Detection of deterministic and probabilistic convective initiation

using Himawari-8 Advanced Himawari Imager data

Sanggyun Lee[1], Hyangsun Han[2], Jungho Im[1,3], Eunna Jang[1], and Myong-In Lee[1]

[1]School of Urban and Environmental Engineering, Ulsan National Institute of Science and Technology (UNIST), Ulsan, 44949, South Korea
[2]Unit of Arctic Sea-Ice prediction, Korea Polar Research Institute, Incheon, 21990, South Korea
[3]Environmental Resource Engineering, State University of New York, College of Environmental Science and Forestry, 13210, Syracuse, NY, USA

Correspondence to: Jungho Im (ersgis@unist.ac.kr)

Abstract. Detection of Convective Initiation (CI) is very important because convective clouds bring heavy rainfall and thunderstorms that typically cause severe socioeconomic damages. In this study, deterministic and probabilistic CI detection models based on decision trees (DT), random forest (RF), and logistic regression (LR) were developed using Himawari-8 Advanced Himawari Imager (AHI) data obtained from June to August 2015 over the Korean Peninsula. We used a total of 12 interest fields that contain brightness temperature, spectral differences of the brightness temperatures, and their time trends to develop CI detection models. While the interest field of 11.2 μm T_B was considered the most crucial to detect CI in the deterministic models and the probabilistic RF model, the trispectral difference, i.e., (8.6-11.2 μm) - (11.2-12.4 μm), was determined as the most important one in the LR model. The performance of the four models varied by CI case and validation data. Nonetheless, the DT model typically showed higher POD, while the LR model produced higher OA and CSI and lower FAR than the other models. The CI detection of the mean lead times by the four models were in the range of 32 - 35 min, which implies that convective clouds can be detected in 30 min advance before precipitation intensity exceeds 35 dBZ over the Korean Peninsula in summer using the Himawari-8 AHI data is possible.

1 Introduction

Atmospheric deep moist convection initiates shallow cumulus clouds, which may continue to grow vertically as cumulonimbus clouds, and this process is so called the convective initiation (CI) (Banacos et al., 2005; Bluestein et al., 1990; Weckwerth and Parsons, 2006). The moist convection appears in a variety of horizontal scales ranging from 1-10 km as individual convective clouds to ~1000 km as mesoscale convective systems (Houze Jr., 2004; Roberts and Lean, 1998; Miyamoto et al., 2013) with heavy rainfall and thunderstorm events (Amorati et al., 2000; Sieglaff et al., 2011; Zuidema, 2003; Haile et al., 2010; Hane et al., 2002; Vondou et al., 2010). The convective events in Northeast Asia often occur during the summer season accompanied by many meteorological hazards such as lightning, floods, and strong winds (Kim and Lee, 2006; Wang et al., 2004). These

Atmospheric
Measurement
Techniques
Discussions

Open Access

EGU

hazards destroy infrastructure in the region and result in huge economic losses. Therefore, it has been desired to forecast CI in Northeast Asia with high accuracy in order to prevent socioeconomic damages caused by the convective events.

The decrease of atmospheric stability drives CI, which is attributed to various weather systems such as large-scale monsoonal fronts, the migration of frontal cyclones, and mesoscale convective systems (Craven et al., 2002; Houze Jr., 2004; Mecikalski

5   and Bedka, 2006). Although such unstable weather systems can increase the potential risk of CI over vast area, they actually trigger CI occupying much smaller area, making it difficult to predict the exact location. CI is characterized by the rapid variation of temperature and the increase of cloud-tops, which can be effectively measured by the brightness temperature (T_B) changes at multispectral channels including visible and infrared (IR) (Mecikalski and Bedka, 2006; Mecikalski et al., 2010). Geostationary satellites carry optical sensors that scan over a few thousand square kilometers with high temporal resolution (~

10   minutes) in the multispectral channels. Therefore, these geostationary satellites can be extremely useful in CI nowcasting. Previous studies developed CI nowcasting algorithms for geostationary satellites by determining a threshold or a range of values of T_B at specific channels, and their spectral and/or temporal differences (Mecikalski and Bedka, 2006; Mecikalski et al., 2008; Walker et al., 2012; Morel and Senesi, 2002; Jewett et al., 2013; Merk and Zinner, 2013; Siewert et al., 2010; Sobajima, 2012; Han et al., 2015). Geostationary Operational Environmental Satellite (GOES) systems and Meteorological

15   Second Generation (MSG) are the representative geostationary satellites operated at the National Oceanic and Atmospheric Administration (NOAA) and European Organization for the Exploitation of Meteorological Satellites (EUMESAT), respectively. These two satellites have forecasted CI using their operational algorithms, i.e., SATellite Convection Analysis and Tracking (SATCAST) and Rapid Development Thunderstorms (RDT), respectively, which are basically based on empirical determination of the thresholds of interest fields in terms of CI development (Mecikalski and Bedka, 2006; Walker

20   et al., 2012; Morel and Senesi, 2002). These algorithms were assessed for CI cases in North America and showed a probability of detection (POD) over 0.8 (80%) and a false alarm rate (FAR) around 0.6 (60%). However, these algorithms for CI detection have not been validated over Northeast Asia.

Several algorithms for detecting CI over Northeast Asia were developed for the Multi-functional Transport SATellite-2 (MTSAT-2) geostationary satellite operated by Japan Meteorological Agency (JMA) and the Communication, Ocean, and

25   Meteorological Satellite (COMS) by Korea Meteorological Administration (KMA). The main instrument of MTSAT-2 is Imager which is composed of a total of 5 channels: a visible channel with 1 km-spatial resolution and 4 infrared channels with 4 km spatial resolution. JMA have developed the algorithm for the detection of CI using MTSAT-2 Imager over Japan area i.e., the Rapidly Developing Cumulus Areas (RDCA) derivation algorithm (Sobajima, 2012). The RDCA algorithm detects CI using several interest fields, but it was validated only during the summer season in 2011 (Sobajima, 2012). The performance

30   of the MTSAT-2 RDCA algorithm needs to be evaluated in different times and circumstances. Han et al. (2015) developed CI detection algorithms for COMS Meteorological Imager (MI) data by determining new rules and thresholds for the interest fields used in the RDCA through machine learning, based on the fact that the characteristics of the spectral channels of COMS MI is similar to those of MTSAT-2. The algorithms were validated for various CI cases over Korea, demonstrating good

Atmospheric
Measurement
Techniques
Discussions

Open Access
EGU

performance with a POD as high as 75.5% and a FAR as low as 46.2%. Himawari-8, launched on 7 October 2014, is one of the geostationary satellites operated by JMA. The primary payload of Himawari-8 is the Advanced Himawari Imager (AHI) which captures the Asia-Pacific region using a 16-channel multispectral imager composed of 3 visible (blue, green, and red) and 13 IR channels with a spatial resolution of 0.5–2 km depending on the spectral channel (Bessho et al., 2016). Himawari-8

5   AHI scans full disk, the whole Earth as seen from the satellite (11000 × 11000 km), every 10 min. The scan interval for the full disk of Himawari-8 AHI is much shorter than that of MTSAT Imager (60 min) and COMS MI (180 min). Furthermore, the spatial and spectral resolutions of Himawari-8 AHI have been substantially improved in comparison to its predecessors. Therefore, Himawari-8 AHI can help enhance the performance of CI detection in Northeast Asia. However, there is no available algorithm for CI detection for Himawari-8 AHI so far. The spectral characteristics of Himawari-8 AHI are

10   comparable to the Advanced Baseline Imager (ABI) integrated onto the Geostationary Operational Environmental Satellite-R series (GOES-R) satellites (Schmit et al., 2005), a series of geostationary satellites of which the first one will be launched in November 2016 and operated by National Oceanic and Atmospheric Administration (NOAA). University of Alabama in Huntsville with NOAA developed a CI detection algorithm for GOES-R using 12 interest fields designed for the spectral bands of ABI (Walker and Mecikalski, 2011; Walker et al,. 2012; Melcikalski et al., 2015). The interest fields of the GOES-R CI

15   algorithm can be directly adopted for Himawari-8 AHI. However, the critical threshold values of the interest fields were empirically determined based on simulations and have not been fully assessed due to the lack of ABI-class datasets. This implies that the criteria of the interest fields should be optimized for use with Himawari-8 AHI data.

The existing literature on nowcasting CI commonly used deterministic approaches which classify clouds into CI and non-CI using the criteria of several interest fields based on simple thresholding approaches (Mecikalski and Bedka, 2006; Mecikalski

20   et al., 2008; Mecikalski et al., 2010; Morel and Senesi, 2002; Sieglaff et al., 2011; Roberts and Rutledge, 2003; Sobajima, 2012). Such deterministic approaches might provide incorrect classification results for unsampled pixels (or objects), especially around the boundaries of CI, increasing FAR of predictions. Probabilistic approaches produce significantly lower FAR than the deterministic ones through the selection of an appropriate probability threshold (Mecikalski et al., 2015). Mecikalski et al. (2015) developed CI nowcasting algorithms combining the interest fields derived from GOES and numerical

25   weather prediction (NWP) model data based on probabilistic approaches. They validated the performance of the probabilistic algorithms for CI cases in the United States, resulting in FAR of 10-18%, which is much lower than the existing deterministic CI detection algorithms for GOES (FAR ~48–60%) (Walker et al., 2012). However, Mecikalski et al. (2015) used fewer satellite-based interest fields than the GOES-R CI algorithm due to the limited number of spectral channels of GOES, which implies that it is unknown if such probabilistic approaches would work for Himawari-8 AHI as well.

30   In this study, CI detection algorithms for Himawari-8 AHI are developed and validated over the Korean Peninsula in East Asia. The objectives of this research were to (1) develop deterministic and probabilistic CI detection algorithms for Himawari-8 AHI data based on rule-based decision trees and random forest approaches and a logistic regression modelling technique, (2) evaluate the CI detection models in terms of performance and efficiency, (3) assess the strengths and weaknesses of the deterministic and probabilistic CI detection models based on CI cases and validation datasets, and (4) examine key predictor

Atmospheric
Measurement
Techniques
Discussions

variables for CI detection. This study extends our previous research in Han et al. (2015), where the COMS MI data were used. One of the main limitations of using COMS MI data in the previous study is its relatively coarse spatial resolution (4km), which is not enough to detect small convective clouds. The himawari-8 AHI used in the present study has higher spatial resolution of 2 km for IR channels. In addition, the higher spectral (i.e., 16 channels) and temporal (i.e., 10 min) resolutions

5   of AHI than COMS MI can significantly improve the forecast skill. The present research proposes not only deterministic approaches but also probabilistic ones for CI detection using Himawari-8 AHI data. Consequently, the use of advanced geostationary satellite data and various modelling techniques is expected to produce better CI forecast performance.

2 Data

10   ### 2.1 Himawari-8 AHI

The specifications of Himawari-8 AHI are summarized in Table 1. Himawari-8 AHI scans three levels of regions: full disk, Japan area, and target area. Full disk images are acquired over the whole Earth as seen from the satellite every 10 min. Japan area images, collected every 2.5 min, cover the northeastern and southwestern areas of Japan (3,000 × 3,000 km). Himawari-8 AHI also scans a target area of 1,000 × 1,000 km every 2.5 min, which is dedicated to monitoring high impact meteorological

15   events. In this study, the full disk images were used even though Himawari-8 AHI scans the Japan area four times more often. This is because the Japan Area images do not cover the upstream side of storm developments in the East China Sea and western part of the Yellow Sea, which are important areas to forecast CI over the Korean Peninsula and Japan. The full disk images obtained for 10 cases of CI from June to August 2015 (Table 2) were used to develop and validate the deterministic and probabilistic CI detection models.

20   ### 2.2 Weather radar echo and lightning data

Rainfall with ≥35 dBZ precipitation intensity measured by weather radar has been known to have significant correlation with the eventual development of cumulonimbus clouds (Mecikalski and Bedka, 2006; Mueller et al., 2003). Therefore, the threshold of ≥35 dBZ precipitation intensity has been widely used as the definition of convective events (Mecikalski and Bedka, 2006; Mecikalski et al., 2008, 2010; Roberts and Rutledge, 2003; Walker et al., 2012). In this study, the first occurrence of

25   rainfall with ≥35 dBZ precipitation intensity was defined as CI. KMA has operated a total of 10 weather radars in South Korea. They have produced Plan Position Indicator (PPI) in which the precipitation echoes measured at a given elevation angle are projected on a plane every 10 min and Constant Altitude PPI (CAPPI) images, which are calculated using several PPI elevations. A 1.5 km CAPPI with mosaic image was used to determine the areas and time of CI occurrences. Since the effective radius of the 1.5 km CAPPI is about 100 km, only the 1.5 km CAPPI echo at each radar within 100 km were used.

Lightning observation data was used as supplementary data for validating the CI detection results, especially for ocean areas. The lightning data were provided by KMA, which operates a ground-based Total Lightning Detection System (TLDS) since 2001. The TLDS has an average accuracy of 90% for lightning detection, with locational accuracy of 500 m over the land and 2 km over the ocean (Kar and Ha, 2003). The TLDS lightning observation data during the same period were used with the

5 ground radar measurements for the validation of the CI detection models.

3 Methods to detect CI using geostationary satellite data

Because of the large similarity in spectral channels between Himawari-8 AHI and GOES-R ABI, one may consider to adopt the GOES-R CI algorithm directly to develop CI algorithms for Himawari-8 AHI. However, as mentioned, the GOES-R CI algorithm uses simple threshold values associated with the interest fields and the values were determined through many

10 experimental simulations in a subjective way. The interest fields and threshold values have not been validated for Himawari-8 AHI. In order to develop more objective CI models for Himawari-8 AHI, rule-based decision trees and random forest machine learning approaches were used as well as a logistic regression model. The interest fields identified in the GOES-R CI algorithm were used as predictor variables in both deterministic and probabilistic approaches. The processing flow diagram of the proposed CI detection approaches is shown in Fig. 1.

15 ### 3.1 Interest fields of GOES-R ABI CI algorithm

The interest fields of the GOES-R ABI CI algorithm (Table 3) were used as a set of predictors to develop the deterministic and probabilistic CI detection models for Himawari-8 AHI. All the interest fields are calculated only from IR channels in order to predict CI using both daytime and nighttime images. T_B measured at 11.2 μm and its time trend represent cloud-top temperature and cloud-top cooling rate, respectively (Mecikalsk and Bedka, 2006; Mecikalski et al., 2010; Walker and

20 Mecikalski, 2011). The interest fields from spectral differences provide information on cloud-top height (cloud depth) and glaciation at the time of image, while those from temporal differences provide information on the rate of vertical cloud-top growth.

Prior to the calculation of the interest fields, cloudless and cirrus regions were masked out from Himawari-8 AHI images using the criteria of the T_B at 11.2 μm-channel < 288.5 K, -3 K < 12.4-11.2 μm-channels difference < 3 K, and -3 K < the trispectral

25 difference < 3 K. Such criteria have been empirically used to identify clear sky and thin clouds in summer by KMA. The first criterion was used to remove land surface and cirrus from AHI images, while the others were employed to remove clear sky areas. To determine the time-dependent interest fields, it is essential to track the motion of cloud objects. A simple temporal overlap object tracking method (Walker and Mecikalski, 2011) was adopted in this study, which uses two consecutively obtained AHI images in order to track cloud objects. This object tracking method has a weakness that fast moving cloud objects

of small size might not be traced. However, the frequent scanning interval and fine spatial resolution of Himawari-8 AHI help mitigate the weakness of the tracking method. The motion of cloud objects was traced from two consecutive 11.2 μm-channel images of AHI and then the time-dependent interest fields were calculated.

3.2 Deterministic and probabilistic approaches for CI detection

5   The event of CI (i.e., CI vs. non-CI) is used as a dependent (i.e., response) variable in the CI detection models based on deterministic and probabilistic approaches. For seven Himawari-8 AHI image data including CI events over the Korean Peninsula, the pixels within cloud objects corresponding to the first occurrence of ≥35 dBZ precipitation intensity were extracted and considered as convective clouds (CI areas). The other clouds were identified as non-CI regions. In order to collect interest field samples to develop CI detection models, each CI and non-CI area was tracked through visual interpretation of

10   the 11.2 μm-channel images obtained 10–60 min before the first occurrence of ≥35 dBZ precipitation intensity, rather than using the temporal overlap object tracking method (Zinner et al., 2008) that was used to calculate the time-dependent interest fields. A total of 3,204 CI reference data (1,324 CI and 1,880 non-CI samples) were extracted from the AHI images and used to train and validate the deterministic and probabilistic CI detection models. From eight percent of the total images (42 scenes), 1,060 CI samples and 1,504 non-CI samples were extracted and used as training data, while 264 CI samples and 376 non-CI

15   samples extracted from the remaining AHI images (18 scenes) were used to validate the models. The models were further validated using 3 additional CI events, which were not used to extract samples.

    In this study, three approaches including decision trees (DT), random forest (RF), and logistic regression (LR) were used for the development of CI detection models. DT and LR were used for deterministic and probabilistic CI detection, respectively. Meanwhile, RF was used for both deterministic and probabilistic detection of CI. DT has been widely used for classification

20   and regression tasks in the remote sensing field (Li et al., 2013; Lu et al., 2014; Kim et al., 2014; Torbick and Corbiere, 2015). See5, developed by RuleQuest Research, Inc. (Quinlan, 2015), was employed to perform the DT-based classification of clouds into CI and non-CI. See5 works by repeatedly splitting samples into two groups of greater homogeneity using an entropy-based parameter to generate a tree (Quinlan, 2015; Jensen and Im, 2007). Pruning to avoid overfitting is often applied when a decision tree is generated. One of the merits of See5.0 compared to other DT algorithms is that the generated tree can be

25   reproduced with multiple if-then rules, which makes it easier to interpret the results than the original tree (Im et al., 2008; Jensen et al., 2008; Rhee et al., 2008; Im et al., 2012; Kim et al., 2015).

    RF uses a bootstrapping strategy from the original training data to produce a series of Classification and Regression Trees (CART) that is a non-parametric decision tree, which produces either classification or regression trees depending on whether the dependent variable is categorical or numerical (i.e., continuous) (Breiman, 2001). The numerous independent trees (e.g.,

30   500, 1000) are grown based on two randomizations including (1) a randomly selected subset of the training samples for each tree and (2) a randomly selected subset of input variables at each node of the tree. This way, RF overcomes the well-known limitation of CART that results are sensitive to the configuration and quality of training data (Lawrence and Wright, 2001;

Rhee et al., 2014). Thus, RF has recently gained popularity in remote sensing classification and regression (Kim et al., 2014; Li et al., 2014; Liu et al., 2015; Lu et al., 2013; Park et al., 2016; Yoo et al., 2012). Two approaches are generally adopted to reach a final conclusion from the independent decision trees including a simple majority voting and weighted majority voting strategy for classification, while a final value is either simply averaged from the results of the multiple regression trees or

5   averaged with weights for regression. In the probabilistic RF, the probability of an event occurrence is calculated using the ratio of voting for CI and Non-CI cases with 500 trees.

LR is one of the statistical regression methods that are used for modelling a categorical dependent variable using independent variables (Hosmer and Lemeshow, 2000). In this study, binary LR, a type of LR technique that deals with only two values for a dependent variable, was used to estimate the probability of CI occurrence. The logistic function is given as follows:

10
$$E(Y) = \frac{1}{1 + \exp\left[-\left(\beta_0 + \sum_{j=1}^{k} \beta_j X_j\right)\right]}$$
(1)

where E is the expected value of the dependent variable Y, k is the number of independent variables and X_j is the value of the j th independent variable, β_0 is the intercept from the linear regression equation, and β_j is the weighting coefficient for the j th independent variable. The logistic function produces E values of [0, 1] which are used as the probability of whether CI will occur or not. DT, RF, and LR provide the relative importance of input variables when developing

15   models such as attribute usage, mean decrease accuracy, and the absolute value of weighting coefficients, respectively. See 5 produces the information on how frequently each variable is used in the results. RF calculates the decrease in accuracy of the model using out-of-bag data through the random permutation of a variable. Therefore, a higher mean decrease accuracy of a variable indicates more contribution of the variable to develop a model. In the LR, the exponentiation of weighting coefficients of the independent variables reflects the relative importance of the variables, which refers to the changes in the odds ratio

20   attributed to an input variable.

A series of typical accuracy metrics were calculated through confusion matrices to assess the performance of the CI detection models, including producer's and user's accuracies, overall accuracy, and kappa coefficients. In addition, the prediction results of the models were further assessed for the 3 cases of CI events over the Korean Peninsula by computing POD, FAR, overall accuracy (OA), and Critical Success Index (CSI) reflecting effects of both POD and FAR as follows (Mecikalski et al., 2015):

25   POD = A/(A+B)   (2)

FAR = C/(A+C)   (3)

OA = (A+D)/(A+B+C+D)   (4)

CSI = (A)/(A+B+C)   (5)

where A is the number of CI objects that correctly detected as CI (i.e., hits), B is the number of CI objects incorrectly classified

30   as non-CI (i.e., misses), C is the number of non-CI objects incorrectly identified as CI (i.e., false alarm), and D indicates correct

Atmospheric
Measurement
Techniques
Discussions

negatives. A through D were counted from the results of the CI detection models using the Himawari-8 AHI images obtained 0–10, 10–20, 20–30, 30–40, and 40–50 min before the first occurrence of ≥35 dBZ intensity from the weather radar data based on CI objects. In order to identify CI objects in the AHI images that were obtained 0-50 min before the CI event occurred, distances from cloud objects to the location of the CI occurrence (i.e., ≥35 dBZ precipitation intensity) were calculated using

5   the atmospheric motion vector (AMV) product hourly generated from COMS MI (as Himawari-8 AHI does not provide AMV yet), assuming that the velocity and direction of moving clouds are constant over 1 h. Mean velocities and directions of the AMV of each cloud object for each case day were used to identify an overall motion vector. The cloud objects with a given direction within a given distance from the location of the first occurrence of ≥35 dBZ precipitation intensity were considered as CI. Overall POD, FAR, OA, and CSI for each CI detection model were computed based on the A-D values of three case

10   days.

For the 3 case days, the lead time for CI detection was calculated using a weighted mean depending on A (i.e., Hits) detected from the AHI images obtained before precipitation intensity exceeds 35 dBZ as follows (Han et al., 2015):

$$\frac{\sum A_t \times n}{\sum n} \quad (t = 0, 10, 20, 30, 40, 50, 60 \text{ min}) \tag{6}$$

where A_t is the number of A counted from the AHI images obtained t minutes before the first occurrence of ≥35 dBZ intensity

15   and n is the number of A_t. The lead time of each CI detection model was determined using the A_t and n of all case days.

4 Performance and validation of CI detection models

4.1 Performance of CI detection models

The box plots of the 12 interest fields generated using CI and Non-CI reference data are depicted in Fig. 2. A line inside a coloured box indicates the median value of the data. The height of the coloured boxes represents the interquartile range of the

20   data, and the 1.5 times interquartile range is shown with the vertical centre lines. The dots above and below the vertical lines are outliers. The p-values in the upper left corner of each box plot were derived from t-test at the 95% confidence level. All the interest fields except for 6.2 - 7.3 μm time trend T_b show p-values below 0.05. In particular, the interest field of 11.2, 6.2 - 11.2, and 6.2 - 7.3 μm T_B showed noticeably low p-values (< 0.001), of which the interquartile ranges of CI and non-CI samples do not overlap each other.

25   The performances of the four CI detection models were assessed using confusion matrices produced from the test dataset (Tables 4-7). Since the probabilistic models produce the possibility of CI occurrence ranging from 0 to 100 %, the pixels with a possibility higher than 50 % were regarded as the predicted CI areas in order to produce the confusion matrices. Among the four CI detection models, the deterministic RF model showed the highest overall accuracy (94.27 %) and kappa coefficient (89.28 %). The better performance of deterministic RF than DT (overall accuracy of 92.67 % and kappa coefficient of 84.84 %)

30   is possibly attributed to randomization strategies of deterministic RF such as bootstrap aggregating and randomized node

optimization which can reduce variance and overfitting in building decision trees. The overall accuracy and kappa coefficient value of the probabilistic RF model (94.27 and 89.98 %, respectively) were similar to those of the deterministic RF model because the same randomization strategies were used in the two models. The LR model showed inferior performance in terms of overall accuracy and kappa coefficient (90.63 % and 80.7 2%, respectively) possibly due to the limited capability of the

5   model to handle the non-linear behavior of the data (Tu, 1996) (Table 6). However, these accuracies using the test dataset could not be generalized to predict real CI cases.

The relative importance of the interest fields that were used for CI detection identified by DT and the deterministic RF are shown in Figs. 3 and 4, respectively. The 11.2 μm T_B, representing cloud top temperature, was identified as the most contributing interest field for the discrimination of the two classes (i.e., CI and non-CI) in both DT and deterministic RF

10   models. This corresponds well with the finding of Mecikalski et al. (2015), which used 25 satellite-based and numerical weather prediction (NWP)-based interest fields to predict CI and identified cloud top temperature as the most contributing satellite-derived variable. The next contributing variables in both DT and deterministic RF models were 13.3-11.2 μm T_B and the trispectral difference time trend (Fig. 3 and 4). The interest field of 13.3-11.2 μm T_B is closely related with cloud top height, which increases as cloud objects evolve into convective ones. Meanwhile, the trispectral difference time trend,

15   representing the temporal variation of the cloud-top glaciation, increases during the growth of convective clouds. The interest field of 6.2-7.3 μm T_B time trend, the only variable showing a p-value greater than 0.05, was defined as the least contributing interest field in the DT-based CI detection model and the second least one in the deterministic RF model. The relative importance of the interest fields for the LR-based CI detection model can be evaluated using the values of the Exp(b) that are listed in Table 8. The variable of trispectral difference, (8.6-11.2 μm)-(11.2-12.4 μm), has the highest value of the Exp(b),

20   2.190, which means the probability that a pixel is an actual CI increases by 2.19 times per 1 K increase in the trispectral difference. The 6.2-11.2 μm time trend (Exp(b) = 0.348) was also identified as a contributing variable in the LR model, which decreases the probability that a pixel is an actual CI by 65.2 % per 1 K decrease in the 6.-11.2 μm time trend. These two important variables in the LR model, however, were not ranked high on the variable importance determined by the DT and RF models. The 6.2-11.2 μm time trend was even identified as the least contributing variable in the RF model.

25   As convective clouds grow vertically in the troposphere (Jorgensen et al., 1989; Trier et al., 2004; Rosenfeld et al., 2008; Sieglaff et al., 2011), time trend variables that represent the vertical growth of clouds might be useful to detect CI. However, the time trend variables among the interest fields here resulted in relatively lower contribution in both DT and RF models. A simple overlap method using two temporally consecutive images was used to determine the temporal change of clouds, instead of AMV that has been widely used to track clouds, because AMV is not yet available for Himawari-8. This may result in

30   somewhat inaccurate estimation of the vertical growth of clouds. Thus, incorporation of AMV from Himawari-8 in the models may improve the performance of CI detection.

Atmospheric
Measurement
Techniques
Discussions

4.2 Validation of three CI cases with ground radar and lightning data

The four CI detection models were applied to the three cases of CI events over the Korean Peninsula and validated using two types of reference datasets (i.e., weather radar and lightning observations). Tables 9 and 10 show the validation metrics of the models based on each reference dataset. Fig. 5-8 show CI areas for the case of CI events on 12 June 2015 predicted using the

5　data collected 0–50 min before CI occurrence by the DT, deterministic RF, probabilistic RF, and LR models, overlaid on a 1.5 km CAPPI image from the case day. For this case day, all models show higher POD, OA, and CSI, but far lower FAR in the validation using the weather radar observations than the lightning. The LR model shows the best performance compared to the other models based on both the radar and lightning observations, producing the highest POD, OA (the second highest in the lightning-based validation), and CSI, and the lowest FAR. Meanwhile, the DT model showed the highest FAR, which would

10　be caused by the well-known overfitting problem of the algorithm (Pal and Mather, 2003).

Fig. 9-12 show CI areas predicted by the four models for the case of CI events on 1 August 2015, overlaid on a 1.5 km CAPPI image. CI objects on this case day were concentrated on Seoul and Won-san Bay (point locations in the figures), which contributed to lower FAR than other case days. Based on the weather radar observations, the DT model showed the best performance in terms of the POD, OA, and CSI. However, the FAR of the DT model was the highest possibly due to the

15　overfitting problem. The LR model showed the lowest FAR, but it was not able to detect CI objects around Won-san Bay. Although the periphery of Won-san Bay showed radar echoes above 35 dBZ, CI objects in the region were not included in the validation dataset due to the limited effective radius of the radar. In the validation using lightning observations, all models produced POD of 100 %, except for the probabilistic RF model, all models produced high POD because locations of lightning greatly coincided with those of CI areas and thus large number of hits was produced. However, the FAR produced using the

20　lightning observations is higher than that using the weather radar. Based on the comparison between the lightning and radar observations, it appears that lightning did not occur for small CI objects.

Fig. 13-16 show CI areas for the case of CI events on 7 August 2015 predicted by the DT, deterministic RF, probabilistic RF, and LR models, respectively, overlaid on a 1.5 km CAPPI image on the day. All models showed better performance in terms of FAR, OA, and CSI based on the weather radar observations than the lightning observations. This is because the number of

25　CI objects detected by the radar is smaller than that by the lightning observations due to the limited effective radius, similar to the case on 12 June. The DT and deterministic RF models detected CI areas around the northwestern Korean peninsula. Such predicted CI areas might be correct despite the lack of lightning observations, but the weather radar data from China Meteorological Administration (CMA) is not available over the region and hence we were not able to confirm whether the CI objects were correctly identified.

30　Fig. 13-16 show CI areas for the case of CI events on 7 August 2015 predicted by the DT, deterministic RF, probabilistic RF, and LR models, respectively, overlaid on a 1.5 km CAPPI image on the day. All models showed better performance in terms of FAR, OA, and CSI based on the weather radar observations than the lightning observations. This is because the number of CI objects detected by the radar is smaller than that by the lightning observations due to the limited effective radius, similar to

Atmospheric
Measurement
Techniques
Discussions

the case on 12 June. The DT and deterministic RF models detected CI areas around the northwestern Korean peninsula. Such predicted CI areas might be correct despite the lack of lightning observations, but the weather radar data from China Meteorological Administration (CMA) is not available over the region and hence we were not able to confirm whether the CI objects were correctly identified.

5 The CI detection models developed in this study showed different performances by CI case and the reference dataset used for validation (c.f., Tables 9 and 10). Overall, the DT model produced higher POD than the other models regardless of the validation data. However, it tends to produce high FAR, yielding widely distributed CI in error. The LR model in general produced higher OA and CSI, and lower FAR. However, the POD of the LR model was low due to a large number of missed CI objects compared to the other models. The two validation datasets, i.e., the weather radar and lightning observations,

10 influenced the assessment of model performance. Since the weather radar sites are located inland, convective clouds over the ocean were out of the detection radius and less likely to be detected. Meanwhile, lightning observations can even detect CI objects over the distant sea, but it is hard to identify the exact location of lightning in clustered-clouds. These limitations in each verification dataset provide the uncertainty in estimating the actual forecast skill of the CI detection models.

Due to the similar number of hits from the four models, the lead time of all four models was around 32 to 35 min. This indicates

15 that CI over the Korean Peninsula can be forecasted using the Himawari-8 AHI images with a usable lead time of 30 to 40 min, which is reasonably comparable with the lead time for CI detection (~30–45 min) in the literature (Han et al., 2015; Mecikalski et al., 2015). AHI images 50 min before CI occurrence were used to detect CI in this study. If the AHI data collected a few hours before CI occurrence were used in the development of CI detection models, a longer lead time could possibly be achieved (refer to the supplementary material as an example).

20 A limitation of this research is in that visible channels, which are critical in identifying cloud tops (Melcikalski et al., 2010), were not used. This was because many of the CI cases over the Korean Peninsula used in this study occurred at night time. The visible channels of Himawari-8 AHI have spatial resolutions of 0.5–1 km which can be used to improve the performance of CI detection models. The use of training samples and validation cases only from summer was another limitation. As the development of convective systems have a clear seasonality (Melcikaski et al., 2010), training samples and validation cases

25 over different seasons should be incorporated to develop robust CI detection models.

5 Conclusions

CI detection models for Himawari-8 AHI data over the Korean Peninsula in East Asia were developed based on DT, RF, and LR. An accuracy assessment of the developed models was conducted using weather radar and lightning observations. The interest field of 11.2 μm T_B, representing cloud top temperature, was identified as the most contributing variable in the

30 deterministic models and the probabilistic RF model. In the LR model, the trispectral difference, i.e., (8.6-11.2 μm)-(11.2-12.4 μm), was identified as the most important one. The developed CI detection models showed varied performance in terms of POD, FAR, OA, and CSI by CI cases and validation datasets. Nevertheless, the DT model produced generally higher POD

than the other models, while the LR model showed higher OA and CSI, and lower FAR. The averaged lead time of the CI detection models was calculated between 32 and 35 min, which means that a 30-min forecast of CI over the Korean Peninsula during summer is possible when using Himawari-8 AHI data.

The overlying method could produce a potential error when calculating the time trend interest fields that represent the vertical

5   growth of convective clouds. AMV might mitigate the error and enhance the availability of time trend variables. Future research includes (1) improving CI detection algorithms using visible reflectance with 0.5 km resolution, (2) increasing training samples and validation cases to reflect diverse convective environments with different seasons, and (3) expanding the period of the Himawari-8 AHI data 2 hours before CI occurrence.

15   **Table 1**. Characteristics of the spectral channels of Himawari-8 AHI

No. Band	Central wavelength (μm)	Bandwidth (μm)	Spatial resolution (km)
1	0.455	0.05	1
2	0.510	0.02	1
3	0.645	0.03	0.5
4	0.86	0.02	1
5	1.61	0.02	2
6	2.26	0.02	2
7	3.85	0.22	2
8	6.25	0.37	2
9	6.95	0.12	2
10	7.35	0.17	2
11	8.60	0.32	2
12	9.63	0.18	2
13	10.45	0.30	2
14	11.20	0.20	2
15	12.35	0.30	2
16	13.30	0.20	2

Atmospheric
Measurement
Techniques

Discussions

Table 2. Convective initiation (CI) cases used to develop and validate the deterministic and probabilistic CI detection
models

ID	Date	Time (hh:mm, UTC)	Usage
1	13 June 2015	09:40	
2	13 June 2015	11:10	
3	16 June 2015	14:30	
4	17 June 2015	10:30	Training dataset
5	4 July 2015	06:30	
6	25 July 2015	00:10	
7	16 August 2015	10:30	
8	12 June 2015	14:30	
9	1 August 2015	19:10	Validation dataset
10	7 August 2015	08:00	

Table 3. Summary of the interest fields to develop convective initiation (CI) detection models used in this study (Walker and
5   Mecikalski, 2011)

ID	Interest field	Contribution
1	11.2 μm T_B	Cloud-top temperature assessment
2	6.2-11.2 μm	
3	6.2-7.3 μm	Cloud-top height relative to tropopause
4	13.3-11.2 μm	
5	12.3-11.2 μm	Cloud-top glaciation
6	8.6-11.2 μm	
7	11.2 μm time trend	Cloud-top cooling rate
8	6.2-11.2 μm time trend	
9	6.2-7.3 μm time trend	Temporal changes in cloud-top height
10	12.3-11.2 μm time trend	
11	(8.6-11.2 μm)-(11.2-12.3 μm)	Cloud-top glaciation
12	(8.6-11.2 μm)-(11.2-12.3 μm) time trend	Temporal changes in cloud-top glaciation

Atmospheric
Measurement
Techniques
Discussions

Table 4. Assessment of the DT model for CI detection using the test data.

Reference / Classification	CI	Non-CI	Sum	User's accuracy
CI	239	21	260	91.92 %
Non-CI	26	355	381	93.18 %
Sum	265	376	641	
Producer's accuracy	90.19 %	91.41 %		
Overall accuracy			92.67 %	
Kappa coefficient			84.84 %	

Table 5. Assessment of the RF model for CI detection using the test data.

Reference / Classification	CI	Non-CI	Sum	User's accuracy
CI	245	11	256	95.70 %
Non-CI	20	365	385	94.81 %
Sum	265	376	641	
Producer's accuracy	92.45 %	97.07 %		
Overall accuracy			94.27 %	
Kappa coefficient			89.98 %	

5

Table 6. Assessment of the probabilistic RF model for CI detection using the test dataset. CI probabilities above 50% considered as a CI.

Reference / Classification	CI	Non-CI	Sum	User's accuracy
CI	242	13	255	94.90 %
Non-CI	23	363	386	94.04 %
Sum	265	376	641	
Producer's accuracy	91.32 %	96.54 %		
Overall accuracy			94.38 %	
Kappa coefficient			88.36 %	

Atmospheric
Measurement
Techniques
Discussions

Table 7. Assessment of the LR model for CI detection using the test dataset. CI probabilities above 50% considered as a CI.

Reference / Classification	CI	Non-CI	Sum	User's accuracy
CI	237	33	270	82.78 %
Non-CI	27	343	370	92.70 %
Sum	264	376	640	
Producer's accuracy	89.77 %	91.22 %		
Overall accuracy			90.63 %	
Kappa coefficient			80.72 %	

Table 8. Exp (b) values from the logistic regression (LR) model, which are odds ratios derived by SPSS at the significance level 95 %.

Interest field	Exp (b)
(8.6-11.2 µm)-(11.2-12.4 µm)	2.190
12.4-11.2 µm time trend	1.456
6.2-11.2 µm	1.146
13.3-11.2 µm	1.059
6.2-7.3 µm time trend	0.979
(8.6-11.2 µm)-(11.2-12.4 µm) time trend	0.979
6.2-7.3 µm	0.923
12.4-11.2 µm	0.802
11.2 µm	0.670
11.2 µm time trend	0.585
8.6-11.2 µm	0.483
6.2-11.2 µm time trend	0.348

5

Atmospheric
Measurement
Techniques
Open Access
Discussions
EGU

Table 9. Validation metrics and averaged lead times based on radar CAPPI data for the deterministic DT and RF, probabilistic RF, and LR models.

		12 June 2015	1 August 2015	7 August 2015	Overall
DT	POD (%)	89.10	87.37	85.1	87.19
	FAR (%)	57.74	18.2	34.42	36.78
	OA (%)	90.59	90.62	83.87	88.36
	CSI (%)	40.17	73.17	58.82	57.39
	Lead time (min)	35	35	34.58	34.86
Deterministic RF	POD (%)	87.12	67.16	85.92	80.07
	FAR (%)	41.72	1.1	27.95	23.59
	OA (%)	88.67	90.58	86.88	88.71
	CSI (%)	53.65	66.7	64.44	61.58
	Lead time (min)	34.8	35.26	34.31	34.73
Probabilistic RF	POD (%)	84.84	75.22	86.61	82.23
	FAR (%)	47.16	3.5	28.9	26.53
	OA (%)	84.65	82.63	86.2	84.49
	CSI (%)	48.27	73.21	64	61.85
	Lead time (min)	34.23	33.23	34.79	34.64
LR	POD (%)	90.19	59.37	69.56	73.04
	FAR (%)	32.84	1	26.15	15
	OA (%)	92.68	81.46	84.1	87.43
	CSI (%)	62.58	59	55.81	61.86
	Lead time (min)	33.6	34.14	32.5	33.81

Atmospheric
Measurement
Techniques
Discussions

Table 10. Validation metrics and averaged lead times based on lightning data for the deterministic DT and RF, probabilistic RF, and LR models.

		12 June 2015	1 August 2015	7 August 2015	Overall
DT	POD (%)	81.48	98.7	87.5	89.66
	FAR (%)	84.61	44.32	46.83	58.59
	OA (%)	44.52	84.9	77.48	69..30
	CSI (%)	14.86	55.67	49.41	39.98
	Lead time (min)	35	35	35	34.39
Deterministic RF	POD (%)	77.41	98.9	78.74	85.29
	FAR (%)	73.03	29.87	35.42	46.11
	OA (%)	48.62	91.12	81.32	74.02
	CSI (%)	25	70.12	54.85	49.94
	Lead time (min)	33.09	35	32.68	34.55
Probabilistic RF	POD (%)	80.35	96.29	79.22	81.29
	FAR (%)	74.28	38.09	39.90	50.76
	OA (%)	45.62	88.11	78.63	70.79
	CSI (%)	24.19	60.46	51.91	45.52
	Lead time (min)	33.18	33.4	34.26	33.61
LR	POD (%)	85.18	98.1	67.36	84.18
	FAR (%)	72.61	14.28	32.63	39.84
	OA (%)	46.64	96.66	81.31	74.87
	CSI (%)	26.13	85.2	50.78	54.21
	Lead time (min)	34.2	33.4	34.26	33.62
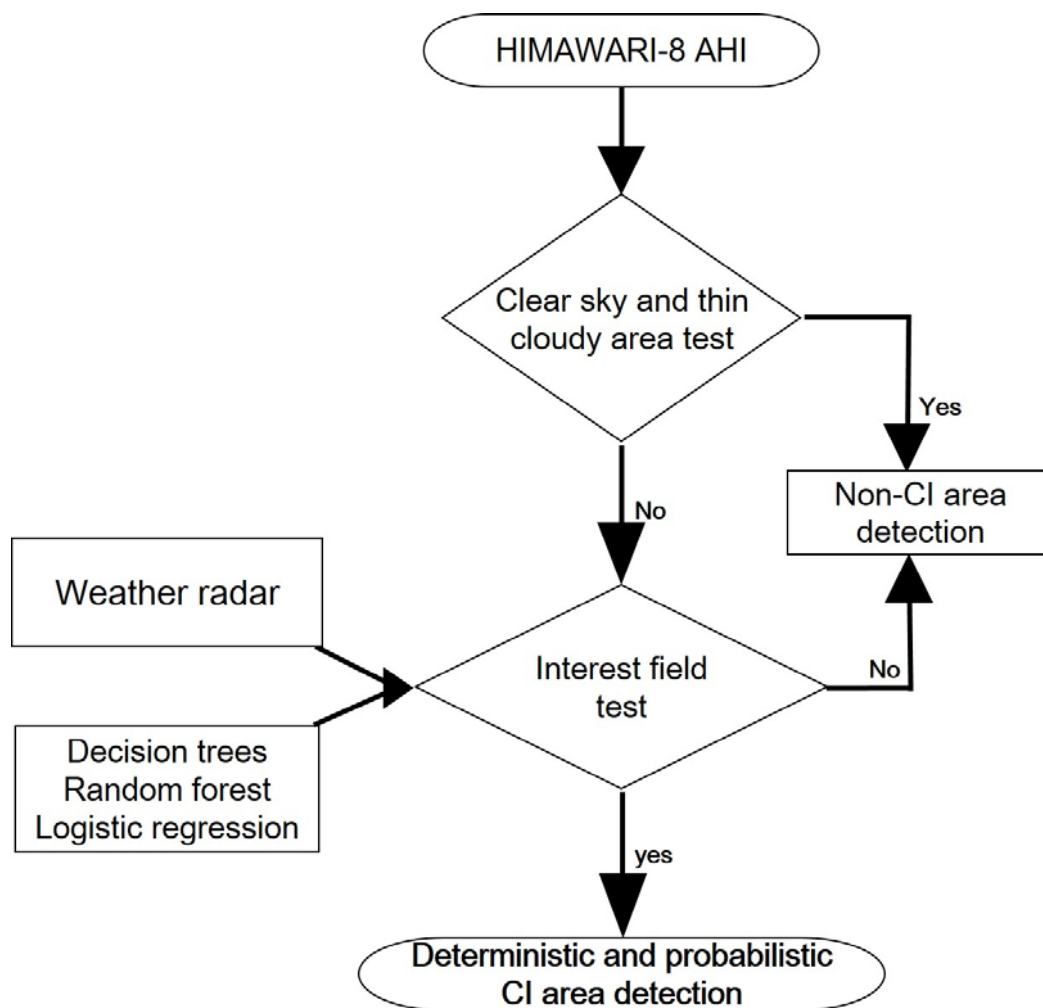
Figure 1. Processing flow of deterministic and probabilistic CI detection based on machine learning and statistical methods.

5

Atmospheric
Measurement
Techniques
Discussions
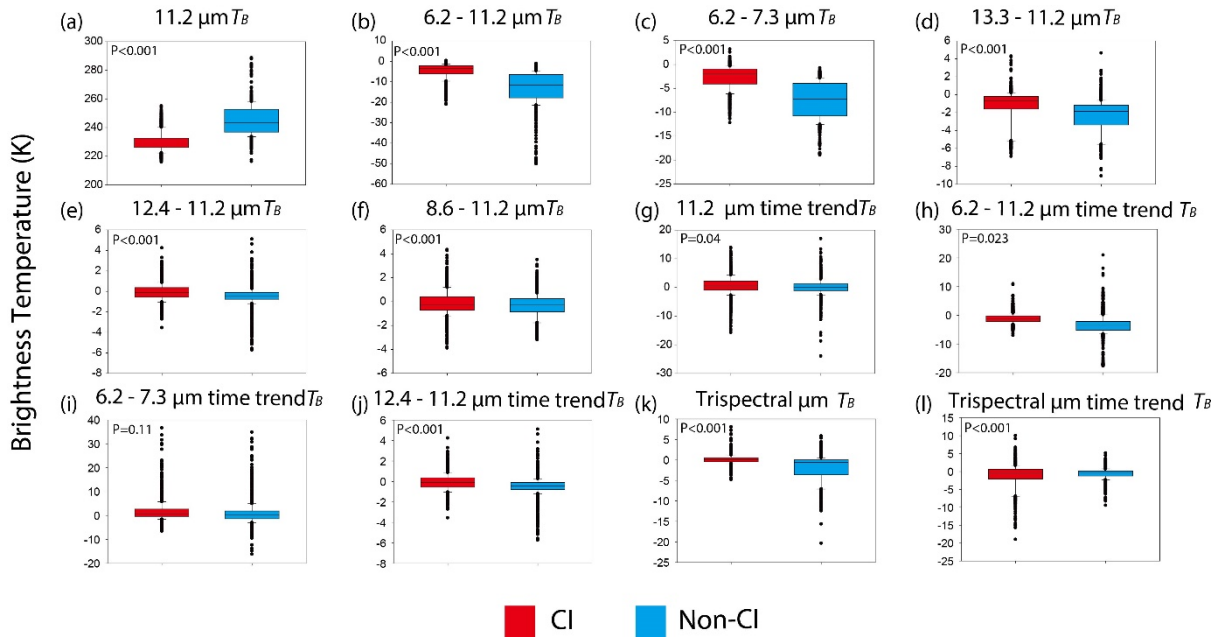
Figure 2. Box plots of 12 input variables (i.e., interest fields) generated based on the reference data used for CI detection models: (a) 11.2 μm T_b, (b) 6.2 – 11.2 μm T_b, (c) 6.2 – 7.3 μm T_b, (d) 13.3 – 11.2 μm T_b, (e) 12.4 – 11.2 μm T_b, (f) 8.6 – 11.2 μm T_b, (g) 11.2 μm time trend T_b, (h) 6.2 – 11.2 μm time trend T_b, (i) 6.2 – 7.3 μm time trend T_b, (j) 12.4 – 11.2 μm time trend T_b, (k) Tri-spectral μm T_b, (l) Tri-spectral μm time trend T_b.

Atmospheric
Measurement
Techniques
Discussions
Open Access
EGU
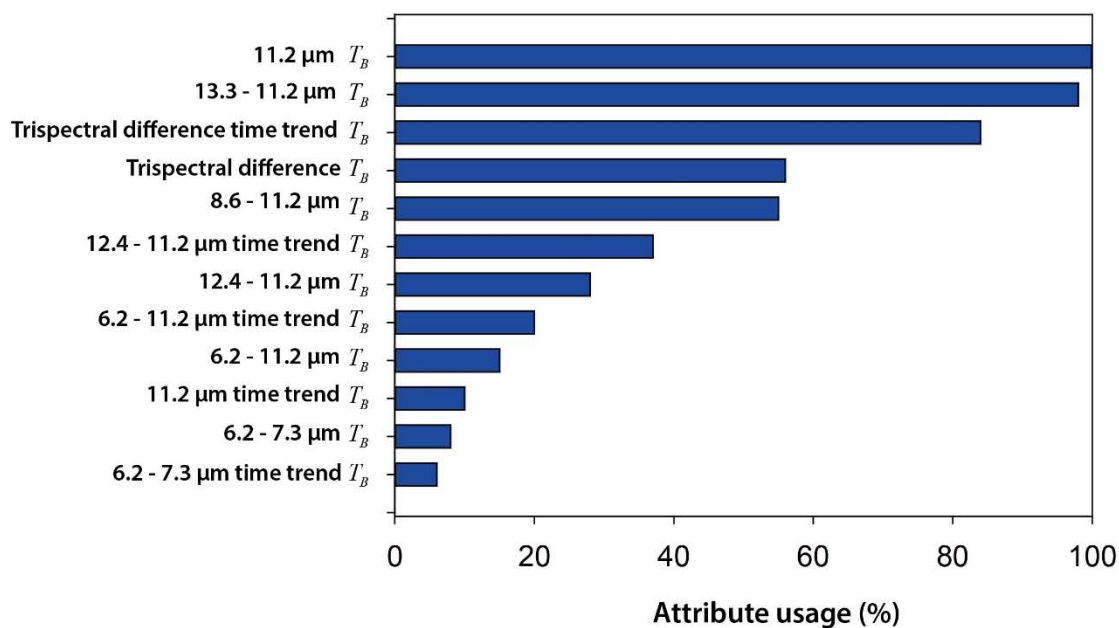
Figure 3. Attribute usage information in percentage by interest field produced in the DT model.

Atmospheric
Measurement
Techniques
Discussions
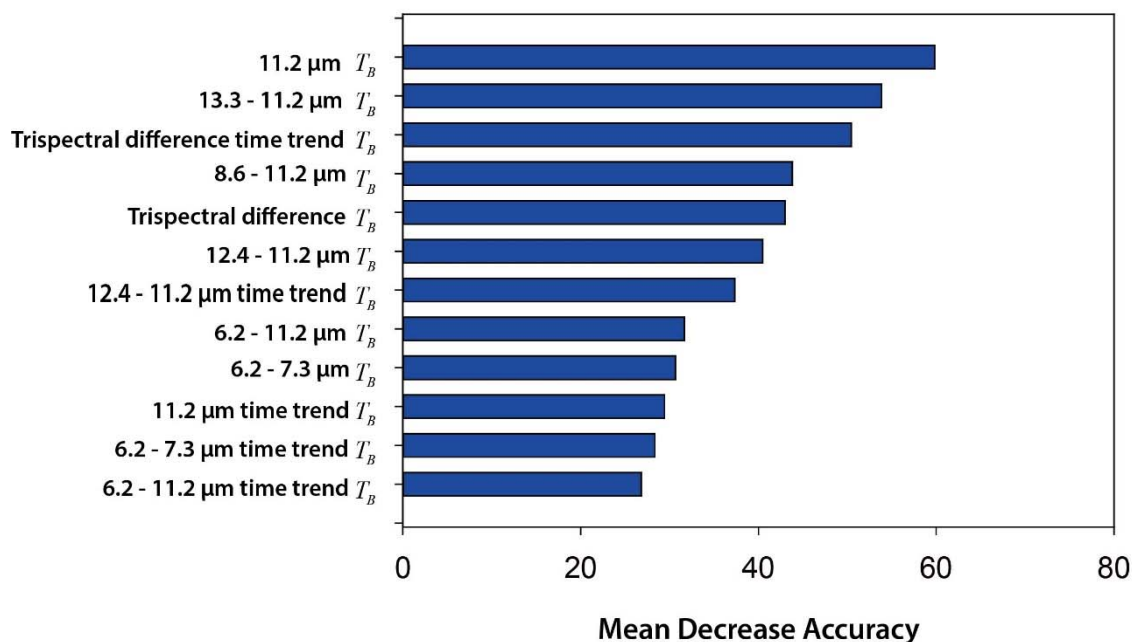
Figure 4. Mean Decrease Accuracy (MDA) in percentage by interest field produced in the RF model. MDA is calculated using out-of-bag (OOB) data when an interest field was randomly permuted. The higher the MDA of an interest field, the more the field contributes to identify CI.

5

10

Atmospheric
Measurement
Techniques
Discussions
Open Access
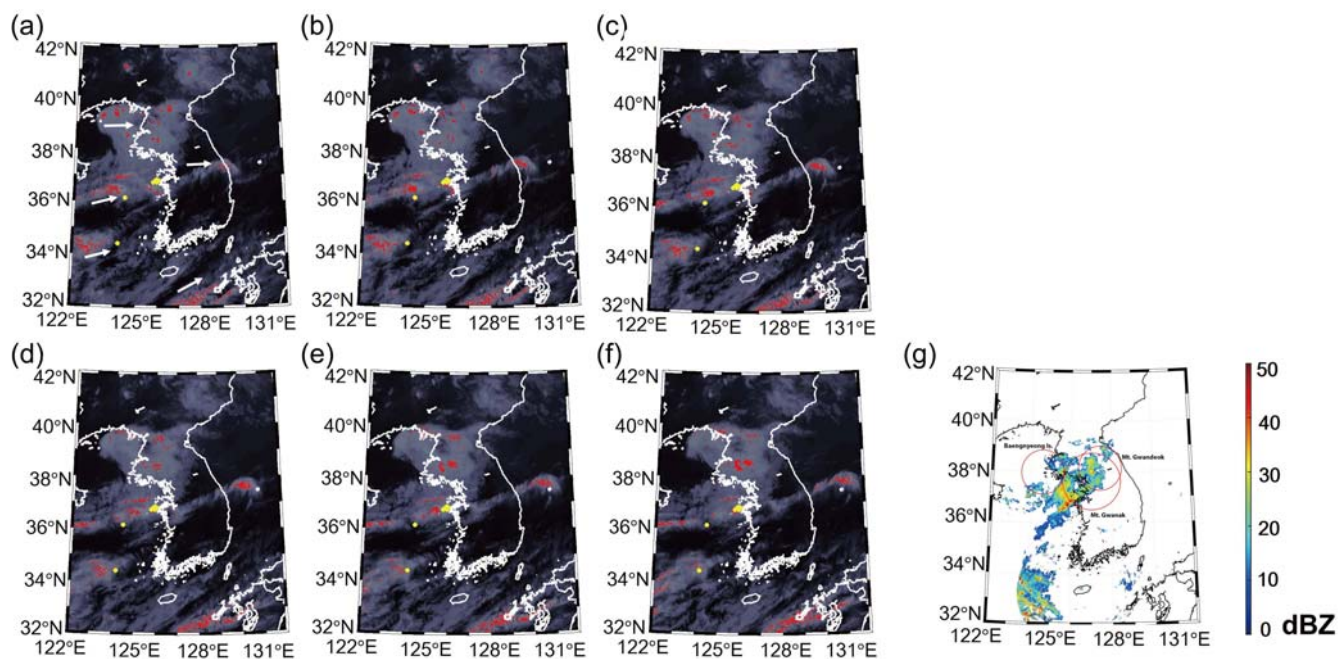
Figure 5. Deterministic CI detection map derived by DT on 12 June 2015 13:40 – 14:30 (UTC): (a) 50 min, (b) 40 min, (c) 30 min, (d), 20 min, (e) 10 min, and (f) 0 min before CI occurrence. (g) is 1.5 km radar CAPPI image at 14:30 (UTC). While predicted CI is in red, the locations of lightning occurrences at 14:30 (UTC) are presented in yellow dots. Red circles in (g) mean the effective radius of each radar site. The averaged direction of cloud objects is shown with white arrows in (a).
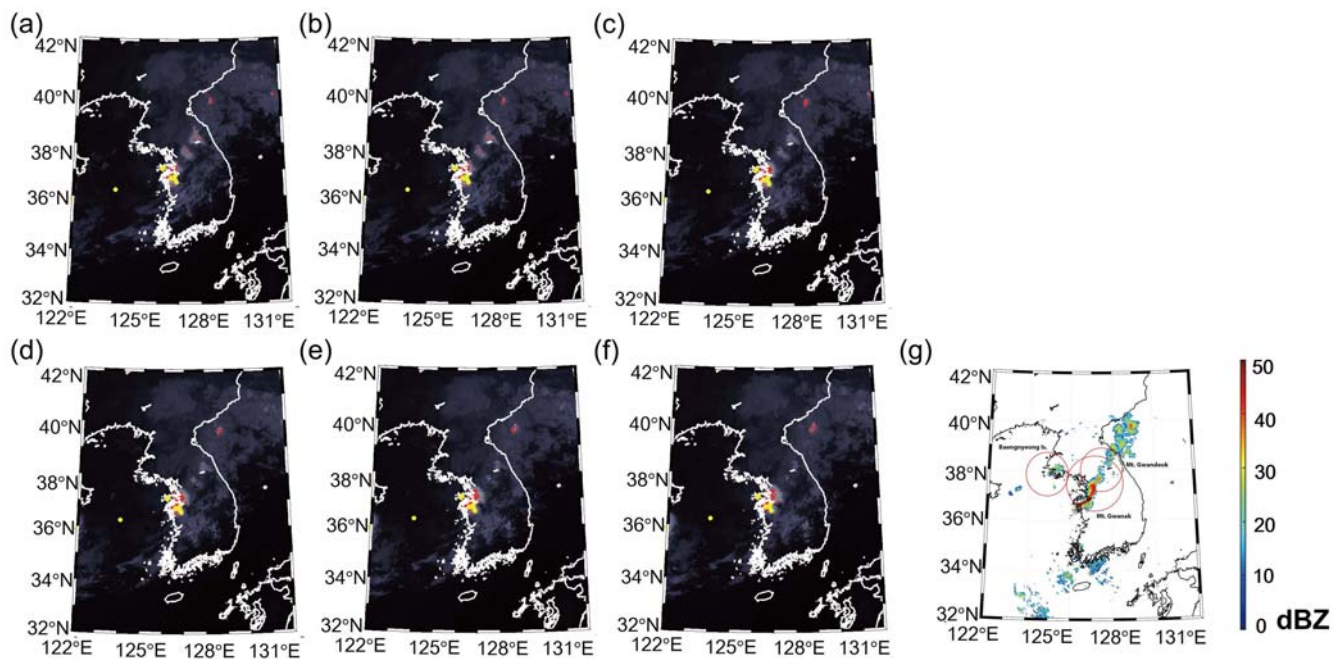
10
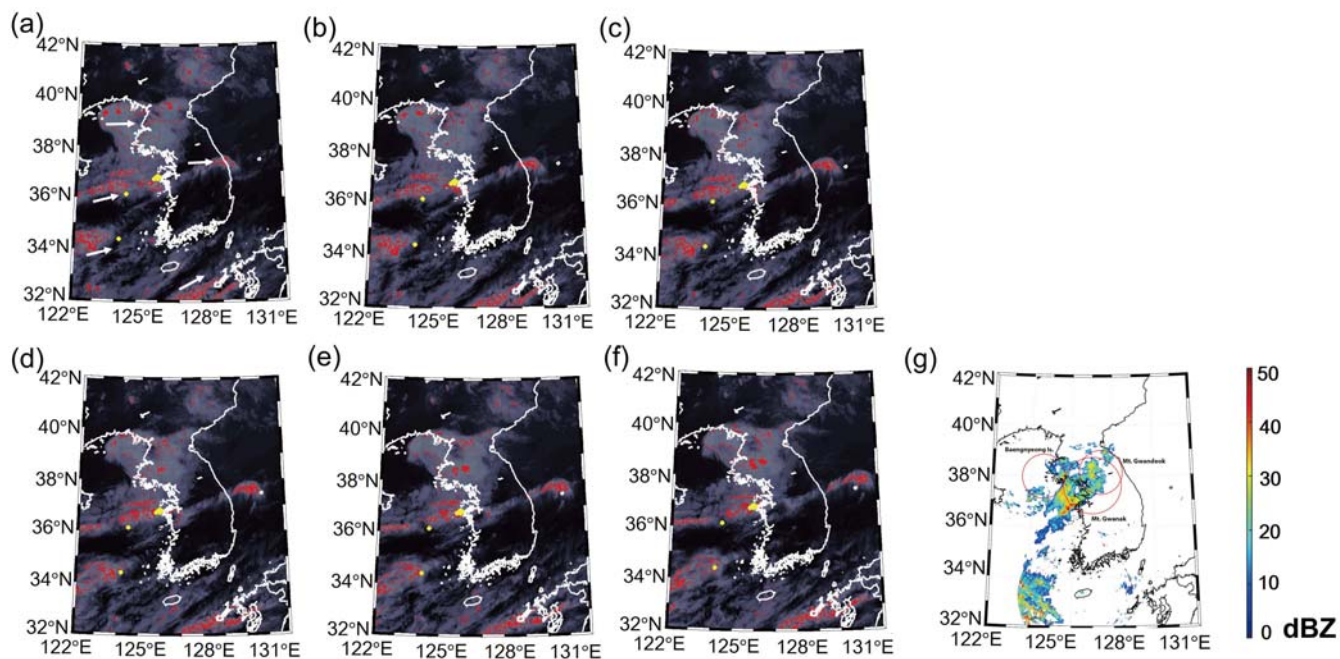
Figure 6. Deterministic CI detection map derived by deterministic RF on 12 June 2015 13:40 – 14:30 (UTC): (a) 50 min, (b) 40 min, (c) 30 min, (d), 20 min, (e) 10 min, and (f) 0 min before CI occurrence. (g) is 1.5 km radar CAPPI image at 14:30 (UTC). While predicted CI is in red, the locations of lightning occurrences at 14:30 (UTC) are presented in yellow dots. Red circles in (g) mean the effective radius of each radar site. The averaged direction of cloud objects is shown with white arrows in (a).
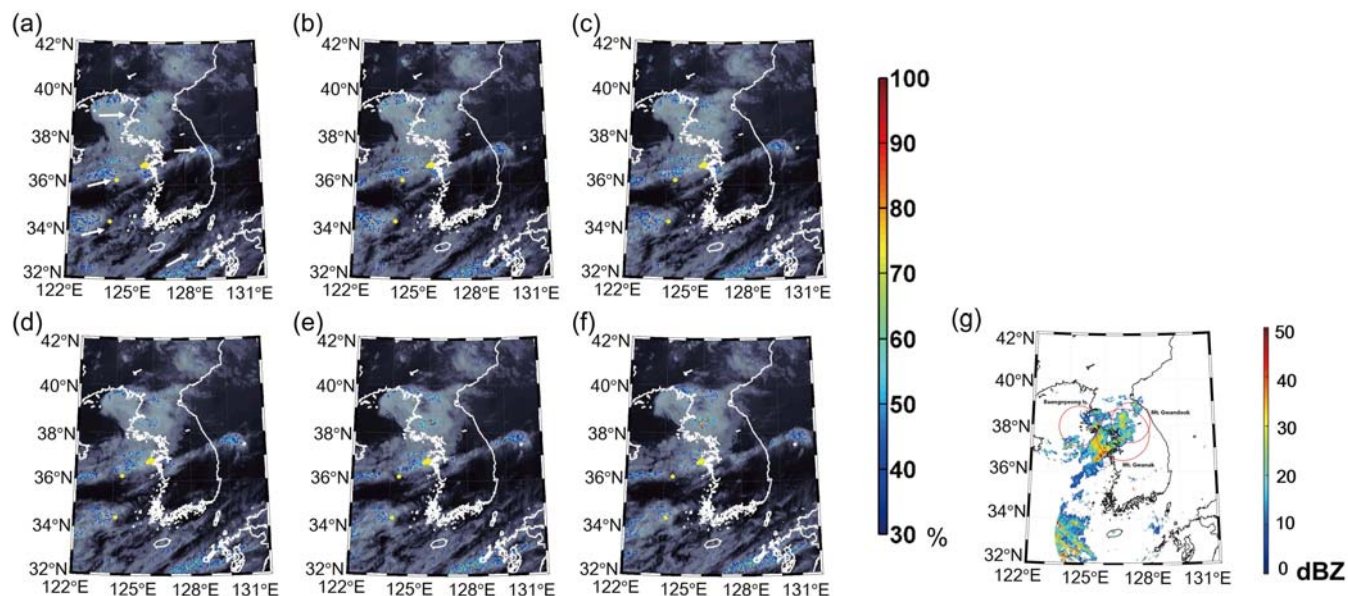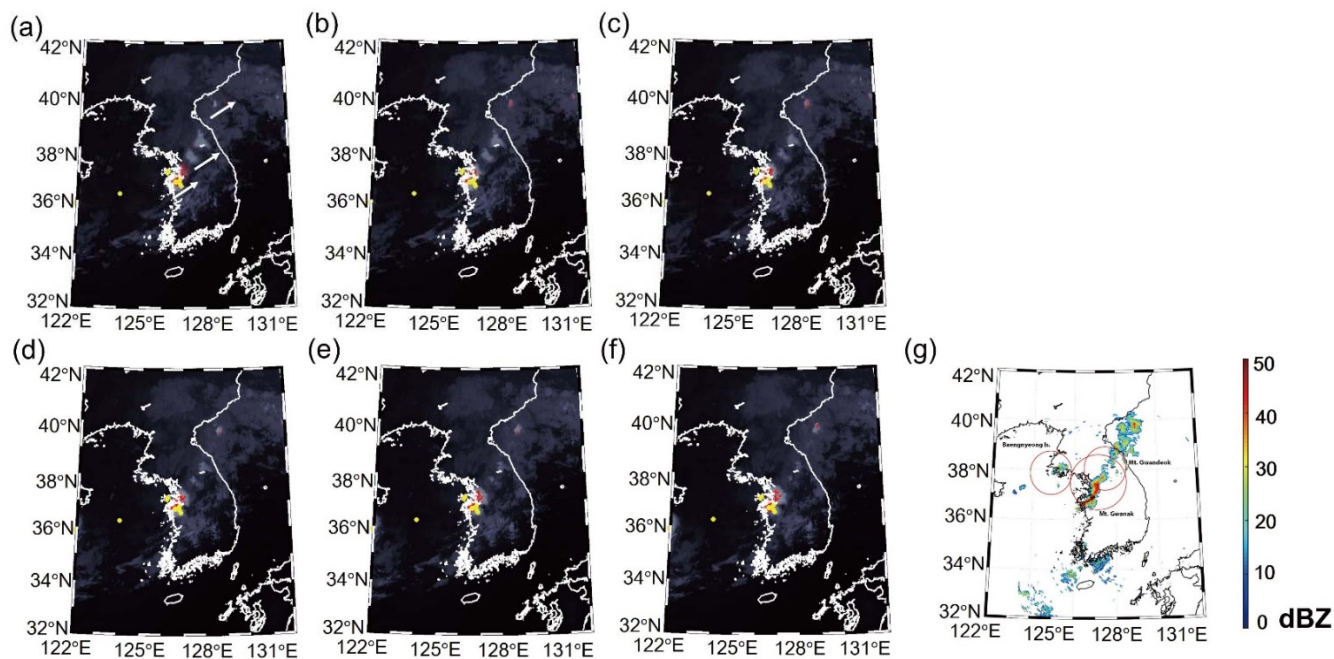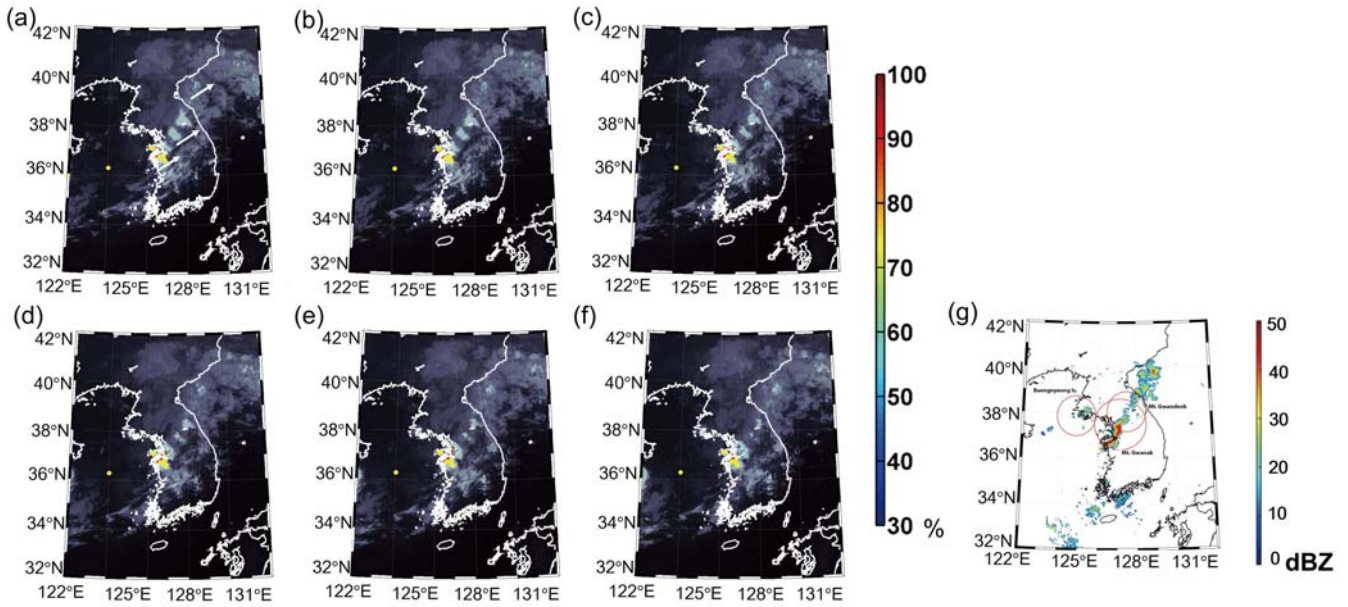
Figure 7. Probabilistic CI detection map derived by RF on 12 June 2015 13:40 – 14:30 (UTC): (a) 50 min, (b) 40 min, (c) 30 min, (d), 20 min, (e) 10 min, and (f) 0 min before CI occurrence. (g) is 1.5 km radar CAPPI image at 14:30 (UTC). While predicted CI is in red, the locations of lightning occurrences at 14:30 (UTC) are presented in yellow dots. Red circles in (g) mean the effective radius of each radar site. The averaged direction of cloud objects is shown with white arrows in (a).
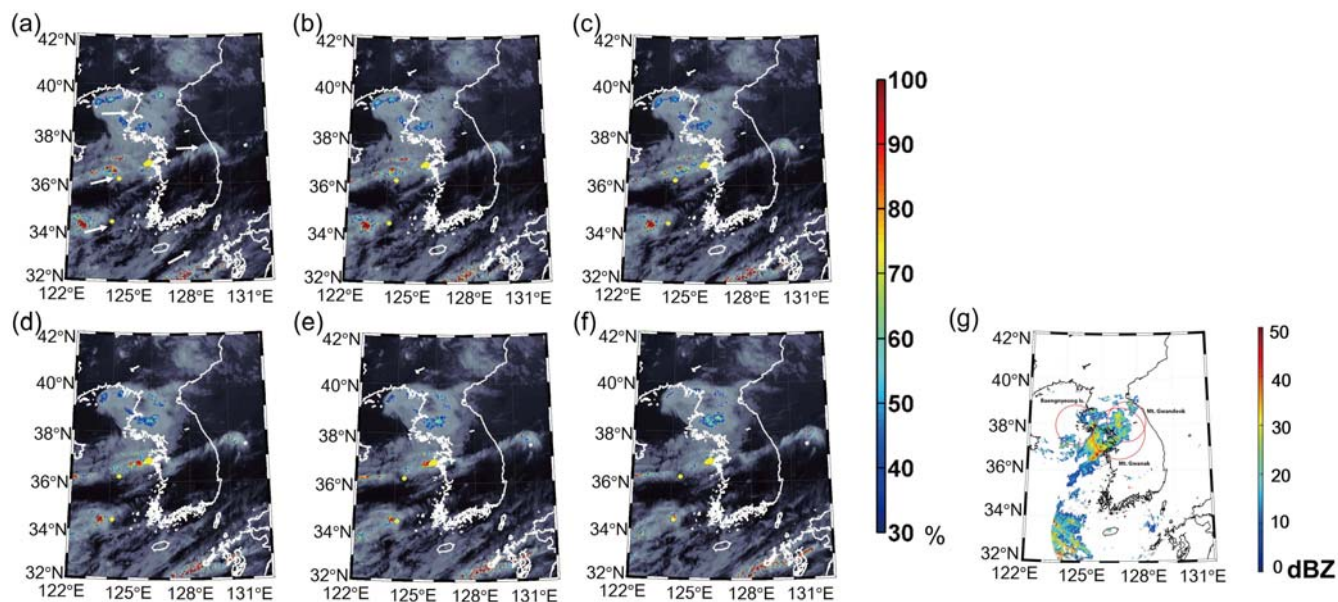
Figure 8. Probabilistic CI detection map derived by LR on 12 June 2015 13:40 – 14:30 (UTC): (a) 50 min, (b) 40 min, (c) 30 min, (d), 20 min, (e) 10 min, and (f) 0 min before CI occurrence. (g) is 1.5 km radar CAPPI image at 14:30 (UTC). While predicted CI is in red, the locations of lightning occurrences at 14:30 (UTC) are presented in yellow dots. Red circles in (g) mean the effective radius of each radar site. The averaged direction of cloud objects is shown with white arrows in (a).

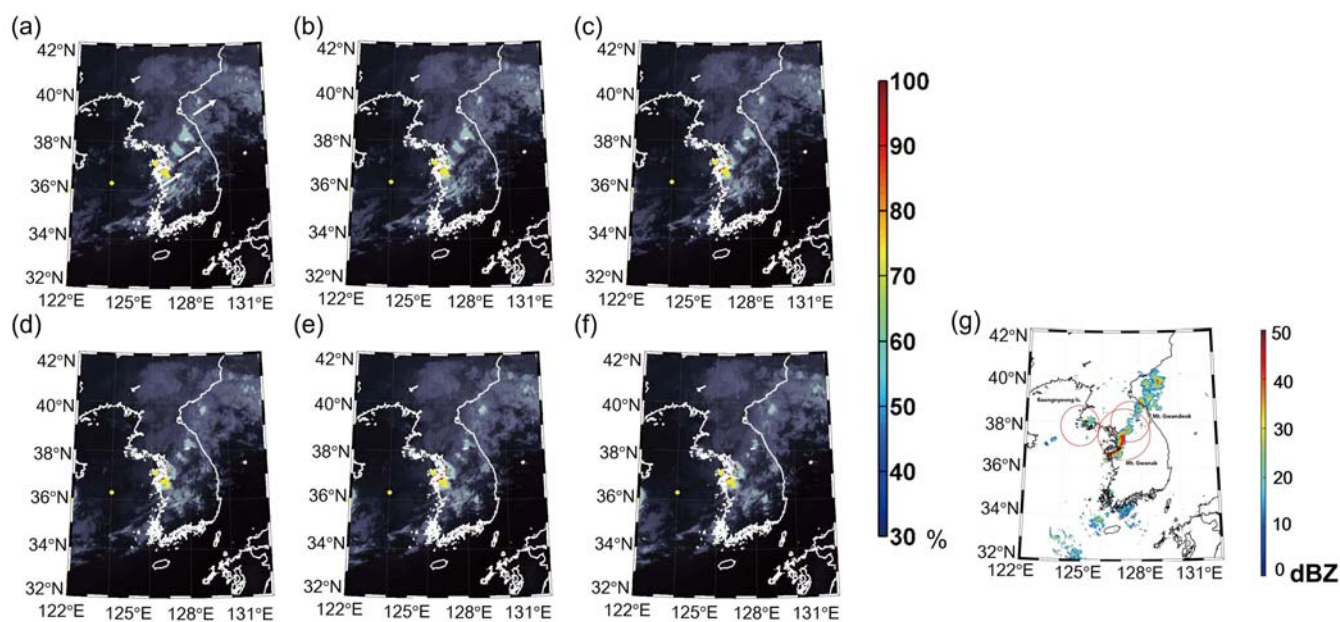Atmospheric
Measurement
Techniques
Discussions

Figure 9. Deterministic CI detection map derived by DT on 1 August 2015 1810 - 1910 (UTC): (a) 50 min, (b) 40 min, (c) 30 min, (d), 20 min, (e) 10 min, and (f) 0 min before CI occurrence. (g) is 1.5 km radar CAPPI image at 19:10 (UTC). While predicted CI is in red, the locations of lightning occurrences at 19:10 (UTC) are presented in yellow dots. Red circles in (g) mean the effective radius of each radar site. The averaged direction of cloud objects is shown with white arrows in (a).

Atmospheric
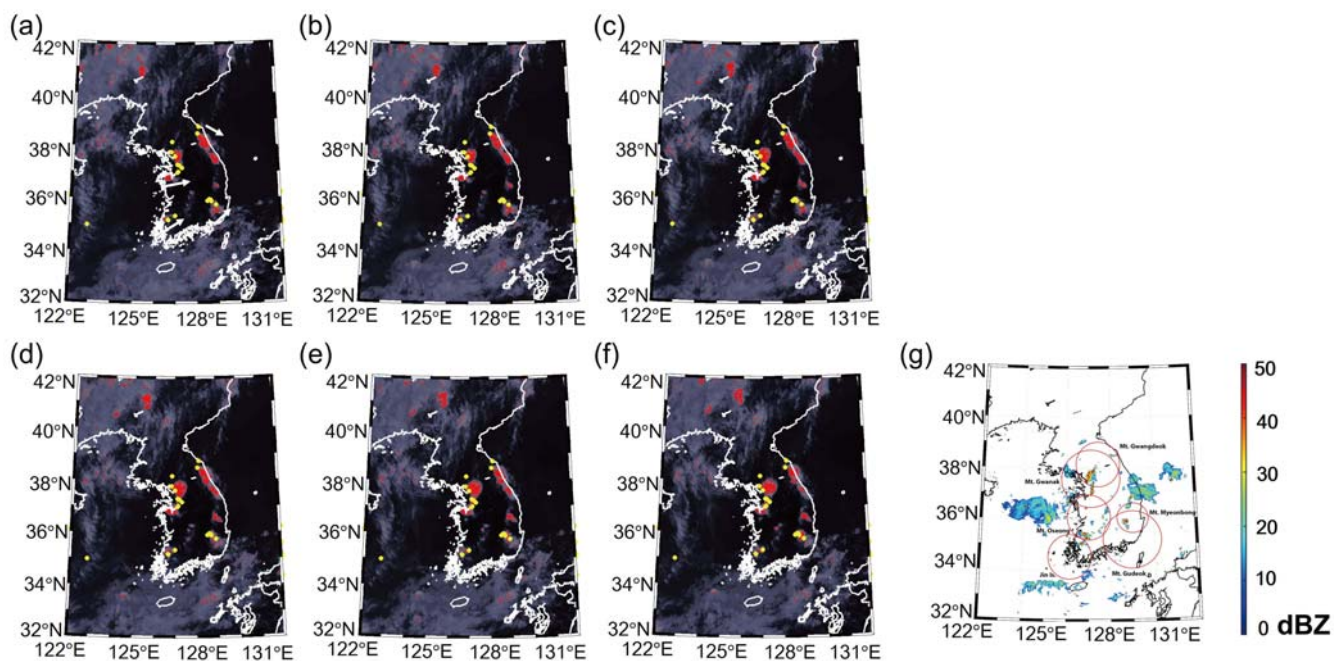Measurement
Techniques
Discussions

Figure 10. Deterministic CI detection map derived by RF on 1 August 2015 1810 - 1910 (UTC): (a) 50 min, (b) 40 min, (c) 30 min, (d), 20 min, (e) 10 min, (f) 0 min before CI occurrence. (g) is 1.5 km radar CAPPI image at 19:10 (UTC). While predicted CI is in red, the locations of lightning occurrences at 19:10 (UTC) are presented in yellow dots. Red circles in (g) mean the effective radius of each radar site. The averaged direction of cloud objects is shown with white arrows in (a).

Atmospheric
Measurement
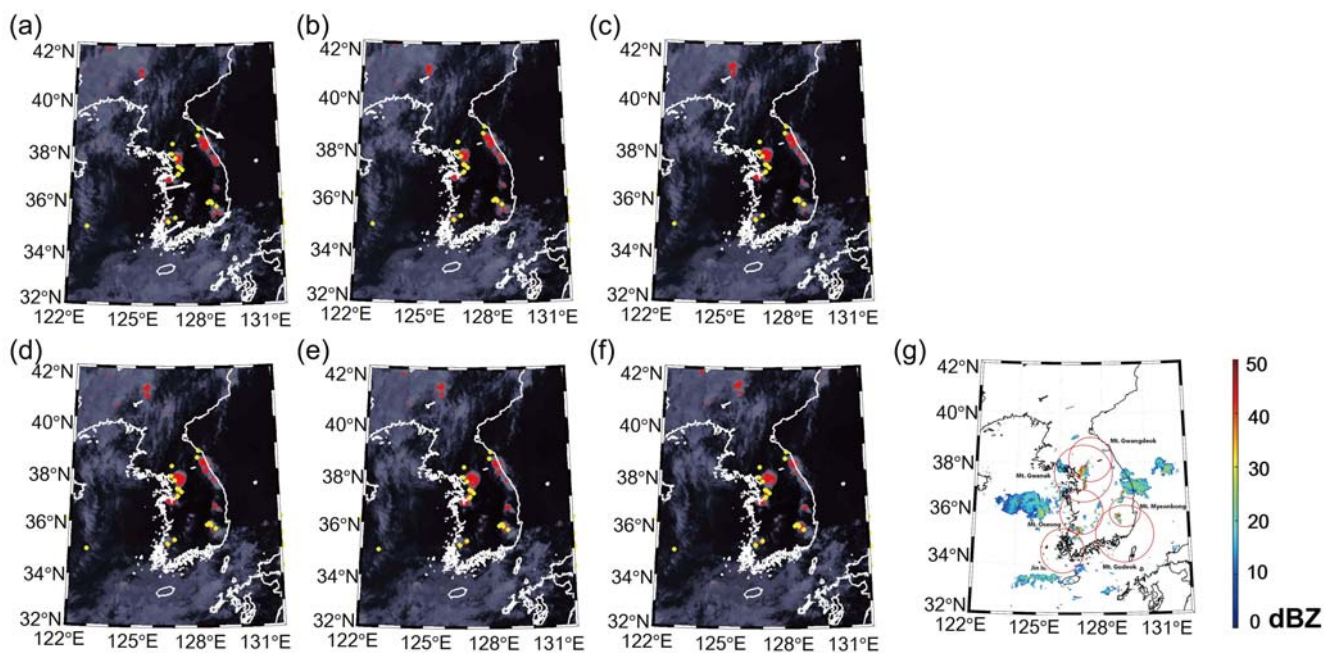Techniques
Discussions

Open Access

EGU

Figure 11. Probabilistic CI detection map derived by RF on 1 August 2015 1810 - 1910 (UTC): (a) 50 min, (b) 40 min, (c) 30 min, (d), 20 min, (e) 10 min, (f) 0 min before CI occurrence. (g) is 1.5 km radar CAPPI image at 19:10 (UTC). While predicted CI is in red, the locations of lightning occurrences at 19:10 (UTC) are presented in yellow dots. Red circles in (g) mean the effective radius of each radar site. The averaged direction of cloud objects is shown with white arrows in (a).
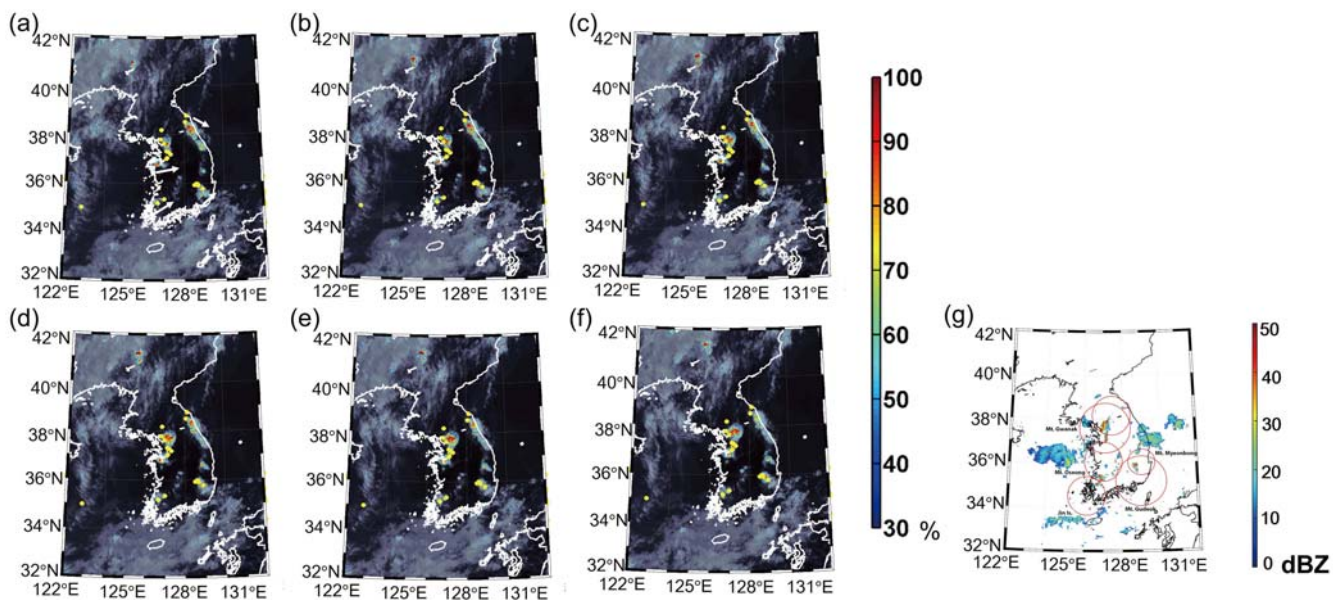
Figure 12. Probabilistic CI detection map derived by LR on 1 August 2015 1810 - 1910 (UTC): (a) 50 min, (b) 40 min, (c) 30 min, (d), 20 min, (e) 10 min, (f) 0 min before CI occurrence. (g) is 1.5 km radar CAPPI image at 19:10 (UTC). While predicted CI is in red, the locations of lightning occurrences at 19:10 (UTC) are presented in yellow dots. Red circles in (g) mean the effective radius of each radar site. The averaged direction of cloud objects is shown with white arrows in (a).

Atmospheric
Measurement
Techniques
Discussions

Figure 13. Deterministic CI detection map derived by DT on 7 August 2015 0710 - 0800 (UTC): (a) 50 min, (b) 40 min, (c) 30 min, (d), 20 min, (e) 10 min, (f) 0 min before CI occurrence. (g) is 1.5 km radar CAPPI image at 08:00 (UTC). While predicted CI is in red, the locations of lightning occurrences at 08:00 (UTC) are presented in yellow dots. Red circles in (g) mean the effective radius of each radar site. The averaged direction of cloud objects is shown with white arrows in (a).
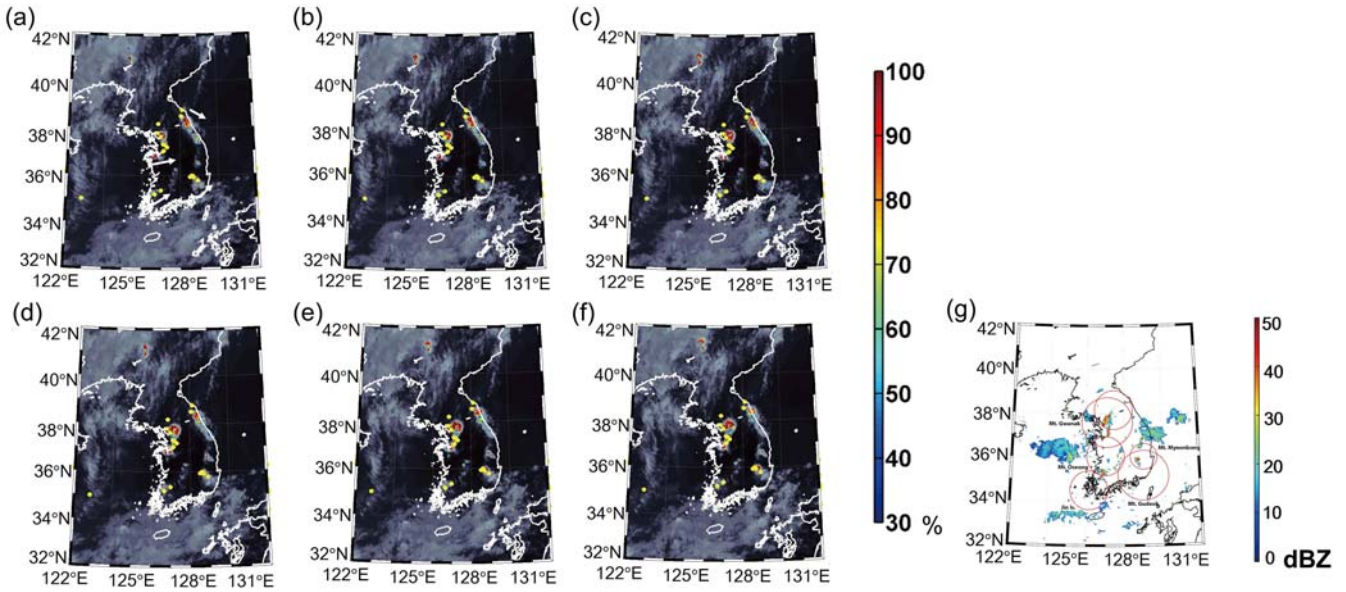
Figure 14. Deterministic CI detection map derived by RF on 7 August 2015 0710 - 0800 (UTC): (a) 50 min, (b) 40 min, (c) 30 min, (d), 20 min, (e) 10 min, (f) 0 min before CI occurrence. (g) is 1.5 km radar CAPPI image at 08:00 (UTC). While predicted CI is in red, the locations of lightning occurrences at 08:00 (UTC) are presented in yellow dots. Red circles in (g) mean the effective radius of each radar site. The averaged direction of cloud objects is shown with white arrows in (a).

10

Figure 15. Probabilistic CI detection map derived by RF on 7 August 2015 0710 - 0800 (UTC): (a) 50 min, (b) 40 min, (c) 30 min, (d), 20 min, (e) 10 min, and (f) 0 min before CI occurrence. (g) is 1.5 km radar CAPPI image at 08:00 (UTC). While predicted CI is in red, the locations of lightning occurrences at 08:00 (UTC) are presented in yellow dots. Red circles in (g) mean the effective radius of each radar site. The averaged direction of cloud objects is shown with white arrows in (a).

Atmospheric
Measurement
Techniques
Discussions

Figure 16. Probabilistic CI detection map derived by LR on 7 August 2015 0710 - 0800 (UTC): (a) 50 min, (b) 40 min, (c) 30 min, (d), 20 min, (e) 10 min, and (f) 0 min before CI occurrence. (g) is 1.5 km radar CAPPI image at 08:00 (UTC). While predicted CI is in red, the locations of lightning occurrences at 08:00 (UTC) are presented in yellow dots. Red circles in (g) mean the effective radius of each radar site. The averaged direction of cloud objects is shown with white arrows in (a).

Atmospheric
Measurement
Techniques
Discussions

References

Amorati, R., Alberoni, P. P., Levizzani, V., and Nanni, S.: IR-based satellite and radar rainfall estimates of convective storms over northern Italy, Meteorol. Appl., 7, 1-18, 10.1017/S1350482700001328, 2000.

Banacos, P. C., and Schultz, D. M.: The Use of Moisture Flux Convergence in Forecasting Convective Initiation: Historical and Operational Perspectives, Weather. Forecast., 20, 351-366, doi:10.1175/WAF858.1, 2005.

Bessho, K., Date, K., Hayashi, M., Ikeda, A., Imai, T., Inoue, H., Kumagai, Y., Miyakawa, T., Murata, H., Ohno, T., Okuyama, A., Oyama, R., Sasaki, Y., Shimazu, Y., Shimoji, K., Sumida, Y., Suzuki, M., Taniguchi, H., Tsuchiyama, H., Uesawa, D., Yokota, H., and Yoshida, R.: An Introduction to Himawari-8/9— Japan’s New-Generation Geostationary Meteorological Satellites, J. Meteorol. Soc. Jpn., 94, 151-183, 10.2151/jmsj.2016-009, 2016.

Breiman, L.: Random Forests, Machine Learning, 45, 5-32, 10.1023/a:1010933404324, 2001.

Craven, J. P., Jewell, R. E., and Brooks, H. E.: Comparison between Observed Convective Cloud-Base Heights and Lifting Condensation Level for Two Different Lifted Parcels, Weather. Forecast, 17, 885-890, doi:10.1175/1520-0434(2002)017<0885:CBOCCB>2.0.CO;2, 2002.

Haile, A. T., Rientjes, T., Gieske, A., and Gebremichael, M.: Multispectral remote sensing for rainfall detection and estimation at the source of the Blue Nile River, Int. J. Appl. Earth Obs. Geoinf., S76-S82, 2010.

Han, H., Lee, S., Im, J., Kim, M., Lee, M.I., Ahn, M.H., and Chung, S.R.: Detection of convective initiation using Meteorological Imager onboard Communication, Ocean, and Meteorological satellite based on machine learning approaches, Remote Sens., 7, 9184–9204, 2015.

Hane, C. E., Rabin, R. M., Crawford, T. M., Bluestein, H. B., and Baldwin, M. E.: A Case Study of Severe Storm Development along a Dryline within a Synoptically Active Environment. Part II: Multiple Boundaries and Convective Initiation, Mon. Weather Rev., 130, 900-920, doi:10.1175/1520-0493(2002)130<0900:ACSOSS>2.0.CO;2, 2002.

Hosmer, D.W., Lemeshow. S.: Applied Logistic Regression. John Wiley and Sons, Inc., New York, pp. 528, 2000.

Houze, R. A.: Mesoscale convective systems, Rev. Geophys., 42, RG4003, 10.1029/2004RG000150, 2004.

Im, J., Jensen, J. R., and Tullis, J. A.: Object-based change detection using correlation image analysis and image segmentation, Int. J. Remote Sens., 29, 399-423, 10.1080/01431160601075582, 2008.

Im, J., Jensen, J., Jensen, R., Gladden, J., Waugh, J., Serrato, M.: Vegetation cover analysis of hazardous waste sites in utah and arizona using hyperspectral remote sensing, Remote Sens., 4, 327–353, 2002.

Jensen, J. R., and Im, J.: Remote Sensing Change Detection in Urban Environments, in: Geo-Spatial Technologies in Urban Environments: Policy, Practice, and Pixels, edited by: Jensen, R. R., Gatrell, J. D., and McLean, D., Springer Berlin Heidelberg, Berlin, Heidelberg, 7-31, 2007.

Jewett, C. P., and Mecikalski, J. R.: Adjusting thresholds of satellite-based convective initiation interest fields based on the cloud environment, J. Geophys. Res-Atmos., 118, 12,649-612,660, 10.1002/2013JD019700, 2013.

Atmospheric
Measurement
Techniques
Discussions

Jorgensen, D. P., and LeMone, M. A.: Vertical Velocity Characteristics of Oceanic Convection, Journal of the Atmos. Sci., 46, 621-640, doi:10.1175/1520-0469(1989)046<0621:VVCOOC>2.0.CO;2, 1989.

Kar, S. K., and Ha, K.-J.: Characteristic Differences of Rainfall and Cloud-to-Ground Lightning Activity over South Korea during the Summer Monsoon Season, Mon. Weather Rev., 131, 2312-2323, doi:10.1175/1520-0493(2003)131<2312:CDORAC>2.0.CO;2, 2003.

Kim, D. H., and Ahn, M. H.: Introduction of the in-orbit test and its performance for the first meteorological imager of the Communication, Ocean, and Meteorological Satellite, Atmos. Meas. Tech., 7, 2471-2485, 10.5194/amt-7-2471-2014, 2014.

Kim, H. W., and Lee, D. K.: An Observational Study of Mesoscale Convective Systems with Heavy Rainfall over the Korean Peninsula, Weather. Forecast., 21, 125-148, doi:10.1175/WAF912.1, 2006.

Kim, M., Im, J., Han, H., Kim, J., Lee, S., Shin, M., and Kim, H.-C.: Landfast sea ice monitoring using multisensor fusion in the Antarctic, GIS. Remote Sens., 52, 239-256, 10.1080/15481603.2015.1026050, 2015.

Kim, Y. H., Im, J., Ha, H. K., Choi, J.-K., and Ha, S.: Machine learning approaches to coastal water quality monitoring using GOCI satellite data, GIS. Remote Sens., 51, 158-174, 10.1080/15481603.2014.900983, 2014.

Lawrence, R.L., Wright, A.: Rule-based classification systems using classification and regression tree (CART) analysis. Photogramm. Eng. Rem. S., 67, 1137–1142, 2001.

Li, M., Im, J., and Beier, C.: Machine learning approaches for forest classification and change analysis using multi-temporal Landsat TM images over Huntington Wildlife Forest, GIS. Remote Sens., 50, 361-384, 10.1080/15481603.2013.819161, 2013.

Li, M., Im, J., Quackenbush, L. J., and Liu, T.: Forest Biomass and Carbon Stock Quantification Using Airborne LiDAR Data: A Case Study Over Huntington Wildlife Forest in the Adirondack Park, IEEE J. Sel. Top. Appl. Earth Obser. Remote Sens., 7, 3143-3156, 10.1109/JSTARS.2014.2304642, 2014.

Liu, T., Im, J., and Quackenbush, L. J.: A novel transferable individual tree crown delineation model based on Fishing Net Dragging and boundary classification, ISPRS J. Photogramme., 110, 34-47, http://dx.doi.org/10.1016/j.isprsjprs.2015.10.002, 2015.

Lu, Z., Im, J., Quackenbush, L. J., and Yoo, S.: Remote Sensing-based House Value Estimation Using an Optimized Regional Regression Model, Photogramm. Eng. Remote Sens., 79, 809-820, 10.14358/PERS.79.9.809, 2013.

Lu, Z., Im, J., Rhee, J., and Hodgson, M.: Building type classification using spatial and landscape attributes derived from LiDAR remote sensing data, Landscape. Urban Plan., 130, 134-148, http://dx.doi.org/10.1016/j.landurbplan.2014.07.005, 2014.

Mecikalski, J. R., and Bedka, K. M.: Forecasting Convective Initiation by Monitoring the Evolution of Moving Cumulus in Daytime GOES Imagery, Mon. Weather Rev., 134, 49-78, 10.1175/MWR3062.1, 2006.

Mecikalski, J. R., Bedka, K. M., Paech, S. J., and Litten, L. A.: A Statistical Evaluation of GOES Cloud-Top Properties for Nowcasting Convective Initiation, Mon. Weather Rev., 136, 4899-4914, 10.1175/2008MWR2352.1, 2008.

Mecikalski, J. R., MacKenzie, W. M., Koenig, M., and Muller, S.: Cloud-Top Properties of Growing Cumulus prior to Convective Initiation as Measured by Meteosat Second Generation. Part I: Infrared Fields, J. Appl. Meteorol. Climatol., 49, 521-534, 10.1175/2009JAMC2344.1, 2009.

Mecikalski, J. R., MacKenzie, W. M., König, M., and Muller, S.: Cloud-Top Properties of Growing Cumulus prior to Convective Initiation as Measured by Meteosat Second Generation. Part II: Use of Visible Reflectance, J. Appl. Meteorol. Climatol., 49, 2544-2558, 10.1175/2010JAMC2480.1, 2010.

Mecikalski, J. R., Williams, J. K., Jewett, C. P., Ahijevych, D., LeRoy, A., and Walker, J. R.: Probabilistic 0–1-h Convective Initiation Nowcasts that Combine Geostationary Satellite Observations and Numerical Weather Prediction Model Data, J. Appl. Meteorol. Climatol., 54, 1039-1059, 10.1175/JAMC-D-14-0129.1, 2015.

Merk, D., and Zinner, T.: Detection of convective initiation using Meteosat SEVIRI: implementation in and verification with the tracking and nowcasting algorithm Cb-TRAM, Atmos. Meas. Tech., 6, 1903-1918, 10.5194/amt-6-1903-2013, 2013.

Miyamoto, Y., Kajikawa, Y., Yoshida, R., Yamaura, T., Yashiro, H., and Tomita, H.: Deep moist atmospheric convection in a subkilometer global simulation, Geophys. Res. Lett., 40, 4922-4926, 10.1002/grl.50944, 2013.

Morel, C., and Senesi, S.: A climatology of mesoscale convective systems over Europe using satellite infrared imagery. I: Methodology, Q. J. Roy. Meteor. Soc., 128, 1953-1971, 10.1256/003590002320603485, 2002.

Mueller, C., Saxen, T., Roberts, R., Wilson, J., Betancourt, T., Dettling, S., Oien, N., and Yee, J.: NCAR Auto-Nowcast System, Weather. Forecast., 18, 545-561, 10.1175/1520-0434(2003)018<0545:NAS>2.0.CO;2, 2003.

Pal, M., and Mather, P. M.: An assessment of the effectiveness of decision tree methods for land cover classification, Remote Sens. Environ., 86, 554-565, http://dx.doi.org/10.1016/S0034-4257(03)00132-9, 2003.

Park, S., Im, J., Jang, E., and Rhee, J.: Drought assessment and monitoring through blending of multi-sensor indices using machine learning approaches for different climate regions, Agr. Forest Meteor., 216, 157-169, http://dx.doi.org/10.1016/j.agrformet.2015.10.011, 2016.

Quinlan, J.R.: Data mining tools See5 and C4.5, version 2.10. Available online at https://www.rulequest.com/see5-info.html (accessed on 10 February 2016), 2015

Rhee, J., Im, J., Carbone, G. J., and Jensen, J. R.: Delineation of climate regions using in-situ and remotely-sensed data for the Carolinas, Remote Sens. Environ., 112, 3099-3111, http://dx.doi.org/10.1016/j.rse.2008.03.001, 2008.

Rhee, J., Park, S., and Lu, Z.: Relationship between land cover patterns and surface temperature in urban areas, GIS. Remote Sens., 51, 521-536, 10.1080/15481603.2014.964455, 2014.

Roberts, N. M., and Lean, H. W.: Scale-Selective Verification of Rainfall Accumulations from High-Resolution Forecasts of Convective Events, Mon. Weather Rev., 136, 78-97, 10.1175/2007MWR2123.1, 2008.

Rosenfeld, D., Woodley, W. L., Lerner, A., Kelman, G., and Lindsey, D. T.: Satellite detection of severe convective storms by their retrieved vertical profiles of cloud particle effective radius and thermodynamic phase, J. Geophys. Res-Atmospheres., 113, D04208, 10.1029/2007JD008600, 2008.

Schmit, T.J., Gunshor, M.M., Menzel, W.P., Gurka, J.J.: Introducing the next-generation Advanced Baseline Imager on GOES-R, B. Am. Meteorol. Soc. 86, 1079–1096, 2005.

Sieglaff, J. M., Cronce, L. M., Feltz, W. F., Bedka, K. M., Pavolonis, M. J., and Heidinger, A. K.: Nowcasting Convective Storm Initiation Using Satellite-Based Box-Averaged Cloud-Top Cooling and Cloud-Type Trends, J. Appl. Meteorol. Climatol., 50, 110-126, 10.1175/2010JAMC2496.1, 2010.

Siewert, C. W., Koenig, M., and Mecikalski, J. R.: Application of Meteosat second generation data towards improving the nowcasting of convective initiation, Meteorol. Appl., 17, 442-451, 10.1002/met.176, 2010.

Sobajima, A.: Rapidly Development Cumulus Areas Derivation Algorithm. Japan Meteorological Agency Algorithm Theoretical Basis Document, Meteorological Satellite Center, Tokyo, Japan, 2012.

Torbick, N., and Corbiere, M.: Mapping urban sprawl and impervious surfaces in the northeast United States for the past four decades, GIS. Remote Sens., 52, 746-764, 10.1080/15481603.2015.1076561, 2015.

Trier, S. B., Chen, F., and Manning, K. W.: A Study of Convection Initiation in a Mesoscale Model Using High-Resolution Land Surface Initial Conditions, Mon. Weather Rev., 132, 2954-2976, 10.1175/MWR2839.1, 2004.

Tu, J. V.: Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes, J. Clin. Epidemiol., 49, 1225-1231, http://dx.doi.org/10.1016/S0895-4356(96)00002-9, 1996.

Walker, J. R., MacKenzie, W. M., Mecikalski, J. R., and Jewett, C. P.: An Enhanced Geostationary Satellite–Based Convective Initiation Algorithm for 0–2-h Nowcasting with Object Tracking, J. Appl. Meteorol. Climatol., 51, 1931-1949, 10.1175/JAMC-D-11-0246.1, 2012.

Walker, J.R., Mecikalski, J.R., 2011. Algorithm theoretical basis document (ATBD) for convective initiation. NOAA NESDIS Center for Satellite Applications and Research, Available online at http://www.nsstc.uah.edu/SATCAST/docs/GOESR%20AWG%20ATBD%20Aviation%20ConvectiveInitiation%20v2.0.pdf (accessed on 10 February 2016).

Wang, C.-C., Chen, G. T.-J., and Carbone, R. E.: A Climatology of Warm-Season Cloud Patterns over East Asia Based on GMS Infrared Brightness Temperature Observations, Mon. Weather Rev., 132, 1606-1629, 10.1175/1520-0493(2004)132<1606:ACOWCP>2.0.CO;2, 2004.

Weckwerth, T. M., and Parsons, D. B.: A Review of Convection Initiation and Motivation for IHOP_2002, Mon. Weather Rev., 134, 5-22, 10.1175/MWR3067.1, 2006.

Vondou, D. A., Nzeukou, A., Kamga, F. M.: Diurnal cycle of convective activity over the West of Central Africa based on meteosat images, Int. J. Appl. Earth Obs. Geoinf., S58-S62, 2010.

Yoo, S., Im, J., and Wagner, J. E.: Variable selection for hedonic model using machine learning approaches: A case study in Onondaga County, NY, Landscape. Urban Plan., 107, 293-306, http://dx.doi.org/10.1016/j.landurbplan.2012.06.009, 2012.

Zinner, T., Mannstein, H., and Tafferner, A.: Cb-TRAM: Tracking and monitoring severe convection from onset over rapid development to mature phase using multi-channel Meteosat-8 SEVIRI data, Meteorol. Atmos. Phys. 101, 191-210,

10.1007/s00703-008-0290-y, 2008.

Zuidema, P.: Convective Clouds over the Bay of Bengal, Mon. Weather Rev., 131, 780-798, 10.1175/1520-0493(2003)131<0780:CCOTBO>2.0.CO;2, 2003.