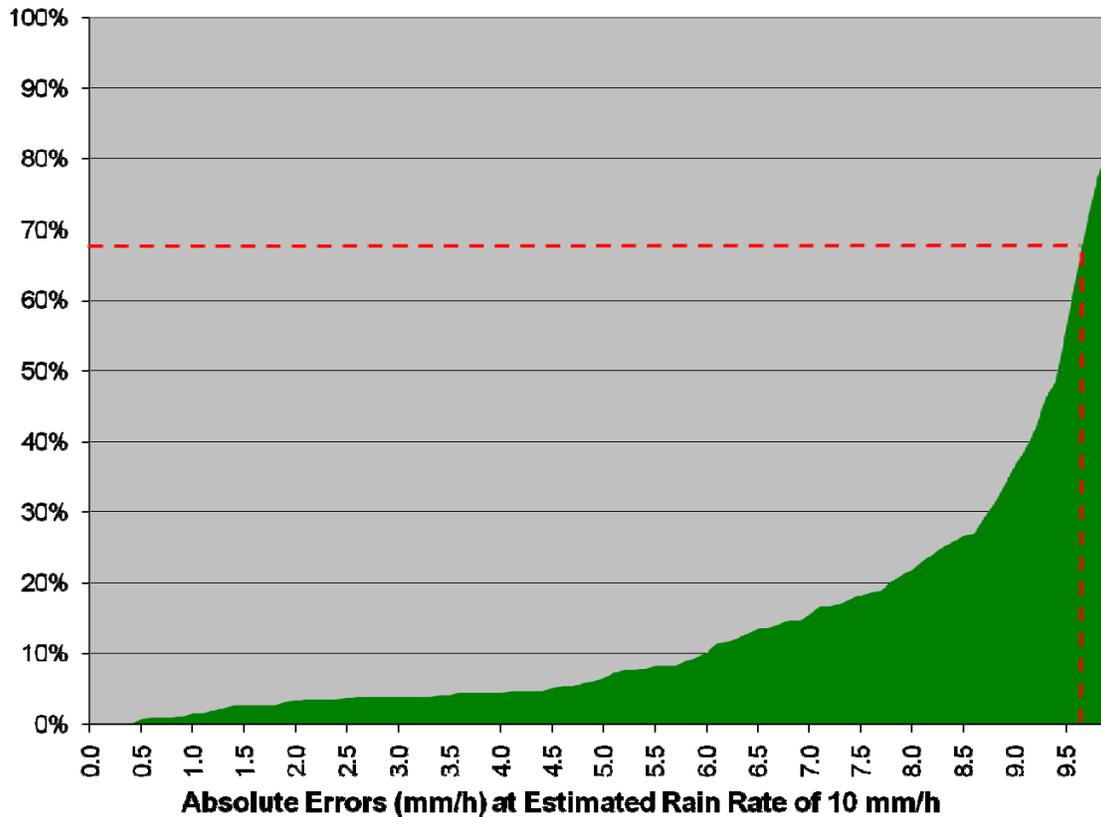


Figure 10. Same as Fig. 7, but vs. Nimrod.

4.2.2 Error Budget

The validation of retrieved rain rates against TRMM PR data for the 5th-9th of January, April, July, and October 2005 plus all of January 2008 is summarized in Table 11 using the “fuzzy” verification described in Section 4.2.1.1. As discussed in the previous section, the precision and accuracy specs are both being met for the TRMM PR. The algorithm does not meet either spec against Nimrod data largely due to the low bias of the algorithm for stratiform rainfall; however, since the accuracy and precision specs are only for convective rainfall this is acceptable. Table 11 summarizes the performance of the algorithm against both TRMM and Nimrod data at the 10 mm/h threshold compared to the F&PS spec.

	Accuracy (mm/h) at 10 mm/h	Precision (mm/h) at 10 mm/h	Number of data points
Vs. TRMM	4.9	8.9	13887
Vs. Nimrod	8.6	9.7	501
Proposed F&PS	6.0	9.0	-----

Table 11. Comparison of Rainfall Rate algorithm validation with F&PS.

5 PRACTICAL CONSIDERATIONS

5.1 Numerical Computation Considerations

The calibration portion of the algorithm creates / updates a series of external files containing matched MW rainfall rates and ABI predictors, and whenever a file is updated, ingests the data into an array and uses L-U (Lower-Upper) matrix decomposition to solve the resulting matrix for calibration coefficients (Eq. 2) and predictor ID's that are stored in a separate external file. The retrieval portion of the algorithm ingests the calibration coefficient files and the ABI predictor fields that are indicated within these files and applies the coefficient files to the predictor fields, resulting in a rainfall rate field on the same grid as the ABI predictors.

The calibration and retrieval portions of the Rainfall Rate Algorithm do not need to be run sequentially; in fact, parallel processing is preferred as long as care is taken to make sure that coefficient files are available whenever needed for the retrieval portion of the algorithm. The only stipulation is that for optimal accuracy the calibration should be updated as frequently as new ABI-MW rain rate matchups are available. Note that since the input microwave data may have a latency of as much as 3 hours, previous ABI data will need to be available for calibration purposes.

5.2 Programming and Procedural Considerations

The Rainfall Rate Algorithm requires knowledge of spatial uniformity metrics that are computed for each pixel using pixels that surround it. Beyond this reliance, the Rainfall Rate Algorithm is purely a pixel by pixel algorithm; no information from previous time periods is required for the retrieval step (though it is needed for the calibration step—see below). Note that although the current requirement for refresh rate (15 min) is longer than the planned ABI refresh rate of 5 min, no temporal averaging is performed to generate the 15-min products; the most recently available ABI image is used to generate the current Rainfall Rate product.

A collection of MW rainfall rates during the previous 2-3 days should be available for use as calibration targets. However, if necessary the rainfall rate algorithm can run using pre-computed calibration coefficients which will be adjusted whenever target MW rainfall rates become available and a sufficient supply of matched data pairs has thus been built up. Furthermore, if the availability of MW rainfall rates is interrupted, the algorithm will continue to produce estimates using the most recently computed calibration coefficients.

5.3 Quality Assessment and Diagnostics

Quality flags will be produced and provided along with the rainfall rate fields, with non-zero values for pixels whose inputs have values outside the acceptable range. These flags are described in detail in Section 3.6. Table 12 lists acceptable range values for the inputs. Note that the minimum values in the table are for computational purposes: values lower than that would produce negative predictor values, which will in turn result in

errors when the nonlinear predictor transformation described in Section 3.5.2 is performed. Note also that

Input ID	Predictor Description	Minimum Value
1	$T_{6.2}$	174 K
2	$S=0.568*(T_{\min,11.2}-217 \text{ K})$	-25 K
3	$T_{\text{avg},11.2}-T_{\min,11.2}-S$	-85 K
4	$T_{7.34}-T_{6.19}$	-30 K
5	$T_{8.5}-T_{7.34}$	-30 K
6	$T_{11.2}-T_{7.34}$	-20 K
7	$T_{8.5}-T_{11.2}$	-30 K
8	$T_{11.2}-T_{12.3}$	-20 K

Table 12. Minimum acceptable values for each algorithm predictor.

The following procedures are recommended for diagnosing the performance of the rainfall retrieval algorithm.

- Periodically image the individual test results to manually identify artifacts or non-physical behaviors.
- Automatically evaluate the time series of the total rainfall area and total rainfall volume and flag excessively large changes for further investigation.

5.4 Exception Handling

The Rainfall Rate Algorithm includes checking the validity of each input ABI band before retrieving a rainfall rate, and a ‘missing’ (-999.0) value is assigned to a pixel if any of the input values are outside the acceptable range. A missing value is also assigned to that pixel if the calibration coefficients are unavailable; this is done by checking the first rain detection predictor ID, the Heidke Skill Score, and the correlation coefficient from the training file for missing values. The bits 1-4 (depending on the predictor; see Table 7 in Section 3.4.3) of the quality flag for that pixel will also be set to 1. The Rainfall Rate Algorithm also expects the Level 1b processing to flag any pixels with missing geolocation or viewing geometry information.

If the microwave data are unavailable, the algorithm will continue to produce estimates of rain rate using the most recently available calibration coefficient tables; performance will degrade slightly as a result but the degradation will be limited. If one or more individual ABI bands used by the algorithm become unavailable, the algorithm will first output missing values (and corresponding quality flags) for any pixels that use the missing ABI band, and then subsequent updates to the calibration coefficients will ignore the missing bands and retrieval will continue as normal. The degree of degradation in performance will depend on the band(s) that are lost.

5.5 Algorithm Validation

Prior to launch, validation efforts will focus on Europe and Africa using SEVIRI data as a proxy for ABI given the previously discussed concerns about using simulated data for rainfall rate validation. The validation data will consist of TRMM PR data over the Tropics and Nimrod ground-based radar data over Western Europe, plus any ground-based radar data from field campaigns that can be obtained. These data sets were described in Section 4.2.1.1. However, it should be noted that ground-based radars have numerous well-documented limitations, so any ground-based radar data used for validation will need to be carefully quality-controlled, including comparisons between radar-derived rainfall total fields and corresponding rain gauges to determine the extent of such errors.

During the pre-launch period, validation tools will also be developed: one set to be used by operations to monitor the performance of the algorithm in real time and identify any anomalies; the second to be used by the algorithm developers to identify systematic algorithm deficiencies, their possible causes, and potential remedies. The former will be transferred to the NOAA / NESDIS Office of Satellite Data Processing and Distribution (OSDPD) while the latter will remain at STAR for use by the algorithm developers and collaborative partners outside STAR.

The post-launch phase will consist of monitoring of the product stream by OSDPD using the aforementioned tools, and close collaboration between STAR developers and the NOAA / NESDIS / OSDPD / Satellite Services Division (SSD) Satellite Analysis Branch (SAB) analysts who are responsible for real-time monitoring of satellite rainfall. They will evaluate the performance of the algorithm both from an “eyeball” perspective of day-to-day performance and from the perspective of systematic behavior of the algorithm as identified using the statistical tools. Modifications to the algorithm to address any deficiencies will then be identified and implemented.

Additional details about algorithm validation can be found in the corresponding Product Validation Plan.

6 ASSUMPTIONS AND LIMITATIONS

The following sections describe the assumptions and limitations of the current version of the Rainfall Rate Algorithm.

6.1 Performance

Several assumptions have been made in developing and estimating the performance of the Rainfall Rate Algorithm. They are listed below, accompanied by proposed mitigation strategies in parentheses.

1. The calibration target (MW) rainfall rates are accurate. (No mitigation possible)
2. The calibration target rainfall rates are available with a reasonably short lag time. Note that in the absence of calibration data, the algorithm will continue to produce retrievals based on the last available set of calibration coefficients. The potential impact on doing so for an extended period of time will be determined via testing.
3. The ABI data have been corrected for parallax prior to retrieval of rainfall rates. Mitigation is in progress via a coordinated parallax correction effort.
4. The available validation data (TRMM for the tropics and Nimrod for Western Europe) provide a sufficiently representative sample for evaluating whether the algorithm will meet spec over GOES-R coverage area. (Investigating additional radar validation data from field campaigns such as NAMMA, and COPS.)
5. The processing system allows for processing of multiple pixels at once for application of the spatial uniformity tests. (No mitigation possible)
6. Striping (i.e., when two or more detectors have slightly different calibrations, producing scan lines that are biased with respect to one another) and spectral shifts are minimal. (No mitigation possible)
7. No data aggregation is performed in time; i.e., if the frequency of ABI imagery exceeds the product refresh rate, only one ABI image will be processed per product. (No mitigation possible)

In addition, a number of limitations in the ability to retrieve rainfall rates from satellite data have been identified and are listed here, along with proposed mitigation strategies:

1. Satellite-based rainfall algorithms generally exhibit much better skill for convective (warm-season) rainfall than for stratiform (cold-season) rainfall, because the relationship between cloud-top temperature and rainfall rate is much stronger for the former than the latter. The inclusion of additional ABI bands

provides some mitigation by implicitly including information about cloud-top properties (particle size and phase); the potential impact of explicitly retrieved cloud top properties from the ABI processing system will be investigated as a mitigation effort.

2. The current version of the algorithm does not account for any influences on precipitation below cloud level; e.g., modulation by topography or evaporation of hydrometeors in dry sub-cloud air. Mitigation of the former is being explored through a GOES-R Critical Path project to develop an orographic correction for the algorithm that accounts for topographically-induced wind flows. The mitigation of other subcloud effects will be investigated by examining the impact of relevant numerical weather model fields (e.g., total precipitable water and/or low-level relative humidity) on precipitation estimates. Note that ABI-derived fields cannot be used for this application since they will not be available for cloudy regions.

Finally, the channel mapping between SEVIRI and ABI has been used in the development and pre-launch validation of the algorithm is shown in Table 13:

ABI Band		SEVIRI Proxy	
Number	Central Wavelength (μm)	Number	Central Wavelength (μm)
8	6.19	5	6.2
10	7.34	6	7.3
11	8.5	7	8.7
14	11.2	9	10.8
15	12.3	10	12.0

Table 13. Channel mapping associated with ABI proxy data from SEVIRI during algorithm development and validation.

6.2 Assumed Sensor Performance

It is assumed that the sensor will meet its current specifications. However, the Rainfall Rate Algorithm will be dependent on the following instrumental characteristics.

- The spatial variation predictors in the Rainfall Rate Algorithm will be critically dependent on the amount of striping in the data. Note that this will affect the retrieval only when any texture-related predictors are among the selected predictors selected by the algorithm.
- Unknown spectral shifts in some channels will affect the BTD calculations and thus compromise some of the predictors. Note that this will affect the retrieval only when any BTD's are among the predictors selected by the algorithm.

6.3 Pre-Planned Product Improvements

A number of potential improvements are being investigated for the “day-2” Rainfall Rate product:

6.3.1 Smoothing along Region Seams

The current version of the algorithm does not attempt to smooth any discontinuities that may occur along the seams between regions (i.e., 30°S, the equator, and 30°N). The code will be modified to overlap the regions somewhat and apply a weighted average to the two overlapping regions to produce a smoother result.

6.3.2 Incorporation of Numerical Model Moisture Fields to Correct for Subcloud Evaporation of Rainfall

The current operational Hydro-Estimator rainfall rate algorithm uses National Centers for Environmental Prediction (NCEP) North American Mesoscale (NAM) model total column precipitable water and mean-layer relative humidity from the lowest third of the troposphere to enhance rainfall rates in moist regions and reduce them in dry regions. Such a correction for the GOES-R Rainfall Rate algorithm is being investigated.

6.3.3 Correction for Orographic Modulation of Rainfall

A GOES-R Critical Path Project in collaboration with Hampton University is working to develop a correction for the orographic modulation of rainfall based on wind, topography, and stability information from an operational numerical weather model.

6.3.4 Incorporation of Time Change Information

Another GOES-R Critical Path Project, in collaboration with City College of New York (CCNY), involves determining Lagrangian (i.e., cloud-following) time changes in cloud properties and using them as predictors in the rainfall algorithm. This is an effort to address the tendency of satellite rainfall algorithms to underestimate rainfall early in the convective cycle.

6.3.5 Incorporation of Retrieved Cloud Microphysics Information

A third GOES-R Critical Path Project, in collaboration with ESSIC involves incorporating retrieved cloud effective radius and cloud liquid water path information in an improved regime classification and as direct predictors in the algorithm.

7 REFERENCES

- Ebert, E. E., 2008: Fuzzy verification of high-resolution gridded forecasts: A review and proposed framework. *Meteor. Appl.*, **15**, 51-64.
- Joyce, R. J., J. E. Janowiak, P. A. Arkin, and P. Xie, 2004a: CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *J. Hydrometeor.*, **5**, 487-503.
- Kuligowski, R. J., 2002: A self-calibrating real-time GOES rainfall algorithm for short-term rainfall estimation. *J. Hydrometeor.*, **3**, 112-130.
- Olson, W. S., C. D. Kummerow, S. Yang, G. W. Petty, W. K. Tao, T. L. Bell, S. E. Lang, D. E. Johnson, and C. Chiu, 2006: Precipitation and latent heating distributions from satellite passive microwave radiometry. Part I: Improved method and uncertainties. *J. Appl. Meteor. Cli.*, **45**, 702-720.