Development of Prototype Algorithms for Quantitative Precipitation Nowcasts From AMI Onboard the GEO-KOMPSAT-2A Satellite

Sukbum Hong, Dong-Bin Shin, Byeonghwa Park, and Damwon So

Abstract—Statistical approaches for quantitative precipitation nowcasts (QPNs) have emerged with recent advances in sensors in geostationary orbits, which provide more frequent observations at higher spatial resolutions. Advanced Meteorological Imager (AMI) onboard South Korea’s second geostationary satellite (GEO-KOMPSAT-2A), scheduled for launch in early 2018, is an example of these sensors. This paper introduces operational prototype algorithms that attempt to produce QPN products for GEO-KOMPSAT-2A. The AMI QPN products include the potential accumulated rainfall and the probability of rainfall (PoR) for a 3-h lead time. The potential accumulated rainfall algorithm consists of two major procedures: 1) identification of rainfall features on the outputs from the GEO-KOMPSAT-2A rainfall rate algorithm; and 2) tracking of these rainfall features between two consecutive images. The potential accumulated rainfall algorithm extrapolates precipitation fields every 15 min. Rainfall rates at each time step are accumulated to yield the 3-hourly rainfall. In addition, the extrapolated precipitation fields at 15-min intervals are used as inputs for the PoR algorithm, which produces the probability of precipitation during the same 3-h period. The QPN products can be classified as extrapolated features associated with precipitation. The validation results show that the extrapolated features tend to meet the designated accuracy for the prototype development stage. We also confirm a tendency for decreasing accuracy of the QPN products with increasing forecast lead time. Mitigating the dependence on lead time may remain a challenge that can be incorporated into the next generation of QPN algorithms.

Index Terms—Geostationary satellite, potential accumulated rainfall, probability of rainfall (PoR), quantitative precipitation nowcast.

I. INTRODUCTION

P R O T O T Y P E algorithms for quantitative precipitation nowcasts (QPNs) have been developed for the Advanced Meteorological Imager (AMI) onboard the second Korean geostationary satellite (GEO-KOMPSAT-2A), which is scheduled for launch in early 2018. The AMI QPN algorithms predict precipitation and probability of rainfall (PoR) for very short time ranges (0–3 h) using only satellite observations. The AMI QPN algorithms are based on statistical and probabilistic approaches that were similarly adopted in the rainfall nowcasting algorithms for the Advanced Baseline Imager onboard the Geostationary Operational Environmental Satellite (GOES)-R in [1]–[3]. Utilization of such satellite data is the focus of this paper, with the aim of improving QPN results.

Precipitation is an exceedingly important meteorological phenomenon because it strongly affects agriculture, industry, and daily life. Moreover, accurate forecasting of precipitation can save human lives and reduce economic losses from extreme weather damage such as thunderstorms, tropical depressions, and cyclones. To obtain the best predictions of precipitation, observations have been made using rain gauges, radar, satellite data collection, and numerical models. Radar is considered the most useful instrument for forecasts over very short time ranges because it provides the high spatial and temporal resolution necessary for precipitation estimates with high accuracy. Radar-based nowcasts currently use extrapolation and statistical analyses such as Tracking Radar Echo by Correlation (TREC) [4] and Thunderstorm Identification Tracking Analysis and Nowcasting (TITAN) [5]. To extend these techniques to satellite-based nowcasting, the AMI QPN algorithms are developed with proxy data from the Spinning Enhanced Visible and Infrared Imager (SEVIRI) onboard the METEOSAT-9 satellite.

The well-known extrapolation techniques introduced by Pierce et al. [6] and Reyniers [7] can be classified into area- and cell-tracking methods. The first strategy divides the latest images of rainfall fields into grids and compares those grids with previous images to detect changes in rainfall patterns. The latter strategy compares the center of individual rainfall fields with previous images of the same area to determine the direction of the rainfall field. Area-tracking methods (e.g., TREC and the McGill Algorithm for Precipitation nowcasting by Lagrangian Extrapolation, i.e., MAPLE [8]) are expected to perform better for larger enveloping rainfall fields. Meanwhile, cell-tracking methods (e.g., TITAN and Storm Cell Identification and Tracking, i.e., SCIT [9]) are optimized in convective rainfall fields. The AMI QPN tracking method is a blend of area- and cell-tracking methods to improve motion vector estimates that utilize the classification of rainfall type based on the AMI rainfall rate algorithm. More details on the AMI rainfall rate algorithm can be referred to in Shin et al. [10].

Another type of precipitation nowcast, precipitation probability, is widely used to predict rainfall rates because it provides
quantitative information in a precise and unambiguous manner [11]. The most popular method for quantifying the PoR is based on statistical analyses of past observations [12]. The great advantage of AMI QPN is conflation of data from two separate analytical strategies, through which the PoR can be more accurately estimated by statistical analysis of 12 consecutive potential rainfalls for the next 3 h. Despite the simple concept of this technique, there may be inherent errors due to incorrect estimations of rainfall rates, computational anomalies in calculating potential accumulated rainfall, linear extrapolation failures resulting from growth or decay of rainfall features, and changes in velocity [13]. However, the PoR based on the AMI QPN algorithm is expected to reduce such errors.

Since the AMI QPN algorithms adopt precipitation data estimated from infrared (IR) observations, it is worthwhile to briefly review the characteristics of IR-based precipitation measurements. IR-based precipitation estimates from satellites are typically resolved by radiation emitted from precipitating clouds that is related to cloud top temperature for optically thick clouds. IR-based estimation assumes that colder cloud top temperatures tend to be associated with higher precipitation rate (e.g., [14]–[16]). As pointed by many investigators, including Scofield and Kuligowski [17], this assumption works better for precipitating clouds with cold top temperatures such as tall convective clouds than for warm or shallow precipitating clouds. Based on the understanding of the different relationships between IR cloud top observations and precipitation, more advanced algorithms utilize the cloud physical properties and microwave-calibrated precipitation data (e.g., [10] and [18]). Precipitation estimation from IR data still has limitations since surface precipitation is indirectly inferred from cloud top observation. However, IR-based precipitation estimates are the only precipitation data available on a continuous basis in remote locations.

This paper presents the prototypes of the AMI QPN algorithms. The data sets for testing these algorithms are described in Section II, and a description of these algorithms is presented in Section III. The results of the AMI QPN algorithms are described in Section IV, followed by a summary of future enhancements to the algorithms.

### II. DATA SET

The development of QPN algorithms for AMI begins with estimations on rainfall rates from SEVIRI onboard the METEOSAT-9 satellite, which has spectral bands of 6.25–12.0 μm and spatiotemporal resolutions of 3 km and 15 min, respectively, similar to those of the AMI. The AMI rainfall rate is estimated from the *a priori* information that is constructed with the IR brightness temperatures from GEO satellites and microwave rainfall data from low-Earth orbiting satellites. This *a priori* database is classified into eight types according to the cloud types and latitudinal bands. Using the brightness temperature difference between IR channels, a threshold is determined to discriminate the cloud types into shallow and not-shallow precipitating clouds. The latitudinal bands are separated into four latitudinal zones: 60°–30° S, 30°–EQ, EQ–30° N, and 30°–60° N. The database is used to invert IR brightness temperatures to the surface rainfall rate based on the Bayesian approach. The Bayesian approach possesses the advantages of simultaneously using multichannel brightness temperatures and using the PoR reserved in the *a priori* database. A detailed description of the AMI rainfall rate algorithm is provided in the theoretical basis document for the algorithm [10].

The outputs from the AMI rainfall rate algorithm utilizing the SEVIRI data are precipitation fields estimated every 15 min, fed into the AMI potential accumulated rainfall algorithm, which, in turn, provides inputs for the AMI probability of the rainfall algorithm. Data for evaluating the AMI QPN algorithms collected from six events during 2012 are listed in Table I. In summary, the AMI potential accumulated rainfall algorithm uses two consecutive rainfall rate images for each event in the selected periods; similarly, the AMI probability of the rainfall algorithm uses 12 consecutive instantaneous potential rainfall images.

### III. DESCRIPTION OF ALGORITHM

The AMI QPN algorithms produce two products, namely, the potential accumulated rainfall and the PoR. The potential accumulated rainfall algorithm provides the 3-h accumulated rainfall for a very-short-range forecast using extrapolation, whereas the PoR algorithm yields the rainfall probability during the same time period through statistical methods.

#### A. Potential Accumulated Rainfall

The potential accumulated rainfall algorithm first identifies rainfall features and computes the motion vectors of the identified rainfall features through extrapolation. The algorithm can then extrapolate future rainfall rates from current and previous rainfall rate outputs, as evidenced by satellite observations. Several investigations, including [19]–[21], have reported that extrapolation-based prediction is more effective than numerical weather model-based prediction for a short lead time of less than a few hours. A primary assumption of the potential accumulated rainfall algorithm is that the rainfall feature is advected by a uniform velocity field, whose rainfall rate remains unchanged because of Lagrangian persistence, as follows [8]:

$$ rr(x, y, t + Δt) = rr(x + Δx, y + Δy, t) $$

where $rr$ is the rainfall rate, $Δt$ is the lead time, and $Δx$ and $Δy$ are the $x$- and $y$-components of the motion vectors, respectively.
The rainfall feature has to be defined as a region of a significant precipitation field that is separated from other regions. The rainfall features are broken into rain pixels and non-rain pixels by the threshold as follows:

\[ rr(x, y) \geq 1 \text{ mm/h}. \]  

(2)

The identification of rainfall features in the prototype algorithm is based on the threshold in (2) and is subsequently enhanced by two-step smoothing filters. The positive impact of smoothing has been demonstrated in previous studies [22], [23]. As the first smoothing step, the rainfall rates in each pixel are substituted by the median rainfall rates in an area of 11 pixels by 11 pixels. For the second step, the average smoothing filter in an area of 11 pixels by 11 pixels is used to enhance the identification of the rainfall features again. After detecting a rainfall feature, the center and size of the feature are determined from the maximum value of rainfall rates in the rainfall features. The identified rainfall feature is masked for removal from the current image, and then multiple rainfall features are detected for multiple iterations following the same procedure.

Once the identification of rainfall features is complete in the current and past satellite images, rainfall features are analyzed to determine their movement. The AMI tracking technique is a combination of area and cell tracking to determine the motion vector of rainfall features. After identifying the rainfall features in the current satellite image, the motion vector is computed for the center of the rainfall features using the cross correlation in (3) between the rainfall features in the current image and those around the search area in the previous image.

\[
C(x, y) = \frac{\sum_{i,j=1}^{M} [(x_{i,j} - \bar{x})(y_{i,j} - \bar{y})]}{\left( \sum_{i,j=1}^{M} (x_{i,j} - \bar{x})^2 \right)^{1/2} \left( \sum_{i,j=1}^{M} (y_{i,j} - \bar{y})^2 \right)^{1/2}}
\]

where \( M \) refers to the maximum diameter of each rainfall feature that has been determined by the ratio of rain pixels to total pixels reaching 0.7 within the rainfall feature or 101 pixels, whichever is smaller. \( x \) and \( y \) are the rainfall rates, and \( \bar{x} \) and \( \bar{y} \) are the averages of \( x \) and \( y \) for the current and previous images, respectively.

A flowchart of the AMI potential accumulated rainfall algorithm is shown in Fig. 1, and user-selectable input parameters for the prototype potential accumulated rainfall algorithm are listed in Table II.

### B. PoR

The probabilistic now cast for AMI derives a precipitation prediction of more than 1.0 mm of rainfall during the next 3 h. The PoR algorithm takes the outputs from the potential accumulated rainfall algorithm as inputs, i.e., 12 instantaneous potential rainfalls. The potential rainfalls can then be regarded as 12 trials that directly convert the PoR in each pixel as follows:

\[
\text{PoR}_1 = \left( \frac{\sum_{t=1}^{12} \alpha_t n_t + N \sum \text{rr}}{12} \right) \times 100\% \]

(4)

where \( n_t \) is a frequency index indicating raining or non-raining pixels for the \( t \)th image. The value of \( n_t \) is defined as \( n_t = 1 \) for rainfall rates \( \geq 1 \text{ mm/h} \), and \( n_t = 0 \) otherwise.

The total count of \( n_t \) for the 12 images of the potential rainfalls then ranges from 0 to 12. In the calculation of the PoR, however, decreasing accuracy of the potential rainfall estimates with increasing time intervals may occur. Moreover, the estimated rainfall rate may have its own particular uncertainty at the lower limit of rainfall intensity. In order to consider both the lower limit uncertainty in the rainfall rate estimates and the degrading quality of the potential rainfall estimates as a function of time intervals, twofold corrections are made on the PoR considering both uncertainty of rainfall rate estimation and

---

**TABLE II**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{Nfr} )</td>
<td>1 mm/hr</td>
<td>Identification</td>
</tr>
<tr>
<td>( \text{med}(x, y) )</td>
<td>11 x 11 pixels</td>
<td>Identification</td>
</tr>
<tr>
<td>( \text{r} )</td>
<td>15 pixels</td>
<td>Identification</td>
</tr>
<tr>
<td>( \text{r}_{\text{max}} )</td>
<td>0.7 or 50 pixels</td>
<td>Identification</td>
</tr>
<tr>
<td>( \text{r}_{\text{search}} )</td>
<td>9 pixels</td>
<td>Tracking</td>
</tr>
<tr>
<td>( \text{Ncell} )</td>
<td>25</td>
<td>Identification &amp; Tracking</td>
</tr>
</tbody>
</table>

(1) is the threshold of rainfall rates, (2) is the median smoothing filter, (3) is the minimum radius of the rainfall features, (4) is the maximum radius of the rainfall features, (5) is the radius of search area in the past satellite image, and (6) is the total number of rainfall features.
weight function for the frequency index. The uncertainty of rainfall rates $N_{r_{rr}}$ is set to 1 if the sum of the potential rainfalls over the entire time span is greater than 0.1 mm; otherwise, it is set to 0. The weight function $a_t$ has been also applied to the PoR equation. The weight can be empirically estimated by the following formula:

$$a_t = \frac{1}{\sum_{t=1}^{12} \frac{1}{\sigma_t}}$$

where $\sigma_t$ is expressed as follows:

$$\sigma_t = \sqrt{\frac{1}{N} \sum_{i=1}^{\max} \sum_{j=1}^{\max} (r_{obs_{i,j}} - r_{for_{i,j}})^2}$$

where $r_{obs}$ is the observed rainfall rate, and $r_{for}$ is the predicted instantaneous potential rainfall. The weight function of the prototype PoR algorithm is calculated from the average of two events on July 28 and 29, 2012. In addition, the potential rainfall estimation contains possible errors due to uncertainties in tracking the motion vector and in defining raining areas. In order to mitigate such problems, the PoR is also estimated for non-raining pixels close to raining pixels. The neighboring pixel of a raining pixel gives the PoR estimated as follows:

$$\text{PoR}_2 = \left( \frac{\langle n \rangle_{15} + \sigma_{15}}{15} \right) \times 100\%$$

where $\langle n \rangle_{15}$ is the average frequency within a 15-pixel radius of a target pixel, and $\sigma_{15}$ is the average standard deviation. The flowchart of the AMI PoR algorithm is shown in Fig. 2.

IV. RESULTS AND VALIDATION

Satellite rainfall rate data estimated from the AMI rainfall rate algorithm were collected for six events, each with a 4-h duration period, during 2012. The distributions of the rainfall rate estimates indicate decaying rainfall features in most cases, except for the July 19 and July 27 events. The highest values, based on the observations, are 48.8 mm/h of rainfall rate and 102.8 mm for 3-h accumulated rainfall. The collected rainfall rate data for six different rain events are used to test the prototype AMI QPN algorithms.

The results are shown in Figs. 3 and 4 for the potential accumulated rainfall algorithm and the PoR algorithm, respectively. Validations of the QPN estimates over the full disk observations of the SEVIRI sensor are performed using scalar accuracy measures, including the correlation coefficient (Corr), bias (Bias), and root mean square error (RMSE). Categorical accuracy measures such as probability of detection (POD), false alarm rate (FAR), and Heidke skill score (HSS) [24] are also adopted for the validation. The retrieved potential accumulated rainfalls are validated with the rainfall data based on SEVIRI observations for the precipitation cases on July 1, 2, 3, 5, 19, and 27, 2012. Fig. 5 presents scatter plots of the retrieved potential accumulated rainfalls versus 3-h accumulated rainfall. Table III summarizes the validation statistics for the retrieved potential accumulated rainfalls with the 3-h accumulated rainfall. The accuracy requirements for the potential accumulated rainfall are 10 and 5 mm for pixels classified as raining at the prototype-development stage and at the final stage, respectively. The current prototype algorithm has a Bias ranging from 0.121 to 0.267 mm and RMSE values in the range of 3.781–7.948 mm. The accuracy of the prototype algorithm based on the ranges of bias and RMSE tend to meet the requirement values at the prototype development stage. Note that all potential accumulated rainfalls are larger than those of the 3-h rainfall accumulation, as shown by the positive biases. This problem may be improved by accounting for growth or decay of rainfall features in a future version of the algorithm. Note also that the forecasting accuracy of the potential accumulated rainfall algorithm decreases with time, as illustrated in Fig. 6. This finding may be due to the limitations of the extrapolation technique and degradation of the accuracy of the potential accumulated rainfall estimates. Table IV summarizes the validation statistics for the retrieved 1-h potential accumulated rainfall and rainfall accumulation. For comparison, TITAN yields a POD value of 0.59 and a FAR value of 0.68 at a forecast time of half an hour [5], but the average values for the AMI potential accumulated rainfall are 0.49 and 0.26 for POD and FAR, respectively. One problem with comparing the results of AMI potential accumulated rainfall with those of TITAN is that there is a difference in methodology. The comparison indicates that the accuracy of the AMI potential accumulated rainfall algorithm is comparable with that of the results from TITAN at a forecast time of 1 h.

A major advantage of PoR is that this method provides a statistical measure to discrete rain/no-rain information. The discrimination of rain/no-rain information appears to be easier than matching a continuous rainfall distribution. For a quantitative comparison of the retrieved PoR and 3-h accumulated rainfall, a pixel with the PoR higher than 50% is considered as a raining pixel. The validation statistics for the PoR are listed in Table V: both scalar and categorical accuracy measures show similar values for the different cases. As the raining and
Fig. 3. Potential accumulated rainfalls derived from the satellite rainfall rates for six different rain events: (a) 1600–1845 UTC July 1, 2012; (b) 1600–1845 UTC July 2, 2012; (c) 1600–1845 UTC July 3, 2012; (d) 1800–2045 UTC July 5, 2012; (e) 1500–1745 UTC July 19, 2012; and (f) 1800–2045 UTC July 27, 2012.

V. SUMMARY AND FUTURE WORK

The prototype algorithms of potential accumulated rainfall and PoR have been introduced for the AMI on GEO-KOMPSAT-2A. The AMI algorithms predict precipitation rates and probability of occurrence for the next 3 h. The AMI algorithms are based on the accuracy of the AMI rainfall rates and assume constant rainfall features in the next 3 h. Evidently, these assumptions are not always true; thus, the challenge of how to improve the precipitation forecast in very short time range remains. In this paper, data from SEVIRI onboard the METEOSAT-9 satellite were used as a proxy for AMI. A total of six events on July 2012 over the full disk were selected for testing the AMI QPN algorithms.

The results also show the positive effect of smoothing filters for identifying the rainfall features and to estimate their motion vector. The positive bias from the results of potential accumulated rainfalls indicates overestimation of precipitation, which provides an important clue about the advection of the rainfall features as they change over time. This challenge can be addressed by accounting for the growth or decay of rainfall features in a future version of the potential accumulated rainfall algorithm. The results of the categorical accuracy measures for both potential accumulated rainfall and PoR show significant drawbacks associated with a pixel-by-pixel comparison. Although the accuracy of measurements decreases over time, this accuracy may be improved by advanced motion vector estimations such as applying the weight function and the Kalman filter. Despite the difficulty in tracking rainfall features due to rapid changes over time, the classification of rainfall types offered by the AMI rainfall rate algorithm can assist with the measurement of motion vectors. The accuracy of the PoR prediction relies not only on the results of potential accumulated
Fig. 4. Same as Fig. 3 but displaying the PoR.

Fig. 5. Scatter plots of the retrieved potential accumulated rainfalls versus corresponding 3-h accumulated rainfalls for the six different rain events. The dashed lines indicate the one-to-one line between forecasts and observations.
rainfall but also on rainfall rates. However, improvement to the accuracy of PoR predictions may be achieved through statistical analysis of rainfall types provided by the AMI rainfall rate algorithm.
Sukbum Hong received the B.S. degree in astronomy from The University of Arizona, Tucson, AZ, USA; the M.S. degree in physics from Rutgers, The State University of New Jersey, New Brunswick, NJ, USA, in 2006; and the Ph.D. degree in physics from George Mason University, Fairfax, VA, USA, in 2012. He is currently a Postdoctoral Researcher with the Department of Atmospheric Sciences, Yonsei University, Seoul, South Korea. His research interests include image processing and machine learning techniques in remote sensing data analysis.

Dong-Bin Shin received the B.S. and M.S. degrees from Yonsei University, Seoul, South Korea, in 1987 and 1989, respectively, and the Ph.D. degree from Texas A&M University, College Station, TX, USA, in 1999, all in atmospheric sciences. He was a Research Professor with the Center for Earth Observing and Space Research, Department of Computational and Data Sciences, George Mason University, Fairfax, VA, USA, and was also with the Department of Atmospheric Science, Colorado State University, Fort Collins, CO, USA, as a Research Associate from 2000 to 2007. He is currently with the Department of Atmospheric Sciences, Yonsei University, where he was first an Associate Professor and has been a Professor since 2007. His research articles have appeared in major scientific journals. His research interests include physically based remote sensing of clouds and precipitation from infrared and microwave observations and radiative transfer modeling. His research is also focused on understanding global and regional hydrological balances.

Prof. Shin is a member of the American Meteorological Society, the American Geophysical Union, and the IEEE Geoscience and Remote Sensing Society.

Byeonghwa Park received the B.S. degree in management information systems from The University of Arizona, Tucson, AZ, USA; the M.A. degree in management from the University of Nebraska–Lincoln, Lincoln, NE, USA; and the M.S. and Ph.D. degrees in computational sciences and informatics from George Mason University, Fairfax, VA, USA. He is currently an Assistant Professor with the Department of Business Statistics, Hannam University, Daejeon, South Korea. His current research interests include big data analysis, data mining, statistical applications, and interdisciplinary areas.

Damwon So received the B.S. degree in atmospheric sciences from Yonsei University, Seoul, South Korea, in 2012, where she is currently working toward the Ph.D. degree with the Department of Atmospheric Sciences.

Her research topics include rainfall rate retrieval using infrared data and characteristic analysis of cloud types from infrared and microwave observations.