NOAA NESDIS
CENTER for SATELLITE APPLICATIONS and RESEARCH

GOES-R Advanced Baseline Imager (ABI) Algorithm Theoretical Basis Document For Rainfall Potential

Robert J. Kuligowski, NOAA/NESDIS/STAR

Version 1.0
September 29, 2010

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

ABI – Advanced Baseline Imager
AIT – Algorithm Integration Team
ATBD – Algorithm Theoretical Basis Document
BADC – British Atmospheric Data Centre
CDF – Cumulative Distribution Function
CLASS – Comprehensive Large Array-data Stewardship System
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
IR – Infrared
LEO – Low-Earth Orbit
LZA – Local solar Zenith Angle
MRD – Mission Requirements Document
MW – Microwave
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
PR – Precipitation Radar
SCaMPR – Self-Calibrating Multivariate Precipitation Retrieval
SEVIRI – Spinning Enhanced Visible Infrared Imager
STAR – Center for Satellite Applications and Research
TRMM – Tropical Rainfall Measuring Mission
QPE – Quantitative Precipitation Estimate

ABSTRACT

This Rainfall Potential Algorithm Theoretical Basis Document (ATBD) contains a high-level description (including the physical basis) of an algorithm for nowcasting 0-3 hour rainfall accumulation at the pixel scale based on rainfall rates derived 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 Potential algorithm, validation efforts, and planned improvements.

1   INTRODUCTION

1.1 Purpose of This Document

The Rainfall Potential Algorithm Theoretical Basis Document (ATBD) provides a high-level description of and the physical basis for the prediction of rainfall accumulation during the next 3 hours from images taken by the Advanced Baseline Imager (ABI) flown on the GOES-R series of NOAA geostationary meteorological satellites—or, more precisely, from rainfall rates derived from current and previous ABI imagery. The rainfall potential is produced as an EDR.

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 ABI 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 specifications of the GOES-R Ground Segment Functional and Performance Specification (F&PS) and Missions Requirements Document (MRD) and to the ATBD for the Rainfall Rate Algorithm.

1.5 Revision History

Version 0.0 of this document was created by Dr. Robert J. Kuligowski of NOAA/NESDIS and its intent was to serve as a draft delivery to the GOES-R AWG Algorithm Integration Team (AIT).

Version 0.1 of this document was created by Dr. Robert J. Kuligowski of NOAA/NESDIS and its intent was to accompany the delivery of the version 1.0 algorithm to the GOES-R AWG Algorithm Integration Team (AIT).

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 version 3 (80%) 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 Potential Algorithm and the requirements it places on the sensor.

2.1 Products Generated

The Rainfall Potential Algorithm produces a field of predicted accumulation of rainfall during the next 3 hours associated with the most recently available GOES imagery and from previous images. In terms of the F&PS, it is responsible directly for the Rainfall Potential product within the Hydrology product sub-type. The Rainfall Potential Algorithm design calls for a quantitative 3-hour rainfall accumulation in mm 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 aside from the official Rainfall Potential product and accompanying quality flags.

Requirement Description	Requirement Value
Name	Rainfall Potential
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 rainfall cases
Vertical Resolution	N/A
Horizontal Resolution	2.0 km
Mapping Accuracy	1.0 km
Measurement Range	0 – 100 mm
Measurement Accuracy	5 mm for pixels designated as raining
Product Refresh Rate / Coverage Time (Mode 3)	15 min
Refreshment Rate / Coverage Time (Mode 4)	5 min
Vendor Allocated Ground Latency	266 sec
Product Measurement Precision	5 mm for pixels designated as raining

Table 1. F&PS Requirements for the Rainfall Potential algorithm.

Note that for pixels outside the 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 potential will be produced for each pixel observed by the ABI with a latitude of less than 60 degrees and a local zenith angle (LZA) relative to the satellite sub-point of 70 degrees. The Rainfall Potential algorithm does not make direct use of any ABI data, but only uses current and previous output from the ABI Rainfall Rate Algorithm. Please refer to the Rainfall Rate ATBD for additional details on the ABI data used in the Rainfall Rate retrievals.

3 ALGORITHM DESCRIPTION

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

3.1 Algorithm Overview

The Rainfall Potential Algorithm derives a prediction of rainfall accumulation during the next 0-3 hours on a pixel level in ABI imagery. The algorithm first predicts future rainfall rates based on extrapolation from current and previous ABI rainfall rates, and then these instantaneous rates are accumulated into total predicted rainfall during the next 3 hours. Additional details on the rainfall rate retrieval can be found in the Rainfall Rate ATBD.

3.2 Processing Outline

The processing outline of the Rainfall Potential Algorithm is summarized in the figure below. The Rainfall Potential algorithm must run on an entire scan sector (typically the full disk) of data since it identifies and extrapolates the motion of 2-D rainfall features in the data. However, an exception may be made if subsectors are used that completely contain any precipitation features of interest; i.e., any precipitation features that are partly outside the scan sector will not be handled correctly by the algorithm.
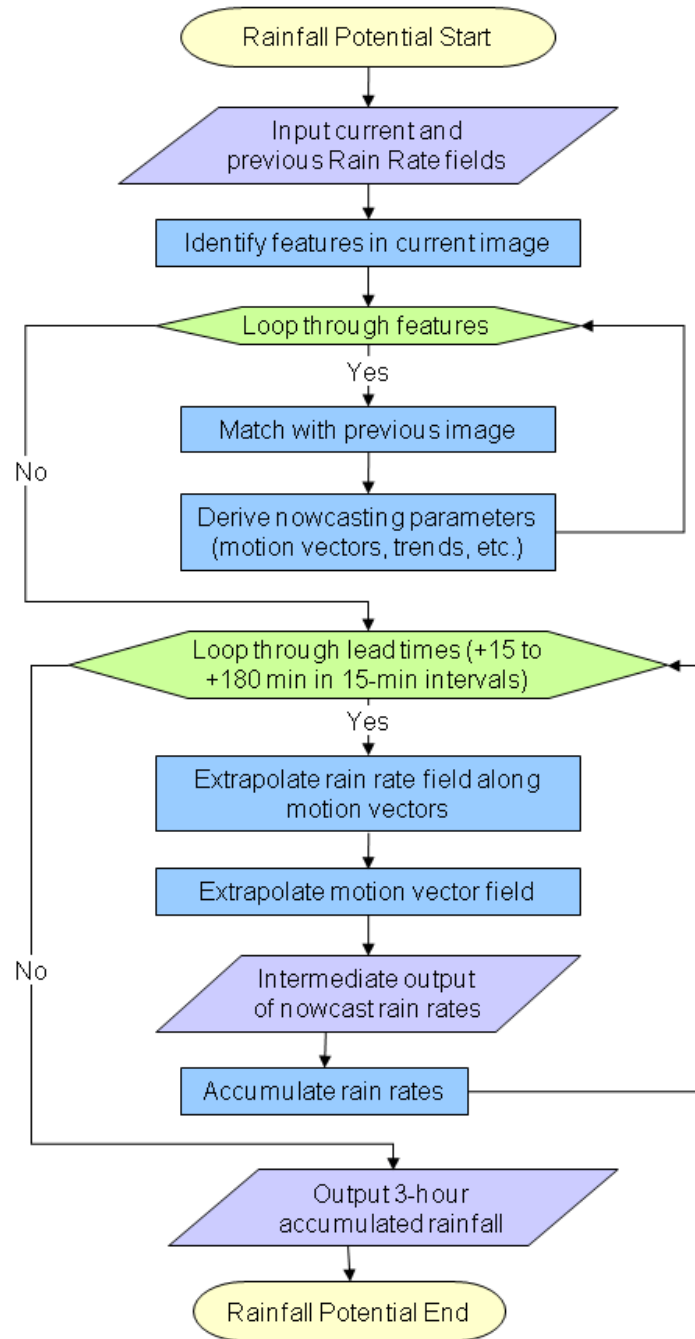
Figure 1. High-level flowchart of the Rainfall Potential Algorithm, illustrating the main processing sections.

3.3 Algorithm Input

This section describes the input needed to compute the rainfall potential. While the rainfall potential is derived for each pixel, the process of identifying, analyzing, and extrapolating cloud features requires information from an entire scan sector.

3.3.1 Primary Sensor Data

The Rainfall Potential algorithm does not directly use any sensor input but relies solely on ABI Rainfall Rate Algorithm output; please refer to the Rainfall Rate Algorithm ATBD for details on the sensory input to that algorithm.

3.3.2 Ancillary Data

The Rainfall Potential Algorithm uses rainfall rates from the current gridded ABI Rainfall Rate product (dynamic ancillary data) and rainfall rates from the most recent previous gridded ABI Rainfall Rate product (static ancillary data).

3.4 Theoretical Description

The prediction of the evolution of rainfall from the current ABI imagery over the next 3 hours (to produce the 3-hour predicted rainfall totals) requires the following steps: identifying rainfall features of interest, determining the current motion of these features, and then using this information to predict the future motion of these features. In addition, an algorithm could also be used to determine time changes in the characteristics of these features and then predict their future evolution; however, this latter step is not performed in the current version of the algorithm described in this ATBD, though such a capability may be added in the future.

3.4.1 Physics of the Problem

In a numerical weather prediction model, the prediction of the atmospheric state is not generally treated on a feature-by-feature basis, and individual fields are not treated individually: rather, the primitive equations are solved at separately at each grid point and the results together comprise the atmospheric state at future time intervals. While this approach is able to account for all of the potential influences on the future state of a precipitation feature, it is also computationally highly complex, and these models themselves are based on assumptions and simplifications that are not always best suited for precipitation forecasting. In fact, numerous authors have demonstrated that a nowcast based on extrapolation of rainfall features based on current and past characteristics can produce forecasts that are more skillful than numerical weather model forecasts for lead times of less than a few hours. At longer lead times, nonlinear factors that cannot be accounted for by simple extrapolation models become more and more important and the numerical model becomes the preferred method of rainfall nowcasting.

However, even though extrapolation techniques are simple relative to numerical weather prediction models, there are still some difficult issues that need to be resolved. With regard to feature identification, precipitation features are seldom well-organized, and it is well-known that smaller-scale features have shorter lifespans than larger-scale features. Consequently, any feature identification technique must successfully discriminate relevant from irrelevant information in defining features that will persist in time long enough for forecasting purposes. Ideally, feature detection will be performed at multiple scales so that the different feature lifetimes at different scales will be accounted for in the extrapolation process.

Another critical question is the determination of feature motion. Since the size and shape of precipitation features typically change with time, it can be difficult to draw a confident association between features on consecutive images, or to confidently determine the direction and speed of motion of the feature.

Finally, there is a question of how to use information about feature properties and motion to predict the location and properties of these features at future time steps in order to derive a 0-3 hour forecast of rainfall. The algorithm selected for nowcasting by the Hydrology Algorithm Team will be described in further detail with regard to its approach to all three of these issues. Additional details beyond the scope of this document can be found in Lakshmanan and Smith (2009) and Lakshmanan et al. (2003, 2009).

3.4.1.1 Feature Identification

In general terms, a feature is defined as a region of significant rain rate that is separated from other regions of significant rain rate. Furthermore, the group of pixels that comprise a feature must meet selected criteria regarding size and intensity (rain rate). The specific assignment of pixels to specific features is accomplished via a clustering technique which is initialized with clusters that are defined by selected rain rate contours. Each pixel is then assigned to a cluster based on two opposing criteria: the difference in intensity between the pixel of interest and the mean intensity of the potential matching cluster (similarity), and the number of pixels neighboring the pixel of interest that do not belong to the potential matching cluster (proximity—or, more specifically, non-proximity).

The algorithm also merges contiguous clusters such that a cluster becomes part of another cluster if they are contiguous and if it's mean is less than that of the adjacent cluster; however, merging is performed only if the combined cluster does not exceed the preset maximum size value. The clustering is performed in repeated iterations until the cluster assignments stop changing (maximum of 50 iterations).

3.4.1.2 Feature Tracking

The first step in feature tracking is associating identified features in consecutive images; however, clustering methods are not guaranteed to identify related pixels in consecutive images since the clusters are determined independently for each image. To avoid this problem, the motion of a particular cluster is estimated by matching a template containing

the cluster over the previous image and determining the mean absolute error as a function of position shift. The centroid of this mean absolute error field is then computed, with each point weighted by how much it differs from the absolute minimum error (the weighting is done to introduce a physically-reasonable degree of smoothness and to avoid influence from pixels that do not actually represent rainfall and do not move—i.e., ground clutter in radar). The location of this centroid is the shift vector for that feature from the previous image. This approach can be problematic when cloud features change significantly from one image to the next; however, the 15-minute time resolution of the ABI should minimize the occurrence of such instances.

Once the motion vectors have been identified for each feature, the vector values are assigned to the respective centroids of each feature and the remaining motion vectors for the entire field are estimated via an inverse distance squared interpolation that is also weighted by the size of the cluster. The resulting spatially distributed vector fields is then smoothed using a Kalman filter.

3.4.1.3 Feature Advection and Evolution

Features are advected forward in time for one time step along the derived wind fields, and then the wind fields themselves are advected forward by one time step; this process continues until the lead time reaches 3 h. The current version of the algorithm does not make any adjustments for growth or decay of precipitation features. This is because previous experiments with linear prediction of intensity did not produce results that were superior to a steady-state prediction. However, growth and decay may be introduced if suitably skilled algorithms can be developed.

3.4.1.4 Rainfall Accumulation

The current rainfall rates from the Rainfall Rate Algorithm plus the extrapolated rainfall rates every 15 minutes out to a lead time of 3 hours are used to produce a 0-3 hour rainfall accumulation using the trapezoidal technique; i.e., every time period is equally weighted except for the 0-hour and 3-hour rainfall rates, which receive half the weight of the others.

3.5 Mathematical Description

3.5.1 Feature Identification

The first step in feature identification is to apply a smoothing filter to the image to enhance the capacity to identify significant rainfall features. Specifically, a non-recursive median smoothing filter is applied; i.e., the value in each pixel is replaced by the median value in the 11x11 neighborhood surrounding the pixel.

The initial feature identification is then performed via a region-growing scheme. After quantizing the image (rounding all values to the nearest integer value of mm/h), the image is scanned and adjacent pixels with identical quantized values are grouped into initial clusters.

Those pixels not assigned to clusters via the region growing scheme are then assigned to the existing clusters based on a cost-minimization scheme. Each pixel xy is assigned a cost of membership in each existing cluster k, and that cost is $E_{xy}(k)$ expressed as:

$$E_{xy}(k) = \lambda d_{m,xy}(k) + (1 - \lambda) d_{c,xy}(k). \tag{1}$$

In this equation, $d_{m,xy}(k)$ is the intensity difference between the pixel of interest (at x,y) and the kth cluster and is defined as

$$d_{m,xy}(k) = \left\| \mu_k - I_{xy} \right\| \tag{2}$$

where μ_k is the mean intensity value for cluster k and I_{xy} is the intensity value (both in mm/h) for the pixel at (x,y). Meanwhile, $d_{c,xy}(i)$ is the discontiguity measure and is given by

$$d_{c,xy}(k) = \sum_{ij \in Nx,y} \left(1 - \delta \left(S_{ij}^n - k \right) \right) \tag{3}$$

where N_{xy} covers the 8 pixels adjacent to the pixel of interest at (x,y) and $\delta(S_{ij}^n\text{-}k)$ is equal to 1 if the neighbor pixel of interest does <u>not</u> belong to cluster k. That is, $d_{c,xy}(i)$ is simply the number of pixels adjacent to the pixel of interest (on 8 sides) that do not belong to the kth cluster. These two criteria are weighted using the weight λ, and the results have been shown to be relatively insensitive to values ranging from 0.2-0.8. (The value used here is 0.4). Once the cost has been computed for each cluster, the pixel is assigned to the cluster with the lowest associated cost.

This technique is applied to the quantized current rain rate field for multiple iterations until the cluster assignments do not change; a maximum number of iterations (50) is included to prevent an infinite loop in case of nonconvergence.

Once this cluster assignment step is completed, a final step is preformed to eliminate excessively small clusters. At three different scales (20, 160, and 480 pixels), the field is examined for clusters below the threshold size for that scale. Clusters that are too small are combined with their most intense neighbor. The final result is actually three cluster fields associated with three different scales.

3.5.2 Feature Tracking

After identifying the clusters in the current rainfall rate image, the next step is to determine how these features have moved since the previous image. Given the inexact nature of the clustering, it has been found that to cluster the previous image and then associate clusters in the two images has not proven to be very robust. Consequently, the clusters are used as a mask for comparing the shifted pixel values between images.

17

Specifically, the pixels in the cluster in the current image are compared to the pixels in the same cluster area in the previous image, and comparisons are also made for offsets associated with speeds of 20 m/s or slower (which would be 9 pixels in 15-min imagery at 2-km spatial resolution). The mean absolute error (MAE) for cluster k is then computed for each pixel shift of Δx and Δy:

$$MAE_k(x + \Delta x, y + \Delta y) = \frac{1}{n_k} \sum_{ny \in k} \left| I_t(x, y) - I_{t-\Delta t}(x + \Delta x, y + \Delta y) \right| \tag{4}$$

where n_k is the number of pixels in cluster k and I_t and $I_{t-\Delta t}$ represent pixel intensity (value) in the current and previous images, respectively. The vertical bars represent absolute difference.

A 2-D field of MAE is created for the offsets within the aforementioned limit. Since this field can be quite noisy, instead of using the minimum MAE value the centroid of the offsets with MAE values within 20% is used as the basis for the motion vector.

Each vector now has a motion vector associated with it—in particular, with its centroid. These cluster motion vectors u_k are used to create a spatially distributed field of motion vectors u(x,y) via interpolation:

$$u(x, y) = \frac{\sum_k u_k w_k(x, y)}{\sum_k w_k(x, y)} \tag{5}$$

where $w_k(x,y)$ is the weight assigned to the motion vector of cluster k and is given as

$$w_k(x, y) = \frac{N_k}{\|xy - c_k\|} \tag{6}$$

where N_k is the total number of pixels in cluster k and the denominator is the Euclidean distance between point (x,y) and the centroid c_k of cluster k; i.e.,

$$\|xy - c_k\| = \sqrt{(x - x_{c_k})^2 + (y - y_{c_k})^2} \tag{7}$$

These motion vectors are computed at all three scales (20, 160, and 480 pixels) for use in feature advection. These motion vectors are then smoothed over time using Kalman filters. A separate Kalman filter is initialized at every pixel in the image and the motion vector computed from the pair of images is treated as the "observed" quantity in the filter. The Kalman filter state model is assumed to be of constant acceleration; i.e., in the equation of motion:

$$v(t + 1) = a * v(t) + b \tag{8}$$

18

where v is the velocity and a the acceleration and b some constant. The corresponding Kalman filter state equations are:

$$x_{k+1} = \Phi_k x_k + w_k \tag{9}$$
$$z_{k+1} = H_k x_k + v_k \tag{10}$$

where **x** is the true velocity vector [v 0], the state transition model Φ_k is the current estimate of the vector [a b] as per Equation (8). **z** is the observed velocity, H_k an identity matrix and w_k and v_k are noise whose covariances are estimated as part of the Kalman filter recursion. The recursive Kalman filter updates are carried out as shown in Figure 2 and the resulting smoothed estimate of velocity is assigned to the pixel.
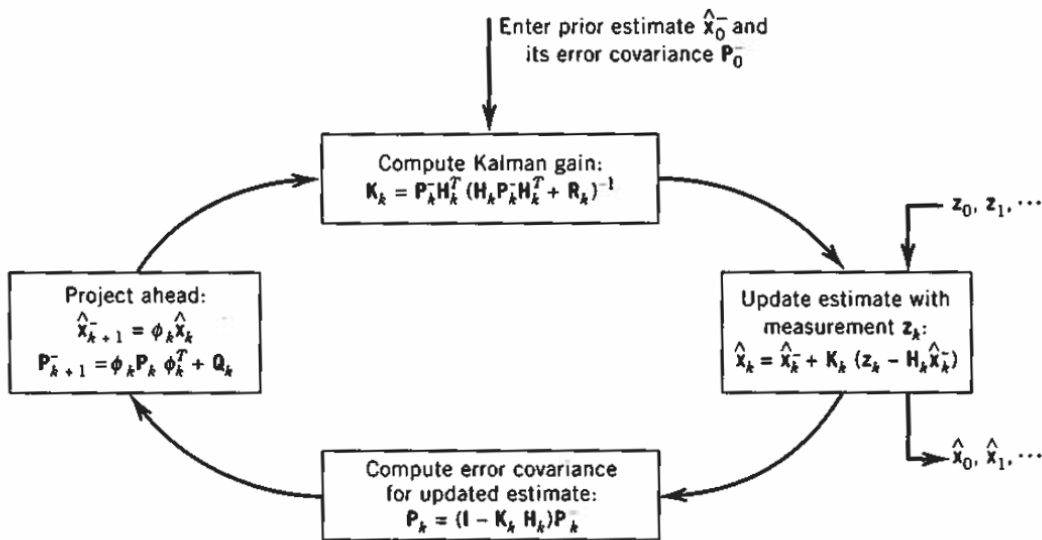
Figure 5.8 Kalman filter loop.

Figure 2: Kalman Filter recursive loop for time smoothing of motion vectors. Figure from Brown and Hwang (1996).

It should be noted that Kalman filtering is used instead of simple time averaging because it allows the weight of past observations to be dependent on their stability; i.e., if the velocity values are highly variable, they are given less weight in the smoothing and the current observation is given more weight, and vice versa.

3.5.3 Feature Advection

This initial motion vector field is then used to extrapolate the rain rate field forward by one time step (15 min for the ABI) by moving each pixel in the current rain rate image along the motion vector for a distance equivalent to one time step. The specific motion vector used is a function of lead time: the finest scale is used for lead times below 30 min; the second-finest for the 30-90 min lead time window; and the coarsest for lead times exceeding 90 min. Once this is done, the motion vector field is used to advect the motion vectors forward in time by one time step to form the motion vector field for the

next time step. Since divergence in the motion field will cause some pixels in the extrapolated rain rate image to not be filled in, the motion vector field from the next time step is used to project backward in time from the location of the empty pixels in the extrapolated image to the corresponding pixel in the current rain rate image. Once this is done, this entire procedure is repeated iteratively at intervals corresponding to the time step (15 min) out to 3 hours. Finally, the current and extrapolated rainfall rates at 15-min intervals are used to produce an accumulation of rainfall for the entire 3-h period.

3.6 Algorithm Output

The final output of this algorithm is a field of rainfall accumulation for the next 3 hours at the same spatial resolution as the ABI IR data. 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 2:

Byte	Bit	Flag	Source	Value
0	0	Rainfall Potential output	RP	1=bad data; 0=OK
	1	Satellite zenith angle block-out zone	SDR	1=zenith angle>70° or lat>60°; 0=OK
	2	Missing value for current rainfall rate	RR	1=bad data; 0=OK
	3	Current rain rate was >100 mm/h but truncated to 100 mm/h	RR	1=rain rate >100 mm/h but truncated at 100 mm/h; 0=rain rate <100 mm/h
	4-7	Not used		

Table 2. Quality flags for the Rainfall Potential product.

In addition, a file of quality information fields will be output, consisting of a gridded file containing the motion vector components at the initialization time followed by the motion vector component and rainfall rate nowcasts out to 3 h lead time at 15-min intervals (Table 3):

Grid	Value	Source	Type	Units
1	U component of motion vector at 0 min lead time	RP	Real*4	m/s
2	V component of motion vector at 0 min lead time at 0 min lead time	RP	Real*4	m/s
3	Rainfall rate nowcast at 15-min lead time	RP	Real*4	mm/h
4	U component of motion vector at 15 min lead time	RP	Real*4	m/s
5	V component of motion vector at 15 min lead time	RP	Real*4	m/s

6-8	Rainfall rate nowcast and motion vector components at 30 min lead time	RP	Real*4	Same as above
9-38	Rainfall rate nowcast and motion vector displacements at lead times of 45, 60, 75, 90, 105, 120, 135, 150, 165, and 180 min lead time	RP	Real*4	Same as above

Table 3. Gridded quality information for the Rainfall Potential product.

Finally, the metadata file will contain the information listed below in Table 4:

Type	Variable
Float	Total rain area (*number of pixels in image with rainfall potential > 1 mm*)
Float	Total rain volume (*total rain in rain area, mm*)
Long	Total number of pixels where retrieval was attempted
Long	Number of QA flag values: *6*
Long	Number of retrievals with QA flag value 0 (*all bits set to 0*)
String	Definition of QA flag value 0: *Good Rainfall Potential retrieval*
Long	Number of retrievals with QA flag bit 0 set to 1
String	Definition of QA flag with bit 0 set to 1: *Bad Rainfall Potential retrieval*
Long	Number of retrievals with QA flag bit 1 set to 1
String	Definition of QA flag with bit 1 set to 1: *Satellite zenith angle block-out zone*
Long	Number of retrievals with QA flag bit 2 set to 1
String	Definition of QA flag with bit 2 set to 1: *Missing value for current rainfall rate*
Long	Number of retrievals with QA flag bit 3 set to 1
String	Definition of QA flag with bit 3 set to 1: *Current rain rate was >100 mm/h but truncated to 100 mm/h*

Table 4. Metadata for the Rainfall Potential product.

4 TEST DATA SETS AND OUTPUTS

4.1 Simulated/Proxy Input Data Sets

As stated previously, the Rainfall Potential algorithm does not directly use any satellite data; rather, it uses as input current and previous output fields from the Rainfall Rate algorithm. For the tests performed here, the Rainfall Rate algorithm used as input SEVIRI observations as a proxy for ABI data, and blended microwave rainfall estimates as a calibration data source. The reader is referred to the Rainfall Rate Algorithm Theoretical Basis Document for additional details on these data sets and on the Rainfall Rate algorithm.

4.2 Intermediate Outputs

4.2.1 Predicted Instantaneous Rainfall Fields

The Rainfall Potential algorithm was applied to test Rainfall Rate ouptut data from the 6th through the 9th of January, April, July, and October 2005 to produce forecasts of rainfall rate every 15 minutes out to 3 hours. These forecasts were then summed into a final 3-hour accumulation as shown in Figure 2.
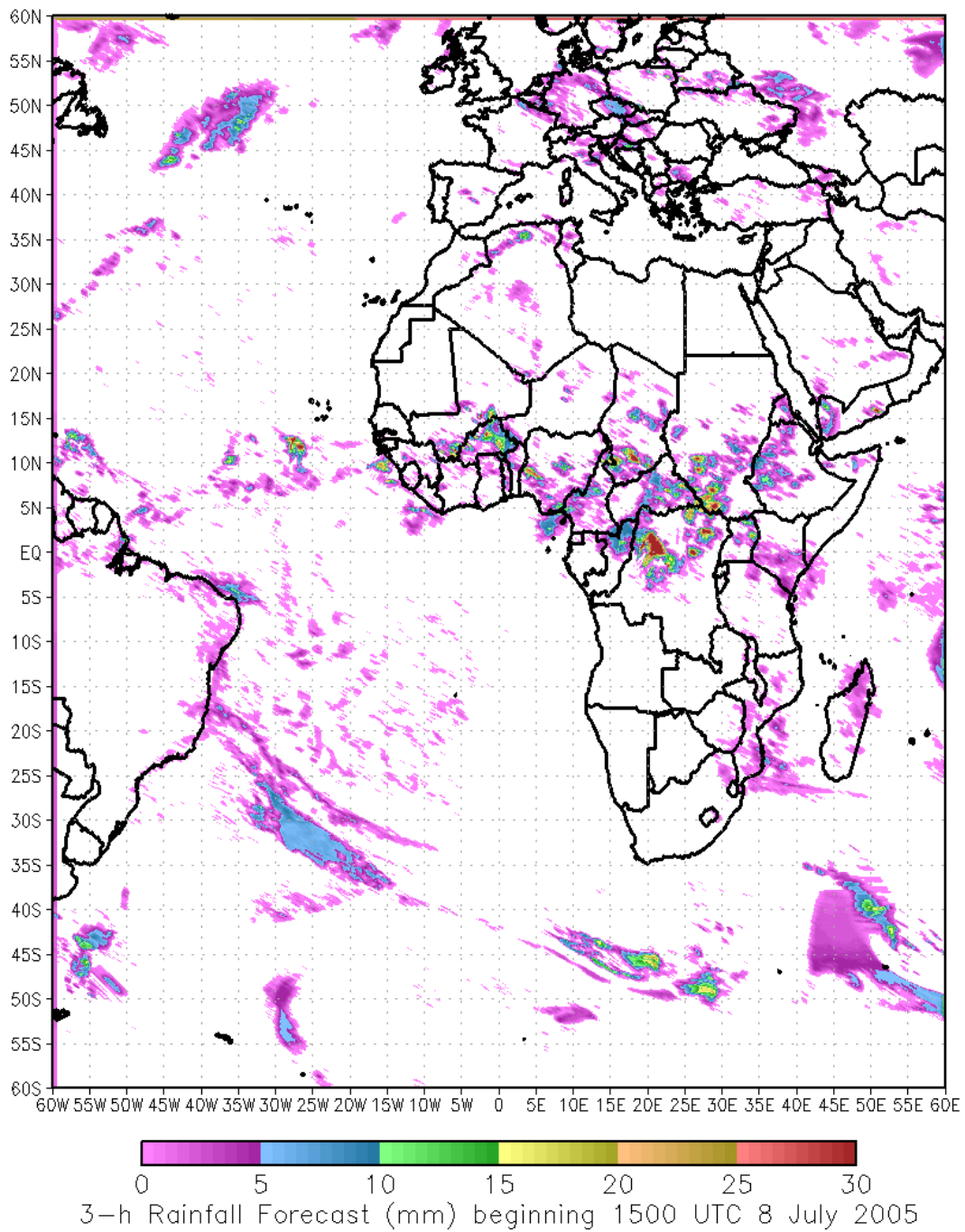
Figure 3. Rainfall Potential for the 3 hours beginning 1500 UTC 8 July 2005 derived from rain rates based on SEVIRI data.

4.3 Algorithm Output Using Proxy Input Data Sets

4.3.1 Precisions and Accuracy Estimates

To estimate the performance and accuracy of the Rainfall Potential Algorithm, we will have to compare the output against available rain gauge data and radar data. However, such data are very difficult to obtain over Europe and Africa. Comparisons will be made against Nimrod radar data over Western Europe, and, if possible, data from the Convective and Orographically-induced Precipitation Study (COPS) and NASA African Monsoon Multidisciplinary Analyses (NAMMA) field campaigns over Europe and West Africa, respectively. This section will present the analysis methodology for estimating the precision and accuracy, followed by the quantitative results in terms of the F&PS specifications.

4.3.1.1 Validation against Nimrod

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). The coverage of these radars is illustrated in Fig. 3.
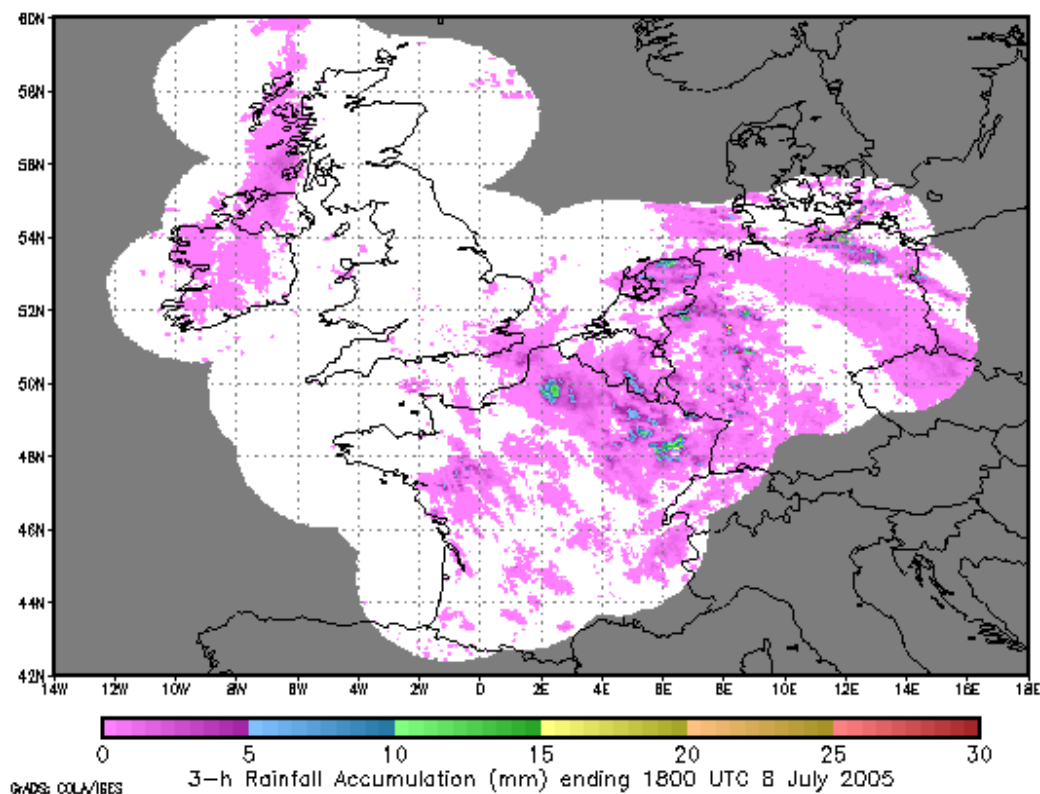

Figure 4. Coverage of Nimrod mosaic radar data.

24

Figure 4 shows a scatterplot of the 3-hour rainfall potential compared to NIMROD data over Western Europe for all 3 months. Note that most of the data fall on the axes rather than on the 1:1 line and that the algorithm exhibits a significant systematic dry bias, as shown by the very small slope of the (dashed) best-fit line (though the significant number of points lying along the y-axis results in an overall bias of approximately 10.9%). However, it should be emphasized that a pixel-by-pixel comparison like this does not account for location errors in a forecast; for instance, some of the points lying on the x- or y-axes would shift toward the 1:1 line if the evaluation were performed at a coarser spatial resolution with less sensitivity to location errors. As Fig. **5** illustrates, the actual forecast fields look better than what might be implied by the scatterplot.
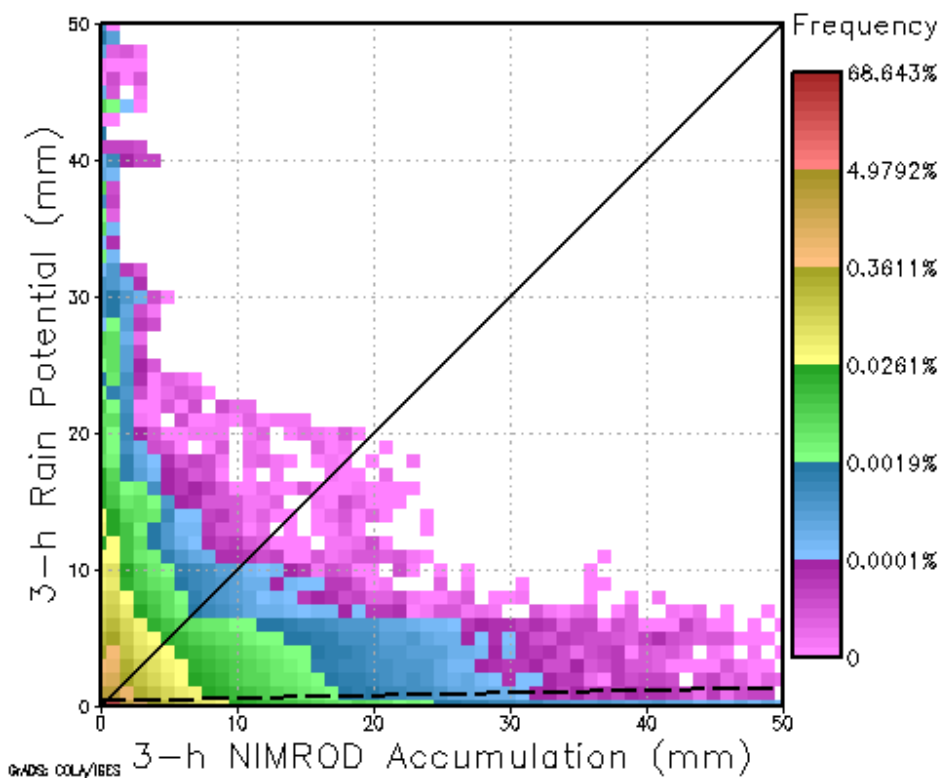
Figure 5. Scatterplot of Rainfall Potential values versus corresponding Nimrod rainfall accumulations for 5-9 April, July, and October 2005. The solid line is the 1:1 line (i.e., forecast = observation); the dashed line is the best-fit line between forecasts and observations.
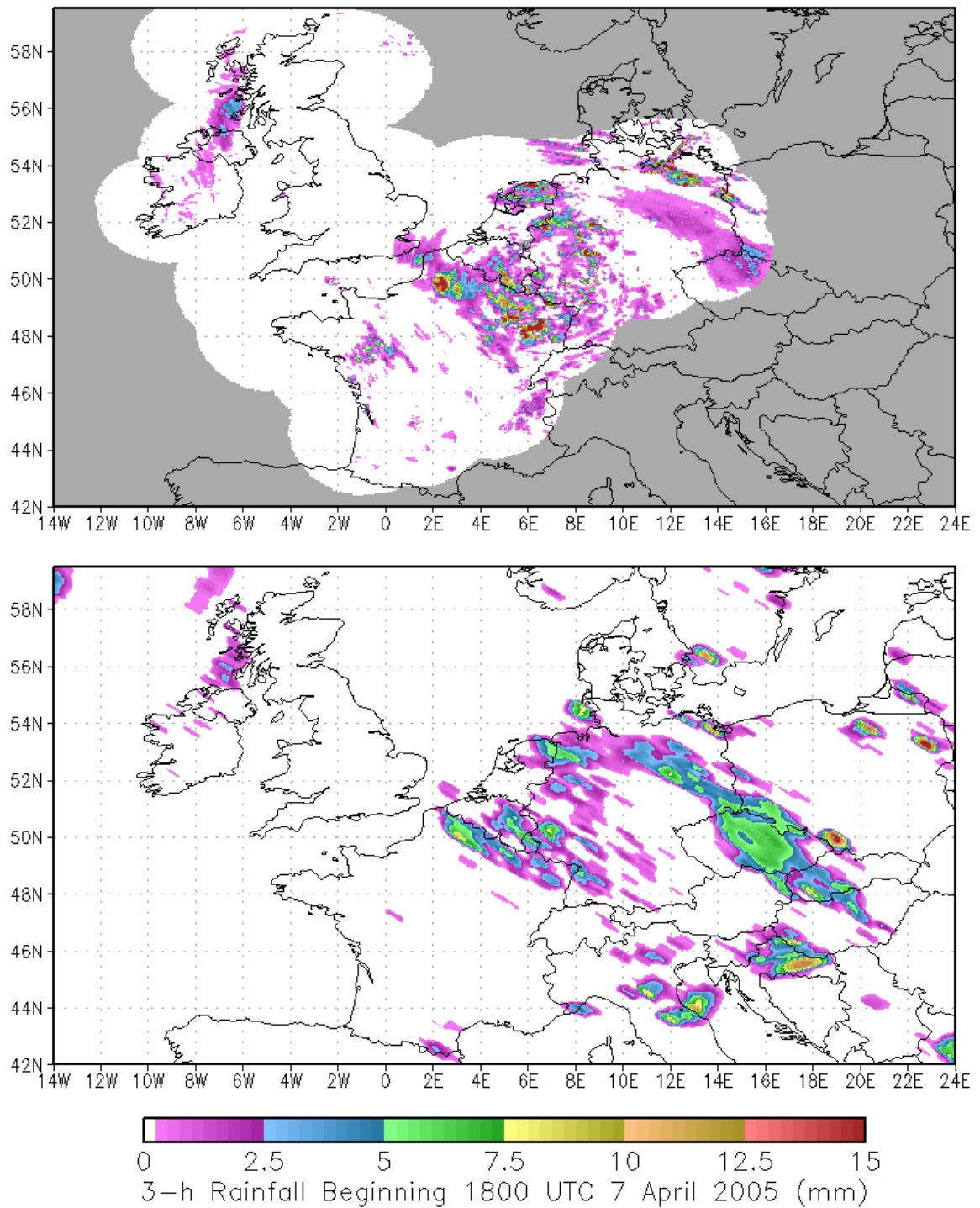
GrADS: COLA/IGES

Figure 6. Plot of Nimrod radar (top) and corresponding 3-hour rainfall potential (bottom) for the 3 h beginning 1800 UTC 7 April 2005.

The skill of the algorithm at discriminating rainfall can be illustrated by the following 2x2 contingency table (Table 5):

		Nimrod	
		< 1 mm	> 1 mm
Rainfall	**< 1 mm**	9,111,941	527,834
Potential	**> 1 mm**	543,199	248,061

Table 5. Contingency table of rainfall discrimination skill of the Rainfall Potential algorithm, using a rain / no rain threshold of 1 mm.

This corresponds to a probability of detection (POD) of 32.0% and a false alarm rate (FAR) of 68.6%. Again, the significant penalties inherent in a pixel-by-pixel evaluation

The specific F&PS precision requirements for the Rainfall Potential algorithm is for a precision of 5 mm. This means that for pixels with nonzero Rainfall Potential values, the corresponding observed value should be within 5 mm of the predicted value 68% of the time. The evaluation of the Rainfall Potential against the precision spec value is illustrated in Fig. 6, which shows the cumulative distribution function (CDF) of the Rainfall Potential errors for 5-9 April, July, and October separately and together. The algorithm meets spec if the CDF curve reaches the 68% value at a value lower than 5 mm, which as Fig. 6 shows is the case except in October.
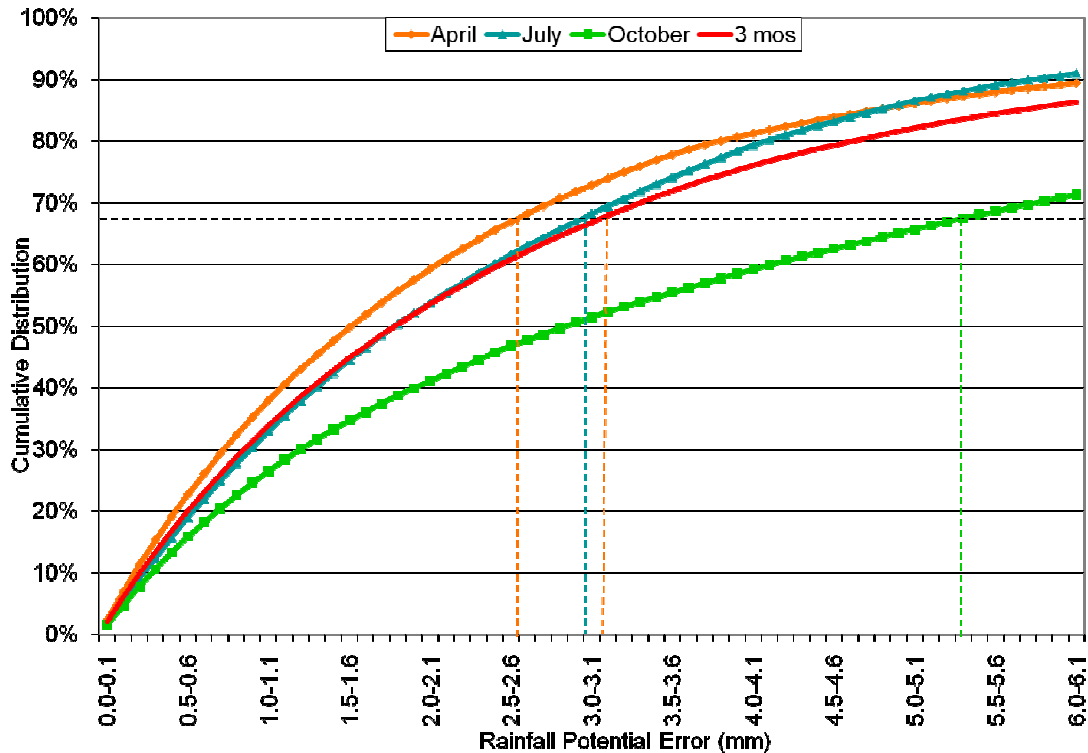
Figure 7. Cumulative distribution function of rainfall potential errors (absolute value of observation minus forecast accumulation) versus NIMROD data over Western Europe for the 5th-9th of April, July, and October 2005, plus the CDF curve for all data combined.

It should be noted that since the validation was restricted to western Europe, the validation statistics may not apply to the tropics. It is not clear how the performance over the tropics may differ; one the one hand, the input Rainfall Rate algorithm should have more skill in the tropics where convective rainfall is predominant. On the other hand, convective rainfall is also more variable in time and thus more difficult to nowcast skillfully using extrapolation methods. Validation over the tropics will be needed to quantitatively determine the difference in performance, but such data are difficult to obtain over the tropical portion of the SEVIRI coverage area.

4.3.2 Error Budget

The validation of the K-Means nowcasts driven by SCaMPR rain rates against NIMROD data for the 5th-9th of April, July, and October 2005 indicates that spec is generally being met over Western Europe. For reference, the accuracy specification refers to bias—the absolute difference between the mean observed and mean estimated rainfall. The precision specification is the 68th percentile of the cumulative distribution function of absolute errors; i.e., 68% of the absolute forecast errors will be below the precision value. It should be noted that these values exclude pixels where less than 1 mm of rainfall was observed and the rainfall potential was likewise less than 1 mm in an effort to reduce statistical variability between relatively dry and wet regions. However, these statistics are still based on a relatively small sample, so additional baseline validation will be

performed as more validation data become available. In the meantime, the results of the initial validation are summarized in Table 6.

	Accuracy (mm/h)	Precision (mm/h)	No. of data points
Vs. NIMROD (Apr)	1.2	2.6	208,895
Vs. NIMROD (Jul)	0.2	3.0	179,835
Vs. NIMROD (Oct)	3.3	5.4	98,199
Vs. NIMROD (3 mo)	0.2	3.2	486,929
F&PS	**5.0**	**5.0**	**-----**

Table 6. Comparison of Rainfall Potential validation with proposed F&PS.

5   PRACTICAL CONSIDERATIONS

5.1   Numerical Computation Considerations

Both the current and immediately previous rainfall rate products must be available to run the Rainfall Potential algorithm.

5.2   Programming and Procedural Considerations

Precipitation areas on the edge of a processing region will be problematic for the algorithm since their full shape and size cannot be known when part of the area is not within the field of view. Consequently, the retrieval should ideally be performed on the full image in order to avoid these problems. The processing of subregions is possible, but great care must be taken to avoid errors induced by precipitation areas that overlap the edges of these subregions.

The code for the Rainfall Potential Algorithm is in C++ and is highly modular to ease upgrades.

5.3   Quality Assessment and Diagnostics

Quality flags will be produced and provided along with the Rainfall Potential 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.

The following procedures are recommended for diagnosing the performance of the Rainfall Potential Algorithm.
- Periodically image the individual test results to look for artifacts or non-physical behaviors.
- Periodically evaluate time series of bias statistics of the algorithm output to identify any anomalous patterns.

5.4   Exception Handling

Quality control flags will be checked and inherited from the input Rainfall Rate fields, including bad data, missing sensor input data, and missing geolocation or viewing geometry information—thus, the algorithm expects the Level 1b processing to flag any pixels with missing geolocation or viewing geometry information. Missing (negative) values will be assigned to any pixel with quality issues or with any missing input values, and the error flags mentioned in Section 5.3 will also indicate these issues.

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

validation. The validation data will consist of Nimrod ground-based radar data over Western Europe, plus any ground-based radar data from field campaigns that can be obtained. This data set was 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 current limitations and assumptions in the current version of the Probability of Rainfall Algorithm.

6.1 Performance

The following assumptions have been made in developing and evaluating the performance of the Rainfall Potential Algorithm. The following list contains the current assumptions and proposed mitigation strategies.

1. The region over which the algorithm has been evaluated (Europe and Africa) represents the meteorological regimes found in the Western Hemisphere, and hence the validation statistics for that region accurately reflect performance in the GOES-R coverage region. (No mitigation possible).
2. The current and previous Rainfall Rate fields are available and accurate. (Please refer to the Rainfall Rate ATBD for details on mitigation of the latter).
3. The algorithm implicitly assumes that no new precipitation cells will form during the nowcast period. (Work with the Convective Initiation product team to incorporate their algorithm output into the Rainfall Potential algorithm).
4. The algorithm implicitly assumes that the strength of the rainfall features will not change with time; i.e., there is no accounting for growth and decay of precipitation. (Investigate the improvement of the growth / decay module in K-Means, which is currently deactivated due to a lack of impact on skill.)

6.2 Assumed Sensor Performance

We assume the sensor will meet its current specifications. However, the Rainfall Potential Algorithm will be dependent on the following instrumental characteristics because of their effects on the antecedent Rainfall Rate Algorithm.

- 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 brightness temperature difference calculations and thus compromise some of the predictors. Note that this will affect the retrieval only when any brightness temperature differences are among the predictors selected by the algorithm.

6.3 Pre-Planned Product Improvements

Work is being performed to optimize the tuning parameters in the algorithm that specific minimum rainfall rate, minimum cluster size, and other values. In addition, the persistence of rainfall features as a function of scale is being studied to determine if the extrapolation of features as a function of scale can be optimized.

7 REFERENCES

Brown, R. G., and P. Y. C. Hwang, 1996: *Introduction to Random Signals and Applied Kalman Filtering, 3rd edition*, Wiley, 496 pp.

Lakshmanan, V., K. Hondl, and R. Rabin, 2009: An efficient, general-purpose technique for identifying storm cells in geospatial images. *J. Ocean. Atmos. Tech.*, **26**, 523-537.

Lakshmanan, V., R. Rabin, and V. DeBrunner, 2003: Multiscale storm identification and forecast. *J. Atmos. Res.*, **67**, 36-380.

Lakshmanan, V., and T. Smith, 2009: Data mining storm attributes from spatial grids. *J. Ocean. Atmos. Tech.*, **26**, 2353-2365.