Development of the Landsat Data Continuity Mission Cloud-Cover Assessment Algorithms

Pasquale L. Scaramuzza, Michelle A. Bouchard, and John L. Dwyer

***Abstract*—The upcoming launch of the Operational Land Imager (OLI) will start the next era of the Landsat program. However, the Automated Cloud-Cover Assessment (CCA) (ACCA) algorithm used on Landsat 7 requires a thermal band and is thus not suited for OLI. There will be a thermal instrument on the Landsat Data Continuity Mission (LDCM)—the Thermal Infrared Sensor—which may not be available during all OLI collections. This illustrates a need for CCA for LDCM in the absence of thermal data. To research possibilities for full-resolution OLI cloud assessment, a global data set of 207 Landsat 7 scenes with manually generated cloud masks was created. It was used to evaluate the ACCA algorithm, showing that the algorithm correctly classified 79.9% of a standard test subset of $3.95\ 10^9$ pixels. The data set was also used to develop and validate two successor algorithms for use with OLI data—one derived from an off-the-shelf machine learning package and one based on ACCA but enhanced by a simple neural network. These comprehensive CCA algorithms were shown to correctly classify pixels as cloudy or clear 88.5% and 89.7% of the time, respectively.**

***Index Terms*—Algorithm, clouds, image classification, Landsat, remote sensing.**

I. Introduction

THE LANDSAT Program is a series of moderate-resolution Earth observing satellites that have provided an archive of multispectral imagery since 1972. With the launch of Landsat 7 in 1999, the Landsat archive became fully global by following the archive strategy called the Long Term Acquisition Plan (LTAP) [1]. The goal of the LTAP is to build a global archive of seasonally refreshed cloud-free scenes. Cloud-cover assessment (CCA) and cloud avoidance is an integral part of the LTAP scheduler, as cloudy scenes must be revisited. With the Enhanced Thematic Mapper (TM) Plus (ETM+) on Landsat 7, an algorithm known as Automated CCA (ACCA) [2] has been used to calculate a scene-wide cloud score for the LTAP [3]. The ACCA cloud score has also been one of the most important considerations for users searching for Landsat data [4].

The next Landsat program is the Landsat Data Continuity Mission (LDCM), which is scheduled for launch in 2012 under the management of the U.S. Geological Survey (USGS). The LDCM will have an advanced pushbroom-type sensor known as the Operational Land Imager (OLI). LDCM is planned to have temporal revisit specifications similar to those of Landsat 7, making it necessary for the program to have an LTAP scheduler with reliable CCA [3], [5]. More accurate cloud assessment will lead to increased utility of not only individual scenes but also those for the global image archive [6].

There are several challenges facing LDCM CCA. First and foremost is the potential absence of a thermal band. Current CCAs for Landsats 5 and 7 rely heavily on the use of thermal data [2]. A requirement for thermal data was left out of the original LDCM specification, but the option for an additional thermal instrument to be flown on the same spacecraft was left open [5]. NASA has exercised that option, and they are developing a thermal instrument known as the Thermal Infrared Sensor (TIRS) to be included on LDCM. TIRS has only a three-year design life, and OLI will have the capability to collect data without TIRS in the event of schedule conflicts or problems with the thermal instrument. This creates a requirement for nonthermal CCA of OLI data in the event that TIRS data are unavailable. This paper addresses that requirement.

The second challenge for LDCM CCA is the creation of full-resolution cloud masks with a cloud/clear designation for each pixel in the image. Masks were not created for previous Landsat data. With Landsat 7, an ACCA score estimating the percentage of cloudy pixels is assigned for each scene and for each image quadrant [4]. For LDCM, every image will have a cloud mask matching the resolution of the reflective bands (30-m pixels) that will distinguish between cloudy and clear pixels.

The third challenge is processing time. The LDCM ground system has a goal of producing user-orderable data within 2 h of acquisition by the satellite, which limits the allowable processing time of all components of the system, including CCA. The program specifications for the LDCM CCA algorithm originally stated that it must run in under 2 min of processing time. This constraint has been relaxed and will certainly be loosened further in the future as the speed of the processing systems improve. It should be kept in mind, however, as a reminder that the run time of the CCA algorithms will impact the throughput performance of the entire LDCM ground system, and thus, it must be as quick as possible.

A fourth challenge is the availability of data. Although the ETM+ spectral bands are analogous to some of the OLI bands, OLI has two additional bands—the coastal aerosol and cirrus bands—that have no analog in previous Landsats but which may be useful for CCA [7]. Additionally, TIRS will provide two bands in the thermal IR that do not correlate well with the Landsat thermal band. Although there are workarounds for all of these issues, for this paper, we have focused our studies

Manuscript received October 26, 2010; revised March 18, 2011 and June 29, 2011; accepted July 17, 2011. Date of publication September 15, 2011; date of current version March 28, 2012. This work was supported by the U.S. Geological Survey Earth Resources Observation Systems Data Center. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

P. L. Scaramuzza is with Stinger Ghaffarian Technologies, Greenbelt, MD 20770 USA (e-mail: pscar@usgs.gov).

M. A. Bouchard and J. L. Dwyer are with the U.S. Geological Survey Earth Resources Observation and Science Center, Sioux Falls, SD 57198 USA (e-mail: mbouchard@usgs.gov; dwyer@usgs.gov).

Digital Object Identifier 10.1109/TGRS.2011.2164087

TABLE I
COMPARISON BETWEEN LANDSAT 7 ETM+ AND LDCM OLI AND TIRS BANDS

ETM+ Band Number	ETM+ Band Wavelengths	OLI / TIRS Band Number	OLI / TIRS Band Wavelengths	Band Name
		Band 1	0.433 – 0.453 μm	Coastal / Aerosol
Band 1	0.450 – 0.515 μm	Band 2	0.450 – 0.515 μm	Blue
Band 2	0.525 – 0.605 μm	Band 3	0.525 – 0.600 μm	Green
Band 3	0.630 – 0.690 μm	Band 4	0.630 – 0.680 μm	Red
Band 4	0.775 – 0.900 μm	Band 5	0.845 – 0.885 μm	Near-IR
Band 5	1.550 – 1.750 μm	Band 6	1.560 – 1.660 μm	SWIR-1
Band 7	2.090 – 2.350 μm	Band 7	2.100 – 2.300 μm	SWIR-2
Band 8	0.520 – 0.900 μm	Band 8	0.500 – 0.680 μm	Pan
		Band 9	1.360 – 1.390 μm	Cirrus
Band 6	10.00 – 12.50 μm			ETM+ LWIR
		Band 10	10.30 – 11.30 μm	TIRS LWIR-1
		Band 11	11.50 – 12.50 μm	TIRS LWIR-2

on Landsat-like data sets, and thus, data from the additional OLI bands and the TIRS instrument are not being considered at this time. Future work will expand on these studies, and it is expected that, after the launch of LDCM, the availability of OLI and TIRS data will enable the creation of new more accurate cloud assessment algorithms (Table I).

The scope of this paper, therefore, is to describe the development of a global algorithm for LDCM CCA, with accuracy similar to those of past Landsat algorithms [4], using only the Landsat-like reflective bands while creating a per-pixel cloud mask, all with only a few minutes of run time.

This paper describes several CCA algorithms that have been designed to work within that narrow scope. That scope should be kept in mind when examining the results. The LDCM CCA masks are intended for use by LTAP and also as a benefit to typical users of Level 1 satellite data for image selection and simple cloud masking. Users of LDCM data who require extremely accurate cloud masks to aid in compositing will require Level 2 cloud masks, which are outside the scope of this paper and of the current USGS scene processing strategy.

II. TEST SET CREATION AND ACCA VALIDATION

The ACCA algorithm is used for cloud assessment of TM and ETM+ data. It was developed by the Landsat Project Science Office at NASA's Goddard Space Flight Center (GSFC). ACCA provides scene-averaged cloud-cover scores which have been validated to within a precision of $\pm 5\%$ for 98% of scenes in the test set studied [8].

When ACCA creates scene-wide cloud-cover scores, it first performs a spectral cloud identification of each pixel to assign preliminary classifications and to collect statistical information about the scene. The spectral classification is performed on the top-of-atmosphere (TOA) reflectance of TM/ETM+ bands 2–5 and the thermal brightness temperature derived from TM band 6 or ETM+ low gain band 6. This pass-1 decision tree can be used to create a per-pixel cloud mask, although many pixels are marked as "ambiguous" in this stage.

Pass-2 of the ACCA algorithm calculates a cloud signature for the scene using aggregate statistics from pass-1. That expected cloud signature is then used to create a thermal threshold test which further classifies the ambiguous pixels. Pass-3 of ACCA then aggregates the pass-1 and pass-2 results and performs a simple hole-filling algorithm. The output of these later passes is a scene-wide cloud-cover score. Because of the reliance upon thermal data and the single score result, pass-2 and pass-3 of the ACCA algorithm are uninteresting from the perspective of the per-pixel masks we intend to create for reflective band LDCM CCA.

Fig. 1 shows the pass-1 ACCA decision tree. The test thresholds are based upon variables such as "B3" for Band 3 reflectance or thresholds based upon ratios of bands. The normalized

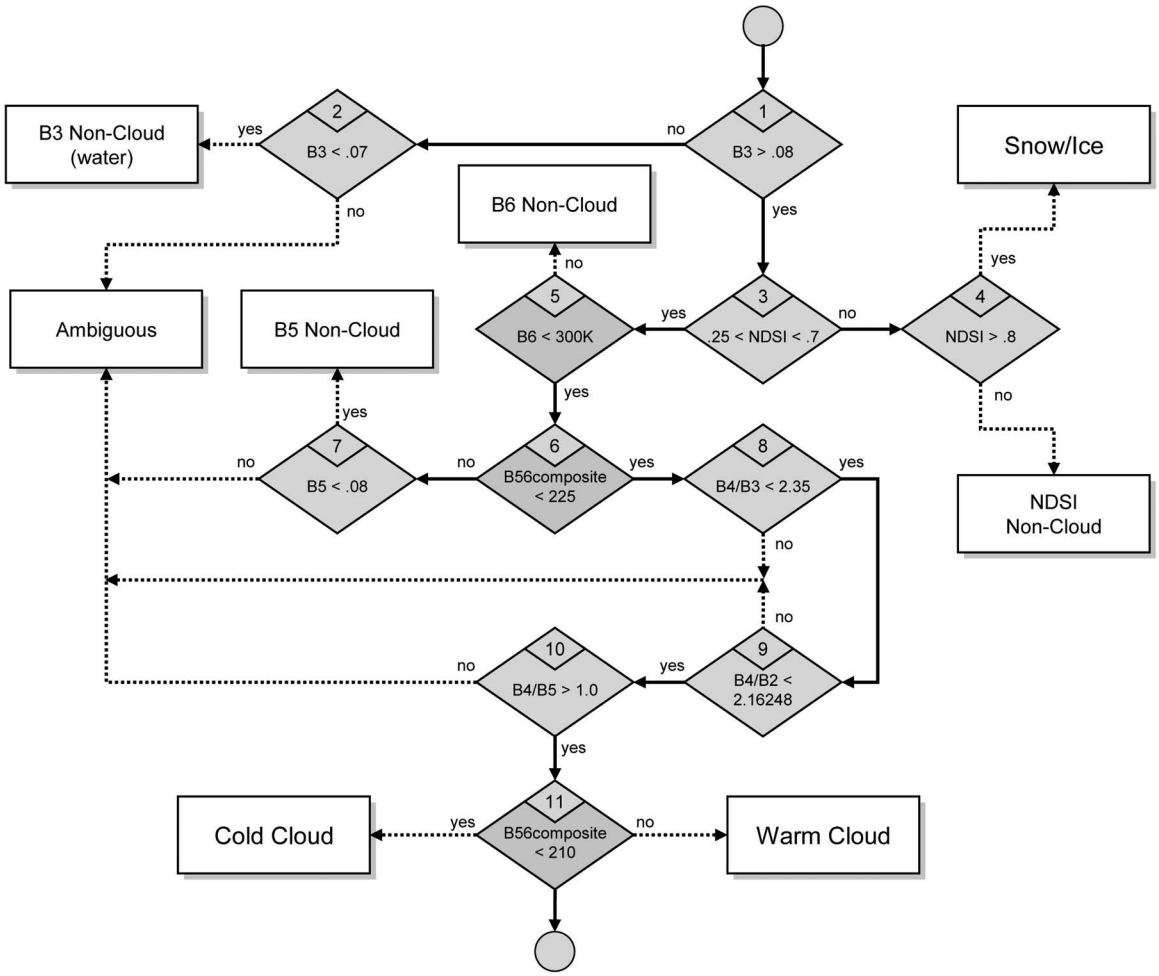


Fig. 1. ACCA pass-1 flowchart.

difference snow index (NDSI) is a normalized difference of bands 2 and 5 [9]. The "B56 composite" test is a threshold test of the value (1-(B5 reflectance))∗(B6 brightness temperature). Full details on ACCA are available in the literature [8] and the Landsat Science Data Users' Handbook [10]. The first challenge to using ACCA for OLI data is that three of its 11 threshold tests use the thermal band. To use pass-1 of the ACCA algorithm as a basis for an LDCM CCA algorithm, it was necessary to validate it and quantify its reliance upon thermal data.

Because the OLI bands 2–7 are similar to ETM+ reflective bands 1–5 and 7, because we needed a Landsat data set with which to validate the ACCA algorithm, and because Landsat data are available for no cost through the USGS Landsat archive, we chose to amass a Landsat 7 data set for use as training and validation data for LDCM CCA algorithms. Another factor that led us to this approach was the availability of a standard test set from NASA GSFC. This data set consists of 212 ETM+ Level 1G images, all taken in the years 2000–2001, and is divided into nine latitude zones with 24 scenes per zone (save for the south polar zone, which has only 20 scenes). Fig. 2 shows the geographic distribution of the GSFC test data set. This test set is the same data set used to validate the Landsat 7 ACCA algorithm [8]. However, that validation was done with manually generated cloud masks resampled to the Landsat browse image size—825 × 750 pixels, with no geometric resampling. The first step necessary to use the GSFC test data as a validation data set was to create full resolution manually generated cloud masks.

A. VCCA Procedure

Three image analysts performed manual assessment of the scenes in the GSFC test set. The Visual CCA (VCCA) process involved opening each full-resolution image as an RGB image in Adobe Photoshop. The bands used for the RGB image varied by the scene and by the analysts' preference. For some images where clouds were indistinct in the reflective bands, the thermal band was resampled to match the reflective bands and used as an overlay on the RGB image. The analysts then used Photoshop functions including (but not limited to) the magic wand tool, the freehand lasso tool, and the Select − > Color Range function to isolate clouds. Two levels of clouds were identified: thick and thin. Cloud pixels were labeled as thin if they were transparent but still visually identifiable as clouds, with an estimated opacity of 50% or more. In some cases where fog, blowing snow, or jet contrails caused confusion between clouds and terrain, the analysts were instructed to label pixels as thin if they were 50% certain that the pixel contained a cloud. In general, these subjective opacity guidelines caused the manual assessment to be conservative in the labeling of thin clouds.

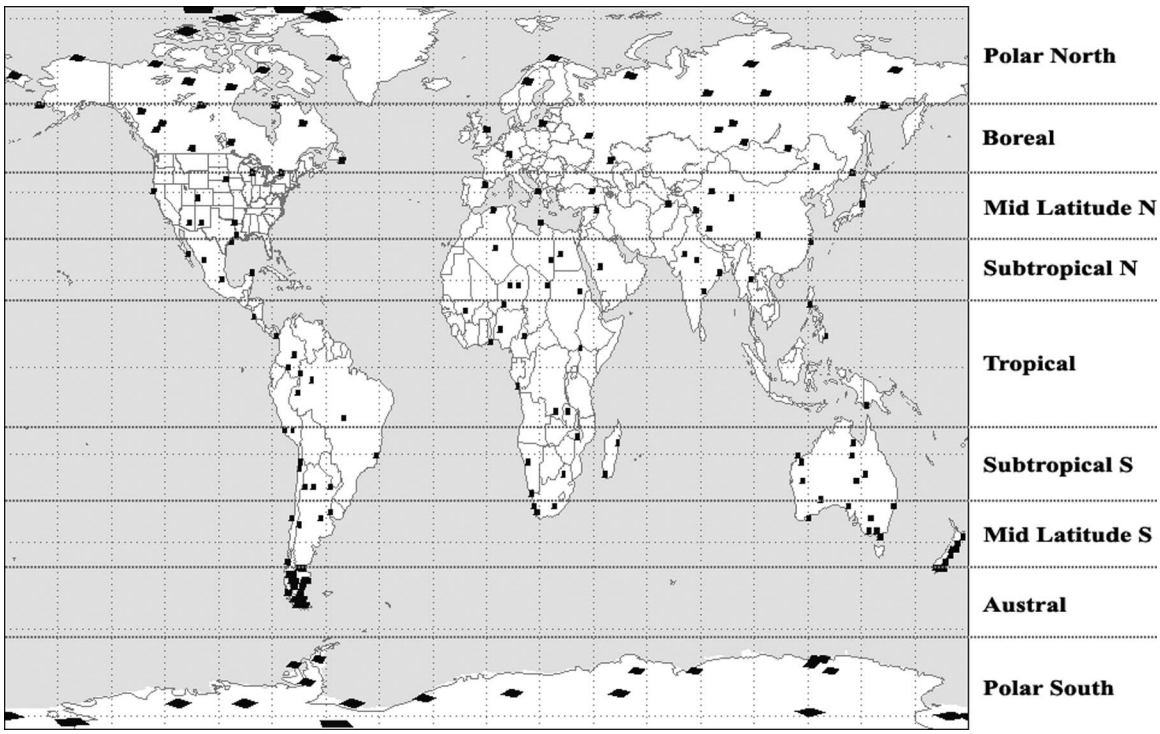
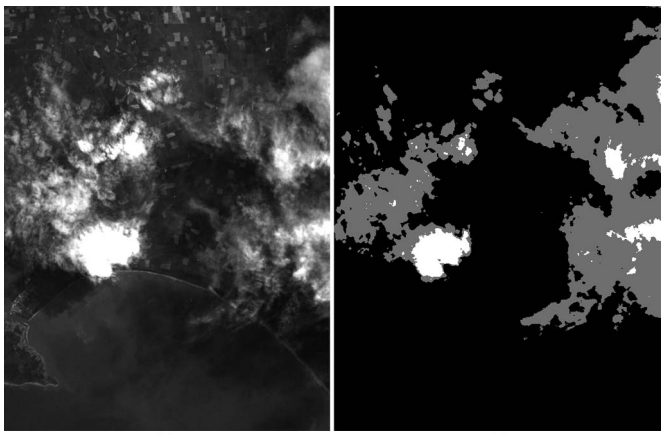


Fig. 2. Geographic depiction of the 212 test images in the GSFC test data set.



(a) (b)

Fig. 3. (a) Band 1 imagery. (b) VCCA cloud mask of the same region (white indicates thick clouds, gray indicates thin clouds, and black indicates clear).

Five scenes were excluded for containing gain change artifacts, a common occurrence on Landsat 7 [8], [11]. A sixth gain change scene was retained as no artifacts were visible. This culling left 207 scenes, each with an associated full-resolution manual truth mask. Fig. 3, image B, shows an example of a VCCA cloud mask.

For a subset of 11 images, all three analysts performed VCCA so that their results could be compared against each other. The mean greatest difference between analysts is 6.75% (Table II). One scene had a large difference (24.5%) in manual CCA due to very thin clouds and fog. Based on this simple analysis, the manual truth images are estimated to have approximately 7% error due to subjective differences in operator methodology.

B. Validation Methodology and ACCA Validation Results

The GSFC data set was split into two subsets, one for training purposes and one reserved for algorithm validation. The subsetting was performed by tagging each image with the following descriptive tags:

heavyclouds	The VCCA score for this scene was greater than 75%.
cloudfree	The VCCA score for this scene was less than 1%.
shadows	Sharp distinct cloud shadows exist in this scene.
lowsun	Solar zenith angle of 65° or greater.

In addition, each image was processed through pass-1 of the ACCA algorithm. The ACCA pass-1 results were then examined, and further discrimination of the data was performed using the following ACCA-based tags:

accafail	ACCA pass-1 differed from the VCCA score by more than 30%.
manyambig	ACCA pass-1 marked more than 30% of the scene pixels as ambiguous.
manyprovs	ACCA pass-1 marked more than 10% of the scene as provisional (or "warm") clouds.
falseclouds	ACCA pass-1 marked more than 10% of clear pixels in this scene as clouds.

After the scenes were tagged, they were partnered with scenes within the same latitude zone with similar tags, and the scenes in each pair were then assigned to separate groups. Scenes with no comparable partner were assigned to a group at random. This created two groups in each latitude zone, with each group containing similar scenes. Collecting these groups together across all the zones gave us one group of 104 images that was designated as the training set and a second group of 103 images that became the validation set.

TABLE II
MANUAL CCA COMPARISONS

Zone	Scene	ACCA Score	Analyst VCCA score A	B	C	Range
Austral	p74_r92	21.7%	20.2%	22.2%	25.5%	5.2%
Boreal	p120_r28	32.2%	24.7%	25.6%	25.7%	1.0%
Boreal	p54_r19	52.6%	52.6%	60.8%	61.1%	8.5%
Mid-latitude_N	p33_r37	4.9%	2.7%	2.9%	2.6%	0.3%
Polar_N	p15_r10	82.1%	93.0%	91.9%	68.6%	24.4%
Polar_N	p35_r15	11.2%	15.6%	21.3%	13.0%	8.4%
Subtropical_N	p139_r45	44.4%	24.1%	17.1%	13.2%	10.9%
Subtropical_S	p113_r75	18.6%	21.7%	25.3%	22.4%	3.6%
Subtropical_S	p232_r79	1.1%	0.5%	0.9%	0.5%	0.5%
Tropical	p170_r57	14.4%	12.3%	19.1%	13.3%	6.8%
Tropical	p4_r61	87.3%	41.2%	42.0%	37.3%	4.7%

TABLE III
CCA ALGORITHM RESULTS OVER THE GSFC VALIDATION DATA SET

	Algorithm Name	Overall Accuracy (% of pixels correct)	Misclassified rate (% of pixels incorrect)	Ambiguous rate (% of pixels ambiguous)	Mis-classified clouds	Mis-classified clears
Reference Algorithms:	ACCA pass 1	79.9%	4.3%	15.8%	7.8%	2.3%
	FT-ACCA	68.7%	4.2%	27.1%	7.9%	2.1%
Primary Algorithms:	AT-ACCA	76.6%	5.5%	17.9%	7.8%	4.2%
	C5 CCA	88.5%	7.3%	4.2%	12.1%	4.5%
	Expanded AT-ACCA (AT-ACCA + gD02)	89.7%	8.5%	1.8%	12.4%	6.3%
Disambiguations:	gD02 (standalone)	84.2%	9.8%	6.0%	18.7%	4.8%

We planned that the validation procedure we would develop would be a pixel-by-pixel comparison of the VCCA cloud mask to the cloud mask created by each CCA. The 103 images in the validation subset constitute $3.95\ 10^9$ pixels, all of which are examined in the validation procedure. (The exact number of pixels in the validation set varies by algorithm, as different algorithms use different bands and pixels are excluded if they contain fill in any of the bands used. The number of pixels used by each algorithm varies by less than 0.3%.)

The pass-1 ACCA algorithm correctly classified 79.9% of pixels in the validation subset. The Overall Accuracy values [12] of the ACCA algorithm validation results are shown in Table III, as a reference to the CCA algorithms being developed for LDCM. The other algorithms listed in Table II will be discussed later in this paper. All of them were validated against the same subset of the GSFC test data.

C. ACCA Results Without the Thermal Band

To examine the importance of the thermal band to ACCA, a version of the algorithm was tested using a fixed brightness temperature of 288 K, approximately the global annual mean. (This is equivalent to removing ACCA test #5, the thermal-only test, and allowing the B56 composite tests #6 and #11 to remain as B5 threshold tests. Any fixed value less than 300 K would have a similar effect.) This "fixed temperature ACCA," or FT-ACCA, was run on the same 103 scene validation subset as the ACCA validation (Table III). This allowed for a comparison of the same ACCA algorithm with and without thermal input. The error arising from the loss of the thermal band is approximately 10%, which agrees with previous studies of the importance of thermal data to the ACCA algorithm [2]. The errors in FT-ACCA did not correlate to any of the

latitude zones in the GSFC test set, leading us to the conclusion that variance in cloud temperature—which occurs globally—is much more important to ACCA than variance in ground surface temperature.

III. LDCM CCA Algorithms

Two different approaches were used to create LDCM CCA algorithms. The first approach involved expanding the ACCA pass-1 algorithm to eliminate its need for thermal data, increase its accuracy, and reduce the number of pixels marked as ambiguous. This was done through empirical modeling of brightness temperature using the TOA reflectance of the reflective bands and by adding a simple neural network to reduce ambiguous results. The second CCA approach was wholly statistical and involved the creation of large decision trees using automated software designed to create models with minimal information entropy.

A. Development of the AT Band

A variant of the ACCA algorithm can be run under the assumption that Landsat 7 band 6 thermal data may be modeled using the reflective bands. The original concept for the Artificial Thermal (AT) band involved the observation that clouds are characterized by higher reflectance and lower temperature than either land or water. Thus, we sought a loose inverse relationship between thermal brightness temperature and TOA reflectance in Landsat 7 band 2, where the peak of the solar spectrum lies. The relationship between the reflective bands and thermal brightness temperature has been studied in past cloud-retrieval methods [13], [14].

For a first "proof of concept" approach, this relationship was examined using a simple in-house genetic algorithm and a readily available test set. The test set used consisted of 90 Landsat 7 scenes selected at random for daily trending in the USGS Image Assessment System [15] during November 2007–January 2008. These data were chosen because they were a random sampling of daytime scenes that was in-house and ready for us to access.

The genetic algorithm was given a seed equation—in the aforementioned example, an inverse relationship with B2. Then, additional arithmetic combinations—addition, division, or multiplication, along with a scaling parameter—of the Landsat 7 reflective bands were added at random until there were 20 permutations of the original equation. Then, the equations were all judged for fitness by measuring the root-mean-square (rms) difference between their output and the low gain band 6 brightness temperature over the test set. The three best performing equations graduated to another iteration of random combinations and fitness tests. After several iterations, the genetic algorithm converged upon one equation with the best fit. Several different seed equations were used, but the inverse relationship with B2 produced the best result.

The best fit equation for AT_{poc}—the "proof of concept" AT band—had the following form:

$$AT_{\mathrm{poc}} = \frac{1}{C_1 B2 + C_2} + C_3 B5 + C_4 \tag{1}$$

where

$B2$ = Landsat 7 band 2 TOA reflectance;
$B5$ = Landsat 7 band 5 TOA reflectance;
$C1$ = 0.1366;
$C2$ = 0.0087;
$C3$ = 36.5441;
$C4$ = 235.4310.

This simple equation, validated over the test set of 90 scenes or approximately 3 billion pixels, had an rms difference from Landsat 7 low gain band 6 brightness temperature of 16.2 K. This is a large difference, but it showed that attempting to model a thermal band using the reflective bands was feasible when considered within the generous constraints of accuracy required for LDCM CCA. The next step was to refine the AT band into a more accurate simulation of thermal data.

With a successful proof of concept, we began a more rigorous modeling approach using the commercial software Cubist, by RuleQuest Research. Cubist is a regression tool using the statistical classifier C5.0, which is similar to the Classification and Regression Tree (CART) methodology for creating decision trees. However, unlike CART, C5.0 is built on the ID3 model of minimizing information entropy in the test subset [16].

Because very large training sets can cause memory errors on computers running Cubist, a subset of training data was selected by randomly sampling 20 000 pixels from each of the 103 images in the GSFC test data set, for a total of 2 080 000 pixels. To avoid artifacts due to saturation, pixels that saturated in any of the reflective bands were discarded. Although this caused us to lose the very bright clouds, it only excluded 18% of the data set, leaving us with 1 706 670 pixels as a training subset for Cubist.

Initial runs of Cubist were run on several parameters including TOA reflectance of the Landsat 7 bands 1–5 and 7, their normalized differences, and the cosine of the scene center solar zenith angle. The least predictive parameters were winnowed out. Cubist was then used to derive two different AT band models: one with a decision tree containing up to 100 rules (the maximum in Cubist) and one constrained to a single rule. In the single rule case, the output of Cubist is a multivariate linear regression model.

The 100-rule AT model had an rms difference of 5.4 K from Landsat 7 low gain band 6 brightness temperature over the 103 image validation subset. However, the 100 rule decision tree created step functions in the AT band [Fig. 4(b)], and there was a concern that these discontinuities would unfairly bias a cloud assessment algorithm and cause artifacts in the cloud mask. The 100-rule model was also significantly slower to run than the single-rule algorithm [Fig. 4(c)].

The single-rule AT algorithm, AT_{cub1}, is

$$\begin{aligned} AT_{\mathrm{cub1}} = {} & -92.7ND(B3,B5) + 261.4ND(B2,B7) \\ & - 48.8ND(B2,B5) - 17.5ND(B4,B2) \\ & - 146.9ND(B1,B7) + 58.7ND(B3,B1) \\ & - 117 * ND(B2,B1) + 172 * CSA * B5 \\ & + 76 * CSA * B4 + 151 * CSA * B3 \\ & - 951 * CSA * B2 + 539 * CSA * B1 + 28 * B7 \\ & - 132 * B5 - 106.2 * B4 - 22.4 * B3 \\ & + 633.1 * B2 - 443.6 * B1 + 302.0986 \end{aligned} \tag{2}$$

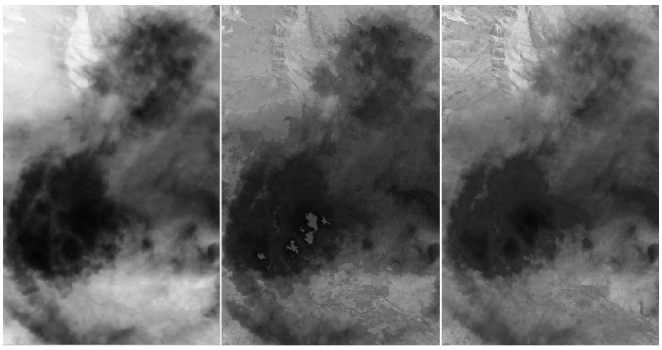


Fig. 4. Examples of the thermal and artificial thermal bands. (a) Landsat 7 low gain band 6 imagery. (b) AT band with a 100-rule decision tree. (c) AT band with a single rule, the ATcub1 formulation, which was chosen for use in AT-ACCA.

where

$ND(x, y)$	normalized difference between x and y; $ND(x, y) = (x - y)/(x + y)$ (for example, $ND(B4, B3)$ is equivalent to NDVI);
CSA	cosine of the solar zenith angle;
Bx	TOA reflectance in Landsat band x; (Note that Landsat 7 bands 1–5 and 7 are expected to correspond to OLI bands 2–7).

This AT band equation gives an rms difference from low gain band 6 brightness temperature of 9.5 K over the $3.95\ 10^9$ pixels in the 103 image validation subset.

Although there may be some collinearity in this model, it should retain predictive power due to the large globally comprehensive training set [18]. Some intermediate models were examined in a search for an optimum solution, but they provided the worst of both models, the slow processing and potential overfitting of the 100-rule model but a compared difference to actual brightness temperature not appreciably superior to that of the single-rule model. Similarly, a bootstrap aggregate approach was considered, which would have reduced overfitting and improved the performance [19], but such an approach would have multiplied the processing time. These enhancements are potential candidates for future work. For the present, the single-rule AT algorithm has been adopted as the AT band for LDCM CCA. All subsequent references in this paper to the AT band use the AT_{cub1} formulation.

It should be stated that the AT band was always intended only for use in an LDCM CCA and that it is not a valid simulation of the Landsat 7 thermal band for general purposes. Although the rms error over the large data set is moderately low, there are local errors. The AT band exhibits errors due to overfitting, which causes small random fluctuations in the middle of the dynamic range and large fluctuations at the extremes. Some examples of AT band errors are shown in Fig. 5. Fortunately, these errors have minimal impact on the ACCA algorithm's threshold tests. Small fluctuations near the threshold will only affect edge cases and very thin clouds, and errors at the extremes—hot objects that are simulated as much too hot or cold objects simulated as far too cold—will not change the outcome of the threshold tests. The AT band is sufficient for threshold-based cloud detection but any other application would require a more robust simulation of the thermal band, which is beyond the scope of this project.

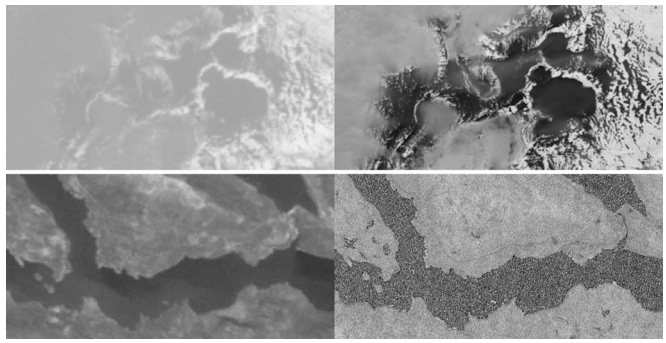


Fig. 5. Examples of AT band overfitting errors. Thermal data are on the left. AT band is on the right.

B. AT-ACCA Validation

With an AT band simulated, it then became possible to develop an AT-ACCA algorithm. The AT-ACCA decision tree (Fig. 6) resembles the ACCA pass-1 decision tree with only two differences. First, the AT band is substituted for band 6 everywhere that band is used in the ACCA pass-1 decision tree. Second, the final discrimination between "warm" and "cold" clouds is discarded—all pixels that reach that branch of the decision tree are labeled as cloudy.

The AT-ACCA algorithm was validated against the entire validation set ($3.95\ 10^9$ pixels in 103 images). The results of the AT-ACCA validation are similar to those of ACCA pass-1, and the overall accuracy [12] for AT-ACCA is shown in Table III. An example of a cloud mask from the AT-ACCA algorithm is shown in Fig. 7(b).

While the results are promising, AT-ACCA designates a large number of pixels as ambiguous. It was decided to add a "disambiguation" algorithm to classify pixels marked as ambiguous by AT-ACCA to increase the accuracy of the overall algorithm. The combination of AT-ACCA plus a disambiguation routine is referred to as "Expanded AT-ACCA."

C. Disambiguation via Threshold Voting Network

To clean up the large population of ambiguous pixels left behind by ACCA and AT-ACCA, we need a disambiguation algorithm. This algorithm should potentially have access to all possible variables in the data, but its output should be a simple classification. It would be desirable for this algorithm to be composed of simple threshold tests to minimize the processing time and for possible future parameterization in the satellite's ground system.

Thus, we want a network of simple logical functions that together map a multidimensional data set into a single output value. This is a working definition of a simple neural network.

With this insight, we envisioned the disambiguation algorithm as a simple McCulloch–Pitts neural network [20] whose threshold logic units (TLUs) are simple threshold tests that

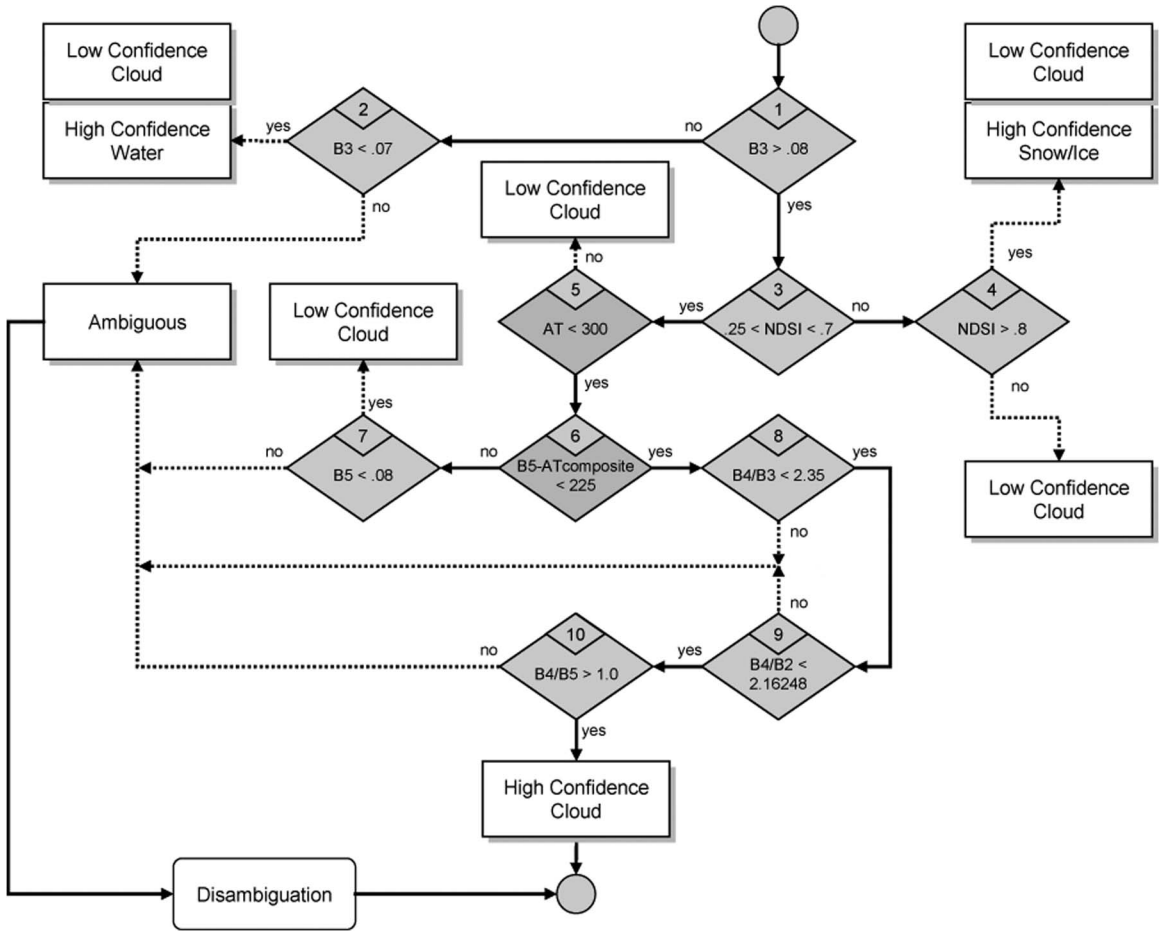


Fig. 6. AT-ACCA flowchart.

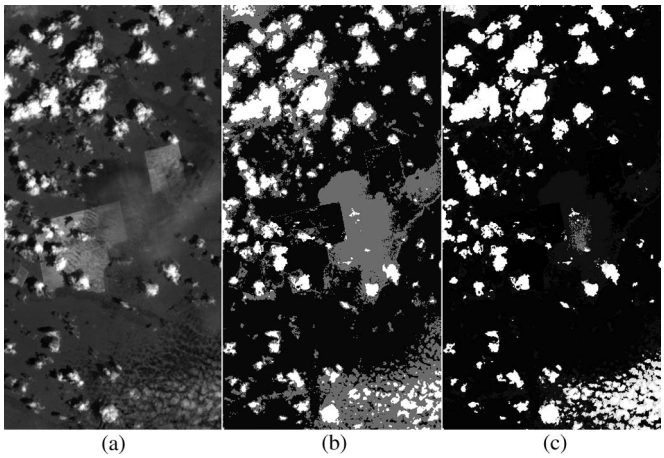


Fig. 7. Example of AT-ACCA cloud masks. (a) Band 5 imagery. (b) AT-ACCA cloud mask of the same region. (c) EXPANDED AT-ACCA cloud mask, with gD02 disambiguation.

are constrained to two outcomes with equal weight (cloudy or clear). The network can be constructed by first compiling all input variables and then winnowing out redundant variable tests. The network can then be reduced to preferred independent maxima by eliminating the TLU nodes that are nonpredictive. Once this is done, the predictive threshold tests then form a sparse network of independent TLUs, which are arranged in an equal-weight voting scheme that will classify the input pixels.

Complex neural networks have been used in the past to create very accurate cloud classification algorithms [21], but the single-layer McCulloch–Pitts network was chosen for its simplicity and speed.

1) Specification of Variables: To construct our neural network, we first began with an initial list of all possible test variable candidates, limiting ourselves to unitless and linear combinations for simplicity. This list included every band, every band normalized to the 6-band brightness, every ratio between two bands, every normalized difference between two bands, and every permutation of the aforementioned multiplied by the cosine of the solar zenith angle. For example, the variables dependent on band 1 were as follows:

$$B1$$
$$B1/N$$
$$CSA * B1$$
$$B1/Bx$$
$$CSA * B1/Bx$$
$$ND(B1, Bx)$$
$$ND(CSA * B1, Bx)$$

where Bx is each other band and N is the 6-band normalization factor

$$N = \sqrt{\sum_x B_x^2}. \tag{3}$$

With duplicates removed by virtue of the fact that $ND(x, y) = -ND(y, x)$, this produced a list of 123 test variable candidates, with 21 distinct one- or two-band combinations. Two additional candidates were added: the AT band and the composite (1-B5)∗AT, which is a threshold test used in AT-ACCA. The composite test was designed to eliminate cold land surface features that have low band 5 reflectance, such as snow and tundra [22].

Each of these variable candidates produced three decision TLUs: two one-sided thresholds and one two-sided threshold. The default hypothesis was chosen to be Cloudy, so that pixels which pass the threshold test would then be classified as Clear.

Node 1: Clear if $V <$ threshold.
Node 2: Clear if threshold $< V$.
Node 3: Clear if $V <$ low_threshold or $V >$ hi_threshold.

The 125 test variable candidates, each with three nodes, gave a total of 375 TLU nodes to be evaluated.

2) Evaluation and Ranking: Evaluation required testing each TLU on a set of training data. Due to computer memory limits, the disambiguation algorithms were trained on a subset created by selecting approximately 77 000 pixels at random from each image in the 103-scene GSFC training set to produce a subset of 7 922 160 pixels. The AT-ACCA algorithm was then run on that subset, and pixels which produced nonambiguous results were discarded. This produced a subset of 1 169 830 pixels, all of which were labeled as ambiguous by AT-ACCA and were used as training data for the disambiguation algorithms.

For each TLU candidate variable, 1000 threshold values were considered, ranging from the minimum to the maximum for that variable. For the dual threshold tests, this resulted in 499 500 threshold pairs considered. The ratio variables such as B1/B2 were clipped to a maximum value of 6 to prevent unbounded values due to data artifacts.

For the single variable tests, 1000 values is oversampling, as there are only 254 possible values in the 8-b inputs, excluding saturations. However, the ratio and normalized difference variables involve two inputs and thus have a range of 254^2 distinct threshold values. However, an initial review of the variable histograms did not show fine scale variance, and hence, 1000 values were chosen as a compromise between full sampling and processing time.

For each of the 1000 threshold values, a performance metric was evaluated for the data in the training set, and the best threshold value for that metric was found. Several metrics were attempted, and all had similar problems. Clear data are overrepresented in the data—the training subset was 63.8% clear, but some individual scenes in the GSFC test set are totally cloud free, which skewed results for all pixels of similar spectra. This caused any measurement based on the strict proportion of pixels correctly classified to favor the clear classification for all pixels. In other words, the optimization process converged on a trivial solution where no clouds would be classified at all. This is a common result when designing machine learning algorithms for image recognition [23].

This drove us to consider a balanced accuracy metric [23], which uses the producer's accuracies of both cloudy and clear pixels averaged without weighting despite the uneven proportion of cloudy and clear pixels in the data

$$Balanced\ accuracy = \frac{P(Pass|Clear) + P(Fail|Cloudy)}{2} \tag{4}$$

where

P(Pass|Clear) producer's accuracy for clear pixels or the probability that a known clear pixel will pass the threshold test;
P(Fail|Cloudy) producer's accuracy for cloudy pixels or the probability that a known cloudy pixel will fail the threshold test; P(Fail|Cloudy) $= 1 - P$(Pass|Cloudy).

As shall be demonstrated shortly, it is desirable to put this metric in terms of misclassification error. Fortunately, the error of omission for any classification is one minus the producer's accuracy [24]; hence, we then have

$$Balanced\ accuracy = 1 - \frac{E(Clear) + E(Cloudy)}{2} \tag{5}$$

where

$E(Clear)$ errors of omission for the clear pixel class; $E(Clear) = 1 - P$(Pass|Clear);
$E(Cloudy)$ errors of omission for the cloudy pixel class; $E(Cloudy) = P$(Pass|Cloudy).

This metric allowed us to find the optimal threshold for each TLU, giving us 375 threshold tests of quantified predictive value. To select the threshold tests that would be included in our network, we needed to establish boundary conditions for $E(Clear)$ and $E(Cloudy)$.

Since our variable candidate list had 23 independent variables in it—6 single band candidates, 15 two-band combinations, and 2 thermal candidates—we expected our final voting scheme to contain 23 TLUs. (We will return to this assumption later.) Each of the TLUs in the network votes on whether a pixel is either cloudy or clear, and the final classification is determined by the accumulation of votes. Clear pixels should receive more votes than cloudy pixels, which ideally will receive no votes at all.

The probability of obtaining exactly n votes from a network with N nodes can be calculated by a simple binomial distribution

$$P(X = n) = \binom{N}{n} p^n (1-p)^{N-n} \tag{6}$$

where

n number of votes received by a pixel;
N total number of votes in the network;
p^n probability that input X will receive a vote from each node.

TABLE IV
EFFECT OF VOTE SEPARATION ON THE TLU CANDIDATE LIST

Vote Separation, n	E(Clear) boundary condition	E(Cloudy) boundary condition	Number of candidate TLUs
2	< 4.89%	< 78.48%	0
3	**< 7.81%**	**< 73.22%**	**121**
4	< 10.97%	< 68.20%	78
5	< 14.32%	< 63.37%	34

Therefore, the probability of n or less votes can be found via the cumulative binomial distribution

$$P(X \leq n) = \sum_{i=0}^{n} \binom{N}{i} p^i (1-p)^{N-i}. \tag{7}$$

This allows us to take known specifications, such as our performance goal and the desired number of votes that separate cloudy from clear pixels, and translate them into boundary conditions for the TLU threshold tests. We want a clear pixel to receive more than n votes in our final network of $N = 23$ tests. We chose to set a performance goal that 90% of pixels are correctly classified. Then, for clear pixels, $P(X \leq n) = 0.10$ and p in (6) is the chance of a clear pixel passing the threshold test or $1 - E(Cloudy)$. For cloudy pixels, $P(X \leq n) = 0.90$ and p is the chance of a cloudy pixel passing the test or $E(Clear)$.

Various vote separations were examined, as shown in Table IV. With a vote separation of $n = 2$, none of the TLU candidates met the boundary criteria. At $n = 3$, the maximum number of TLUs were qualified as candidates, with 121 out of 375 TLU threshold tests passing the boundary conditions of $E(\text{Clear}) < 7.81\%$ and $E(\text{Cloudy}) < 73.22\%$. These error boundaries were chosen as the criteria for accepting the TLUs into our network.

At this point, any or all of the 121 optimized TLUs could have been used in a classification algorithm. We further winnowed down the nodes by excluding redundant band combinations, retaining the threshold test for each band combination with the highest balanced accuracy. For example, if both $B1/B5$ and $ND(1,5)$ were valid tests, only the test with the best accuracy was retained. This excluded all but 16 tests, which are shown in Table V.

Note that seven band combinations failed to meet the error boundary criteria and are thus unrepresented—B4; B7; B1 and B2; B4 and B5; B4 and B7; and both of the AT band candidates. These combinations do not produce tests that are relevant enough to include in our disambiguation algorithm. Some of these failed combinations are used by ACCA; however, recall that the training data for this algorithm were data already identified as ambiguous by AT-ACCA and are thus not a true representation of the global data set. For reference purposes, Table VI shows the reflective band ACCA test thresholds run on the same ambiguous training set.

3) Voting Rules: To determine the final voting rules for the disambiguation algorithm, we must revisit the assumptions we have made. We expected 23 independent tests—one for every band or band combination in the variable candidates—and predicted that those tests would separate clear and cloudy pixels by more than three votes. However, only 16 tests met the error boundary criteria; thus, we need to re-examine our vote separation.

To determine the separation empirically, several trials were run over the ambiguous pixels in the training subset. In each trial, pixels were classified as Cloudy, Ambiguous, or Clear based on a lower threshold value $V1$ and an upper threshold value $V2$

$$\begin{aligned} &Pixel\ is\ Cloudy\ if && Votes <= V1 \\ &Pixel\ is\ Ambiguous\ if && V1 < Votes < V2 \\ &Pixel\ is\ Clear\ if && Votes >= V2 \end{aligned} \tag{8}$$

where $V1 = (0, 1)$, $V2 = (1, 2, 3, 4)$, and $V2 > V1$.

This classification was then compared with the manually generated truth masks to produce an accuracy score for each trial. The results are shown in Table VII.

All of these results are usable; even the weakest classifier (gD04) correctly sorts two-third of the ambiguous pixels. Selecting the "best" classifier is a matter of subjective optimization. This is a benefit in an operational system like LDCM, as it provides parameters by which future data analysts can fine-tune the classification algorithm for best performance. For an initial disambiguation algorithm, however, our primary goal has been to reduce, but not necessarily to eliminate, ambiguous pixels. This leads us to the trials that classify a small but nonzero number of pixels as ambiguous: gD02, gD13, and gD24. Of those, gD02 has the highest accuracy, and hence, the initial parameters of the disambiguation algorithm were chosen to be $V1 = 0$ and $V2 = 2$. In the final LDCM CCA system, these will be calibration file parameters that can be easily changed should we decide to adjust the performance of the disambiguation routine.

The results of the combined Expanded AT-ACCA algorithm, run on the entire $3.95\ 10^9$ pixel validation set, are shown in Table III. The Expanded AT-ACCA algorithm uses AT-ACCA as an initial classifier, with any ambiguous pixels then sent to gD02 for further classification. This will be implemented as part of the LDCM CCA system. An example of the Expanded AT-ACCA CCA is shown in Fig. 7(c).

Table III also shows the results of the gD02 algorithm on the validation set, with an accuracy that compares well with the other primary algorithms under study. While this indicates that gD02 may be valuable as a stand-alone CCA algorithm, this was not our purpose in developing it. For an operational system like the LDCM CCA system, it is preferable to use algorithms with some history of use (such as ACCA) or which have been developed using accepted methods and software (such as C5.0). The gD02 algorithm is intended as an adjunct algorithm for disambiguation of AT-ACCA results, nothing more. However, further study of the neural network cloud assessment methodology might prove interesting, if outside the scope of this paper.

TABLE V
TESTS CHOSEN FOR AT-ACCA DISAMBIGUATION

Parameter	Low Threshold	High Threshold	Misclassified clouds P(Pass|Cloudy) E(Clear)	Misclassified clears 1-P(Pass|Clear) E(Cloudy)	Accuracy $1-\frac{E(Clear)+E(Cloudy)}{2}$
B1	0.140	--	7.76%	43.62%	74.31%
B2	0.111	--	7.62%	49.56%	71.41%
B3	0.093	--	7.71%	52.17%	70.06%
B5/nfac	0.087	0.481	7.76%	65.60%	63.32%
B3/B1	0.640	1.034	7.81%	67.99%	62.10%
ND(CSA*B1,B4)	-0.454	0.262	7.78%	68.83%	61.70%
ND(B1,B5)	-0.138	0.716	7.81%	55.94%	68.13%
CSA*B1/B7	0.736	3.914	7.81%	71.79%	60.20%
B3/B2	0.810	1.075	7.76%	63.71%	64.26%
ND(B2,B4)	-0.404	0.160	7.74%	68.26%	62.00%
ND(B2,B5)	-0.186	0.716	7.81%	56.50%	67.84%
ND(B2,B7)	-0.018	0.754	7.71%	66.58%	62.86%
ND(CSA*B3,B4)	-0.566	-0.016	7.80%	66.30%	62.95%
ND(B3,B5)	-0.232	0.692	7.81%	62.02%	65.09%
ND(B3,B7)	-0.030	0.738	7.80%	68.19%	62.01%
ND(B5,B7)	-0.050	0.300	7.73%	70.21%	61.03%

TABLE VI
REFLECTIVE BAND ACCA TEST PERFORMANCES ON AMBIGUOUS DATA SET

Parameter	Low Threshold	High Threshold	Misclassified clouds P(Pass|Cloudy) E(Clear)	Misclassified clears 1-P(Pass|Clear) E(Cloudy)	Accuracy $1-\frac{E(Clear)+E(Cloudy)}{2}$
B3	0.08	--	4.10%	61.90%	67.00%
ND(B2,B5)	-0.25	0.7	0.40%	78.70%	60.45%
B5	0.08	--	2.30%	91.70%	53.00%
B4/B2	2.16248	--	90.60%	30.50%	39.45%
B4/B3	2.35	--	87.50%	33.20%	39.65%
B4/B5	--	1.0	77.70%	26.80%	47.75%

TABLE VII
VOTING RESULTS BY TRIAL TO DETERMINE VOTE THRESHOLDS $V1$ AND $V2$

Trial	V1	V2	Total Correct	Total False	Total Ambiguous	Misclassified clouds E(Clear)	Misclassified clears E(Cloudy)	Accuracy $1-\frac{E(Clear)+E(Cloudy)}{2}$
gD01	0	1	78.03%	21.97%	0.00%	34.11%	15.08%	75.40%
gD02	**0**	**2**	**74.69%**	**18.55%**	**6.76%**	**24.67%**	**15.08%**	**80.13%**
gD03	0	3	71.33%	16.74%	11.93%	19.68%	15.08%	82.62%
gD04	0	4	67.16%	14.97%	17.87%	14.79%	15.08%	85.07%
gD12	1	2	78.11%	21.89%	0.00%	24.67%	20.32%	77.51%
gD13	1	3	74.75%	20.09%	5.16%	19.68%	20.32%	80.00%
gD14	1	4	70.58%	18.32%	11.11%	14.79%	20.32%	82.45%
gD23	2	3	76.55%	23.45%	0.00%	19.68%	25.58%	77.37%
gD24	2	4	72.38%	21.67%	5.95%	14.79%	25.58%	79.81%
gD34	3	4	74.15%	25.85%	0.00%	14.79%	32.12%	76.54%

TABLE VIII
MISCLASSIFICATION COSTS USED FOR C5.0 CCA DEVELOPMENT

	Clear	Thin Cloud	Thick Cloud
Clear	-	0.75	1.0
Thin Cloud	0.75	-	0.5
Thick Cloud	1.0	0.5	-

D. C5 CCA

The AT-ACCA algorithm described previously was based on the ACCA algorithm, and its disambiguation algorithm was developed from statistical principles. As a third option for developing a LDCM CCA, we chose to use the statistical classifier C5.0, by RuleQuest Research. As aforementioned, C5.0 is an off-the-shelf regression tool based on the ID3 model of minimizing information entropy, which can create stand-alone decision trees from a set of training data.

Very large training data sets can cause memory errors with C5.0; thus, for C5 CCA development, a training subset was created by selecting approximately 77 000 pixels at random from each image in the 103-scene GSFC training set discussed in Section II-B. This produced a training subset of 7 922 160 pixels, all of which were converted to TOA reflectance. The C5.0 software was given the training data and a list of variable candidates similar to the list presented in the Threshold Voting algorithm aforementioned, with the exclusion of the AT band and all variables derived from it. The AT band was not used in any way for C5 CCA development. Part of the C5.0 output is a list of variables used and the percentage of training pixels whose classification depends on the value of that variable. Variables that are relevant to only a small fraction of the training set can be excluded from use in the decision tree.

An initial attempt to make a CCA decision tree with C5.0 was disappointing but instructive for two reasons. First, using a large number of variables caused the decision trees to expand to gigantic proportions. To prevent this expansion in subsequent runs, the variable list was truncated to only the most relevant variables of the initial run. Second, many of the variables typically used for cloud detection were nonpredictive due to the effects of saturation. This was most evident over vegetation (which saturates in Landsat 7 band 4) and desert regions (which may saturate in bands 5 or 7) [25]. With this knowledge, we further divided the training subset into saturated and nonsaturated subsets. For each pixel, if any reflective band was saturated (reached a value of 255 in the level 1 product), that pixel was assigned to the saturated subset. All other pixels were assigned to the nonsaturated subset.

For training data with several classifications, the C5.0 software calculates the errors of its fit using a matrix of misclassification costs. The initial run was performed with the default costs of 1.0: Every error in classification was given equal weight. To enhance performance of the algorithm, the C5.0 misclassification cost matrix was changed to the values in Table VIII. These costs allow the classifier some leeway in the definition of thin clouds.

TABLE IX
SPECIFICATIONS OF THE C5.0 RUNS

Subset	Pixels in Training subset	Minimum pixels per leaf (-m)	Leaves in Output Tree
Non-Saturated	6,484,262	2900	245
Saturated	1,437,898	500	70

In addition, it was decided that each C5.0 decision tree should be limited to less than 255 "leaves"—final outcomes for which a classification is assigned. This decision was made both to limit the potential run time of the algorithm and to aid in development, as an 8-b debug image containing each pixel's ultimate decision leaf could then be made. The size of the decision tree was limited by using the "$-m$" pruning option of C5.0, which allowed us to specify the minimum number of pixels in each leaf. Of course, a large minimum population for each leaf causes the model to lose accuracy by underfitting the data, while very small minimum populations will invite overfitting and will create complex trees that are slow to run. The balance between accuracy and complexity was achieved through trial and error over many runs of the C5.0 software.

The final variable list was trimmed to use only the 16 most relevant variables:

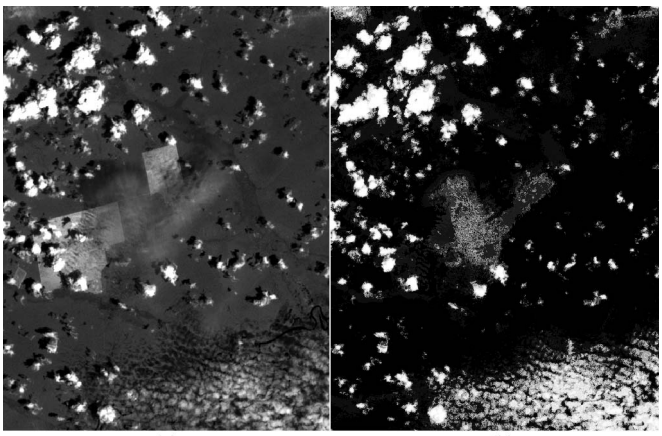
1) Bx (for all $x = \{1, 2, 3, 4, 5, 7\}$);
2) $CSA * Bx$;
3) B4/B2, B4/B5
4) ND(2, 5);
5) ND(1, 7).

Note that ND(2,5) is NDSI [9], which is used in ACCA. ND(1,7) is a normalized difference of bands 1 and 7; C5.0 ranked it as highly relevant in the initial training subset.

Table IX shows the two classification subsets, the "$-m$" values used in each, and the number of leaves in the final output decision trees.

These two decision trees were then combined into one algorithm. The first step in the C5 CCA algorithm is to determine if saturations exist in any band. Then, the appropriate (saturated or nonsaturated) decision tree is called.

Each final output leaf in the decision trees reports the total number of pixels in each class that arrive at that leaf. Using these totals, confidence values were set for each leaf by taking the ratio of cloudy pixels (both thin and thick) to the total number of pixels that pass through that leaf. Thus, the output of the C5 CCA algorithm is a pixel map of cloud confidence values that range from 0 (clear with 100% certainty) to 1 (cloudy with 100% certainty). These confidence values come from a discrete set of 315 values (245 nonsaturated + 70 saturated) and hence appear continuous in an 8-b C5 CCA cloud mask. An example of a C5 CCA cloud mask is shown in Fig. 8, where white indicates high-confidence clouds and black indicates zero confidence or clear. Shades of gray form a linear confidence scale of cloudiness. The thick clouds throughout the image are masked correctly as high-confidence clouds. In the center of this image are also some thin clouds over highly variable terrain, and the C5 CCA assigns a degree of uncertainty to these clouds as well.



(a) (b)

Fig. 8. Example of C5 CCA cloud mask. (a) Band 5 imagery. (b) C5 CCA cloud mask of the same region.

To validate the C5 CCA algorithm, the confidence values were mapped to three values: Clear (confidence of less than 0.35), Ambiguous (confidence between 0.35 and 0.65, inclusive), and Cloudy (confidence greater than 0.65). This mapping was compared with the $3.95\ 10^9$ pixel validation subset. The results are compared with other CCA algorithms in Table III.

The C5 CCA correctly classified 88.5% of pixels in the validation subset and thus meets our goal of outdoing ACCA without the need of a thermal band. However, because its thresholds are very specific, it may be more sensitive to the differences in spectral bandwidth between ETM+ and OLI. The algorithm is also very complex, which may make it difficult to maintain. Development of the C5 CCA algorithm, including validation against data from instruments other than the Landsat 7 ETM+, will continue.

IV. Integrated LDCM CCA System

The LDCM CCA system is intended to be a modular collection of algorithms whose weighted output can be merged together to create a CCA mask for any OLI scene. Each CCA algorithm will be run separately and will create its own intermediate CCA mask. When all the CCA algorithms have finished processing, a merge procedure will combine the separate intermediate masks into the final CCA mask, to be delivered to the user with the level 1 product.

As of this writing, only two cloud-cover algorithms—Expanded AT-ACCA and the C5 CCA—have been selected for inclusion in the LDCM CCA system. A version of ACCA that uses TIRS thermal data is also planned, but the LDCM-specific elements of that algorithm have not yet been defined. The hope is that more algorithms will be selected before or after launch of LDCM.

The format of the final CCA mask which will be distributed with Level 1 LDCM scenes is still in development, but it is certain that it will support several classes (cloud, cirrus, possibly snow, and water), and each class will be scored with a two-bit confidence level:

1) 00 No confidence level set.
   (Used for fill or for class not reported.)
2) 01 Low confidence.
   ($<$ 0.35 confidence of class in this pixel.)
3) 10 Mid confidence.
   (0.35–0.65 confidence.)
4) 11 High confidence.
   ($>$ 0.65 confidence.)

The merge procedure must therefore take the confidence values from each CCA algorithm, allow weighting of each algorithm's class performance, and create a three-level confidence score for each pixel in the image.

The decision was made to use a simple weighted voting mechanism as a merge procedure. Thus, for each pixel, the following sums are calculated:

$$high_k = \sum_{j=1}^{N} \begin{matrix} W_{j,k} & \text{if } C_{j,k} > 0.65 \\ 0 & \text{otherwise} \end{matrix}$$

$$mid_k = \sum_{i=1}^{N} \begin{matrix} W_{j,k} & \text{if } 0.35 \leq C_{j,k} \geq 0.65 \\ 0 & \text{otherwise} \end{matrix}$$

$$low_k = \sum_{i=1}^{N} \begin{matrix} W_{j,k} & \text{if } C_{j,k} < 0.35 \\ 0 & \text{otherwise} \end{matrix} \tag{9}$$

where

- j algorithm j of N; (Currently, $N = 2$: Expanded AT-ACCA and C5 CCA);
- k class k;
- $W_{j,k}$ fractional weights assigned to each algorithm and class;
- $C_{j,k}$ pixel's class k confidence value from algorithm j.

With these sums, the final score of the class (S_k) is found by simple comparison:

$$S_k = \begin{cases} \text{High if } high_k > low_k \text{ and } high_k > mid_k \\ \text{Mid if } mid_k \geq low_k \text{ and } mid_k \geq high_k \\ \qquad \text{or if } high_k = low_k \\ \text{Low if } low_k > high_k \text{ and } low_k > mid_k. \end{cases} \tag{10}$$

This leaves the question of what the class-dependent weights should be for each algorithm. Of the two algorithms developed so far neither has a clear advantage in thick cloud classification, so they will be given equal weights for that class. AT-ACCA can be configured to provide rudimentary water and snow classifications; if those classes are included in the final CCA mask, then the C5 CCA algorithm will be assigned weights of zero for those classes. Thus, the development of weights for the LDCM CCA system is currently a trivial problem. It will become more complicated (and more necessary) if and when additional CCA algorithms are added to the system.

Future algorithms will be added to the modular path, where they will create their own intermediate masks that will be merged into the final CCA mask. Processing time constraints may limit the number and complexity of algorithms added to the CCA system, but already three algorithms are planned; a Cirrus detection algorithm using the OLI cirrus band, an implementation of ACCA using the TIRS thermal data, and

a cloud shadow detection algorithm, which has been an oft-requested feature from the community.

V. Conclusion

The results given here have shown that an OLI cloud-cover algorithm is possible, with a per-pixel mask and without the need for a thermal band, with performance that meets or exceeds that of CCA algorithms delivered with level 1 products in the past.

To create CCA algorithms for LDCM, we first obtained the data set from GSFC that had been used in the development of the Landsat 7 ACCA algorithm. We created manual cloud masks for this data set and divided it into training and validation subsets. This allowed us to quantify the performance of pass-1 of the Landsat 7 ACCA algorithm as correctly classifying 79.9% of the $3.95\ 10^9$ pixel validation subset. A version of ACCA that does not depend on the thermal band, FT-ACCA, was validated as 68.7% correct.

Once a standard test set was available, we used three different approaches to create new CCA algorithms. The first approach is an extension of the Landsat 7 ACCA algorithm, using a synthetic band known as the AT band as a replacement for the thermal band. The AT band was shown to have an rms error of ± 9.5 K, and that led to an AT-ACCA algorithm that correctly classified 76.6% of the validation data.

To reduce the number of pixels marked as "ambiguous" by AT-ACCA, a second approach was made using a simple McCulloch–Pitts neural network derived from statistical metrics of the ambiguous pixels in the training subset. This network correctly classified 84.2% of the GSFC validation data by itself, which demonstrates its usefulness, but when added onto the AT-ACCA algorithm as a disambiguation algorithm, it brought the Expanded AT-ACCA performance to 89.7% correctly classified pixels.

A third approach used an off-the-shelf statistical classifier package known as C5.0 to generate a large decision tree classifer. The C5 CCA algorithm correctly classified 88.5% of the GSFC validation data.

These two algorithms—C5 CCA and Expanded AT-ACCA—will form the initial basis of LDCM CCA. A voting mechanism has been created to allow several CCA algorithms to contribute to the cloud mask of LDCM products, and plans are for additional algorithms to be implemented in the future, whether developed in-house or donated from the community.

Acknowledgment

The authors would like to thank Dr. R. Irish [Goddard Space Flight Center (GSFC)] and the GSFC Science Office for providing the GSFC test data set, Dr. D. Steinwand [U.S. Geological Survey (USGS)], Dr. E. Fosknight (USGS), T. Maeirsperger [Stinger Ghaffarian Technologies (SGT)], and Dr. D. Roy [San Diego State University (SDSU)] for their feedback and advice.

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

References

[1] T. Arvidson, J. Gasch, and S. Goward, "Landsat 7's long-term acquisition plan—An innovative approach to building a global imagery archive," *Remote Sens. Environ.*, vol. 78, no. 1/2, pp. 13–26, Oct. 2001.

[2] B. Hollingsworth, L. Chen, S. Reichenbach, and R. Irish, "Automated cloud cover assessment for Landsat TM images," in *Proc. SPIE Conf.*, vol. 2819, *Imaging Spectrometry II*, 1996, pp. 170–179.

[3] T. Arvidson, S. Goward, J. Gasch, and D. Williams, "Landsat-7 long-term acquisition plan: Development and validation," *Photogramm. Eng. Remote Sens.*, vol. 72, no. 10, pp. 1137–1146, Oct. 2006.

[4] R. Irish, "Landsat 7 automatic cloud cover assessment in "Algorithms for Multispectral, Hyperspectral, and Ultraspectral Imagery VI"," in *Proc. SPIE*, 2000, vol. 4049, pp. 348–355.

[5] J. R. Irons and J. G. Masek, "Requirements for a Landsat data continuity mission," *Photogramm. Eng. Remote Sens.*, vol. 72, no. 10, pp. 1102–1108, Oct. 2006.

[6] J. Ju and D. Roy, "The availability of cloud-free Landsat ETM+ data over the conterminous United States and globally," *Remote Sens. Environ.*, vol. 112, no. 3, pp. 1196–1211, Mar. 2008.

[7] Landsat Data Continuity Mission Announcement Brochure. [Online]. Available: http://landsat.gsfc.nasa.gov/pdf_archive/LDCM2010_4web.pdf

[8] R. R. Irish, J. L. Barker, S. N. Goward, and T. Arvidson, "Characterization of the Landsat-7 ETM+ automated cloud-cover assessment (ACCA) Algorithm," *Photogramm. Eng. Remote Sens.*, vol. 72, no. 10, pp. 1179–1188, Oct. 2006.

[9] G. Riggs, D. Hall, and V. Salomonson, "A snow index for the Landsat thematic mapper and moderate resolution imaging spectroradiometer," in *Proc. IGARSS*, Aug. 8–12, 1994, vol. 4, pp. 1942–1944.

[10] NASA Landsat Project Science Office, Landsat 7 Science Data Users' Handbook, Appendix A. [Online]. Available: http://landsathandbook.gsfc.nasa.gov/handbook/handbook_htmls/chapter12/htmls/ACCA_Appendix.html

[11] B. Markham, S. Goward, T. Arvidson, J. Barsi, and P. Scaramuzza, "Landsat-7 long-term acquisition plan radiometry—Evolution over time," *Photogramm. Eng. Remote Sens.*, vol. 72, no. 10, pp. 1129–1135, Oct. 2006.

[12] R. Congalton and K. Green, *Assessing the Accuracy of Remotely Sensed Data*, 2nd ed. Boca Raton, FL: CRC Press, 2009, pp. 56–61.

[13] A. H. Goodman and A. Henderson-Sellers, "Cloud detection and analysis: A review of recent progress," *Atmos. Res.*, vol. 21, no. 3/4, pp. 203–228, May 1998.

[14] G. S. Pankiewicz, "Pattern recognition techniques for the identification of cloud and cloud systems," *Meteorological Appl.*, vol. 2, no. 3, pp. 257–271, Sep. 1995.

[15] J. Storey, R. Morfitt, and P. Thorson, "Image processing on the Landsat 7 image assessment system," in *Proc. Amer. Soc. Photogramm. Remote Sens. Annu. Conf.*, Portland, OR, May 17–21, 1999, pp. 743–758.

[16] J. R. Quinlan, "Improved use of continuous attributes in c4.5," *J. Arti. Intell. Res.*, vol. 4, no. 1, pp. 77–90, Jan. 1996.

[17] R. Kohavi and J. R. Quinlan, "Decision-tree discovery," in *Handbook of Data Mining and Knowledge Discovery*, W. Klosgen and J. M. Zytkow, Eds. New York: Oxford Univ. Press, 2002, pp. 267–276.

[18] J. Rawlings, S. Pantula, and D. Dickey, *Applied Regression Analysis: A Research Tool*. New York: Springer-Verlag, 1998, pp. 445–446, 457–458.

[19] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996.

[20] R. Rojas, *Neural Networks—A Systematic Introduction*. Berlin, Germany: Springer-Verlag, 1996, pp. 30–38.

[21] J. Lee, R. C. Weger, S. K. Sengupta, and R. M. Welch, "A neural network approach to cloud classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 28, no. 5, pp. 846–855, Sep. 1990.

[22] D. K. Hall, G. A. Riggs, and V. V. Salomonson, "Development of methods for mapping global snow cover using moderate resolution imaging spectroradiometer data," *Remote Sens. Environ.*, vol. 54, no. 2, pp. 127–140, Nov. 1995.

[23] M. Sokolova, N. Japkowicz, and S. Szpakowicz, "Beyond accuracy, F-score and ROC: A family of discriminant measures for performance evaluation," in *Proc. ACS Australian Joint Conf. Artif. Intell.*, vol. 4304, *Lecture Notes in Computer Science*, 2006, pp. 1015–1021.

[24] L. Janssen and F. van der Wel, "Accuracy assessment of satellite derived land-cover data: A review," *Photogramm. Eng. Remote Sens.*, vol. 60, no. 4, pp. 419–426, 1994.

[25] T. Arvidson, R. Irish, B. Markham, D. Williams, J. Feuquay, J. Gasch, and S. N. Goward, "Validation of the Landsat 7 long-term acquisition plan," in *Proc. Pecora 15*, Bethesda, MD, 2002, pp. 1–13.

Pasquale L. Scaramuzza received the B.S. degree in physics from Drexel University, Philadelphia, PA, in 1990 and the M.S. degree in physics from Temple University, Philadelphia, in 1993.

He has worked as a Radiometric Analyst with the Landsat program at the U.S. Geological Survey Earth Resources Observation and Science Center, Sioux Falls, SD. He is currently a Scientist with Stinger Ghaffarian Technologies, Greenbelt, MD, and is a Contractor for the Landsat Data Continuity Mission (LDCM). He is involved with the creation of calibration and classification algorithms to be used with LDCM.

Michelle A. Bouchard received the M.S. degree in geography from South Dakota State University, Brookings, in 2009.

She has worked on cloud-cover assessment and Landsat time-series analysis of glaciers for the Landsat Data Continuity Mission Science Office. She is currently a Scientist with ASRC Research and Technology Solutions, U.S. Geological Survey Earth Resources Observation and Science Center, Sioux Falls, SD, modeling land-use change with the LandCarbon project.

John L. Dwyer received the M.S. degree in geological sciences from the University of Colorado, Boulder.

He is currently a Physical Scientist with the U.S. Geological Survey (USGS) Earth Resources Observation and Science Center, Sioux Falls, SD. He is currently the USGS Project Scientist for the Landsat Data Continuity Mission Ground System. He has more than 25 years of experience with remote sensing and geospatial data applications in mapping land cover and land-cover change, assessing mineral and energy resources, and studying glacier dynamics.