NASA-CR-201472

# SEMI-ANNUAL PROGRESS REPORT

Period: January-June 1996

"The Effects of Cloud Inhomogeneities Upon Radiative Fluxes, and the Supply of a Cloud Truth Validation Dataset"

Submitted by

Institute of Atmospheric Sciences
South Dakota School of Mines and Technology
501 E. St. Joseph Street
Rapid City, SD 57701-3995

to

National Aeronautics and Space Administration
Goddard Space Flight Center
Contract NAS5-31718
Greenbelt, MD 20771

## 1. Work Accomplished During the Report Period

This work is directed towards the development of algorithms for the ASTER science/instrument teams. Special emphasis is being placed on a wide variety of cloud optical property retrievals, and especially retrievals of cloud and surface properties in the polar regions.

## 2. Research Activities

### 2.1 Cloud Algorithms

#### 2.1.1 ASTER Polar Cloud Mask

During this reporting period the first stage of the ASTER Polar Mask Classifier (i.e., the cluster based approach or preclassification stage) was converted from IDL to C. It also was integrated into the Paired Histogram Method (PHM) classifier (which was already coded in C). It was tested on 84 Landsat TM scenes. This integrated classifier provided for improved classification results over both classifiers separately and decreased the processing time for the PHM classifier from 60 minutes to 20 minutes for a scene comprised of approximately 12 million pixels. It was tested on two scenes from which we have not extracted any samples. The integration was achieved by making the PHM classifier the main program and the first stage the called program. The first stage is comprised of two modules - the first derives the adaptive threshold values for two key features and the second implements rules based on those thresholds. The PHM calls the first module once and the second module for each pixel or feature vector in the image. The output from the second module of the first stage, if possible, reduces the class ambiguity from 10 classes to 2 to 4 classes. The PHM classifier then is used to resolve the ambiguity among the remaining classes. For example, the output from the second module of the first stage might be encoded to indicate that the pixel classification is either wet ice, ice/snow, or shadowed ice/snow but is not any other class. The PHM classifier then only runs class comparison tests for those 3 classes but not for any of the others. Therefore, the PHM classifier only runs 6 comparison tests instead of 45. Another version of the classifier was also developed in which the PHM was supplanted by backprop neural networks. Eleven backprop neural networks were trained on each possible combination of outputs from the preclassification stage. The results from this classifier appear to be as good as those from the PHM based version of the classifier. During the next reporting period we will be testing both of these integrated classifiers on additional Landsat circumpolar scenes that we received recently.

We continue testing of a hierarchical neural network (HNN). We are in the process of evaluating whether the network is superior to other techniques (i.e., the Paired-Histogram Method (PHM) classifier and the fuzzy logic classifier) in the classification of specific classes and if it might be useful in the second stage of the ASTER Polar Cloud Mask classifier. We have found that, when applying the classifier to the labeled sample set, that the classification accuracies are, in general, superior to those obtained from any other classifiers tested to date. Confusion matrices were generated which show a comparison of the classification results with the actual or known classes of the labeled samples. The elements of each confusion matrix are normalized to percent of the total number of test samples for the class tested. The results are presented in 3 confusion matrices for each of the northern and southern latitude data sets. The first matrix shows the accuracy of the classifier in its most important role - as a cloud mask. It is a 2 by 2 matrix in which all the classification results for the clear classes and the cloud classes have been accumulated together. For example, the value in the first row and column show the percent of samples from cloud classes classified as one of the cloud classes (thin cloud over ice/snow, water, or land, and thick cloud) while the value in the second row and column show the percent of samples from clear classes classified as one of the clear classes (water, slush/wet ice, ice/snow, land, shadowed land, and shadowed ice/snow). The off-diagonal elements, row 1 - column 2 and row 2 - column 1, show the errors; that is, they show the percent of samples from cloud classes classified as one of the clear classes and the percent of clear classes classified as one of the cloud classes, respectively. The second matrix shows the classification results for those samples from clear classes correctly classified as one of the clear classes while the third matrix shows the classification results for those samples from cloud classes correctly classified as one of the cloud classes. These 2 matrices depict the accuracy of the classifier in its secondary role in distinguishing among clear classes or cloud classes. The diagonal elements in these 2 matrices indicate the classification accuracies for each class while the off-diagonal elements indicate the percent inaccuracies or "confusion" of the classifier. The accuracy over all classes for each of these matrices is shown below the matrix (i.e., the percent of correctly classified samples). In the case of the second matrix for the cloud classes, the main distinction is thin or thick cloud and, if the sample is from a thin cloud, the distinction is the underlying surface (water, land, ice/snow). In the case of the third matrix for the clear classes, the main distinction is between land and some phase of water (liquid, frozen, melting) and shadowed and unshadowed. Six confusion matrices follow, 3 each for the northern and southern latitude data. These results are based on randomly selected samples from a pool of approximately 1 million labeled pixel samples.

Northern Latitude
Cloud vs. Clear

| | Cld | Clr |
|---|---|---|
| Cld | 97.3 | 2.54 |
| Clr | 2.7 | 97.5 |

Total: 97.4

Cloud Classes

| | 4 | 5 | 6 | 7 |
|---|---|---|---|---|
| 4 | 93.0 | 0.3 | 4.5 | 2.3 |
| 5 | 2.6 | 94.2 | 8.0 | 0.0 |
| 6 | 2.8 | 5.5 | 86.9 | 0.4 |
| 7 | 1.6 | 0.0 | 0.6 | 97.2 |

Total: 92.1

| | Clear Classes | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 8 | 9 | 10 |
| 1 | 96.3 | 7.4 | 0.0 | 0.4 | 7.1 | 9.3 |
| 2 | 2.4 | 82.3 | 0.5 | 0.0 | 5.5 | 0.8 |
| 3 | 0.1 | 3.4 | 96.9 | 0.8 | 5.3 | 0.9 |
| 8 | 0.6 | 0.4 | 0.9 | 95.7 | 15.4 | 27.6 |
| 9 | 0.3 | 6.4 | 1.7 | 0.1 | 63.9 | 6.3 |
| 10 | 0.5 | 0.2 | 0.0 | 3.0 | 2.9 | 55.1 |

Total: 90.4

Southern Latitude
Cloud vs. Clear

| | Cld | Clr |
|---|---|---|
| Cld | 94.3 | 2.7 |
| Clr | 5.7 | 97.3 |

Total: 95.93

| | Cloud Classes | | |
|---|---|---|---|
| | 4 | 5 | 7 |
| 4 | 83.8 | 9.2 | 12.8 |
| 5 | 8.1 | 83.7 | 21.1 |
| 7 | 0.4 | 0.9 | 65.1 |

Total: 82.5

| | Clear Classes | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 8 |
| 1 | 92.3 | 3.6 | 0.1 | 0.8 |
| 2 | 7.3 | 87.9 | 6.1 | 4.9 |
| 3 | 0.0 | 5.3 | 91.2 | 2.9 |
| 8 | 0.4 | 3.3 | 2.7 | 91.4 |

Total: 90.3

During the last reporting period we developed a new method for assessing the accuracy of any classification algorithm we are testing. Until now, our only method for assessing the accuracy, quantitatively, was through analysis of confusion matrices. The confusion matrices are derived from the results of applying a given algorithm to the labeled samples and, therefore, do not provide a quantitative measure of the accuracy of the algorithm when applied to a specific scene. Until now, the scene classification accuracy has been estimated subjectively. The ideal method for determining this accuracy would be to have a manually (human) classified mask for each and every scene against which the algorithm derived mask could be compared. Since it is not practical to do this, we developed a validation tool, to be used by human analysts, that provides output that can be used to estimate this accuracy. Through a random process, 16 small subregions within a given scene are selected (by the computer). The 16 subregions are derived from 16 uniformly partitioned regions within the scene. The size of the subregion is such that a human analyst can visually determine the fraction of each class present and is currently 16 by 16 pixels. The analyst is able to display any band or 3-band overlay of any 3 bands to augment his determination of the classes present in the subregion. To date, 2 analysts have performed this random manual classification process on 12 scenes (6 south of 60S and 6 north of 60N) for a total of 384 classified subregions (in term of fractional presence of a class). The statistics from these 384 manually classified subregions were compiled. The statistics for these same regions when using the PHM and HNN classifiers were also compiled. When comparing the results obtained between the 2 analysts, the agreement in cloud and no cloud classification was only at the 90 percent level. If the mean classification results for the 2 analysts are averaged and compared to each of the 2 classifiers (PHM and HNN), the agreement in cloud and no cloud classification is 83 and 84 percent, respectively. These preliminary results indicate that, depending on which analyst the algorithm classification results are compared to, the accuracy could range anywhere from 83 to 93 percent. Results obtained from confusion matrices indicate cloud and no cloud classification accuracies of 95-96 percent. However, as expected, these results suggest that the overall scene classification accuracy is less and is probably between 85-90 percent. In addition, the certainty of the classification accuracy is probably only 5 percent. Since it is difficult to present the results from this process in a concise manner, we are now trying a variation on this methodology. Instead of using a 16 by 16 region, we are now using an 8 by 8 region and instructing the human analyst to indicate the dominant class in the region, although more than one might be present. When comparing these results to the classifier the same 8 by 8 region will be checked in the classifier mask for the dominant class. This comparison will provide for a binary result (correct/incorrect or agree/disagree). The results then can be presented concisely in a confusion matrix format. We plan to use more analysts, collect more samples and continue to compile statistics on classification accuracies using this methodology. The results from this process will be included in a paper describing the Polar Cloud Mask algorithm, which we are currently preparing for journal submission. We also are examining the results from this process to determine if some classification algorithms are superior to others for specific cloud or no cloud subclasses.

For the purpose of testing our classification algorithms for robustness, we parsed our northern and southern latitude data into training and testing subsets. In the past our testing set of samples was always different from our training set of samples; however, the testing and training sets were taken from the same pool of samples. The samples were parsed into two pools in which one pool was derived from the samples extracted from one set of scenes while the other pool was derived from the samples from the balance of the scenes. The testing samples were drawn from one pool and the training samples from the other pool. During this reporting period we tested the PHM on these testing and training sets. When clustering the results into 4 classes (water, frozen water, clouds, and land) the classification accuracy decreased by approximately 4 percent. However, the overall accuracy for all ten classes (water, slush/wet ice, ice/snow, thin cloud over ice/snow, thin cloud over water, thin cloud over land, thick cloud, land, shadow on ice, and shadow on land) decreased 20 to 25 percent, depending on the random set of samples chosen. We also tested a Mahalanobis classifier on the same sample sets and observed the same relative decreases in performance. A backprop neural network was trained and tested on this same data set. As indicated above the PHM classifier decreased in clear/cloud accuracy about 5 percent from 95% to 90%. The neural network classifier did not manifest a decrease in accuracy for clear/cloud classification and remained around 95%. The within clear and cloud accuracies for the PHM classifier decreased 15 to 20 percent. Again the neural network results did not decrease significantly. It appears that the neural network classifier is more robust when applied to this data set. Upon further examination we discovered that the feature distributions (histograms) were significantly different for some of the classes between the testing and training sets. We are presently uncertain as to why the backprop neural networks performed better here but during the next reporting period we plan to determine if that result is manifested in the full scene classification masks.

We participated with Dorothy Hall and George Riggs of NASA Goddard in a joint conference paper that was presented at the Eastern Snow Conference in Williamsburg, VA during the first week of May. The topic of the paper is a comparison of the results from their SNOMAP and ICEMAP classifiers (to be applied to MODIS) with our ASTER polar cloud mask classifier. George Riggs selected two scenes on which to conduct the comparison and sent that data to us. We processed the two scenes and obtained classification masks. We exchanged classification results and made comparisons. The paper submission has also been completed. The comparisons indicate relatively good agreement. The main failing of the SNOMAP algorithm was to misclassify some cloudiness as snow. The main failing of the ASTER polar cloud mask algorithm was to misclassify some multi-layer and thick cloud as thin cloud over ice/snow. This pointed up a new concern for our classification scheme. In the past, we had placed most of our emphasis on detecting cloud pixels correctly but much less emphasis on the accuracy of classifying the underlying surface and, in fact, we were only including the within cloud classification as a point of interest. To preclude a potential user from placing too much faith in the within cloud classification (for example, someone trying to construct a snow map), the product information needs to be very explicit that the clear and cloud accuracy is significantly higher than the within clear and within cloud classification accuracy.

During this reporting period we finished transferring the 500 plus scenes data set of subsampled Landsat TM data from Rich Irish. Since the imagery in this dataset does not have header records, we are currently implementing a technique to derive solar zenith angle from path and row. We plan to start testing the aforementioned integrated classifier on this dataset during the next reporting period; however, the classification accuracy may be less than adequate since our classifier has been trained on polar imagery and this dataset is mostly nonpolar.

During the last reporting period the ASTER Polar Cloud Mask Validation Plan was prepared and submitted to the ASTER project office. Validation of the algorithm is to be performed through four mechanisms. The first will be by applying the algorithm to labeled samples. The labeled samples will be partitioned into two scene groups. One group will be used for training the classifier and the other group will be used for testing. The results from testing will be accumulated in confusion matrix form which shows, for each set of samples corresponding to a specific class, the percent or fraction of the samples classified into each of all possible classes. The diagonal elements of the matrix indicate the accuracy of the classifier for each individual class. Three matrices will be constructed showing the results for clear/cloud, within clear class accuracy, and within cloud class accuracy. The results from this kind of analysis tend to overestimate the accuracy of the classifier for a given class or class group (such as clear or cloud) by a few percent (in the range of 1 to 7 percent, depending on the class) since labeled samples are generally selected from spectrally homogeneous regions with unambiguous identity. However, the matrices provide a quantitative result which can be tracked as new scenes and samples are added to the data set. The second method is qualitative. Color coded classification masks will be examined by human experts who will asses the overall performance of the classifier and determine which types of misclassifications occur most frequently. The primary assessment will be the accuracy of the classifier in distinguishing between clear and cloud classes. The third method is similar to the first. In the first method the human expert selects the samples. In this procedure the computer randomly selects small subregions within the image and the human expert estimates the most dominant class in the subregion. The results from this process will be presented in confusion matrix form also. The classification accuracy estimates from this process should be lower than those from the first method and be a better estimate of the accuracy of the classifier. This method will not be used as extensively as the first as it requires more than one expert and it is more tedious to classify randomly selected samples. The fourth method involves the comparison of classification results with surface based observations. Whenever available human observations, and lidar and other measurements of cloud fraction or presence are available, they will be used to validate the cloud detection capability of the classifier. Ron Welch attended the Validation Workshop held at GSFC on 8-10 May 1996 at which the validation plans were reviewed.

During this reporting period, a Quality Assurance Plan was also prepared and submitted to the project office. The plan includes both nonspecific or "pass through" QA information as well as specific product QA information. The product specific QA information derived from the classifier will be stored in the 2 least significant and 4 most significant digits of the QA plane. Those 2 least significant digits will be encoded as follows:

| Dec | Binary | Meaning |
|---|---|---|
| 0 | 00 | Certainty measures derived from the classification algorithm indicate a "high" confidence in the result |
| 1 | 01 | Certainty measures derived from the classification algorithm indicate "less than high" (moderate) confidence in the result |
| 2 | 10 | Same as binary 01 |
| 3 | 11 | Certainty measures derived from the classification algorithm indicate a relatively "low" confidence in the result (i.e., the feature vector is highly ambiguous between 2 or more classes); however, the result is possibly correct |

The 4 most significant digits will be encoded as follows:

| Dec | Binary | Meaning |
|---|---|---|
| 9 | 1001 | This pixel has been discovered to be bad during the production of this product due to out of range feature value(s) (e.g., VNIR or SWIR band reflectance values less than 0 or greater than 2.0)<br>Note: Other six codes not used. |
| 1 | 0001 | After processing, this pixel is now deemed suspect due to marginally out of range feature value(s) (e.g., Band 14 brightness temperature greater than 310 K for a particular geographic region and season)<br>Note: Other five codes not used. |

Work continues on a journal submission describing the application of the paired histogram method to Landsat TM polar scene classification. Work also continues on a similar paper for the Hierarchical Neural Network. The same sample sets and scenes are being used so that a comparison can be made.

A conference paper describing the ASTER polar cloud mask is also in preparation which is to be submitted during the first week of July to the International Symposium on Optical Science, Engineering, and Instrumentation, SPIE's Annual Meeting, held 4-9 August 1996 in Denver, CO. An oral presentation will be made on August 6th during the Infrared Spaceborne Remote Sensing IV Session.

Ron Welch traveled to Pasadena, CA during the month of June to attend the next ASTER Science Team meeting.

### 2.1.2 Simulation of 3-D Cloud Effects

We completed the conversion of our Monte Carlo photon transport model from IDL to C. Initial tests indicate a 100-fold improvement in speed. This will enable us to run much larger numbers of photon trajectories than we did in the past. The intent is to increase the number of photon trajectories sufficiently to generate radiance patterns that can be compared to those obtained from our analytical Picard Iterative Method for 3-D radiative transfer. During the next reporting period we plan to make comparison runs between the Monte Carlo model and the Picard Iterative method for some simple 3-D geometric shapes.

### 2.1.3 Cloud Base Height Retrievals

The paper entitled "Estimation of Cirrus and Stratus Cloud Height Using Landsat Imagery" by Yasushi Inomata and Ronald M. Welch, appeared in the March issue of the *Journal of Applied Meteorology*. The paper describes a technique for estimating the height of clouds with thin and/or ill defined edges using 2-D cross correlation.

# REPORT DOCUMENTATION PAGE

*Form Approved*
*OMB No 0704-0188*



| 1. AGENCY USE ONLY (Leave blank) | 2. REPORT DATE | 3. REPORT TYPE AND DATES COVERED |
|---|---|---|
| | 10 July 1996 | Semi-annual, 1 Jan 96-30 Jun 96 |

**4. TITLE AND SUBTITLE**
The Effect of Cloud Inhomogeneities Upon Radiative Fluxes, and the Supply of a Cloud Truth Validation Dataset

**5. FUNDING NUMBERS**
C - NAS5-31718

**6. AUTHOR(S)**
Ronald M. Welch

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**
Institute of Atmospheric Sciences
S.D. School of Mines and Technology
501 E. St. Joseph Street
Rapid City, SD 57701-3995

**8. PERFORMING ORGANIZATION REPORT NUMBER**
SA-9

**9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)**
National Aeronautics and Space Administration
Goddard Space Flight Center
Greenbelt Road
Greenbelt, MD 20771

**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**
IN-46-CR

**11. SUPPLEMENTARY NOTES**

**12a. DISTRIBUTION/AVAILABILITY STATEMENT**
Unlimited

**12b. DISTRIBUTION CODE**

**13. ABSTRACT** *(Maximum 200 words)*

The ASTER polar cloud mask algorithm is currently under development. Several classification techniques have been developed and implemented. The merits and accuracy of each are being examined. The classification techniques under investigation include fuzzy logic, hierarchical neural network, and a pairwise histogram comparison scheme based on sample histograms called the Paired Histogram Method. Scene adaptive methods also are being investigated as a means to improve classifier performance. The feature, arctan of Band 4 and Band 5, and the Band 2 vs. Band 4 feature space are key to separating frozen water (e.g., ice/snow, slush/wet ice, etc.) from cloud over frozen water, and land from cloud over land, respectively. A total of 82 Landsat TM circumpolar scenes are being used as a basis for algorithm development and testing. Numerous spectral features are being tested and include the 7 basic Landsat TM bands, in addition to ratios, differences, arctans, and normalized differences of each combination of bands.

A technique for deriving cloud base and top height is developed. It uses 2-D cross correlation between a cloud edge and its corresponding shadow to determine the displacement of the cloud from its shadow. The height is then determined from this displacement, the solar zenith angle, and the sensor viewing angle.

**14. SUBJECT TERMS**
ASTER, Landsat TM, polar cloud mask, scene classification, neural network, cloud base height

**15. NUMBER OF PAGES**
10

**16. PRICE CODE**

| 17. SECURITY CLASSIFICATION OF REPORT | 18. SECURITY CLASSIFICATION OF THIS PAGE | 19. SECURITY CLASSIFICATION OF ABSTRACT | 20. LIMITATION OF ABSTRACT |
|---|---|---|---|
| Unclassified | Unclassified | Unclassified | |