

Cloud Detection from Satellite Imagery: a Comparison of Expert-Generated and Automatically-Generated Decision Trees

Smadar Shiffman
Computational Sciences Division
NASA Ames Research Center / QSS
Moffett Field, CA 94303

Abstract

Automated cloud detection and tracking is an important step in assessing global climate change via remote sensing. Cloud masks, which indicate whether individual pixels depict clouds, are included in many of the data products that are based on data acquired on-board earth satellites. Many cloud-mask algorithms have the form of decision trees, which employ sequential tests that scientists designed based on empirical astrophysics studies and astrophysics simulations. Limitations of existing cloud masks restrict our ability to accurately track changes in cloud patterns over time. In this study we explored the potential benefits of automatically-learned decision trees for detecting clouds from images acquired using the Advanced Very High Resolution Radiometer (AVHRR) instrument on board the NOAA-14 weather satellite of the National Oceanic and Atmospheric Administration. We constructed three decision trees for a sample of 8km-daily AVHRR data from 2000 using a decision-tree learning procedure provided within MATLAB[®], and compared the accuracy of the decision trees to the accuracy of the cloud mask. We used ground observations collected by the National Aeronautics and Space Administration Clouds and the Earth's Radiant Energy Systems S'COOL project as the gold standard. For the sample data, the accuracy of automatically learned decision trees was greater than the accuracy of the cloud masks included in the AVHRR data product.

1. Introduction

Understanding the role of clouds in the current climate is prerequisite for predicting future climate change due to human activities (Wielicki et al., 1995). The National Oceanic and Atmospheric Administration (NOAA) polar-orbiting satellites provide observations of the earth's oceans, lands, and atmosphere, which are used by scientists to study long-term weather patterns and for weather forecasting. The satellites carry a suite of instruments that measure parameters of the Earth's surface, atmosphere, and cloud cover. For example, the

NOAA-14 satellite carries the Advanced Very High Resolution Radiometer (AVHRR) instrument. The data acquired by the satellites is packaged in a variety of data products of different spatial, temporal, and spectral resolutions. The data products are distributed via the Goddard earth sciences distributed active archive center.

Cloud masks have been included in data products produced from satellite radiometry since its early days. Scientists use cloud masks both to identify surface and atmospheric data of compromised quality due to cloud interference and to describe clouds and their properties. Cloud masks indicate, for a given location on the earth, the presence or absence of clouds, and measure characteristics of clouds when those are present. The cloud-masks are computed from measured reflectance and emission values using algorithms that scientists designed based on acquisition parameters, on simulated clear-sky and cloud characteristics for a variety of surface and atmospheric condition, and on analysis of ambiguous manifestation of different physical phenomena, for example, similar reflectance values for snow, ice and clouds. The algorithms employ sequential threshold tests to arrive at a decision about the presence of clouds or about cloud composition (Stowe et al., 1999, Trepte et al., 2002). Thus, the algorithms are essentially decision trees. The limitations of existing cloud masks (Stroeve, 2002) generated motivation for on-going research to develop improved cloud detection and characterization algorithms.

Cloud detection and characterization is a challenging task. Cloud-detection algorithms must disambiguate clouds and other entities that have similar manifestation as clouds. Entities whose appearance in satellite imagery may be similar to that of clouds differ from region to region. In the polar region, clouds and snow/ice are difficult to differentiate because all three entities are reflective in the visible wavelengths and demonstrate little contrast in the thermal infrared. Sun glitter may interfere with cloud detection in the tropics

due to spatially unresolved water bodies, or recent rainfall. Over volcanic areas clouds and volcanic ash may appear similar in the visible wavelengths. Over the desert, clouds and dust may appear similar. Over forests, clouds and fire may appear similar. Terrain shadows may also interfere with cloud detection in the tropics.

Scientists used a variety of machine-learning approaches to process remote sensing data, for example, neural networks (Promcharoen et al., 1999), Bayesian classification (Murtagh et al., 2003), support vector machines (Lee et al., 2004), genetic algorithms (Davis et al., 2001), and decision trees (Hansen et al., 1996). The results of these approaches range from promising preliminary results to validated algorithms that are deployed in high-level remote sensing data products (Zhan et al., 2002). The goal of this work was to explore the benefits of automatically-learned decision trees for cloud detection, and to determine whether decision trees that are based on functional relationships between sensed data that were determined theoretically performed better than decision trees that were based on the sensor data alone. We compare cloud detection results for the CLAVR expert-generated decision tree (Stowe et al., 1999), which is currently deployed in the NOAA-14 AVHRR daily 8km global data products, to results of three automatically-learned decision trees based on various degrees of physics modeling. The next section presents the cloud mask that is produced with an expert-generated decision tree, and discusses the limitation of the cloud mask. The section also lists the challenges in evaluating the results of cloud detection methods. Section 3 describes the methods we used, Section 4 reports the results of our work, and Section 5 includes a discussion of our findings. Section 6 presents our conclusions.

2. Background

The NOAA-14 AVHRR daily 8km global data product includes 12 scientific datasets (SDSs), each of which incorporates a measured parameter, flag, or computed parameter within a single plane. The SDSs are: a normalized difference vegetation index, the CLAVR cloud mask, a quality control flag, scan angle, solar zenith angle, relative azimuth angle, surface reflectance in the visible wavelengths (channel 1), surface reflectance in the near-infrared wavelengths (channel 2), and surface brightness temperature in the thermal infrared wavelengths (channels 3-5), and acquisition day and time (Agbu and James, 1994). The CLAVR

algorithm performs a series of threshold and uniformity tests on a 2x2 array of pixels, and classifies pixels as either *clear*, *mixed*, or *cloudy*. The values used for each test are either retrieved channel values, or functions of retrieved values that incorporate acquisition parameters and estimates of emitted radiances (Stowe et al., 1999). The thresholds used for the tests were derived empirically or via simulations of a variety of cloud-surface-daytime observation conditions.

The sequential decision process in CLAVR is designed to discriminate between clouds first by their gross characteristics, and then by their subtle characteristics. The algorithm ensures that pixels that fail all the tests have a very small probability of having radiatively significant clouds. The algorithm includes tests that are designed specifically to resolve previously encountered ambiguities, for example, ambiguities due to reflectance greater than 44% in channel 1 or channel 2 for snow, ice, or sun glint. CLAVR includes four decision trees, each for one of daytime land scene, daytime ocean scene, nighttime land scene, and nighttime ocean scene. Figure ## (a) displays the decision tree for daytime land scenes.

The CLAVR cloud mask has several limitations [### Check if Stroeve can be referenced here###]. First, the mask assumes that there is a representative sample of clear pixels in each image, however, this assumption does not hold when clouds are persistent at a single pixel coordinate. Second, the CLAVR algorithm may produce different results for a given pixel based on the neighborhood to which the pixel belongs. Because the class that the algorithm assigns to a given pixel depends on the uniformity of neighboring pixels, the class of a pixel may differ if the pixel is grouped with pixels on the left, right, above, or below. The ability of CLAVR to differentiate between clouds and other entities that appear as clouds in AVHRR images is limited.

The evaluation of cloud masks is difficult because there is no gold standard to which to compare the cloud mask. Researchers estimate the quality of cloud masks by comparing their agreement with masks produced by human analysis or by other algorithms. Stowe and colleagues (Stowe et al., 1999) compared the results of CLAVR to classification results of a human-expert analyst. In general, for small cloud amounts, CLAVR overestimated fractional amounts by 0.1 compared to the analysis interpretation, and for large cloud amounts, CLAVR underestimated the cloud amount by about 0.1. The evaluation showed larger errors for certain geographical locations and seasons. In a recent study, Thomas and Heidinger (Thomas and Heidinger, 2004)

showed that cloud amounts that resulted from recent improvements in CLAVR were in agreement with cloud amounts from established satellite-derived cloud climatologies.

The CLAVR algorithm is similar to automatically-learned decision trees in that it employs a structure of sequential threshold-tests. However, the test sequence and thresholds in CLAVR were derived by scientists via theoretical and empirical analysis of specific AVHRR data (radiances from individual channels, or acquisition parameters), and not via analysis of the data space as a whole. In the next section we describe our work on learning decision trees from AVHRR data and comparing the cloud-masks that they produce to ground observations performed by humans in multiple locations around the earth.

3. Methods

We obtained ground observations from the NASA Langley Atmospheric Sciences Data Center CERES S'COOL Project (Chambers et al. 2004). High-school students from many geographical locations around the world recorded the S'COOL observations and reported the observation using a well defined protocol whose goal was to provide sufficient information for validation of measurements taken by the Clouds and the Earth's Radiant Energy System (CERES) instruments on NASA's Earth Observing System satellites. Among the recorded data were date and time of observation, longitude and latitude, cloud observations at low, mid and high altitudes (categorical

variables: cloud type, visual opacity; ordinal variables: percent cloud cover), and surface cover characteristics.

We selected all observations that were available for the year 2000. We then retrieved 8km daily AVHRR data that matched in acquisition date and in longitude and latitude. We excluded from this dataset all points for which the data quality flag indicated out-of-range values or processing errors (about 20 percent of the data points) and obtained 2869 data points. We used the S'COOL project *cloud-present/no-cloud-present* observations for a given date and location as the gold standard, for labeling training data and for comparing to results of classified test data. In addition, we compared the S'COOL project observations to the AVHRR cloud masks for the retrieved data points.

We performed three experiments with the AVHRR data that we selected. The experiments differed in the set of variables that constituted the input to the decision-tree learning procedure. Experiment I included the variables: observation identification, the radiances of channels 1 through 5, and the binary label that indicated the presence or absence of clouds obtained from the S'COOL data. Experiment II included the same variables as those of Experiment I, as well as the latitude, longitude, the acquisition parameters: scan angle, solar azimuth angle, and relative azimuth angle, day of year and time of acquisition. Experiment III included the same variables as those of Experiment II, as well as three additional computed variables that are used within the CLAVR daytime-land algorithm (Stowe et al., 1999). Table 1 describes how the computed variables relate to sensed data and to acquisition parameters.

Table 1 Variables computed from sensed data and acquisition parameters

Test Name	Test Description	Variables	Function
Reflectance ratio cloud test (RRCT)	Examines the ratio of Channel 1 and Channel 2 reflectance	R_1 - Channel 1 reflectance R_2 - Channel 2 reflectance	$RRC = \frac{R_2}{R_1}$
Channel 3 albedo test (C3AT)	Extracts the reflectance component of the mixed Channel 3 signal	Spacecraft- dependent coefficients: a, b, c, d D_0 - Earth sun distance D - mean Earth-sun distance B - Planck blackbody radiance function v_0 - Channel 3 central wave number T_i - observed equivalent blackbody temperature in channel i	$C3A = \frac{3.14159\Delta R_3 \cdot 100}{\cos(Z_0)(D_o / D)^2}$ $\Delta R_3 = B(T_3) - B(T_{3e})$ $T_{3e} = -\left[\frac{(b/a)T_4 + (c/a)T_5 + d/a}{1}\right]$

		T_{3e} —estimated brightness temperature for Channel 3 due to emission only S_3 —Channel 3 filtered solar irradiance at normal incidence and mean Earth-sun distance	
Four-minus-five test (FMFT)	Examines the Channel 4 – Channel 5 brightness temperature difference	T_4 – Channel 4 brightness temperature T_5 – Channel 5 brightness temperature	$FMF = T_4 - T_5$

For each experiment we ran 100 trials, in which we partitioned the data into a training set and a test set in a random manner (9:1 ratio of training to test set size). For each trial we used the training set to learn a decision tree using the *treefit* procedure available within the MATLAB[®] statistics toolbox. We then classified the data in the corresponding tests set as *clear* or *cloudy* using the decision tree. We compared the classification results to the S’COOL observations for matching date and location. We computed the mean classification mismatch between the results of the decision trees and the S’COOL observations, and between the CLAVR cloud mask and the S’COOL observations¹, and ran a two-sided paired t-test to determine if there was a significant difference between rate of classification mismatches for CLAVR and for each of the three decision trees.

4. Results

The mean and standard deviation classification mismatches for each trial appear in Table 2. A two-sided paired *t*-test showed that the differences in mean misclassification mismatch between CLAVR and each of the decision trees was significant $p < 0.025$ in all three cases. The differences in the mean classification mismatch between each pair of decision trees were not significant.

Table 2 Summary statistics for classification mismatch in the 100 trials of each experiment.

Method and experiment\Statistic	Mean	Standard Deviation
CLAVR I	0.216921	0.023423
Decision Tree I	0.159299	0.020823
CLAVR II	0.214516	0.021428
Decision Tree II	0.157303	0.021721
CLAVR III	0.211762	0.021566
Decision Tree III	0.155276	0.018263

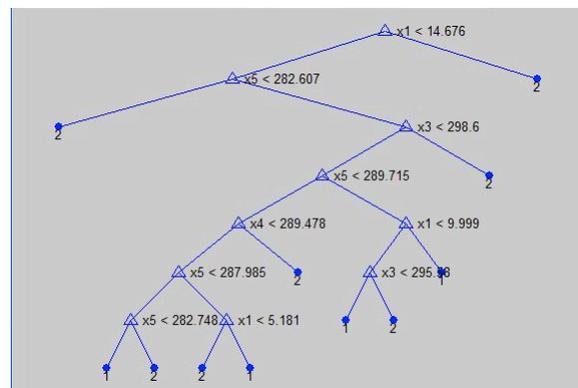


Figure 1 Decision tree...

¹ Although both CLAVR and the S’COOL project utilize an ordinal scale for characterization of cloud amount, the scales are not identical and mapping one scale to the other can be done in more than one way. Consequently, we mapped the CLAVR scale to a binary variable with values *clear* and *cloudy*, which we then could compare to the S’COOL project binary variable with values *cloud-present/no-cloud-present*.

1. Discussion
 - 1.1 Sampling bias
 - 1.2 Quality of gold standard data
 - 1.3 Challenges in evaluation in general
 - 1.4 Binary classification vs. multicategory
 - 1.5 Similarities in expert-generated tree and automated tree
 - 1.6 Implication of no difference among trees
2. Conclusion

Acknowledgements

These data were obtained from the NASA Langley Research Center Atmospheric Sciences Data Center as part of the CERES S'COOL Project."

References

Wielicki, B.A., Harrison, E.F., Cess, R.D., King, M.D., Randall, D.A., Mission to Planet Earth: Role of Clouds and Radiation in Climate, Bulletin of the American Meteorological Society, vol.76, Issue 11, pp.2125-2154, 1995.

Stowe, L. L., P. A. Davis, and E. P. McClain, 1999: Scientific Basis and Initial Evaluation of the CLAVR-1 Global Clear/Cloud Classification Algorithm for the Advanced Very High Resolution Radiometer. *J. Atmos. & Oceanic Tech.*, 16, 6, 656-681.

J. Stroeve, "Assessment of greenland albedo variability from the avhrr polar pathfinder data set," *Journal of Geophysical Research*, vol. 33, pp. 989–1034, 2002.

A.B. Davis, S.P. Brumby, N.R. Harvey, K. Lewis Hirsch, and C.A. Rohde, Genetic refinement of cloud-masking algorithms for the multi-spectral thermal imager (MTI)" in Proc. IGARSS 2001, Sydney, Australia, 9-13 July 2001.

Trepte, Q. Z., P. Minnis, and R. F. Arduini, Daytime and nighttime polar cloud and snow identification using MODIS. *Proc. SPIE Conf. Optical Remote Sens. of Atmosphere and Clouds*, Hangzhou, China, Oct. 23-27, 2002.

Hansen, M., Dubayah, R., & DeFries, R. (1996). Classification trees: an alternative to traditional land cover classifiers. *International Journal of Remote Sensing*, 17, 1075– 1081.

F. Murtagh, D. Barreto and J. Marcello, Decision boundaries using Bayes factors: the case of cloud masks. *IEEE Transactions on Geoscience and Remote Sensing*, 14, 2952-2958, 2003.

Lee, Y., Wahba, G., Ackerman, S. A., Cloud classification of satellite radiance data by multicategory support vector machines, *Journal of Atmospheric and Oceanic Technology* Vol. 21 No. 2 pp. 159-169, 2004.

Promcharoen, S., Rangsanseri, Y., Suwit Ongsomwang, S., Jarupatt, J., Supervised Classification of Multispectral Satellite Images using Fuzzy Logic and Neural Network, Proceeding of the 20th Asian Conference on Remote Sensing, November 22-25, 1999, Hong Kong, China.

Zhan, X., Sohlberg, R.A, Townshend, J.R.G, DiMiceli, C., Carroll M.L., Eastman, J.C., Hansen, M.C., DeFries, R.S. Detection of land cover changes using MODIS 250 m data. *Remote Sensing of Environment* (2002), pp. 336–350.

Agbu, P.A., and M.E. James. 1994. The NOAA/NASA Pathfinder AVHRR Land Data Set User's Manual. Goddard Distributed Active Archive Center, NASA, Goddard Space Flight Center, Greenbelt.

Thomas, S., A. K. Heidinger, 2004: Comparison of NOAA's Operational AVHRR Derived Cloud Amount to other Satellite Derived Cloud Climatologies. *in press to the Journal of Climate*.

Chambers, L.H, Costulis, P.K., Young, D., and Rogerson T.M., Students as Ground Observers for Satellite Cloud Retrieval Validation (Invited Presentation), 13th Conference on Satellite Meteorology and Oceanography, Sep 20-23, 2004, pp. Norfolk, VA.