A New Scheme for MODIS Cloud Classification

Bormin Huang[1*], Hao Chen[1], Steven A. Ackerman[1], W. Paul Menzel[2]

[1]Cooperative Institute for Meteorological Satellite Studies
University of Wisconsin-Madison, Madison,
[2]Office of Research and Applications, NOAA/NESDIS, Madison, Wisconsin

Abstract: - A new technique called the local region of influence (LROI) scheme for supervised cloud classification of the Moderate Resolution Imaging Spectroradiometer (MODIS) is proposed. The classification of each observation is performed within the LROI, where the center of each class is calculated as a weighted average of its training class members with respect to each new observation. The probability of each class is assigned to each observation. The proposed LROI scheme is applied to the MODIS radiances observed from the scenes of clear skies, ice clouds, or water clouds. The classification results are compared with those from the maximum likelihood (ML) classification method, the multicategory support vector machine (MSVM) and the operational MODIS cloud mask algorithm. The lowest misclassification error rates show the advantage of the LROI scheme.

Key-Words: - Cloud classification, Local region of influence (LROI), maximum likelihood classification, multicategory support vector machine (MSVM), MODIS

1 Introduction

The MODIS (or Moderate Resolution Imaging Spectroradiometer) is a key instrument aboard the NASA EOS Terra and Aqua satellites. It provides 12-bit high radiometric sensitivity in 36 infrared and visible bands ranging in wavelength from 0.4 μm to 14.4 μm, with high spatial resolution from 250 m to 1 km. Terra MODIS and Aqua MODIS are viewing the entire Earth's surface every 1 to 2 days, acquiring data that will improve our understanding of global weather and climate.

The high spatial resolution of MODIS provides valuable cloud information. Clouds play a vital role in the earth's radiation budget and climate change, and can exist in the form of single or multilayer cloud system within a sensor's field of view [1]. The representation of cloud cover in numerical models has long been recognized as a potential uncertainty in climate prediction [2]. Thus, classification of cloudy and clear scenes is important. Several cloud classification schemes have been developed before. The MODIS cloud mask algorithm use simple thresholds on individual bands of radiances or brightness temperatures for clear-sky and cloudy discrimination, as described in [3], [4], and [5]. The well-known maximum likelihood (ML) scheme [6] [7] is an unsupervised, global classification approach. The multicategory support vector machine (MSVM) for classifying the MODIS observations has been recently proposed, which handles unequal misclassification costs and nonrepresentative samples in a principled way [8]. In this paper we propose the local region of influence (LROI) scheme for supervised cloud classification. The potential of the LROI scheme as an effective tool is illustrated by comparing the misclassification error rates with the the MODIS cloud mask algorithm, the ML classification, and MSVM.

Section 2 provides a description of the MODIS radiance data used in our classification experiments. Section 3 details the proposed classification scheme and the procedures to perform this scheme. Section 4 shows the cloud classification results and the misclassification error rates comparison with other classification schemes. Section 6 summarizes the paper.

2 Data Description

In this study, 1536 MODIS scenes over the Gulf of Mexico in July 2002 were classified as clear, ice, or water cloud by a satellite expert. There are 256 clear scenes, 952 ice clouds and 328 water clouds identified. Each of the three categories were randomly divided in half. The first halves are used as a training set which consists of 256 clear skies, 952 ice clouds and 328 water clouds, leaving the second halves as a test set. The training and test data sets are identical to the ones adopted by Lee *et al.* (2004) in the MSVM classification study [8].

It was observed that no single channel gives a clear separation of the three categories. Most of the misclassifications occurr in the overlap among categories. The composite variables involving two channels have been conventionally used to better distinguish among categories. The pairs of $R_{\text{channel 2}}$ vs. $\log_{10}(R_{\text{channel 5}} / R_{\text{channel 6}})$ have been used in the MSVM classification study [8], which appear to be informative as judged by the scatterplot in Figure 1. The same pairs are adopted in this study.

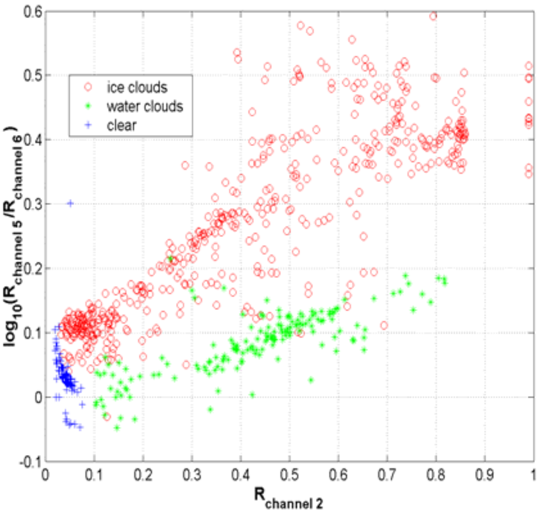

Fig. 1. Scattereplot of $\log_{10}(R_{\text{channel 5}} / R_{\text{channel 6}})$ vs. $R_{\text{channel 2}}$.

The nonconvex boundary of each category in the scatterplot implies that the unsupervised global maximum likelihood (ML) classification approach [6] [7] is not a good candidate for dealing with the MODIS cloud classification. Figure 2 illustrates the classification results of the ML approach.

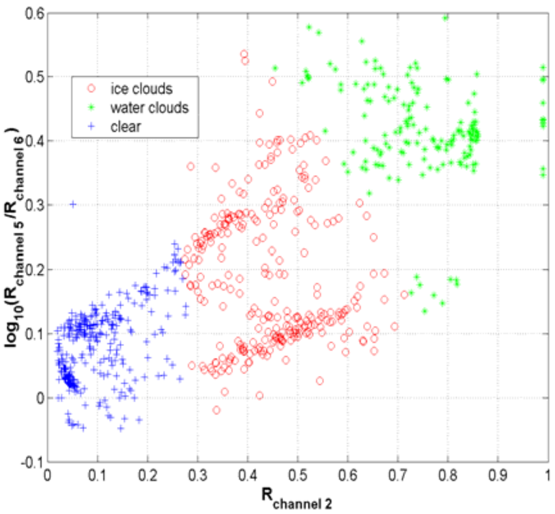

Fig. 2. Scattereplot of $\log_{10}(R_{\text{channel 5}} / R_{\text{channel 6}})$ vs. $R_{\text{channel 2}}$ after the ML classification.

Table 1 shows the distribution of the predicted category vs. the true category after the ML classification

Table 1 Distribution of predicted category vs. true category after the ML classification. The misclassification error rate is 481/768 = 62.63%.

True category	Predicted category			Total
	Clear sky	Ice clouds	Water clouds	
Clear sky	128	0	0	128
Ice clouds	173	150	153	476
Water clouds	32	123	9	164

As comparing Fig. 2 with Fig. 1, it is expected that the unsupervised ML classification scheme will yield high misclasslification error rate, which is 62.63% from Table 1.

3 New Local Classification Scheme

The local region of influence (LROI) scheme is proposed here for supervised cloud classification of MODIS observations. The classification of each observation is performed within the LROI, where the center of each class is calculated as a weighted average of its training class members with respect to each new observation. The probability of each class is assigned to each observation.

Specifically, the LROI scheme consists of the following three steps:

1). Given an observation in the test set and its LROI, find all the neighboring members in the training set whose distances are within the LROI. If there is no training data in the LROI, keep increasing the radius of LROI.

2) Compute the center of each class within LROI as the weight average of all training members of the same categories with respect to each new testing member:

$$\vec{r}_{class_k} = \frac{\sum\limits_{\forall j \in class_k} \vec{r}_{ij} * e^{-|\vec{r}_{ij}|^2 / \min\left(|\vec{r}_{ij}|^2\right)}}{\sum\limits_{\forall j \in class_k} e^{-|\vec{r}_{ij}|^2 / \min\left(|\vec{r}_{ij}|^2\right)}}$$

where $\vec{r}_{ij}$ is the distance vector from the testing sample to the training sample, and $\vec{r}_{class_k}$ is the distance vector from the testing data to the center of the class k. The vector $\vec{r}_{class_k}$ is computed as a weighted sum of the distance vectors $\vec{r}_{ij}$ from the testing sample to the training sample. An exponential weighting function is used for each vector $\vec{r}_{ij}$ such that the training sample with a smaller distance will have a larger weight. In this way, we can reduce the effects of training samples that are located farther away from the testing sample.

3) Calculate the probability of each cloud type and classify the data type. In this step, the probabilities of the testing samples for each cloud type are calculated by

$$P_{class_k} = \frac{e^{-|\vec{r}_{class_k}|^2 / \min\left(|\vec{r}_{class_k}|^2\right)}}{\sum_{k=1}^{no.\,of\ classes} e^{-|\vec{r}_{class_k}|^2 / \min\left(|\vec{r}_{class_k}|^2\right)}}$$

An exponential function of the class vectors $\vec{r}_{class_k}$ is used to calculate the probability of each class for the testing sample. In this study the testing sample will be assigned to the category for which it has the maximum probability.

4 Results

The MODIS training data set includes several misclassified isolated ice cloud samples, which are individually surrounded by the water clouds samples. The cloud classification experiments have been performed using the training data sets with or without the misclassified isolated samples removed.

1) Using the complete training set including the misclassified isolated samples, the LROI scheme yields the classification boundaries depicted in Fig. 3.

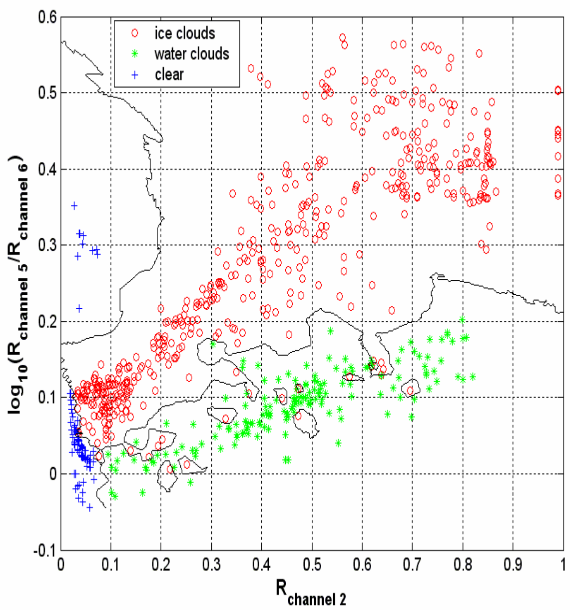

Fig. 3. Classification boundaries on the complete training set based on the LROI scheme.

The classification boundaries projected derived from the training set are projected on the testing set as shown in Fig. 4. As seen, except a few misclassified isolated training samples the training set is overall representing the testing set.

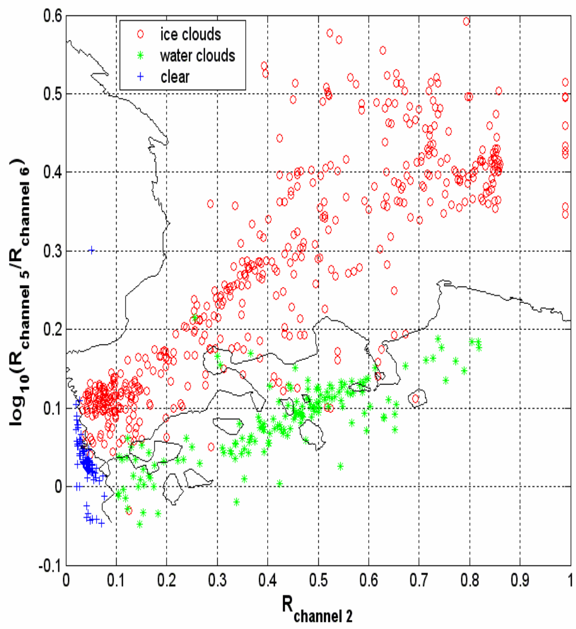

Fig. 4. Classification boundaries on the complete testing set based on the LROI scheme.

The distribution of the predicted category vs. true category is shown in Table 2. The misclassification error rate is 4.30%. Most errors occur in the prediction of ice clouds and water clouds. The classification boundary of ice clouds and water clouds is relatively complex and thus the errors between these two types are relatively high (12+18=30). Note that the misclassification error rate of this LROC scheme is only 4.30%, which is lower than the 4.6875% of MSVM from Lee *et. al.*, and the 18% of the MODIS cloud mask algorithm [7], and the 62.63% of the ML approach in Table 1.

Table 2 Distribution of predicted category vs. true category using the LROI classification. The misclassification error rate is 33/768 = 4.30%.

True category	Predicted category			Total
	Clear sky	Ice clouds	Water clouds	
Clear sky	127	1	0	128
Ice clouds	2	456	18	476
Water clouds	0	12	152	164

2) With the removal of misclassified isolated samples in the training data as a procedure of quality control,

the LROI scheme applied to the training set yields the classification boundaries as shown in Fig. 5.

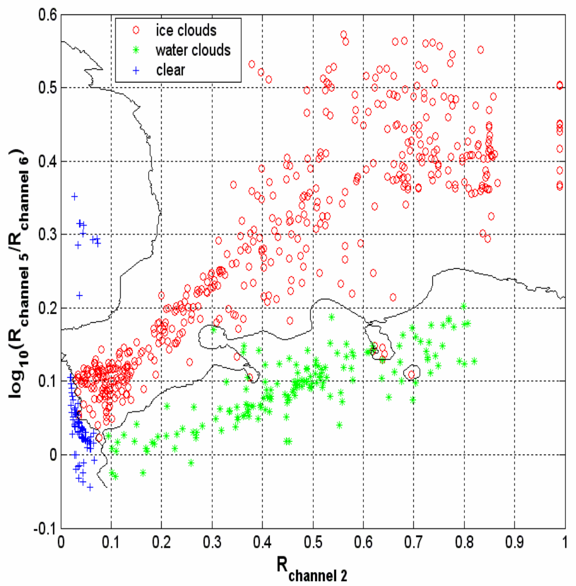

Fig. 5. The LROI-derived classification boundaries on the training set with the misclassified isolated samples removed.

The classification boundaries derived from the training set with the misclassified isolated samples removed is projected on the testing set as shown in Fig. 6.

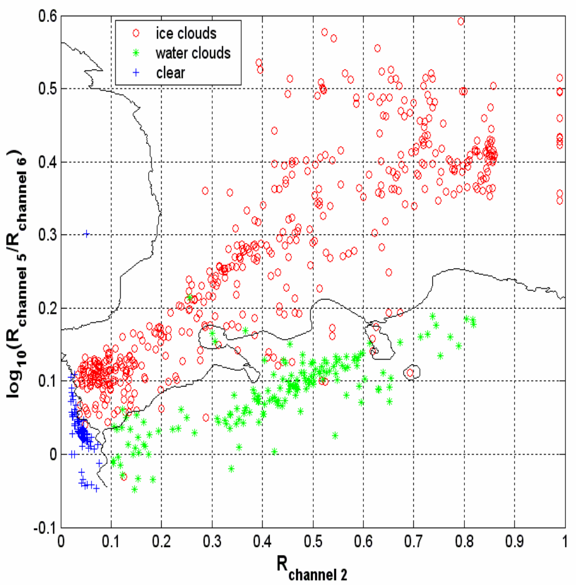

Fig. 6. Classification boundaries from Fig.5 projected on the testing set.

Table 3 shows the distribution of the predicted category vs. true category using the LROI scheme after the misclassified isolated samples removed. The misclassification error rate drops to 3.26 %.

Table 3 Distribution of predicted category vs. true category using the LROI classification. The misclassification error rate is 25/768 = 3.26%.

True category	Predicted category: Clear sky	Ice clouds	Water clouds	Total
Clear sky	127	1	0	128
Ice clouds	2	456	18	476
Water clouds	0	4	160	164

6 Conclusion

Valuable cloud information is provided by the MODIS instrument due to its high spatial resolution. A new supervised classification scheme is proposed that uses a local region of influence (LROI) to perform the classification. A comparison of this scheme is done with the MODIS cloud mask algorithm, the ML classification method, and the MSVM. Results show that the misclassification error rate of the proposed scheme is lower than those obtained with the other schemes testifying to the robustness of the LROI scheme.

References:

[1] B. Huang, H-L Huang and W.L. Smith, " Generalized Radiative Transfer Equation for Multilevel Clouds," in *Proc. SPIE*, vol. 4891, pp. 201-210, 2002.

[2] G. G. Mace, and S. Benson-Troth, "Cloud-layer overlap characteristics derived from long-term cloud radar data," *J. Climate*, 15, 2505-2515, 2002.

[3] K. Strabala, S. Ackerman, and P. W. Menzel, "Cloud properties inferred from 8-12 μm data," *J. Applied Meteorology*, 33, 212-229, 1994.

[4] S. Ackerman, K. Strabala, P. W. Menzel, R. Frey, C. Moeller, and L. Gumley, "Discriminating clear sky from clouds with MODIS," *J. Geophysical Research* **103**, 32, 141-32, 157, 1998.

[5] S. Platnick, M. King, S. Ackerman, P. W. Menzel, B. Baum, J. Riedi, and R. Frey, "The MODIS cloud products: Algorithms and examples from Terra," *IEEE Trans. Geosci. Remote Sens.*, 41, 459-473, 2003.

[6] V. Haertel and D. A. Landgrebe, "On the classification of classes with nearly equal spectral response in remote sensing hyperspectral image data," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, pp. 2374-2385, 1999.

[7] Li, J., W. P. Menzel, Z. Yang, R. A. Frey, and S. A Ackerman, "High spatial resolution surface and cloud type classification from MODIS multispectral band measurements," *J. Appl. Meteor.*, 42, 204-226, 2003.

[8] Y. Lee, G. Wahba, and S. Ackerman, " Classification of Satellite Radiance Data by Multicategory Support Vector Machines," *J. Atmos. Oceanic Tech.*, 21, pp. 159-169, 2004.