Testing Two Cloud Removal Algorithms for SSM/I

Perry J. Hardin, Ryan R. Jensen, David G. Long and Quinn P. Remund.
Microwave Earth Remote Sensing Laboratory, 447 CB. Brigham Young University, Provo, UT 84602.
Tel (801) 378-6062, FAX (801) 378 – 5110, perry_hardin@byu.edu

Abstract – The ability to monitor and map change in tropical forest regions is critical for the study of both carbon dioxide exchange and global climate. Remote sensing provides a very cost-effective and efficient method to monitor and map rainforest extent. In particular, moderate-resolution spaceborne sensors such as the Special Sensor Microwave/Imager (SSM/I) provide the ability to monitor large geographic areas such as the Amazon Basin with great frequency. However, despite their sophistication, passive sensors such as the SSM/I are unable to "see" through heavy clouds. This research was designed to test the effectiveness of two simple algorithms used to remove the effects of cloud cover on SSM/I data. The study area was the Amazon Basin. The approach taken in this research was to subtract original SSM/I imagery from the algorithm-processed imagery on a pixel-by-pixel basis. This was done for each of the SSM/I bands. These difference images were then examined statistically against rainfall data acquired from 321 stations in the Amazon Basin. Based on correlation analysis, it appears that the two algorithms are very effective in removing cloud contamination from SSM/I data. However, their effect varied by SSM/I band and polarization.

INTRODUCTION

The capability to map change in tropical forests and woodlands is essential to the study of global climate change. Satellite-borne remote sensors are particularly cost effective for this monitoring task. However, despite their sophistication, passive sensors such as the Special Sensor Microwave / Imager are unable to sense the earth's surface through heavy precipitation and clouds. Given the prevalent nature of cloud cover and rain that characterize equatorial rainforests, the removal of cloud cover from radiometer imagery has become an important preprocessing task prior to any landcover mapping.

Two simple algorithms have been used in removing cloud cover from raw SSM/I image data. Both examine each image pixel through a period of time that includes multiple overpasses. The goal of both algorithms is to produce an average apparent brightness image with clouds removed.

Using the modified maximum average (MMA) method, [1] the several pixel values for a given row and column image cell are placed in an array. The raw mean of this pixel vector is then calculated. Next, a subset of pixels is extracted from the complete vector by comparing each pixel in the vector to the mean. If the pixel value is above the average, it is retained. Subsequently, the highest value of this subset is eliminated and the average of the remaining pixel subset is then computed. This is the MMA average for the pixel.

The second highest (SH) algorithm [2] uses the same starting pixel vector, sorts the vector, and then designates the second highest value as the cloud-free brightness temperature estimate. The logic behind this approach is simple. Clouds generally lower the brightness temperature, and in consideration of the possibility that the highest value in the vector is an artifact of processing, the second-highest value would be a logical candidate as an uncontaminated measurement.

The goal of this research was to compare the behavior of these two algorithms in an area of Brazil where daily rainfall data was available for the entire overpass period of the sensor.

METHODS AND DATA

Imagery

The source imagery used in this study consisted of all seven bands of SSM/I imagery acquired between Julian days 245 and 264, 1992. Considering both ascending and descending passes together, each pixel had the potential to be measured from five to ten times throughout the 21 day period. The study area for the project was limited to the Amazon Basin of Brazil. Figure 1 shows the 85GHz vertically polarized SSM/I band of central South America.

For each of the seven bands individually, these raw overpass images were combined into three time-composite images. The first was a simple average image – no rainfall removal was attempted. The second and third images were created respectively using the MMA and SH algorithms described above.

In order to isolate the cloud cover and precipitation in the scene, two other datasets were created from these 21 composites. The first was the difference between the average image and the SH image, determined on a pixel-by-pixel basis. The second dataset was calculated the same way using the MMA image. It was these 21 difference images that were used in further analysis.

Weather data

As part of an unrelated project conducted by the University of California, Santa Barbara, daily rainfall data were collected from 321 weather stations in the Brazilian Amazon River Basin. For this study, the variables extracted from the 321 weather stations were 1) total days of rain during the overflight period and 2) total depth of rain during the overflight period.

Using the latitude and longitude tags associated with each station, the two rainfall variables were geographically registered with the 21 MMA and SH difference images. Interpolation was done between the weather stations to estimate the rainfall variable values across the study area.

This process created 1) a single total rainfall depth map and 2) a single total days of rainfall map that corresponded to the area covered by the difference images.

Statisical analysis

Simple statistics were calculated to determine the relationship between the 21 difference images and the two rainfall variable surface maps. In particular, correlation analysis was used to determine the strength of the relationship between:

- Total days of rainfall and the MMA difference images
- Total days of rainfall and the SH difference images
- Rainfall depth and the MMA difference images
- Rainfall depth and the SH difference images

The metric used to assess goodness was the coefficient of determination (usually designated in the literature as R^2). It is interpreted [3] as the percentage of variance in the dependent variable (a rainfall variable) attributable to the independent variable (an algorithm). Depending on the context of the analysis, higher coefficients have different interpretations.

- When comparing the two algorithms (using identical bands / polarizations), a higher correlation indicates a greater removal of cloud cover.
- When comparing coefficients across the various bands / polarizations with respect to a single algorithm, a higher correlation indicates a stronger linear relationship between cloud attenuation and the frequency / polarization.

RESULTS

As shown in Table 1, both the SH and MMA algorithms generated images which had a higher mean brightness than their non-corrected counterparts. Likewise, variance was lower in the cloudless images than in the originals.

Tables 2 and 3 show the results from the correlation analysis relating the rainfall variables to the SH and MMA difference images. For both algorithms, the relationship between the rainfall variables and the difference images increases with frequency. For any band where both polarizations were available, there is also a slightly stronger relationship between the rainfall variables and the vertically polarized band. Overall, the strongest relationship is in the 85 GHz vertically polarized band whereas the weakest relationship was between the rainfall variables and the longer wavelength, horizontally polarized 19 GHz band.

In general, the tabled values for the MMA algorithm are higher than corresponding values for the SH algorithm. Although the difference is not dramatic, this seems to indicate that that the MMA provides better cloud removal.

Perfect correlation between the rainfall data and the difference images would have been indicated by a coefficient of variation reaching 100%. However, several factors in the data set would preclude perfect correlation even if the algorithms removed 100% of the cloud contamination. First, the rainfall data only recorded days of rainfall and daily rainfall amounts. It did not monitor days of cloud cover. In other words, the rainfall data is an imperfect surrogate for cloud cover. Worse still, the original rainfall data only contained summary data for an entire 24 hour period, not for the precise moment of satellite imagery acquisition. Furthermore, other factors such as soil moisture, plant water content, and even forest water inundation are unaccounted for in the analysis. Given these critical issues, it is noteworthy that the relationships between the rainfall variables and difference images are as strong as they are.

CONCLUDING COMMENTS

While both the MMA and SH algorithms may appear to remove the effects of cloud cover to a reasonable degree, the question of bias requires comment. When many overpasses are used to determine the average brightness value of a given pixel, the SH algorithm would be expected to shift the average brightness temperature higher than the MMA algorithm. This is empirically supported by the data in Table 1. For all possible frequency / polarization possibilities, the SH estimates were higher than the MMA estimates.

It could also be expected that the MMA algorithm should be a better estimator of the true average because it does not exclude as many measurements as the SH alternative. Preliminary results of simulations being conducted at the time of this writing indicate this is expectation is warranted.

While the MMA appears to be superior to the SH alternative, it operates without tolerance – for any given pixel vector, it discards all values below the mean. This is done with the tacit assumption that cloud cover contaminates all (and *only*) values below the mean of the vector. This is not true – at the extremes, either all or none of the pixels in a vector *may* be contaminated. In the former case, good data is needlessly eliminated. In the latter case, the contamination remains undetected.

Currently we are conducting research into other algorithms for cloud removal [4] with particular emphasis on a hybrid algorithm that executes the MMA logic only in areas where contamination is likely. That contamination is detected by the magnitude of the variance within the pixel vector – the variance within the pixel vector is larger when contamination is present. In the hybrid algorithm, the MMA logic is prohibited when the variance in the vector is below a certain threshold.

REFERENCES

1. D.G. Long and D.L. Daum. 1998. Spatial Resolution Enhancement of SSM/I Data. *IEEE Trans. Geosci. Remote Sensing*, 36: 407 – 417.
2. B.J. Choudhury and C.L. Tucker. 1987. Satellite Observed Seasonal and Inter-Annual Variation of Vegetation over the Kalahari, The Great Victoria Desert, and the Great Sandy Desert: 1979 – 1984. *Remote Sensing of Environment.* 23:233-241.
3. J.F. Hair, Jr., R.E. Anderson, R.L. Tatham, and W.C. Black. 1998. Multivariate Data Analysis. 5th ed. Upper Saddle River, New Jersey: Prentice Hall. p 143.
4. D.G. Long, Q.P. Remund, and D.L. Daum. 1999. A Cloud-Removal Algorithm for SSM/I Data. . *IEEE Trans. Geosci. Remote Sensing*, 37: 1 – 8.

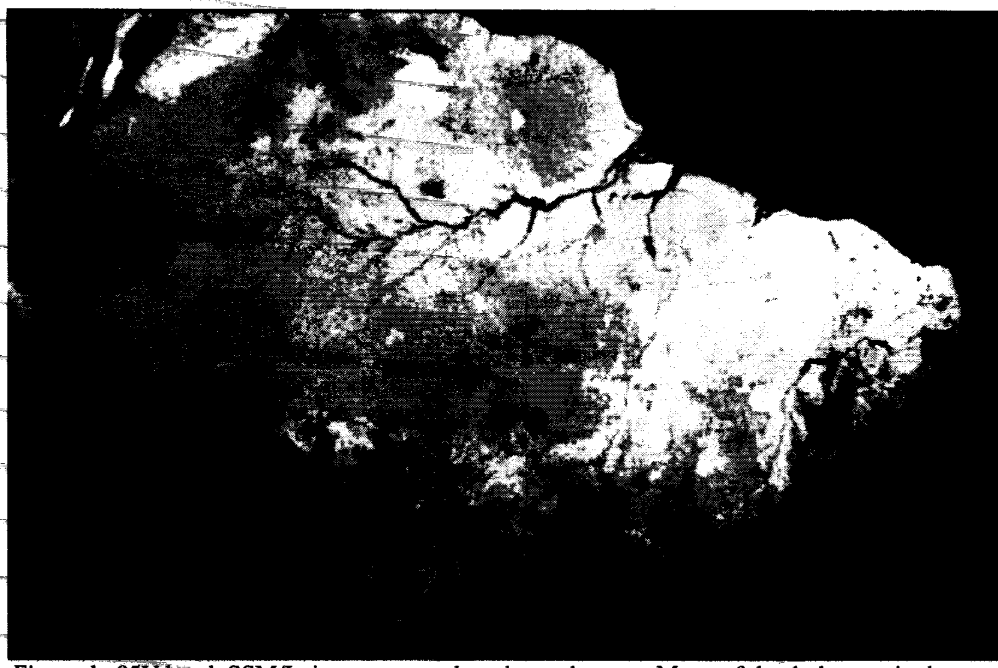

Figure 1. 85V band, SSM/I picture centered on the study area. Many of the dark spots in the Amazon basin are clouds.

Cloud removal algorithm

Band	Second highest		MMA estimate		None	
	Mean	Stdv	Mean	Stdv	Mean	Stdv
19H	284.8	9.2	283.5	9.6	281.7	10.2
19V	288.2	5.5	287.2	5.6	285.8	5.8
22V	287.1	3.2	286.2	3.3	284.8	3.3
37H	281.1	9.3	279.9	9.8	275.0	10.8
37V	284.3	5.2	283.3	5.4	281.9	5.7
85 H	286.6	3.1	285.4	3.4	283.3	4.4
85 V	287.6	1.9	286.6	1.9	284.9	2.6

Table 1. Overall image brightness values. Units are Kelvin. Rainfall removal has increased the mean brightness and decreased the variance in the images.

Rainfall variable

Band	Total days of rainfall	Total depth of rainfall
19V	12%	17%
19H	8%	15%
22V	15%	30%
37V	36%	42%
37H	28%	37%
85V	49%	50%
85H	45%	48%

Table 2. Coefficient of variation relating rainfall variables with the Second Highest (SH) difference Image

Rainfall variable

Band	Total days of rainfall	Total depth of rainfall
19V	20%	25%
19H	18%	19%
22V	22%	36%
37V	36%	44%
37H	25%	37%
85V	52%	55%
85H	47%	50%

Table 3. Coefficient of variation relating rainfall variables with the Modified Maximum Average (MMA) difference Image.