DEVELOPING THE CLOUD DETECTION ALGORITHM FOR COMS METEOROLOGICAL DATA PROCESSING SYSTEM

Chu Yong Chung, Hee Kyo Lee, Hyun Jung Ahn, Myoung Hwan Ahn, Sung Nam Oh

Remote Sensing Research Laboratory/Meteorological Research Institute, KMA, cychung@metri.re.kr

ABSTRACT ... Cloud detection algorithm is being developed as major one of the 16 baseline products of CMDPS (COMS Meteorological Data Processing System), which is under development for the real-time application of data will be observed from COMS Meteorological Imager. For cloud detection from satellite data, we studied two different algorithms. One is threshold technique based algorithm, which is traditionally used, and another is artificial neural network model. MPEF scene analysis algorithm is the basic idea of threshold cloud detection algorithm, and some modifications are conducted for COMS. For the neural network, we selected MLP with back-propagation algorithm. Prototype software of each algorithm was completed and evaluated by using the MTSAT-1R and GOES-9 data. Currently the software codes are standardized using Fortran90 language. For the preparation as an operational algorithm, we will setup the validation strategy and tune up the algorithm continuously. This paper shows the outline of the two cloud detection algorithm and preliminary test result of both algorithms.

KEY WORDS: COMS, CMDPS, cloud detection, threshold technique, artificial neural network

1. INTRODUCTION

Communication, Ocean and Meteorological Satellite (COMS), planned for launch in 2008 will be the first Korean multi-purpose geostationary satellite. The development of systems for the meteorological mission sponsored by the Korea Meteorological Administration (KMA) consists of payloads, ground system, and data processing system (METRI, 2004). The program called COMS Meteorological Data Processing System (CMDPS) has been initiated for the development of data processing system to support the COMS operational meteorological application. The major function of CMDPS is the derivation of the baseline meteorological parameters from the calibrated and geo-located level 1B data. The planned baseline products consist of 16 parameters such as the analysis of special weather phenomena such as the yellow sand event in addition to the standard derived products from the current geostationary data. Additional function of CMDPS includes the development of calibration monitoring, upgrade, and validation mechanism of the baseline products (METRI, 2005).

The current baseline products consist of 16 products, which can be categorized into scene analysis, surface information, cloud information, water vapour information, environmental information, and atmospheric motion vectors. The very beginning of the baseline products generation is the cloud detection. The results from cloud detection play key role in the determination of product type, whether it is a cloudy or clear sky product, and all the necessary basic information for the consequent products. Thus, accuracy of the cloud detection affects on the accuracy of all of the products and quality information of the cloud detection will be provided.

Cloud detection scheme is now under development on two separate algorithms, which are threshold technique based algorithm for the operational purpose and artificial neural network based algorithm for backup or experimental. The cloud detection algorithm of MSG (Meteosat Second Generation), which is EUMETSAT Meteorological Products Extraction Facility (MPEF) scene analysis algorithm (EUMETSAT, 2003), is the basic idea of COMS threshold based cloud detection. For artificial neural network cloud detection algorithm, we selected multi-layer perceptrons (MLP) with back-propagation algorithms. Detail design of each algorithm is given in section 2. Section 3 gives preliminary results and comparison between two algorithms. This paper is concluded in section 4.

2. CLOUD DETECTION

2.1 Background

Accurate objective and automatic identification of clouds in satellite imagery is important. The detection of clouds is not only an essential first step in a cloud-type classification or in the retrieval of other cloud parameters. It is also essential first to mask out any cloud-contaminated field of view (FOV) in order to retrieve both surface and atmospheric parameters (Dybbroe et al., 2005). Most operational schemes use threshold method for detecting cloud, which contains several separate and independent threshold tests, for example, the Moderate Resolution Imaging Spectroradiometer (MODIS) cloud mask (Ackerman et al. 1998), the updated APOLLO Kriebel et al. 2003), NWCSAF cloud detection (Dybbroe et al., 2005) and many others. Another approach of cloud detection is to use the expert system, such as artificial neural networks. This requires a large amount of training data which contains input variable and true cloud detection results. Yhann and Simpson(1995)'s approach, which is Bayesian methods like neural networks, is an example for this approach.

2.2 Threshold technique based algorithm

The COMS main cloud detection algorithm is being developed based on MPEF scene analysis algorithm, which is one of the most advanced cloud detection algorithm by using threshold technique for the data observed from the geostationary satellite. But major modification is required to adapt to COMS data, because there are many differences between MSG SEVIRI (Spinning Enhanced Visible and Infrared Imager) and COMS MI (Meteorological Imager) which are the number of channels, observing area, different auxiliary data, and so on. Advantages of this algorithm are as follows. Dynamic threshold values are used for each test which are based on the predicted clear sky brightness temperature from previously derived clear sky radiance or calculated using radiative transfer model (RTM). Whole possible tests are applied to decide whether the corresponding pixel is cloud contaminated, including one or dual channel threshold tests and consistency tests. Quality flag of the test results will be provided for additional information.

Cloud detection algorithm is composed of 7 steps. On the first step, the clear sky brightness temperatures for 4 infrared channels are predicted by using clear sky radiance product of the previous repeat cycle and RTM simulated data. On step 2, solar zenith angle check is performed for each pixel to determine whether corresponding pixel is for day, dawn or night. This result will be used to define the channels and the threshold tests which are required to be used in the following steps. Channel data availability and quality are checked for each pixel using the information from the data preparation to determine which of the channel data can be used on step 3. On step 4, the data of the current repeat cycle is compared with those from the previous repeat cycle to determine if the data has changed to a certain extent on not. If no change is detected, it is assumed that the cloud detection scheme will provide the same result as the previous one and skip all of following steps for that pixel. For COMS algorithm, this step is considered to be removed after evaluating the time performance. Next step is threshold determination. The thresholds are determined to be used for each of the tests by using the clear sky brightness temperature predicted in the first step, clear sky reflectance, bidirectional reflectance distribution function, and some parameters including margin, slope and offset, and so on. Series of threshold tests are performed on step 6. On the final step, cloud detection results and their quality flags are determined by using all the results of each test. These are summarized in figure 1.

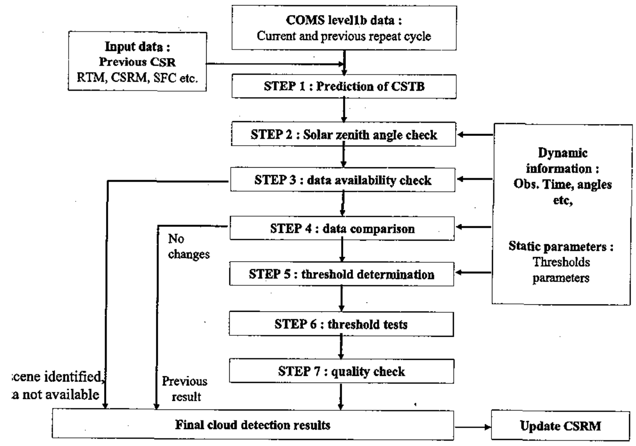

Figure 1. Flow chart for the cloud detection algorithm

2.3 Artificial neural network algorithm

For the backup and experimental purpose, artificial neural networks are studied for the COMS cloud detection. Detail description of this algorithm was presented by Ahn et al. (2005), so we delineates briefly here. The learning algorithm, the network architecture, the input variables and the training dataset play important roles for development of an accurate and reliable model. For the COMS neural network cloud detection algorithm, MLP methods are selected which is one of the most common neural networks model and has been widely used to solve cloud detection problems. The input layer contains nodes for each input variable, and the out layer represents the target variable. Determination of the proper number of input variables and hidden layers and the optimized weights are the key for the model development.

In this study, the number of input layer nodes is set to 12, which contains 5 channel observation data, position information (latitude, longitude), time (season, julian day), angles (satellite and solar zenith angle), and surface information (land or sea). The number of hidden layer is fixed by 1, and the number of hidden nodes is tested from 1 to 50 by a trial-and-error approach. Now we use 40 nodes in hidden layer. At last, the number of output layer is the only one, that is, cloud detection result, clear or cloudy.

The back-propagation algorithm which minimizes difference between the estimated outputs and true outputs is used for MLP model training. A schematic diagram for this model is shown in figure 2. The initial weights given randomly and the estimated outputs are obtained by running the model. The estimated output is compared with the true output from training datasets. If the difference is larger than preset threshold value, the weights are updated. These processes are repeated until the error is within the acceptable value.

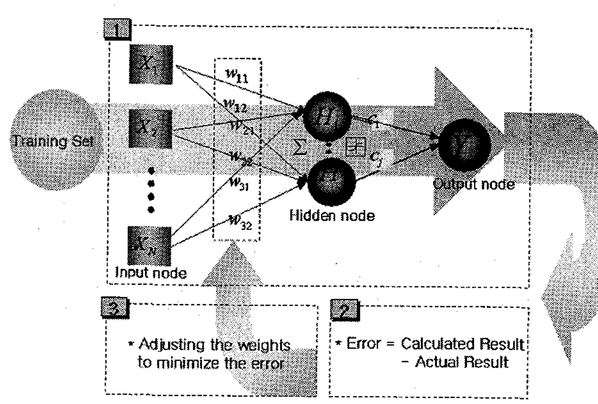

Figure 2. Schematic diagram for Multi-Layer Perceptrons model by back-propagation

3. ALGORITHM SOFTWARE DEVELOPMENT AND PRELIMINARY RESULT

Prototype cloud detection algorithm software based on threshold-technique was developed by using the MTSAT-1R observing data, which has similar imaging channels to COMS. The prototyping activities provide information about the scientific background for the cloud detection algorithm and verify that the algorithm is scientifically correct. For the prediction of clear sky brightness temperature in the first step, RTTOV (Radiative Transfer Models for TOVS) radative transfer code is used and GDAPS (Global Data Assimilation and Prediction System, KMA) T426 analysis fields are used as an input of RTM. And some parameters and thresholds were determined for the solar zenith angle check, channel availability check, and threshold tests. However, it should be noted that these values will be modified or updated with more studies on long term data processing and validation. To apply the COMS data before and after launch, also they should be changed.

An example of developed threshold-based cloud detection result is shown in figure 3. (a) and (b) are full-disk visible and infrared imageries for 0533UTC, October 31, 2005, observed from MTSAT-1R, respectively. Cloud detection result is shown in figure 3c, in which white corresponds to cloudy region and black corresponds to clear. Quality flag is also shown in figure 3d, of which value has 5 steps, 100% clear (white), 75% clear, 50% cloudy, 75% cloudy and 100% cloudy (black), generated from the combination of each individual test result. We can see that cloud detection results are reasonable in visual aspects for within plus minus 65-degree area of latitude and longitude from the satellite nadir point, which is described in user requirement document (KMA, 2004).

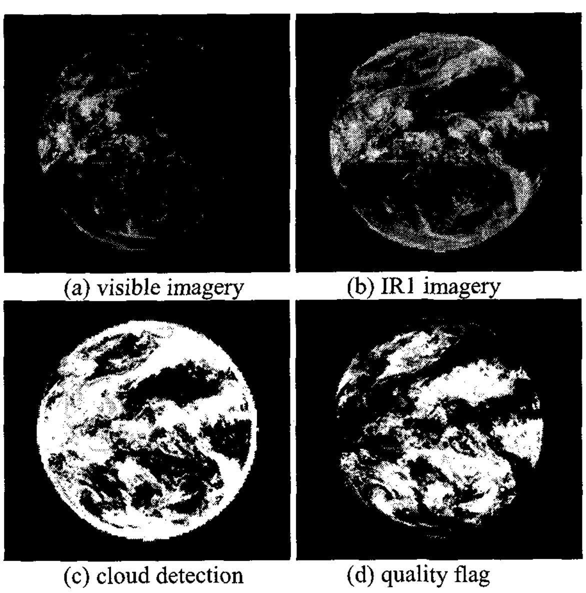

Figure 3. An image example of cloud detection results based on threshold-technique on 0533 UTC October 31. 2005, MTSAT-1R imageries for (a) visible, (b) infrared, (c) cloud detection and (d) quality flag of cloud detection.

In case of neural network algorithm, it was developed by using the GOES-9 data, which has also similar channels to COMS. For the training of neural network model, true cloud information on whether the corresponding pixels are cloud-contaminated or not should be necessary. In the preparation of training dataset, each GOES-9 scene is analysed and cloud detection is performed visually by varying thresholds using TeraScan software. Using these well cloud detected data, we selected more than 510,000 pixel training data which have uniform distribution. Training dataset was collected from the GOES-9 data during about 1 year, September, 2004 to July, 2005. Training dataset consists of 13 variables, 12 inputs and 1 output. With this training dataset, MLP with back-propagation neural network model were trained. In figure 4, we show a real image example of cloud detection result of neural network (c), together with visible (a), infrared (b) imageries. Threshold based cloud detection and the difference between neural network and threshold algorithm are also shown in (d) and (e), respectively. To evaluate the cloud detection tendencies, threshold based cloud detection algorithm, which was developed by using MTSAT-1R

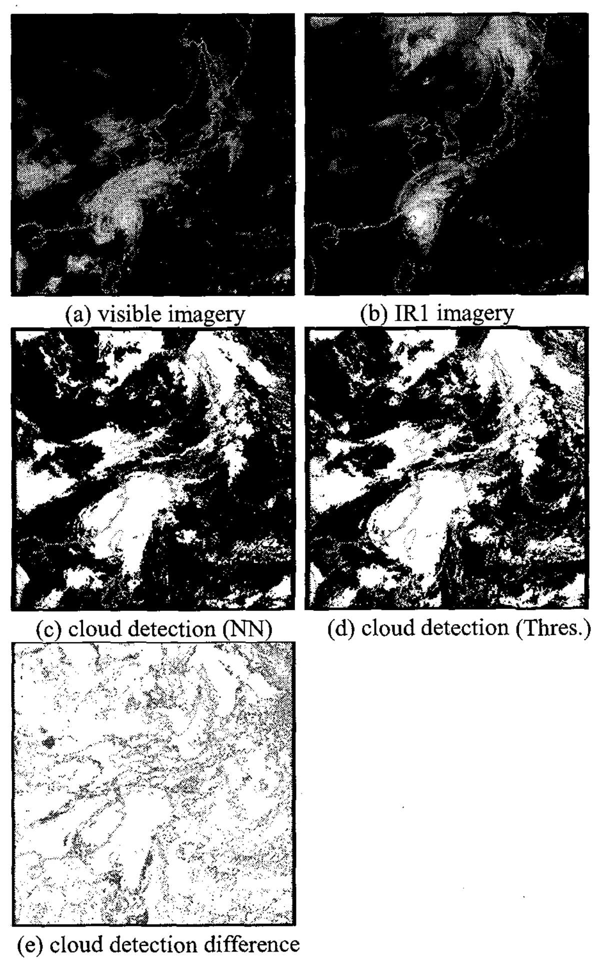

Figure 4. An example of cloud detection results on 0325 UTC, October 25, 2004, GOES-9. Imageries for (a) visible, (b) infrared, (c) threshold cloud detection result, (d) neural network cloud detection result, and (e) difference between (c) and (d).

data, was modified to apply the GOES-9. Major modification was held in the threshold determination parameter change. We can see that both cloud detection algorithm detect cloud area similarly, but threshold based algorithm over-detected cloud most part of cloud edge (red in figure 4e). From this, we can assume that threshold values were set strictly and also that uniformity test derived this result. The agreement between two algorithm results of this case is about 88 %.

However the cloud detection from satellite image data seems to be easy, we can aware that cloud detection is a difficult task without any standard criteria. More strictly do we set the threshold values, the more cloud could be detected, and what is worse, even the clear sky region can be considered as cloudy pixel. Otherwise, too loose thresholds produce the opposite result. So in development of the threshold based cloud detection algorithm, it is a key task how we set the threshold values properly.

For the neural network model, it depends mostly on the training data. So it is important that how the training data be prepared well, before the neural network studies are conducted.

Thus, validation and verification process is needed to complete the algorithm software. Visual inspection is an important first step in validating any cloud detection algorithm. The analyst uses knowledge of and experience with cloud and surface spectral properties to identify obvious problems. However, visual inspection provides poor quantitative evaluation. More quantitative validation can be attained through direct pixel-by-pixel comparison with collocated ground and instrument platform-based observations, such as lidar. While this approach provides quantitative accuracy, it possesses the problem that the two measurement systems often observe different cloud properties (Baum et al., 1995). For the future study, we will prepare the validation data and objective statistical validation method. Two different algorithms developed in this study will be validated and upgraded. Also for the preparation of real time operation, software code standardization and integration with whole CMDPS products algorithm software will be conducted.

4. CONCLUSIONS

For the cloud detection of COMS data processing, the algorithm is designed and being developed in two different ways. One is threshold technique based cloud detection algorithm, which is traditionally used for operational purpose, and another is neural network based algorithm, which is recently taken an interest. Two algorithms were studied theoretically and evaluated in the practical application. We estimated that both algorithms showed suitable cloud detection performance.

Prototype software was developed and currently the software codes are standardized using Fortran90 language. For the preparation as an operational algorithm, we will setup the validation strategy and update and upgrade the algorithm using standardized codes, such as tune up the threshold values, modification of automatic quality flag generation scheme for the threshold based one. Neural network study is continued to use the MTSAT-1R data, such as effective training dataset production, model optimization, and so on.

References from Journals:

Ackerman, S. A., K. I. Strabala, W. P. Menzel, R. A. Frey, C. C. Moeller, and L. E. Gumley, 1998, Discriminating clear sky from clouds with MODIS, *J. Geophys. Res.*, **103**, 32, 141-157.

Ahn, H. J., M. H. Ahn, and C. Y. Chung, 2005, Application of neural network for the cloud detection from geostationary satellite data, in proc., Int. Sym. Remote Sensing (ISRS2005), Jeju, Korea.

Baum, B. A., T. Uttal, M. Poellot, T. P. Ackerman, J. M. Alvarez, J. Intrieri, D. O'C. Starr, J. Titlow, V. Tovinkere, and E. Clothiaux, Satellite remote sensing of multiple cloud layers, J. Atmos. Sci., 52, 4210-4230, 1995.

Bybbroe, A., K. G. Karlsson, and A. Thoss, 2005, NWCSAF AVHRR cloud detection and analysis using dynamic thresholds and radiative transfer modelling. Part I: Algorithm description, *J. Appl. Meteo.*, **44**. 39-54

EUMETSAT, MSG Ground Segment: MPEF Algorithm Specification Document, pp203

Krievel, K. T., G. Gesell, M. Kastner, and H. Mannstein, 2003, The cloud analysis tool APOLLO: Improvements and validations. *Int. J. Remote Sens.*, 24, 2389-2408.

KMA, 2004, User requirement of Meteorological data for COMS, pp50.

METRI/KMA, 2004, Development of Meteorological Data Processing System of Communication, Ocean, and Meteorological Satellite(I), pp421

METRI/KMA, 2005, Development of Meteorological Data Processing System of Communication, Ocean, and Meteorological Satellite(II), pp846

Yhann, S, R., and J. J. Simpson, 1995: Application of neural networks to AVHRR cloud detection. *IEEE Trans. Geosci. Remote Sens.*, 33, 590-604.

Acknowledgements

This study is supported by the "Development of Meteorological Data Processing System of COMS" of KMA.