Conference on Satellite Meteorology and Oceanography, 28 Jan - 2 Feb 1996, Atlanta, GA by the AMS, Boston, MA

P3.26

THE EOS CERES GLOBAL CLOUD MASK

T. A. Berendes* and R. M. Welch

Institute of Atmospheric Sciences
South Dakota School of Mines and Technology
Rapid City, SD 57701-3995

Q. Trepte

Lockheed Engineering and Sciences Corporation
Hampton VA 23666.

C. Schaaf

Air Force Geophysics Laboratory
Hanscom AFB, MA 01731.

B.A. Baum

NASA Langley Research Center
Hampton, VA 23681-0001.

1. INTRODUCTION

To detect long-term climate trends, it is essential to produce long-term and consistent data sets from a variety of different satellite platforms. With current global cloud climatology data sets, such as the International Satellite Cloud Climatology Experiment (ISCCP) or CLAVR (Clouds from Advanced Very High Resolution Radiometer), one of the first processing steps is to determine whether an imager pixel is obstructed between the satellite and the surface, i.e., determine a cloud "mask." A cloud mask is essential to studies monitoring changes over ocean, land, or snow-covered surfaces. As part of the Earth Observing System (EOS) program, a series of platforms will be flown beginning in 1997 with the Tropical Rainfall Measurement Mission (TRMM) and subsequently the EOS-AM and EOS-PM platforms in following years. The cloud imager on TRMM is the Visible/Infrared Sensor (VIRS), while the Moderate Resolution Imaging Spectroradiometer (MODIS) is the imager on the EOS platforms. To be useful for long term studies, a cloud masking algorithm should produce consistent results between existing (AVHRR) data, and future VIRS and MODIS data. The present work outlines both existing and proposed approaches to detecting cloud using multispectral narrowband radiance data.

Clouds generally are characterized by higher albedos and lower temperatures than the underlying surface. However, there are numerous conditions when this characterization is inappropriate, most notably over snow and ice. Of the cloud types, cirrus, stratocumulus and cumulus are the most difficult to detect. Other problems arise when analyzing data from sun-glint areas over oceans or lakes, over deserts, or over regions containing numerous fires and smoke. The cloud mask effort builds upon operational experience of several groups that will now be discussed.

2. HERITAGE ALGORITHMS

The CERES cloud masking algorithm (Baum *et al.* 1994) will rely heavily upon a rich heritage of both NASA and NOAA experience with global

* *Contributing Author's Address:* Institute of Atmospheric Sciences, South Dakota School of Mines and Technology, 501 East Saint Joseph Street, Rapid City, SD 57701-3995.

data analysis. Initial algorithm design will incorporate the approaches used by ISCCP (International Satellite Cloud Climatology Project) (Rossow and Garder 1993), CLAVR (Clouds from AVHRR) (Stowe *et al.*, 1991), and SERCAA (Support of Environmental Requirements for Cloud Analysis and Archive). The ISCCP algorithms are based upon two channels, one in the visible wavelength region and one in the infrared. The CLAVR approach uses all five channels of the AVHRR instrument. The CLAVR multispectral threshold approach incorporates narrowband channel difference and ratio tests, including dynamic threshold specification with clear-sky radiation statistics.

The SERCAA (Gustafson *et al.* 1994) algorithm is operational at the Phillips Laboratory, Hanscom Air Force Base, and uses all five AVHRR radiometric channels. The SERCAA is sponsored jointly by the Department of Defense, Department of Energy, and Environmental Protection Agency Strategic Environmental Research and Development Program.

The International Satellite Cloud Climatology Project (ISCCP) cloud masking algorithm is described by Rossow (1989, 1993), Rossow *et al.* (1989) and Seze and Rossow (1991a). Data are used from the narrowband VIS (0.6 micron) and the IR (11 micron) channels. The ISCCP algorithm is based on the premise that the observed VIS and IR radiances are caused by only two types of conditions, 'cloudy' and 'clear', and that the ranges of radiances and their variability that are associated with these two conditions do not overlap (Rossow and Garder 1993). As a result, the algorithm is based upon thresholds, where a pixel is classified as "cloudy" only if at least one radiance value is distinct from the inferred "clear" value by an amount larger than the uncertainty in that "clear" value. The uncertainty can be caused both by measurement errors and by natural variability. The "threshold" for cloud detection is the magnitude of the uncertainty in the clear radiance estimates. This algorithm is constructed to be "cloud-conservative," minimizing false cloud detections but missing clouds that resemble clear conditions.

The NOAA CLAVR algorithm (Phase I) uses all five channels of AVHRR to derive a global cloud mask (Stowe *et al.*, 1991). It examines multispectral information, channel differences, and spatial differences and then employs a series of sequential decision tree tests. Cloudfree, mixed (variable cloudy) and cloudy regions are identified for 2x2 GAC pixel arrays. If all four pixels in the array fail all the cloud tests, then the array is labeled as cloud-free (0% cloudy); if all four pixels satisfy just one of the cloud tests, then the array is labeled as 100% cloudy. If 1 to 3 pixels satisfy a cloud test, then the array is labeled as mixed and assigned an arbitrary value of 50% cloudy. If all four pixels of a mixed or cloudy array satisfy a clear-restoral test (required for snow/ice, ocean specular reflection, and bright desert surfaces) then the pixel array is reclassified as "restored-clear" (0% cloudy). The set of cloud tests is subdivided into daytime ocean scenes, daytime land scenes, nighttime ocean scenes and nighttime land scenes. Subsequent phases of the CLAVR cloud mask, now under development, will be incorporated as modifications become available.

SERCAA, Support of Environmental Requirements for a Cloud Analysis and Archive is the prototype for the US Air Force's new global cloud analysis model. SERCAA makes use of a number of algorithms tailored to sensors on both the polar orbiting and geostationary meteorological satellite platforms. The resulting cloud masks are determined at sensor pixel resolution rather than a common grid. These algorithms have been extensively tested at various global locations.

Unfortunately, existing approaches have limitations, notably in detecting cloud shadows, cloud over snow- or ice covered surfaces, clouds in sunglint areas, fires and smoke from biomass burning, and dust storms over deserts. An improved global cloud mask appropriate for the EOS timeframe is discussed, and examples will be shown from application of the cloud mask to existing AVHRR data.

3. NEW METHODOLOGY

Two new classification methods are used in this study. Both methods use a set of pairwise decisions to classify samples. Most classification methods utilize a small number of features due to their multidimensional nature. For those methods, it is not feasible to use more than 10-20 features.

The feature vector used in this study consists of 164 spectral and textural and pseudo-textural features. The spectral features are created from the original 5 channels of the AVHRR data and a sixth channel which is the reflectance of channel 3. Spectral ratios, differences, arctangents and various other functions are computed. Grey level difference vector (GLDV) textural features are computed over a 7x7 mask. Pseudo textures are computed over a 3x3 neighborhood.

The pairwise classifiers select a subset of features from the feature vector. The selected features are optimal for distinguishing between pairs of classes. This reduces the size of the final feature vector to approximately 20 - 30 features. The final size of the feature vector is determined by the number of tests performed for each pair of classes.

3.1 Training

The classifiers were trained using the Satellite Imagery Visualization System (SIVIS). SIVIS allows the user to visualize large satellite images and select samples from the data. Representative samples of 23 different classes were selected from over 40 scenes. The scenes are chosen from three main climate regimes; polar, Middle-Eastern desert and South American rainforest during the burning season. A separate classifier is constructed for each regime.

3.2 Paired Rule Classifier

The paired rule classifier takes the 164 element feature vector and creates a new feature vector consisting of the original 164 features and all possible ratios of the form A/B. This results in a feature vector with 13694 features.

Next, the divergence for each pair of classes is computed for each feature. For a given feature F, the divergence between class i and class j is defined as:

$$DIV(F)_{i,j} = | m_i - m_j | / (s_i + s_j),$$

where m_i and m_j are the means for classes i and j, and s_i and s_j are the standard deviations for those classes.

A list of features, sorted by decreasing divergence, is constructed for each pair of classes. In this study, the 5 features with highest divergence are chosen for each class pair. A threshold then is determined for each of the chosen features. For a given feature F, the threshold T is used as a test for discriminating between the two classes i and j as follows:

$$T = m_i + s_i (m_j - m_i) / (s_i + s_j).$$

If F < T, then the test returns class i, if F > T the test returns class j. Each class has 5 tests in this study. A histogram tabulates the number of tests satisfied by each class. After all tests for each class have been completed, the class with the highest histogram count is chosen.

3.3 Paired Histogram Classifier

The paired histogram classifier takes the 164 element feature vector and constructs histograms for each pair of classes. For feature F, two histograms I_F and J_F are created for classes i and j. The histogram ranges are scaled to accommodate the minimum and maximum values of both classes and discretized into 256 bins. The histogram values for each class are normalized by the number of samples in the class. These paired histograms are analyzed and sorted, based upon overlap and divergence. Overlap, O, is defined as follows:

$$O(F)_{i,j} = \sum_{x=1}^{256} I_F(x) J_F(x)$$

Divergence is defined above.

The three features with the lowest overlap and highest divergence for each pair of classes are chosen. For a chosen feature F, the paired histograms I_F and J_F are used for discriminating between class i and j.

The histograms can be used as discriminators in a variety of ways. The method chosen for this study considers the histograms as range specifiers for the class pairs. Each of the 256 bins of the normalized histograms I_F and J_F are compared and the following rules are applied:

if $I_F(x) > J_F(x)$ then	if $I_F(x) < J_F(x)$ then
$I_F(x) = 1$	$I_F(x) = 1$
$J_F(x) = 0$	$J_F(x) = 1$

where x = 1.256

The paired histograms I_F and J_F now represent ranges in feature F which correspond to classes i and j respectively. The three features used for each class pair produce three pairs of histograms which are used for classification. This process is repeated for each pairwise combination of classes.

The following procedure is used to classify a sample. First, calculate the 164 feature vector for the sample. Next, compute the histogram bin for each set of paired histograms. Retrieve the value at each computed histogram bin and increment a voting histogram by that value. Finally, examine the voting histogram and assign the class with the highest value.

4. RESULTS

The paired rule and paired histogram methods were both used on three areas of interest; polar, desert, and South America. These areas were chosen because they are particularly difficult areas to classify. Results are preliminary at this time, but both classifiers are performing well for all of the scenes analyzed so far. In particular, the South American classifier is able to detect smoke from biomass burning. The desert classifier can detect desert, dust storms and some sunglint areas in the ocean. The polar classifier is able to differentiate between clouds and ice. We are very encouraged by our results with these difficult classes.

Color photographs of the classification results will be presented at the conference.

5. FUTURE WORK

The classification algorithms are undergoing continuing revision and enhancement. Work is also continuing on developing better features for classification.

The training sample database is constantly expanding to include more representative samples of each class and more comprehensive coverage of the earth.

6. ACKNOWLEDGEMENTS

This work was performed as part of the EOS Clouds and Earth's Radiant Energy System (CERES) program. Two of us (Todd A. Berendes and Ron M. Welch) were supported under National Aeronautics and Space Administration contract NAS1-19077.

7. REFERENCES

Baum, B. A., R. M. Welch, P. Minnis, L. L. Stowe, J. A. Coakley, Jr, and colleagues, 1994: Imager clear-sky determination and cloud detection. Clouds and the Earth's Radiant Energy System (CERES) algorithm theoretical basis document. Subsystem 4.1. Postscript version available via the World Wide Web at http://asd-www.larc.nasa.gov/ATBD/ATBD.html.

Gustafson, G. B., R.G. Isaacs, R. P. d'Entremont, J.M. Sparrow, T.M. Hamill, C. Grassotti, D.W. Johnson, C.P. Sarkasian, D.C. Peduzzi, B.T. Pearson, V.D. Jakabhazy, J.S. Belfiore, A.S. Lisa, 1994: Support of environmental requirements for cloud analysis and archive (SERCAA): Algorithm descriptions.. Technical Report PL-TR-94-2114, Phillips Laboratory, Air Force Materiel Command, Hanscom Air Force Base, MA 01731, 100 pp.

Neu, T.J., R.G. Isaacs, G.B. Gustafson, and J.W. Snow, 1994: Improved Cloud Analysis for CDFSII through the SERCAA Research and Development Program. Seventh Conference on Satellite Meteorology and Oceanography, Monterey, CA, 6-10 Jun 1994, American Meteorological Society, Boston, MA, 239-242.

Rossow, W. B., 1989: Measuring cloud properties from space. A review. *J. Climate*, **2**, 201-213.

Rossow, W. B. and L. C. Garder, 1993: Cloud detection using satellite measurements of infrared and visible radiances for ISCCP. *J. Climate*, **6**, 2341-2369.

Seze, G., and W. B. Rossow, 1991: Time-cumulated visible and infrared radiance histograms used as descriptors of surface and cloud variations. Int. J. Remote Sen

Stowe, L. L., E. P. McClain, R. Carey, P. Pellegrino, G. Gutman, P. Davis, C. Long, and S. Hart, 1991: Global distribution of cloud cover derived from NOAA/AVHRR operational satellite data. *Adv. Space Res.*, **11**, 51-54.

Stowe, L. L., S. K. Vemury, and A. V. Rao, 1994: AVHRR clear sky radiation data sets at NOAA/ NESDIS. *Adv. Space Res.*, **14**, 113-116.