

Improved Cloud Detection for Daytime AVHRR Scenes over Land

James J. Simpson* and Jason I. Gobat*

Accurate cloud detection in Advanced Very High Resolution Radiometer (AVHRR) data over land is a difficult task complicated by spatially and temporally varying land surface reflectances and emissivities. The AVHRR Split-and-Merge Clustering (ASMC) algorithm for cloud detection in AVHRR scenes over land provides a computationally efficient, scene-specific, objective way to circumvent these difficulties. The algorithm consists of two steps: 1) a split-and-merge clustering of the input data (calibrated channel 2 albedo, calibrated channel 4 temperature, and a channel 3 – channel 4 temperature difference), which segments the scene into its natural groupings; and 2) a cluster-labelling procedure that uses scene-specific, joint three-dimensional adaptive labelling thresholds (as opposed to constant static thresholds) to label the clusters as either cloud, cloud-free land, or uncertain. The uncertain class is used for those pixels whose signature is not clearly cloud-free land or cloud (e.g., pixels at cloud boundaries that often contain subpixel cloud and land information that has been averaged together by the integrating aperture function of the AVHRR instrument). Results show that the ASMC algorithm is neither regionally nor temporally specific and can be used over a large range of solar altitudes. Sensitivity of the segmentation and labelling steps to the choice of input variables also was studied. Results obtained with the ASMC algorithm also compare favorably with those obtained from a wide range of currently used algorithms to detect cloud over land in AVHRR data. Moreover, the ASMC algorithm can be adapted for use with data to be taken by the Moderate Resolution Imaging Spectrometer-Nadir (MODIS-N).

INTRODUCTION

The accurate separation of clouds from land in satellite imagery [e.g., the Advanced Very High Resolution Radiometer (AVHRR), the Geostationary Observing Earth Satellite (GOES), LandSat] is a critical step in the analysis of these data for use in land-based applications (Ohring et al., 1985; Sellers et al., 1990). Historically, efforts to address this important issue have used three types of analysis: 1) maximum compositing techniques; 2) expert knowledge-based techniques; and 3) multispectral statistical approaches.

Maximum compositing techniques filter the data so that the highest normalized vegetation index (NDVI) is retained for a given pixel within a given compositing period (Tarpley et al., 1984; Justice et al., 1985). NOAA uses this approach to produce operational NDVI products.

Several studies (e.g., Taylor et al., 1985; Spanner et al., 1990) combine the expert knowledge of a trained analyst with interactive display methods to identify cloudy pixels in a scene. Although this approach can provide properly cloud-screened images, it has several disadvantages: 1) it is labor intensive, 2) it is not statistically reproducible by other investigators, and 3) it requires detailed knowledge and considerable experience in photo interpretation. Thus, such an approach is not well-suited to process the large data sets required for regional- and continental-scale studies of land-based global change processes [e.g., International Geosphere and Biosphere Program (IGBP), the BOREAL Ecosystem-Atmospheric Study (BOREAS), and the Global Energy and Water Cycle Experiment (GEWEX/GCIP)].

More recently, several investigators have used a variety of statistical procedures applied to multispectral subsets of AVHRR data to detect clouds in scenes over land. For example, Gutman et al. (1987) used a two-step process to detect clouds in reduced spatial resolution (e.g., GAC) AVHRR data. First, temperature and albedo thresholds perform coarse cloud screening. Then, fine cloud screening is done using standard deviation criteria

* Digital Image Analysis Laboratory, University of California, San Diego

Address correspondence to Dr. James J. Simpson, Scripps Satellite Oceanography Center, Scripps Institution of Oceanography, Digital Image Analysis Laboratory, University of California, San Diego, La Jolla, CA 92093-0237.

Received 30 November 1994; revised 28 July 1995.

developed from the coarsely screened images. Gutman (1991) modified this procedure by taking a minimum of standard deviations of visible albedo over 1 month for each pixel in the scene and adding an albedo increment in which the increment is dependent on the minimum value of standard deviation. This procedure assumes that in areas of higher natural variability in reflectance, the uncertainty in the albedo estimate is greater. Gutman (1992) classified daytime AVHRR data (cloudy versus clear) using a combination of threshold values for spatial variability tests, a dynamic cloud filter designed using the assumption that the absolute value of the near-infrared (IR) to visible reflectance ratio correlates negatively with surface temperature and climatology. Choudhury (1990) formed monthly composites, based on data taken within a range of scan angles between $\pm 30^\circ$ of nadir, to minimize the effects of aerosols and clouds in land applications. Kaufman and Sendra (1988) removed cloudy regions in scenes by masking pixels with high wavelength-independent reflectance or low infrared radiative temperature. Subpixel clouds, which cannot be removed by this procedure, are considered part of the aerosol layer and treated as such.

This article develops a new algorithm for accurate cloud detection in AVHRR scenes over land. The algorithm, called the AVHRR Split-and-Merge Clustering (ASMC) cloud-detection algorithm, uses an iterative, nested, partitional, hierarchical clustering procedure to segment the AVHRR scene into its natural divisions and then a joint, three-dimensional, adaptive thresholding procedure to label the segments as cloudy, cloud-free, or uncertain. For the reader's convenience, this article is divided into six sections, including the present introduction. The second section contains details of the ASMC algorithm. A brief description of the data used in this study is given in the third section. Representative results (fourth section) for tropical, mid-latitude temperate and desert images confirm the regional independence of the new algorithm. The fifth section discusses possible input variants to the ASMC process and their effects on segmentation, labelling, and performance issues. It also provides comparisons with other approaches, discusses application of the ASMC process to MODIS-N data, and gives limitations of the new algorithm (e.g., snow versus cloud discrimination). The last section provides a summary for the article.

The work presented here differs in several important respects from our previously published related work. The clustering procedure, although similar to that developed by Gallaudet and Simpson (1991), has been extended to include arbitrary geophysical inputs (e.g., albedo, temperature, temperature difference). Gradient of albedo data and postprocessing spatial uniformity tests used by Simpson and Gobat (1995a) to cloud screen GOES data over land are not used in the present AVHRR case because of the better spectral and finer spatial resolution of AVHRR data compared to that of

GOES data. Thus, the adaptive labelling criteria used herein differ substantially from those used with GOES data, a result consistent with the International Satellite Cloud Climatology Project's (ISCCP) experience (Rossow et al., 1985) that the selection of the "best" analysis technique is constrained by the properties (e.g., spectral characteristics, spatial resolution) of the available satellite data. Moreover, the labelling procedure has been generalized to include multiple decision surfaces that allow for a statistically accurate segmentation of the scene into cloudy, cloud-free, and ambiguous or mixed regions. Likewise, the improved spectral and spatial characteristics of the AVHRR data (compared to GOES), coupled with the improved AVHRR specific adaptive labelling scheme allow us to segment traditionally very difficult scenes (e.g., complex desert regions) without the use of "a priori" mask information as was required in the GOES case (see Simpson and Gobat, 1995a). Finally, the discussion section includes a detailed comparison of results obtained with variations of the ASMC procedure based on possible, different data inputs available in AVHRR data as well as detailed comparisons between the ASMC procedure and several other cloud-detection techniques used with AVHRR data over land [e.g., albedo and thermal models, texture models, and NOAA's CLAVR-I algorithm (Stowe et al., 1991)].

THE AVHRR SPLIT-AND-MERGE CLUSTERING CLOUD-DETECTION ALGORITHM

The AVHRR Split-and-Merge Clustering algorithm uses high spatial resolution channel-2 visible albedo, infrared temperature derived from channel-4 data, and a channel 3 minus channel 4 temperature difference. Although channel-1 data may give better contrast between land and cloud than channel-2 data, channel-2 data are preferred over channel-1 data for this application because channel-1 data are much more affected by time-and-space varying atmospheric processes (e.g., Rayleigh and aerosol scattering) than are channel-2 data. A detailed discussion of this point, including results from radiative transfer model computations, is given by Simpson and Gobat (1995b, in particular their figure 9). Prefiltering the channel-3 data with a Wiener filter to remove noise (Simpson and Yhann, 1994) generally has not proved necessary in this application based on tests that used both raw and filtered channel-3 data to form the required channel 3 minus 4 temperature difference (see Discussion). Extremely noisy channel-3 data, however, may require preconditioning with the Wiener filter methods. The algorithm has two steps: 1) a split-and-merge clustering procedure, and 2) an adaptive-labelling procedure. Each step is described separately below.

Split-and-Merge Iterative Clustering

Clustering refers to any type of method that attempts to automatically partition a given data set by identifying

the natural groupings of the data within a specified feature space by optimizing a chosen clustering criterion (Pairman and Kittler, 1986). Hence, this approach is dependent on the global properties of the data. Clustering attempts to make all the elements of a given cluster as similar as possible (i.e., minimize within-cluster variance) and simultaneously make individual clusters as distinct as possible (i.e., maximize between-cluster variance). Scatter matrices used here provide a computationally convenient way in which to formulate this minimization problem. Moreover, Gallaudet and Simpson (1991) have shown that minimizing the within-cluster scatter matrix simultaneously maximizes the between-cluster scatter matrix if total variance is to be conserved.

We define n_d as the number of bands in the input image. Then we can describe the input data as N n_d -dimensional column vectors where N is the number of valid pixels in the input image. Each column vector corresponds to a given pixel location within the input image and each component of the column vector corresponds to a single spectral or derived value at that location. For the ASMC algorithm, $n_d = 3$; each vector is defined as

$$y = \begin{bmatrix} a_2 \\ T_4 \\ \delta \end{bmatrix} \quad (1)$$

where a_2 is the channel-2 albedo, T_4 is the surface temperature derived from channel 4, and δ is the temperature difference between channels 3 and 4, subject to the constraint that negative differences are mapped to 0.0.

The clustering operation proceeds as follows:

1. Distribute the initial K_i cluster centers, m_k , equidistant throughout the range of the input data

$$m_k = y'_{min} + \left(\frac{k-1}{K_i-1} \right) (y'_{max} - y'_{min}) \quad (2)$$

where the maximum and minimum vectors of the entire data set, y'_{max} and y'_{min} , respectively, are defined as the component-wise maximum and minimum of the n_d spectral or derived inputs.

2. Assign each input vector, y , to the nearest cluster center by computing the Euclidean distance from the vector to each cluster center

$$d_k = (y - m_k)^T (y - m_k) \quad k = 1 \dots K \quad (3)$$

(where K is the total number of clusters) and assign the vector to the cluster that corresponds to the minimum d_k . In Eq. (3) and elsewhere superscript T denotes the transpose operation.

3. Compute the between-cluster scatter matrix, S_b

$$S_b = \sum_{k=1}^K n_k (m_k - m)(m_k - m)^T \quad (4)$$

where n_k is the number of vectors in cluster k , and m is the mean vector of the entire input data

$$m = \frac{1}{N} \sum_{k=1}^K n_k m_k \quad (5)$$

4. Check for convergence based on the trace of the between-cluster scatter matrix, $Tr(S_b)$. $Tr(S_b)$ is a measure of between-cluster variance. Convergence occurs when there is negligible change in this measure from one iteration to the next, that is, if

$$\left| \frac{Tr(S_b) - Tr(S'_b)}{Tr(S_b)} \right| < tol \quad (6)$$

then we stop the clustering process and move on to the labelling stage. In the convergence check above, S'_b represents the between-cluster scatter matrix from the previous iteration and tol is a predefined numerical tolerance.

5. For each cluster k , compute the distance between the component-wise cluster maximum and minimum vectors,

$$d_m = (y_{max} - y_{min})^T (y_{max} - y_{min}) \quad (7)$$

where the cluster minimum and maximum vectors are defined by a procedure analogous to that used for the image minimum and maximum vectors in (2); they are the component-wise cluster minimum and maximum vectors. If d_m is greater than the splitting threshold (T_s), split the cluster by reassigning each vector to one of two new clusters, based on whether it is closer to the maximum or minimum.

6. Repeat step 5 until nothing is split during that step.
7. For every pair of clusters, compute the distance, d_b , between the cluster mean vectors

$$d_b = (m_i - m_j)^T (m_i - m_j). \quad (8)$$

If d_b is less than the merging threshold (T_m), combine all the vectors in clusters i and j into a single cluster.

8. Repeat step 7 until nothing is merged during that step.
9. Based on the current K cluster centers, repeat the process beginning at step 2.

The process outlined above uses four predetermined parameters: the number of initial clusters (K_i), the splitting threshold (T_s), the merging threshold (T_m), and the convergence tolerance (tol). All four are relatively easy to determine because, as Gallaudet and Simpson (1991) have shown, the clustering procedure is largely insensitive to the values chosen. Specifically, Gallaudet and Simpson (1991, figure 8) show the sensitivity contours of the T_s and T_m thresholds to: a) within-cluster variance, b) number of split and merge opera-

tions in the first iteration of the clustering procedure, c) the percentage of pixels rejected as cloud, d) the total number of iterations required for convergence to a stable cluster population, and e) the final number of clusters at convergence. Moreover, in appendix C of that same article, a detailed discussion of the sensitivity of the clustering procedure to the initial number of clusters (K_i) and to the convergence criteria (tol) is given. Because the ASMC clustering procedure is a slight adaptation of that used by the original PCTSMC algorithm of Gallaudet and Simpson (1991), the same rules apply for selection of input parameters to the clustering procedure of the ASMC algorithm. Thus, for the ASMC algorithm, these parameters are specified as follows. K_i , the initial number of clusters, was chosen as 30 because this number is small enough that the algorithm still performs acceptably fast and large enough that the clustering result is generally stable for any number larger than this. The splitting and merging thresholds are defined by Gallaudet and Simpson (1991, their appendix C) as

$$\beta(y'_{max} - y'_{min})^T(y'_{max} - y'_{min}), \quad (9)$$

that is, as some percentage of the distance between the input data set maximum and minimum vectors. β was chosen as 0.01. The convergence tolerance was taken as 0.05. This means that if the trace of the between-cluster scatter matrix changes by less than 5% from one iteration to the next, then the algorithm will accept the current set of clusters and proceeds to the labelling stage.

The method of clustering just described combines both partitional and hierarchical approaches. It consists of a partitional clustering algorithm augmented by a splitting-and-merging step at each iteration. Combining a partitional with a hierarchical method has several advantages over the use of either method alone (see Gallaudet and Simpson, 1991 and the references contained therein).

Labelling

Once the data have been segmented into their natural classes, a labelling algorithm is used to label each cluster (and thus all the pixels in that cluster) as either cloudy or clear. The labelling criteria for ASMC are based on an adaptively determined decision plane that splits the three-dimensional input space into clear and cloudy regions. Figure 1a illustrates the three-dimensional vector space of the problem (albedo, temperature, $T_3 - T_4$ temperature difference) as the dashed outline; the plane that splits this space into clear and cloudy regions is drawn as the solid triangle.

Given n final clusters from step 1 of the ASMC algorithm, we define α_{min} as the minimum albedo component of the n cluster mean vectors. Similarly, δ_{min} is the minimum $T_3 - T_4$ temperature difference and T_{max} is the

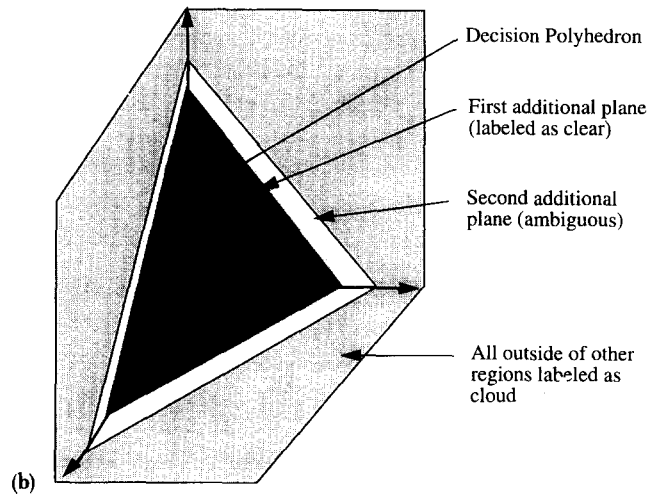
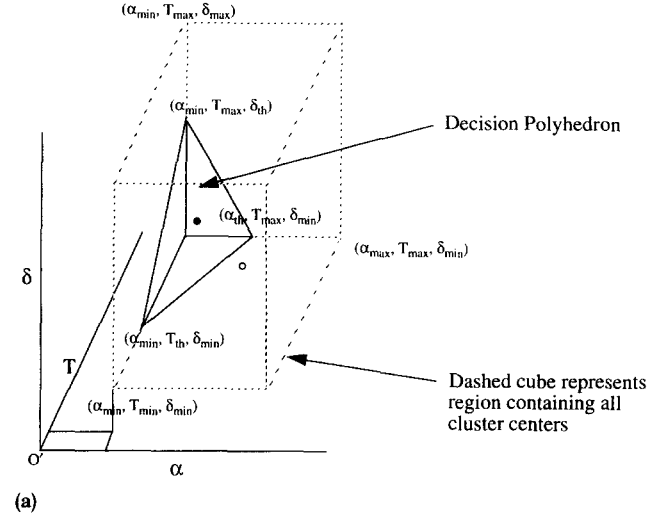


Figure 1. a) Definition of the decision plane used to separate cloudy from cloud-free land pixels in the labelling step of the ASMC algorithm. The closed circle represents a cloud-free cluster inside the decision polyhedron, whereas the open circle represents a cloudy cluster outside the polyhedron. b) Definition of the ambiguous domain used in the labelling step of the ASMC algorithm. All symbols are defined in the text.

maximum temperature component. The scene-specific adaptive labelling threshold values, α_{th} , δ_{th} , and T_{th} , are determined from an adaptive threshold of the appropriate component of the input data of the entire image using the following procedure: 1) for a given input component, select the image mean as an initial threshold; 2) use this value to divide the image into two groups having new means m_1 and m_2 ; 3) define a new threshold as the average of m_1 and m_2 ; and 4) iterate until the adaptively computed threshold remains constant. The complementary roles of the split and merge thresholds used to segment the image and the adaptive labelling thresholds used to label the segments (cloudy versus cloud-free) are discussed in the next section.

Given a cluster mean vector with components $(\alpha_0,$

δ_0, T_0), the cluster is labeled as cloud if it falls outside the polyhedron defined by the adaptive labelling thresholds and the extrema of the cluster means (Fig. 1a). Algebraically, a cluster will be labeled as cloud if the inner product of the cluster mean vector and the adaptively determined decision plane is positive (i.e., if the normal vector from the decision plane to the cluster center points outward from the polyhedron). If we write the equation of the plane as

$$ma + nT + \delta - (\delta_{\min} + ma_{\min} + nT_{th}) = 0 \quad (10a)$$

where m and n are defined as

$$m = \frac{(\delta_{th} - \delta_{\min})}{(a_{th} - a_{\min})} \quad (10b)$$

and

$$n = \frac{(\delta_{\min} - \delta_{th})}{(T_{\max} - T_{th})} \quad (10c)$$

Then the signed distance, d_s , from the plane to the cluster center is

$$d_s = \frac{ma_0 + nT_0 + \delta_0 - \delta_{\min} - ma_{\min} - nT_{th}}{\sqrt{m^2 + n^2 + 1}}. \quad (10d)$$

It is important to emphasize that eqs. 10a and 10d provide a cluster-specific labelling rule for determining if a given cluster falls outside the decision polyhedron ($d_s > 0$) and hence is a cloud (Figure 1a). Eqs. 10a and 10d do not compute a single static threshold for the entire scene.

Complementary Roles of Clustering and Labelling Procedures

The clustering procedure (step 1 of the ASMC algorithm) uses cluster split (T_s)- and merge (T_m)-thresholds to partition the original spectral and derived input space (α_2, T_4, δ) into a set of homogeneous regions in cluster space. The adaptive-labelling thresholds ($a_{th}, \delta_{th}, T_{th}$) partition the cluster space into the geophysical domains of interest (cloudy, cloud-free). The cluster thresholds are determined solely from the spectral and derived properties of the specific image under study using the dynamic computational procedures defined in detail in appendix C of Gallaudet and Simpson (1991). The adaptive-labelling thresholds are computed as defined in the section entitled "Labelling."

If the adaptive-labelling thresholds were used singularly, then they would behave very much like the traditionally used static thresholds. Use of the decision polyhedron, however, allows us to apply the individual adaptive thresholds in a novel simultaneous sense (novel and simultaneous as opposed to the more traditional simple "and" or "or" logical conditions). Moreover, the joint adaptive-labelling thresholds are applied to the clusters and not to individual pixels as generally has

been done. The variance associated with even small subsets of adjacent pixels can give rise to noise in a statistical sense, which often causes single and/or logically linked sequences of static thresholds to break down. This problem is resolved by the initial clustering step of the ASMC algorithm because the joint-labelling criteria are only applied to statistically homogeneous structures (i.e., the clusters) whose statistical signatures are more clearly indicative of cloud or cloud-free conditions than are the signatures of individual pixels.

The difference between traditional thresholding of individual pixels and the use of clustering [with splitting (T_s)- and merging (T_m)-thresholds] to segment the scene into homogenous natural groupings, followed by cluster labelling (with adaptive labelling thresholds a_{th}, δ_{th} , and T_{th}) is clearly shown in Figure 2. These data are for a 1024×1024 pixel² region of western Europe (Table 1, Image 7) and thus the data set contains about one million pixels minus the excluded ocean pixels ($\sim 41.9\%$ Table 5). Figures 2a, 2b, and 2c show these data subsampled every 10th line by every 10th sample (solely to make the illustration legible) and plotted as T_4 versus a_2 , $T_3 - T_4$ versus a_2 and $T_3 - T_4$ versus T_4 scatterplots, respectively. Each of these scatterplots corresponds to a subdomain of the input vector space used to define the decision polyhedron in Figure 1. Examination of these scatterplots shows no natural breaks in the data with which to infer a static set of thresholds. Figures 2d, 2e, and 2f show the results obtained by using a set of static bispectral thresholds (e.g., albedo, temperature) to classify the data. Pixels shown in gray correspond to clouds in the scene as identified by the static thresholds, whereas pixels shown in black correspond to cloud-free land pixels. Figures 2g, 2h, and 2i show analogous plots of the cluster-mean values for the 11 clusters produced by the ASMC clustering (Step 1) for this image (Table 3). The locations of the adaptive-labelling thresholds constructed by the second step of the ASMC procedure also are shown in these panels as solid triangles on the axes of these scatterplots. Figures 2j, k, and l show the results obtained using the ASMC process, which combines dynamic scene-specific clustering (Fig. 2g, h, and i) with dynamic adaptive labelling of the clusters. Whereas the locations of the cluster centers within the spectral and derived input space are shown in Figures 2g, h, and i, the actual locations of pixels within the dynamically labeled clusters are shown in Figures 2j, k, and l. Again, gray points correspond to clouds over land as identified by ASMC and black corresponds to cloud-free land. Note carefully that the boundary between the two sets [gray pixels (cloudy), black pixels (cloud-free land)] does not have a simple shape. Static thresholds, however, can be successful only if such boundaries are simple (i.e., formed by simple perpendicular lines). The ASMC decision polyhedron is much more capable of analyzing such complex decision surfaces

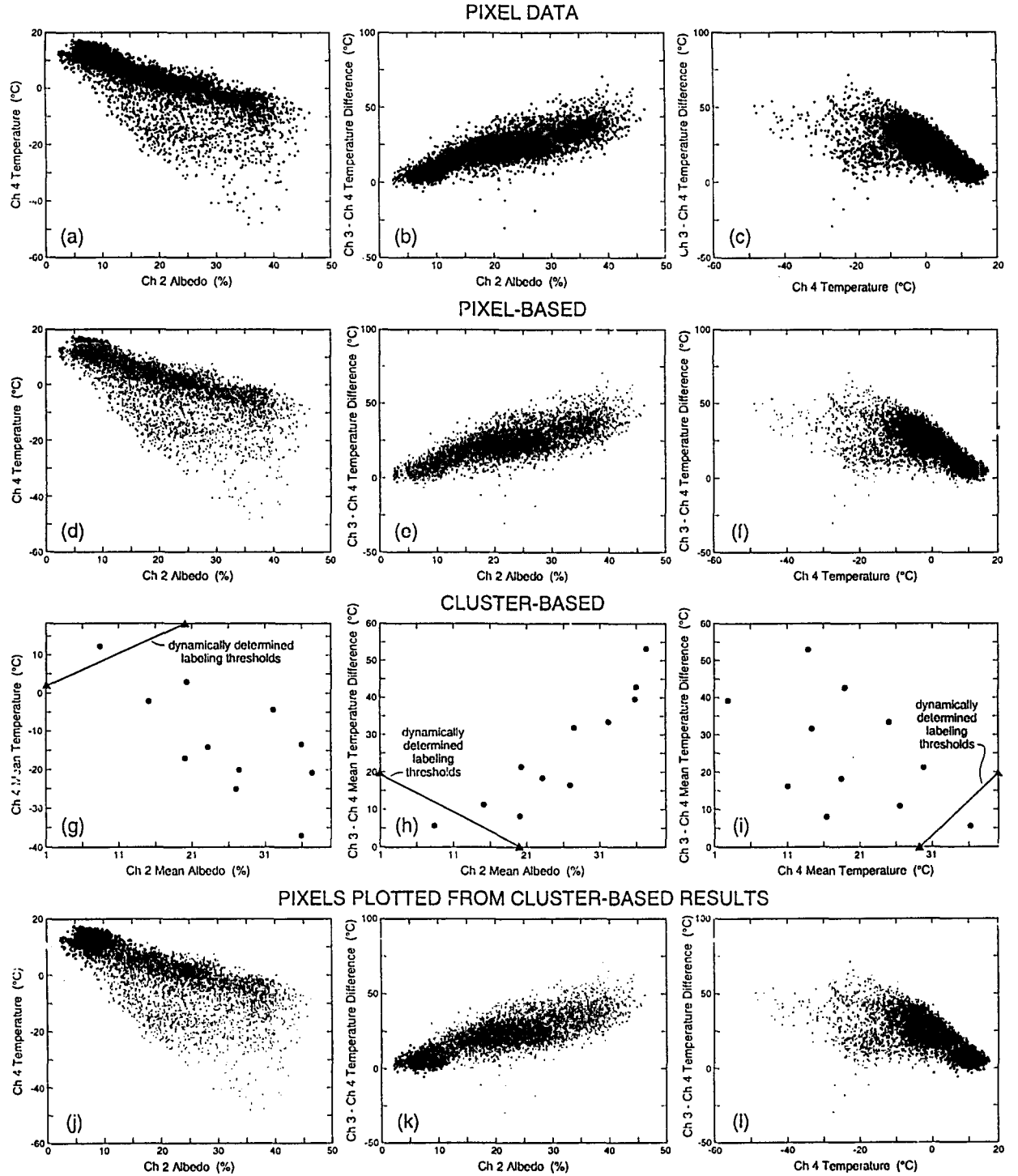


Figure 2. Scatterplots of T_4 vs. a_2 , $T_3 - T_4$ vs. a_2 and $T_3 - T_4$ vs. T_4 in panels a, b, and c, respectively. Data are from Image 7 of Western Europe (Table 1) and decimated every 10th line and every 10th sample for visualization only. Panels d, e, and f show results of cloud-detection process using static thresholds. Black corresponds to cloud-free land and gray to clouds. Analogous plots, but for the 11 clusters (Table 3), produced by the segmentation step of the ASMC algorithm for Image 7 are shown in panels, g, h, and i. Mean cluster centers are shown as solid circles. The adaptively determined labelling thresholds for the image are shown as small solid triangles, and d) panels j, k, and l show individual pixel classifications that resulted from the ASMC procedure based on clustering and adaptive labelling. Again, black is clear land and gray corresponds to cloud. See section entitled “Complementary Roles of Clustering and Labelling Procedures” for details.

than are static threshold methods. (See the ASMC algorithm section of the Discussion.)

The importance of the ASMC joint, multispectral approach to cloud detection also is supported by other work using very different techniques. Specifically, Yhann and Simpson (1995) show scatterplots of TOVS-derived total precipitable water vapor as a function of $T_3 - T_4$ temperature differences. These plots can produce nonlinear decision surfaces, especially as the total precipitable water vapor exceeds 30 gm cm^{-2} . Yhann and Simpson (1995) also performed many neural network experiments on cloud detection in satellite scenes and found that the accuracy of the neural network classification increased from 83% to 91% when channel-2 albedo was used, in conjunction with the $T_3 - T_4$ temperature differences, as inputs to the neural network. Simpson and Keller (1995), using fuzzy-logic techniques to separate cloud from sea ice, found that neither the $T_3 - T_4$ temperature difference nor the separation of the emitted from the reflected component of the T_3 signal was robust enough to segment clouds from sea ice and from cloud-free ocean. The accuracy of the fuzzy-logic classification, however, greatly improves when channel-1 and channel-2 data are added to the fuzzy-logic procedure. In general, such a joint, multispectral approach (e.g., ASMC, neural networks, fuzzy logic) is preferred for the classification of complex satellite scenes.

The Uncertainty Criteria

In addition to the binary clear/cloud classification defined by the sign of the distance from the decision plane to cluster center (Eq. 10d), the labelling scheme also can make use of the actual distance from the decision plane and use it to attach a reliability estimate to the resulting label. Thus, a cluster with a large negative distance is interpreted as clear land with a high degree of certainty; a large positive distance is interpreted as cloud with a high degree of certainty. The smaller the distance (either positive or negative), the less certain is the final label (either cloud or clear land). Because of the uncertain labels associated with cluster that lie very close to the decision plane, additional planes (parallel to the original) can be added to the decision rules to define zones of ambiguity or zones where very uncertain labels are likely to occur (Figure 1b). The definition of these additional planes depends on the application (i.e., recovering as much clear land as possible, ensuring that clear land labels are assigned with a high degree of certainty, or identifying regions of possible subpixel cloud contamination).

For the purposes of this article, two additional planes were defined. The first, just outside of the original decision plane effectively increases the size of the polyhedron in which clear land labels should be assigned.

The second additional surface is even further beyond the original plane and the area between this plane and the first additional plane defines a region where ambiguous labels are assigned. The actual placements of these planes are defined as 5% and 12% of the maximum corner-to-corner distance beyond the original decision plane (where the corner-to-corner distance refers to the dashed cube in Fig. 1a). With the addition of these two planes, the labelling rules can be summarized as follows:

$$D = \sqrt{(a_{\max} - a_{\min})^2 + (T_{\max} - T_{\min})^2 + (\delta_{\max} - \delta_{\min})^2}$$

$$\begin{aligned} d_s < 0.05D &\rightarrow \text{clear} \\ 0.05D \leq d_s \leq 0.12D &\rightarrow \text{ambiguous} \\ d_s > 0.12D &\rightarrow \text{cloud} \end{aligned} \quad (11)$$

where d_s is the signed distance defined by Eq. 10d, and D is the corner-to-corner distance of the cube that completely encloses all of the cluster-mean vectors. Thus, if d_s is negative or a small positive number less than $0.05 \times D$, the cluster is labeled cloud-free land.

Clusters labeled as ambiguous are examined in more detail using additional physical criteria. Specifically, any cluster that either has a mean temperature greater than 20°C , the climatological value for cloud-free land (Stowe et al., 1991), or has a mean albedo less than the mean albedo of previously assigned cloud-free clusters is reassigned from the ambiguous to the clear-land label. Clusters not so reassigned retain the ambiguous label in the final classification.

The above labelling criteria worked well for all AVHRR scenes used in this study. These labelling criteria also proved very suitable over highly reflective desert regions, which traditionally have been difficult to cloud screen properly. Thus, the ASMC algorithm generally does not use information from an a priori mask in its segmentation and labelling steps as does the GSMC algorithm for GOES data. This difference largely results from two major differences between AVHRR and GOES data: 1) the resolution of AVHRR thermal data is 1 km, whereas that of GOES thermal data is 8 km and often resampled to 4 km; and 2) AVHRR channel-3 data provide unique capabilities for cloud detection not offered by the GOES thermal channels. Information from an a priori mask, however, can be incorporated into the ASMC process should it be judged desirable.

DATA AND PREPROCESSING

AVHRR data were selected from the archive of the Scripps Satellite Oceanography Center (SSOC). Data over the western United States and Mexico were captured directly by SSOC. Images from Europe, Africa, Australia, South America, and Asia were captured locally at high resolution [High Resolution Picture Transmission (HRPT) mode] by NOAA field stations and subse-

Table 1. Geographic, Temporal, Satellite, and Climate Information for the Images Shown in This Study

Image #	Location	Center Latitude	Center Longitude	# of Pixels	Date	Local Time	Satellite	Climate
1 (AR7733)	Baja	27 N	113 W	512 ²	Mar. 15, '91	14 26 24	noaa11	Hot arid
2 (AR2589)	Gobi	45 N	110 E	512 ²	Apr. 11, '82	14 00 15	noaa7	Hot arid
3 (AR2695)	N. Africa	30 N	5 E	512 ²	Oct. 2, '82	13 35 40	noaa7	Hot arid
4 (AR3981)	Somalia	10 N	49 E	512 ²	Oct. 9, '84	15 24 40	noaa7	Hot arid
5 (AR8987)	Grt. Lakes	45 N	85 W	1024 ²	Aug. 11, '92	14 58 00	noaa11	Humid mid-latitude
6 (AR5510)	Rockies	46 N	112 W	512 ²	Sep. 7, '86	14 38 00	noaa9	High altitude
7 (AR2750)	W. Europe	46 N	2 W	1024 ²	Oct. 19, '82	13 33 40	noaa7	Temperate marine
8 (AR2688)	Italy	40 N	10 E	1024 ²	Oct. 9, '82	14 53 40	noaa7	Dry subtropical
9 (AR3982)	S. Africa	24 S	25 E	512 ²	Oct. 9, '84	15 54 40	noaa7	Semi-arid tropical
10 (AR4046)	G. Britain	54 N	4 E	512 ²	Oct. 3, '84	15 28 31	noaa7	Temperate marine
11 (AR2377)	S. America	25 S	57 W	512 ²	May 19, '82	14 31 50	noaa7	Wet & dry tropical
12 (AR2293)	S. America	25 S	57 W	512 ²	Apr. 22, '82	14 55 35	noaa7	Wet & dry tropical
13 (AR3980-2)	Pakistan	23 N	70 E	512 ²	Oct. 9, '84	16 40 40	noaa7	Semi-arid tropical
14 (AR3982-2)	Mozamb.	18 S	37 E	512 ²	Oct. 9, '84	15 54 40	noaa7	Wet & dry tropical
15 (AR7948)	Oregon	45 N	120 W	512 ²	Sep. 14, '91	14 43 30	noaa11	Semiarid mid-latitude

quently archived at SSOC. Data from the Great Lakes region of the U.S. were supplied by NOAA's National Operational Hydrologic Remote Sensing Center in Minneapolis, MN. The general geographical location of each image, its center latitude and longitude, size, date, local time, satellite, and climate characteristics are given in Table 1. Climate characteristics were taken from the World Scene Climate Regions analysis (pp. 306-307) of the International Atlas by Rand McNally (1977). Albedo and temperature statistics for these images prior to cloud screening are given in Table 2. These images provide examples of both highly variable atmospheric moisture conditions and highly variable land forms: 1) dry desert areas (e.g., Baja California, Sahara Desert), 2) moist tropical environments (e.g., Amazon basin, India, Northeast Africa), and 3) highly variable temperate mid-latitude regions (e.g., the Rocky Mountains, Great Lakes

region, France). Data were ingested at full spatial resolution from archive tapes and calibrated to geophysical units (Lauritson et al., 1979 and updates). A total of 125 images were analyzed. Space precluded a detailed presentation of results of all 125 images. Therefore, only representative results for all the climates and land forms studies are presented here.

RESULTS

In this section, cloud masks produced by the ASMC algorithm for a wide variety of terrestrial environments are presented. Each image consists of four panels. Panel a corresponds to calibrated albedo with low values mapped to dark shades and high values mapped to lighter shades. Panel b corresponds to calibrated temperatures for AVHRR channel 4. Dark shades are

Table 2. Image Albedo and Temperature Statistics Before Cloud Detection

Image Reference Name	Minimum Albedo (%)	Maximum Albedo (%)	Mean Albedo (%)	Standard Deviation Albedo (%)	Minimum Thermal (°C)	Maximum Thermal (°C)	Mean Thermal (°C)	Standard Deviation Thermal (°C)
AR7733	0.66	44.66	11.73	5.32	-29.75	38.94	24.04	10.37
AR2589	0.01	98.07	13.34	11.56	-86.73	36.82	4.50	22.36
AR2695	9.13	51.46	26.02	6.33	-53.38	45.00	16.25	24.33
AR3981	1.43	58.51	11.46	5.27	-59.04	39.76	26.18	12.40
AR8987	1.16	80.30	18.23	7.63	-95.67	34.53	8.97	12.28
AR5510	1.28	77.60	14.40	9.65	-72.54	41.45	7.82	21.55
AR2750	1.86	48.79	18.24	10.31	-50.20	18.98	2.88	10.42
AR2688	1.54	65.25	13.71	6.69	-57.11	44.18	13.39	14.76
AR3982	1.32	59.48	19.58	10.25	-71.53	37.27	-2.59	25.76
AR4046	1.22	51.24	10.63	5.06	-56.52	12.78	-14.60	19.02
AR2377	1.22	33.29	7.36	1.64	5.19	26.30	19.26	1.87
AR2293	1.75	65.03	15.18	6.27	-50.31	37.95	10.81	8.80
AR3980-2	1.54	39.70	8.44	3.34	-76.05	45.21	29.05	10.96
AR3982-2	1.32	45.58	7.12	3.56	-4.07	36.15	25.30	5.96
AR7948	0.76	29.77	7.10	2.06	-42.89	42.56	25.49	8.12

Ocean and cloud over ocean pixels are excluded from these statistics.

Table 3. Image Number with SSOC Archive Reference in Parenthesis

Image #	Location	Total Number of Clusters	Number of Cloud Clusters	Number of Clear Clusters	Number of Ambiguous Clusters	Number of Ambiguous to Cloud	Number of Ambiguous to Clear
1 (AR7733)	Baja	9	7	2	0	N/A	N/A
2 (AR2589)	Gobi	14	13	1	0	N/A	N/A
3 (AR2695)	N. Africa	7	5	2	0	N/A	N/A
4 (AR3981)	Somalia	12	11	1	0	N/A	N/A
5 (AR8987)	Grt, Lakes	8	7	1	0	N/A	N/A
6 (AR5510)	Rockies	10	9	1	0	N/A	N/A
7 (AR2750)	W. Europe	11	10	1	0	N/A	N/A
8 (AR2688)	Italy	15	12	3	0	N/A	N/A
9 (AR3982)	S. Africa	9	8	1	0	N/A	N/A
10 (AR4046)	G. Britain	15	13	1	1	0	0
11 (AR2377)	S. America	7	6	1	0	N/A	N/A
12 (AR2293)	S. America	11	9	1	1	0	1
13 (AR3980-2)	Pakistan	11	8	1	2	0	0
14 (AR3982-2)	Mozamb.	12	9	1	2	0	2
15 (AR7948)	Oregon	10	6	2	2	0	2

Geographical location, total number of clusters produced by the ASMC algorithm, number of cloud clusters, number of cloud-free (clear) clusters, number of ambiguous clusters and the number of ambiguous clusters relabeled to cloud and cloud-free (clear) classes. N/A means not applicable.

mapped to warm temperatures and lighter shades are mapped to colder temperatures. Panel c is identical to panel a except that the ASMC cloud mask has been overlaid. Yellow in the mask of panel c indicates confirmed clouds as determined by the ASMC algorithm, whereas pink indicates an ambiguous assignment. Panel d contains the cluster distribution produced by the first step of the ASMC algorithm. The number of colors in the bar code of this panel indicates the number of final clusters in the segmented image. Note, this can vary greatly from image to image. A statistical profile for the cluster distributions produced by the ASMC procedure for each image studied is given in Table 3. Some images contain a mixture of land and water (e.g., oceans, large lakes or rivers). For these images water is highlighted

in dark blue and lake shores or coastlines are designated in red. Water and coastline were identified using Geographical Information Systems (GIS) data and morphological transformations (Simpson, 1992). Water and coastline pixels are not used in the ASMC analysis. For a given image, the same gray scale stretch is used for panels a and c. Statistics for each of the images after cloud detection with the ASMC procedure are summarized in Table 4.

Desert Regions

Image 1 (Fig. 3) shows a large segment of Baja California and the Mexican mainland along with sections of the Pacific ocean and Gulf of California (dark blue) and was taken on 15 March 1991. The albedo (Fig. 3a) and

Table 4. Statistics, after Cloud Detection with the ASMC Procedure, are Given for the Images Used in This Study

Image #	Mean Land Albedo (%)	Mean Land Thermal (°C)	Mean Cloud Albedo (%)	Mean Cloud Thermal (°C)	Mean Ambiguous Albedo (%)	Mean Ambiguous Thermal (°C)
1 (AR7733)	9.90	28.55	18.38	7.60	N/A	N/A
2 (AR2589)	8.30	16.22	25.67	-24.16	N/A	N/A
3 (AR2695)	22.58	31.67	32.93	-14.73	N/A	N/A
4 (AR3981)	9.73	31.57	18.15	5.41	N/A	N/A
5 (AR8987)	13.64	17.53	25.08	-3.80	N/A	N/A
6 (AR5510)	9.81	20.38	23.85	-18.01	N/A	N/A
7 (AR2750)	8.36	12.29	24.78	-3.35	N/A	N/A
8 (AR2688)	10.47	21.67	20.00	-2.64	N/A	N/A
9 (AR3982)	9.42	24.74	25.89	-19.60	N/A	N/A
10 (AR4046)	6.26	5.49	13.66	-31.02	10.01	-4.79
11 (AR2377)	7.04	19.58	10.66	15.94	N/A	N/A
12 (AR2293)	11.70	16.21	19.72	3.78	N/A	N/A
13 (AR3980-2)	7.53	32.99	18.49	-7.49	11.72	13.25
14 (AR3982-2)	6.10	27.04	15.08	11.70	N/A	N/A
15 (AR7948)	6.97	26.46	9.91	5.74	N/A	N/A

thermal structures (Fig. 3b) of this image show several different types of cloud structures. The first step of the ASMC algorithm identified nine distinct clusters in the scene (Figure 3d), and the adaptive-labelling scheme was effective in unambiguously separating cloudy from cloud-free land pixels (Fig. 3c). Thus, no pink-colored pixels are shown in Figure 3c.

Image 2 (Fig. 4) is centered over the Gobi Desert of Asia and was taken on 11 April 1982. There are no discernible water formations in this scene. The albedo data (Fig. 4a) and the thermal data (Fig. 4b) contain more complicated structure than that shown in the Baja California scene (Fig. 3). In fact, the ASMC segmenter produced 14 separate clusters (Fig. 4d) compared to 9 for the Baja California scene (Fig. 3d). Again, the adaptive-labelling scheme accurately labeled all the major cloud structures and no ambiguous areas were identified. Moreover, the labeler was able to discriminate between clouds and bright underlying terrain in the left-hand part of the image using its temperature and temperature-difference criteria.

Image 3 (Fig. 5) is over North Africa and was taken on 2 October 1982. Although this scene has complex cloud structure (Fig. 5a, 5b), it is simpler to classify than either of the previous two images because of the high contrast between the cloud-free land mass and the clouds. Thus, a comparatively simple segmentation of only seven clusters was produced (Fig. 5d). The accuracy of the corresponding cloud mask is clear (Fig. 5c).

Image 4 (Fig. 6) is a scene over Somalia taken on 9 October 1984. The albedo (Fig. 6a) and thermal (Fig. 6b) data show that this is a relatively complex scene to segment. A total of 12 clusters was produced by the ASMC segmenter (Fig. 6d). The real difficulty with this scene, however, is the presence of both large- and small-scale cloud structures in the data as well as the occurrence of both highly reflective arid land surfaces (Somalia) and moist tropical regions (interior regions of the Horn of Africa). The labeler was successful in identifying these complex structures.

Temperate Regions

Image 5 (Fig. 7) shows a region of the United States and Canada in August 1992 that includes the Great Lakes. Different types of clouds are discernible in the albedo (Fig. 7a) and thermal data (Fig. 7b). This was a relatively simple image to segment, only eight clusters appear in the final segmented image (Fig. 7d). The ASMC labeler unambiguously identified the cloud structure.

Image 6 (Fig. 8) shows a region over the Rocky Mountains taken on 7 September 1986. The reflectance data (Fig. 8a) and the thermal data (Fig. 8b) show complex structure over much of the scene (Fig. 8d). Cloud-free land pixels, however, were correctly identified (Fig. 8c).

Image 7 (Fig. 9) contains regions of France, Spain, England, and Portugal taken on 19 October 1982. Parts of the Atlantic Ocean, Mediterranean Sea, and English Channel also are in this scene. Eleven individual clusters (Fig. 9d) were identified by the ASMC segmenter and the labelling (Fig. 9c) appears accurate.

Image 8 (Fig. 10) contains a region of the Mediterranean Sea and parts of Italy, France, Yugoslavia, North Africa, and the islands of Sicily, Sardinia, and Corsica. Data were taken on 9 October 1982. This was a complex scene to segment (15 clusters in Fig. 10d) because of the broad ranges of albedo (Fig. 10a), temperature (Fig. 10b) and viewing angles (not shown). Nonetheless, the ASMC segmenter and labeler accurately identified the cloud structures in this complex scene. Moreover, the cloud-free desert region of North Africa was correctly classified.

Image 9 (Fig. 11) shows a region of South Africa taken on 9 October 1984. The albedo (Fig. 11a) and thermal (Fig. 11b) data show a variety of complex cloud structures. The segmentation (Fig. 11d) and the cloud mask (Fig. 11c) accurately isolate the cloud structures in the data.

Image 10 (Fig. 12) shows the British Isles and part of Ireland. This image was taken on 3 October 1984. These data are representative of a temperate marine environment. Figures 12a and 12b show very complex cloud structure of different types. Fog may occur in part of the coastal domains. This complex image resulted in 15 distinct clusters (Fig. 12c). The final classification contains several regions of ambiguous pixels (pink) as well as identified clouds (yellow).

Tropical Regions

Image 11 (Fig. 13) shows a region of the Amazon drainage taken on 19 May 1982. Small irregular cloud structures are seen in both the albedo (Fig. 13a) and thermal (Fig. 13b) data. The segmentation for this scene is relatively simple and produced only seven distinct clusters (Fig. 13d). The final cloud mask (Fig. 13c) correctly labeled the clouds.

Image 12 (Fig. 14) shows the same region of the Amazon drainage as shown in Fig. 13 but it was taken on 22 April 1982. Cloud structure (Fig. 14a, b) in this scene is more varied and complex than that shown in Figure 13a, b. Now 11 clusters are needed to segment the data (Fig. 14d) and the final cloud mask accurately identifies these more complex clouds (Fig. 14c).

Image 13 (Fig. 15) shows a complex region of Pakistan and India taken on 9 October 1984. This region contains the Gulf of Kutch, which has numerous marshlands. Eleven clusters constitute the final segmentation (Fig. 15c). Labelling, however, produced a few ambiguous clusters in some regions of the image because the bright clouds generally are warm and moist. Thus, their thermal structure is somewhat ill-defined. The cloud

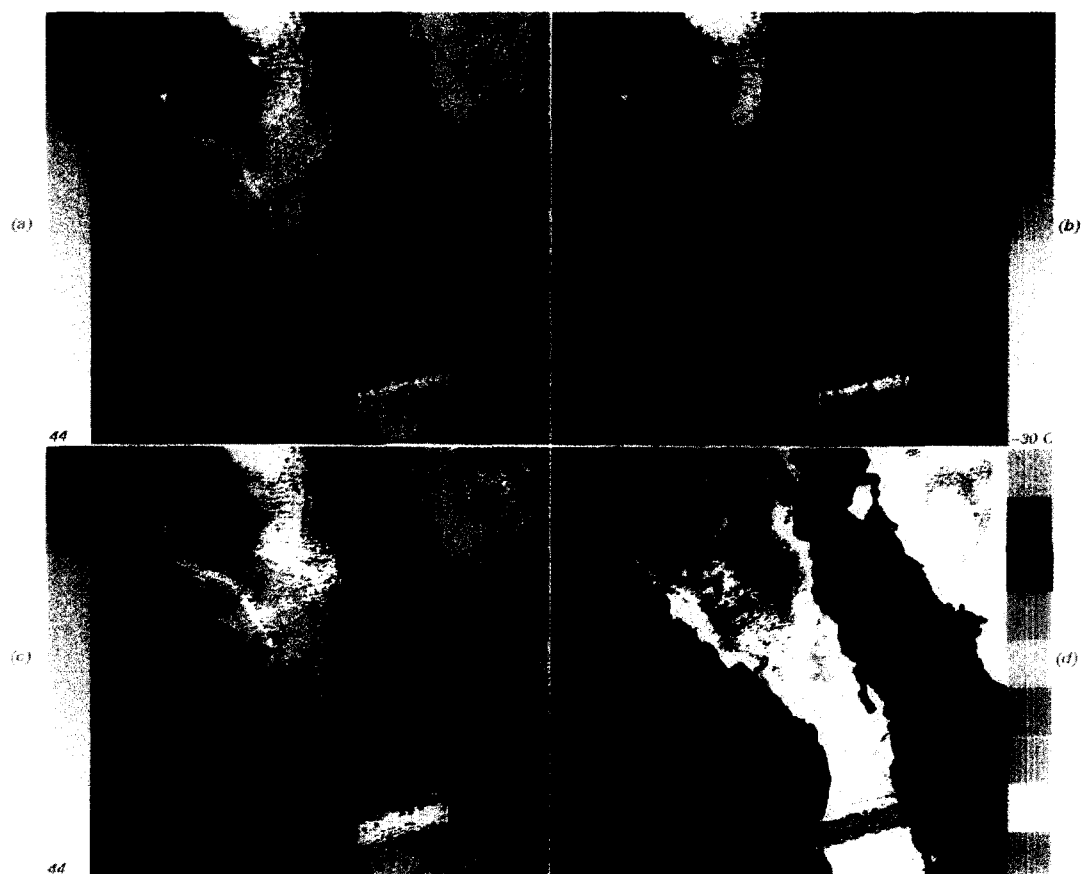


Figure 3. An AVHRR image showing parts of Baja California and the Mexican mainland, a) albedo, b) temperature, c) the ASMC cloud mask with yellow pixels unambiguously marked as clouds and pink pixels tagged as ambiguous. Note, no ambiguous pixels were found in this scene. d) The segmentation produced by the ASMC clustering procedure. Water bodies, when present, are shown in blue, and red designates either shore or coastline. Water and coastline pixels, however are not used in the ASMC analysis. (Image Reference #AR7733.)

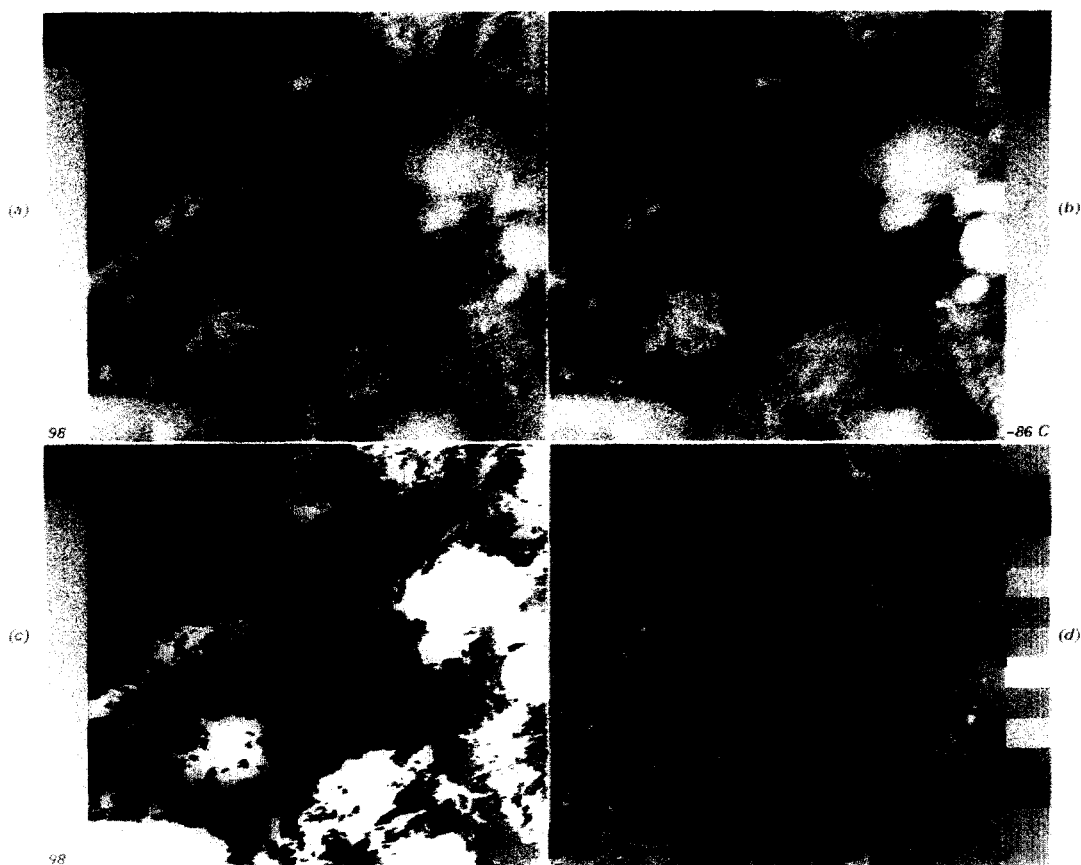


Figure 4. Analogous to Figure 3, except over the Gobi desert. (Image Reference #AR2589.)

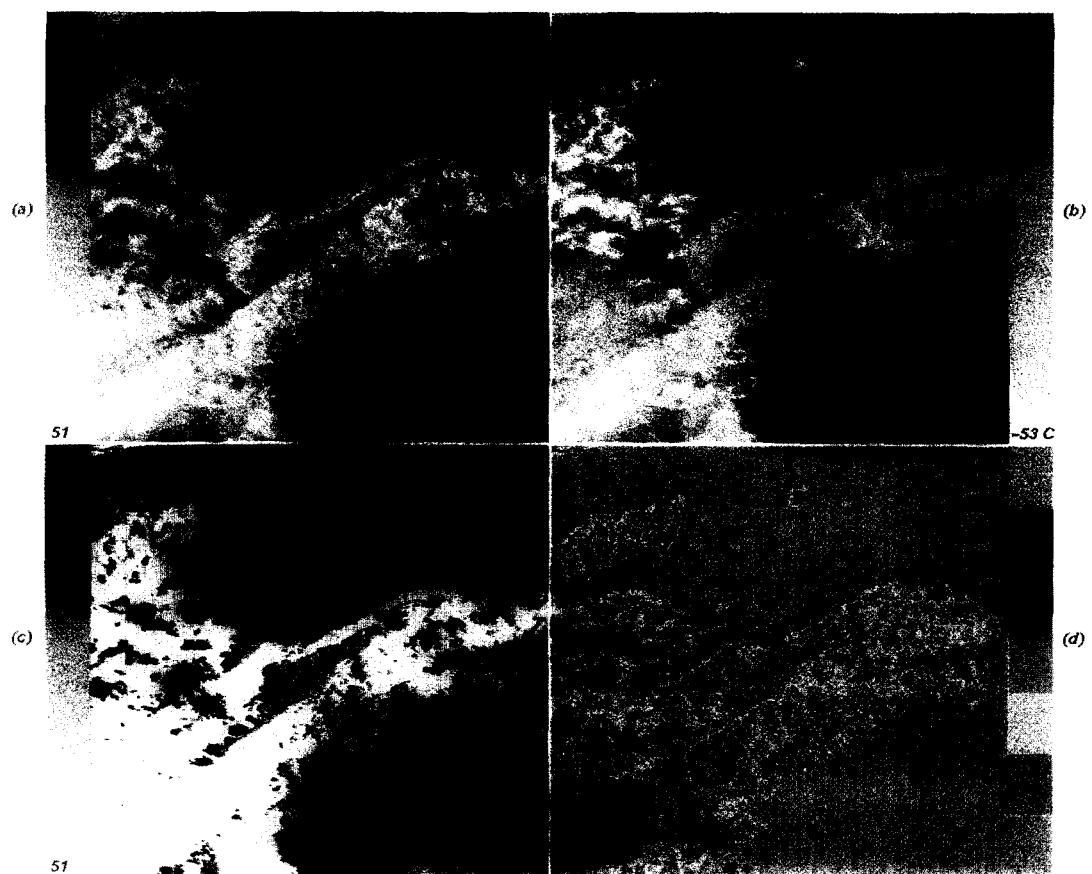


Figure 5. Analogous to Figure 3, except over North Africa. (Image Reference #AR2695.)

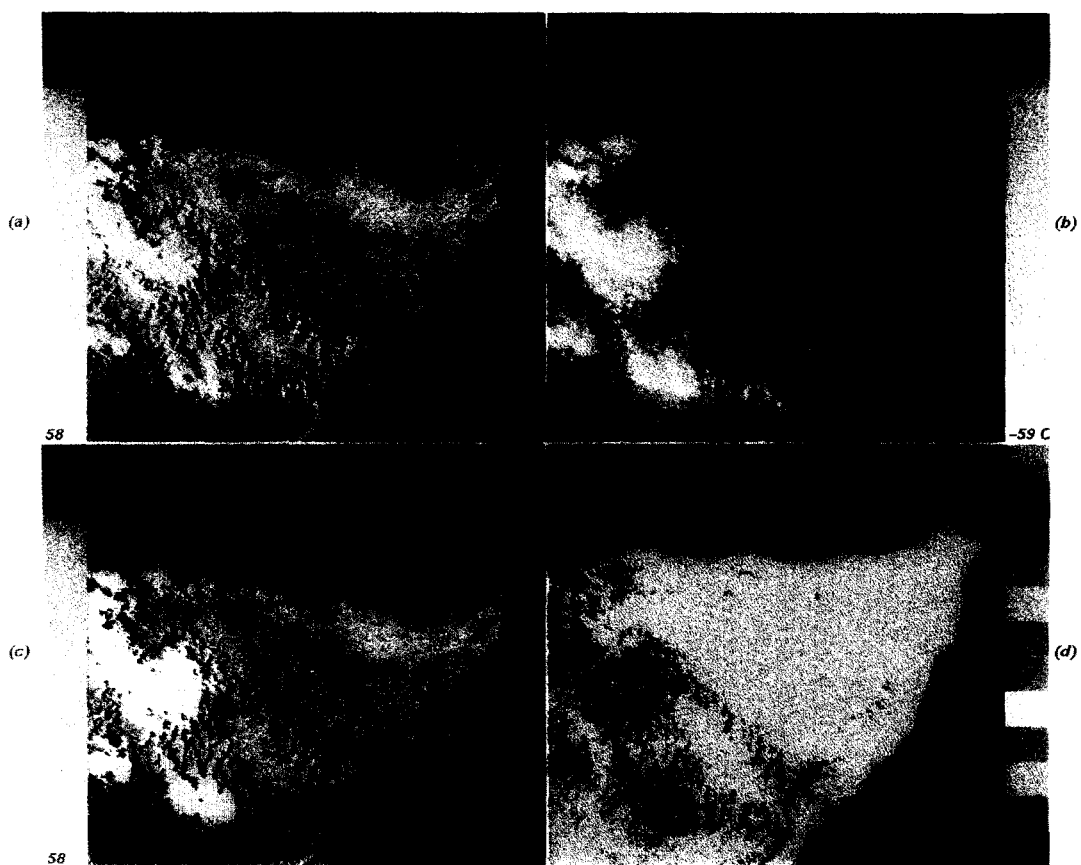


Figure 6. Analogous to Figure 3, except over part of Somalia. (Image Reference #AR3981.)

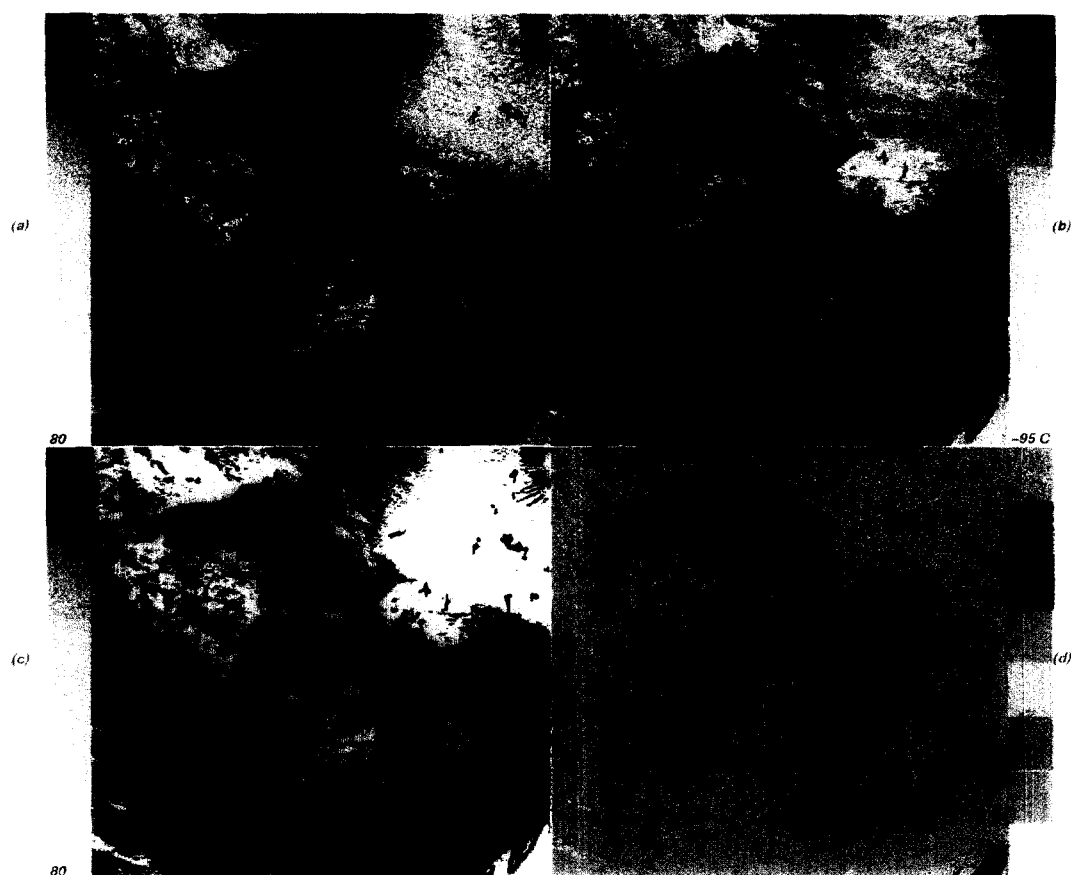


Figure 7. Analogous to Figure 3, except over the Great Lakes region of the United States and Canada. (Image Reference #AR8987.)

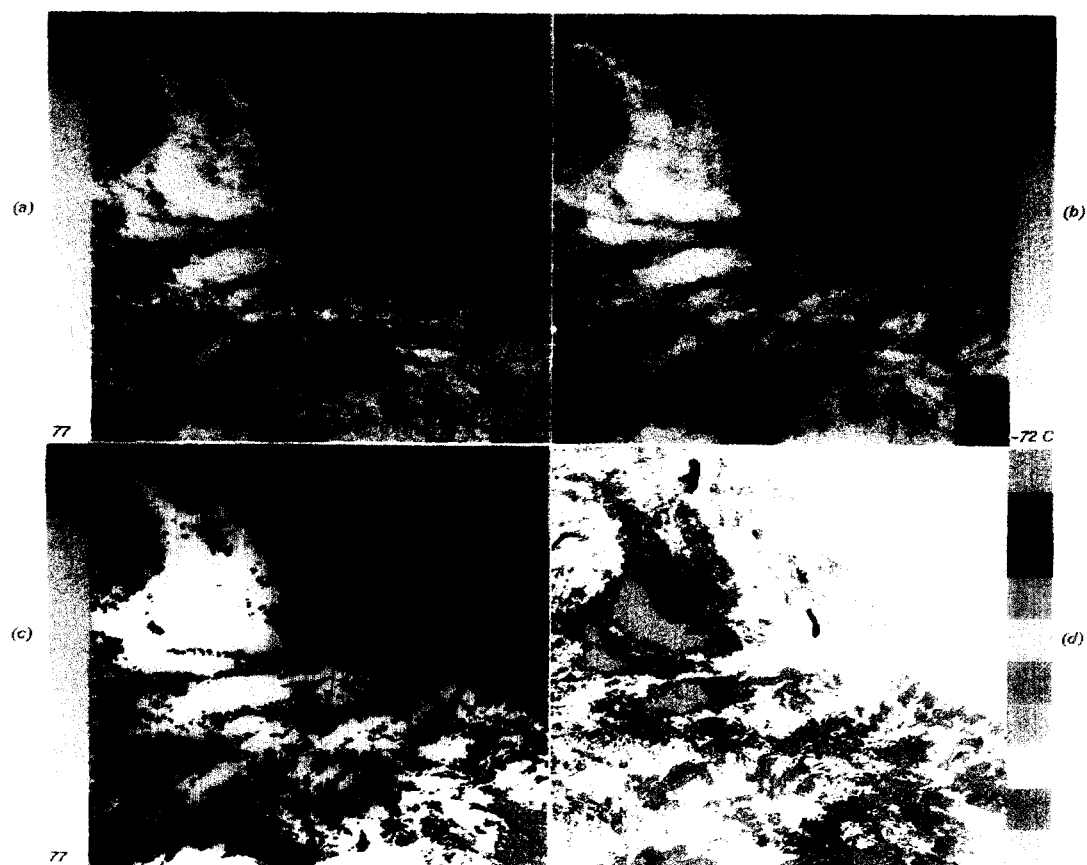


Figure 8. Analogous to Figure 3, except over the Rocky Mountains. (Image Reference #AR5510.)

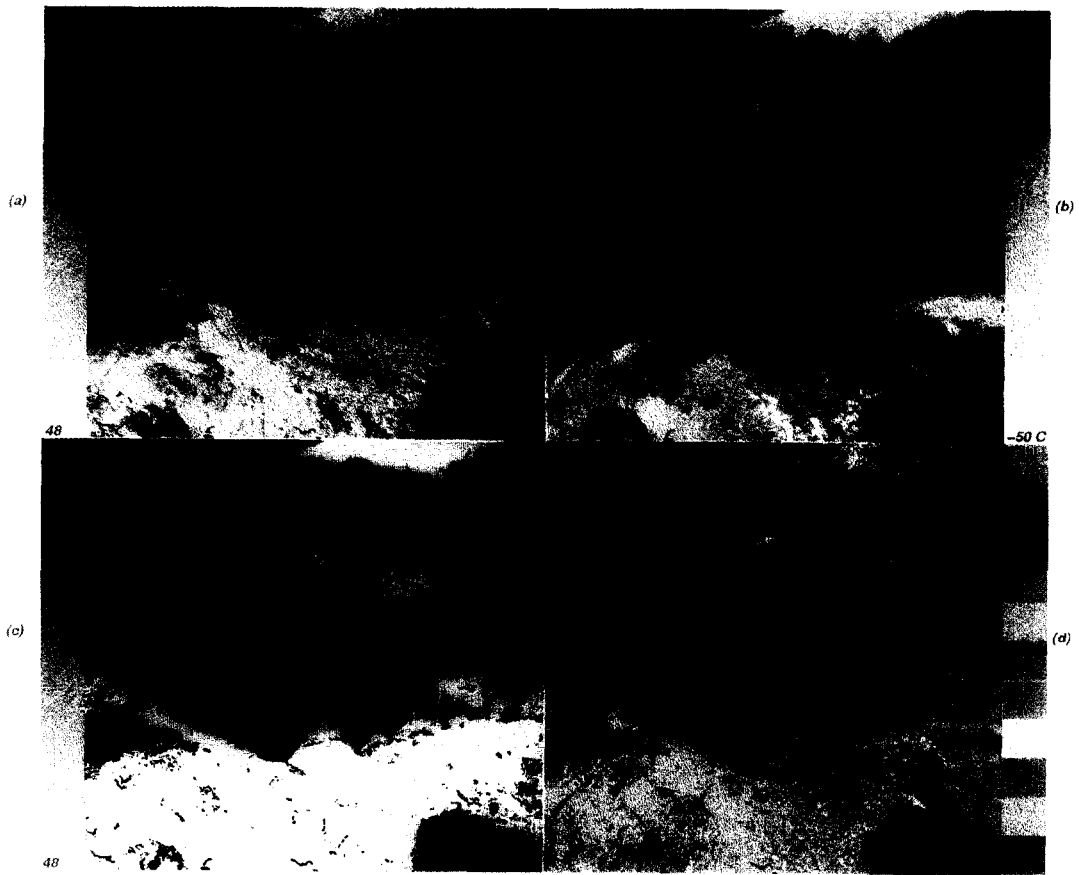


Figure 9. Analogous to Figure 3, except over a region of western Europe. (Image Reference #AR2750.)



Figure 10. Analogous to Figure 3, except over a region of the Mediterranean basin. (Image Reference #AR2688.)

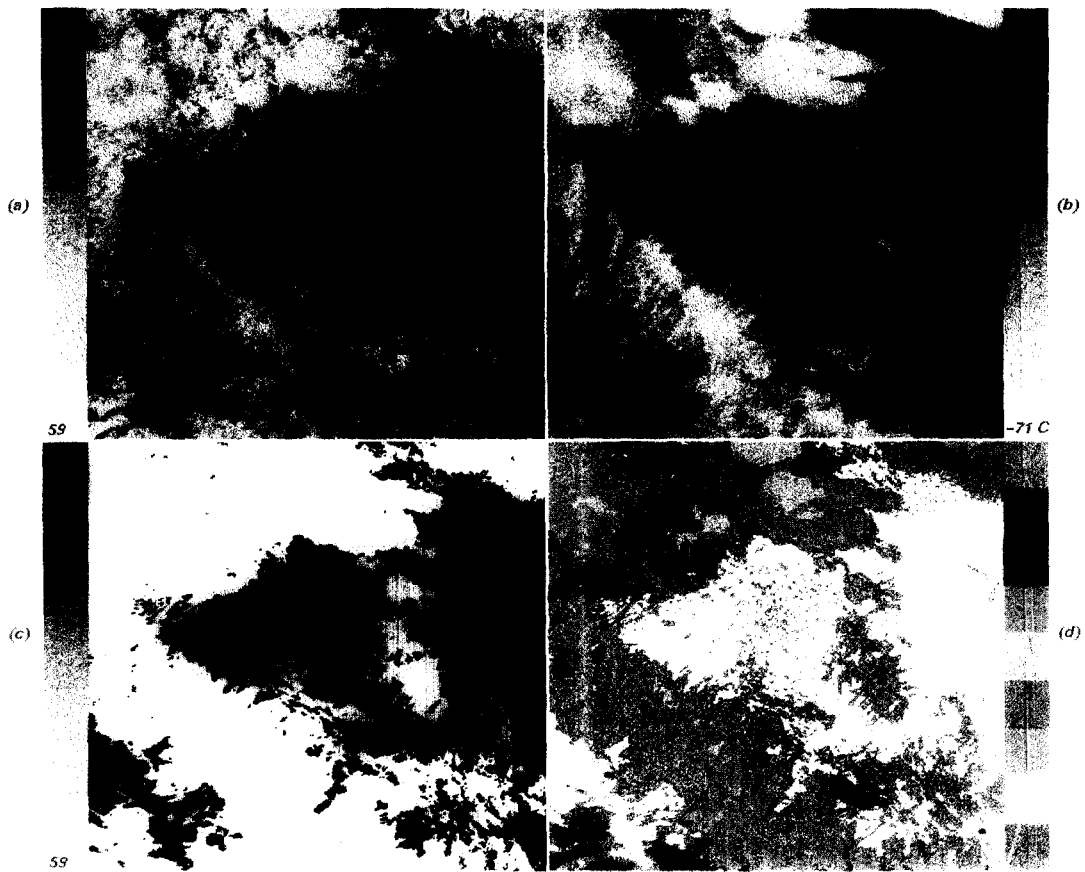


Figure 11. Analogous to Figure 3, except over a region of South Africa. (Image Reference #AR3982.)

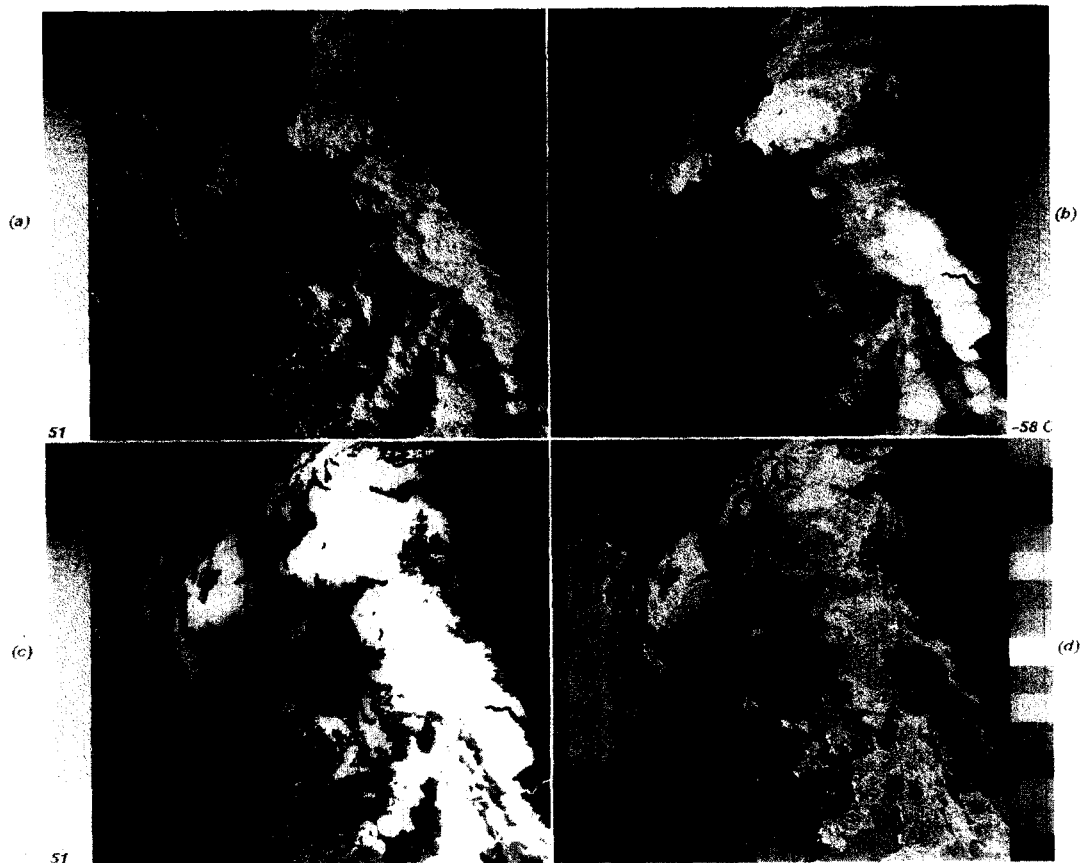


Figure 12. Analogous to Figure 3, except of the British Isles and part of Ireland. Ambiguously labeled pixels in this scene are shown in pink. (Image Reference #AR4046.)

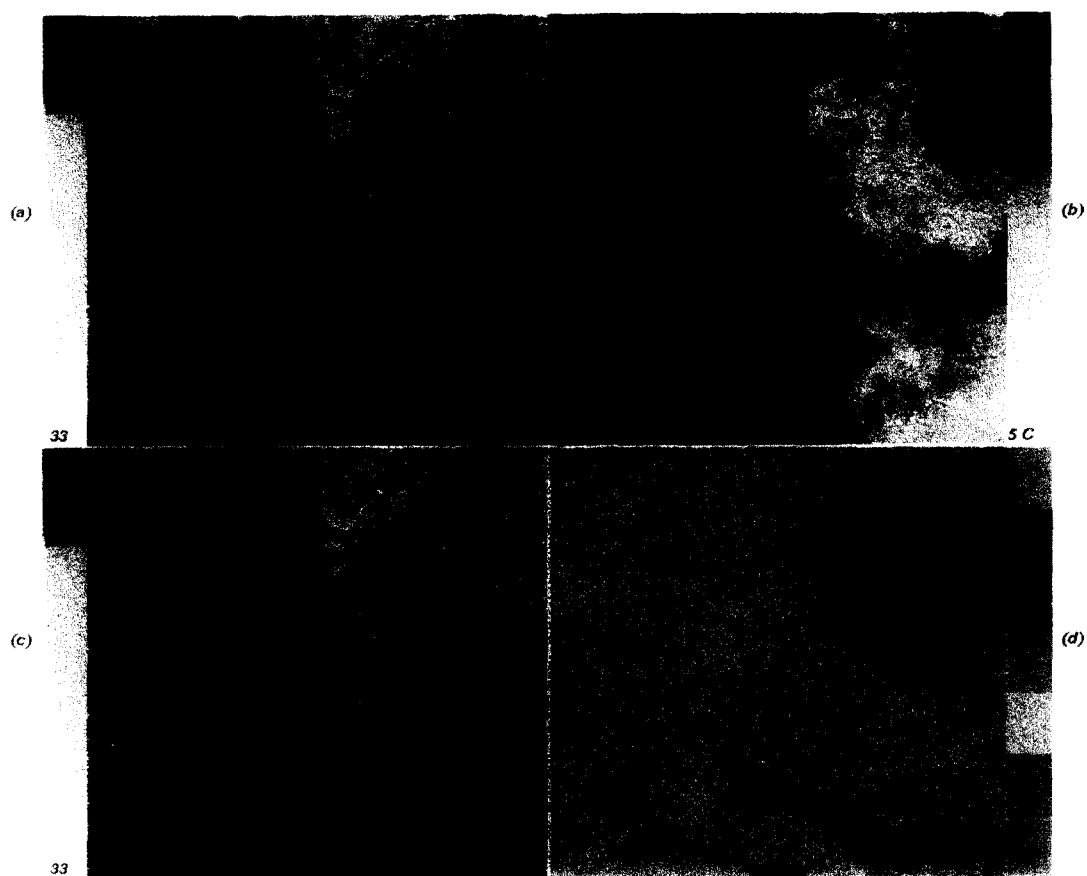


Figure 13. Analogous to Figure 3, except of a region of the Amazon Basin. (Image Reference #AR2377.)

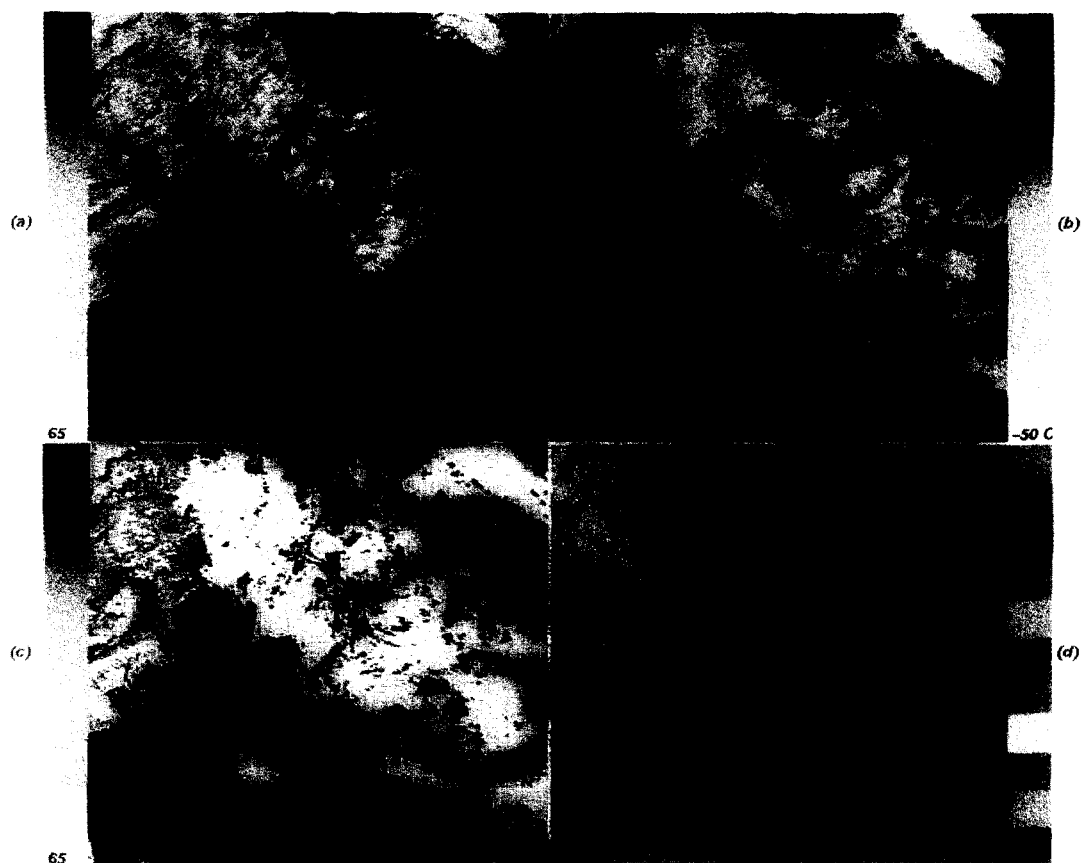


Figure 14. Analogous to Figure 3, except of the same region of the Amazon basin as in Figure 13 but at a different time. (Image Reference #AR2293).

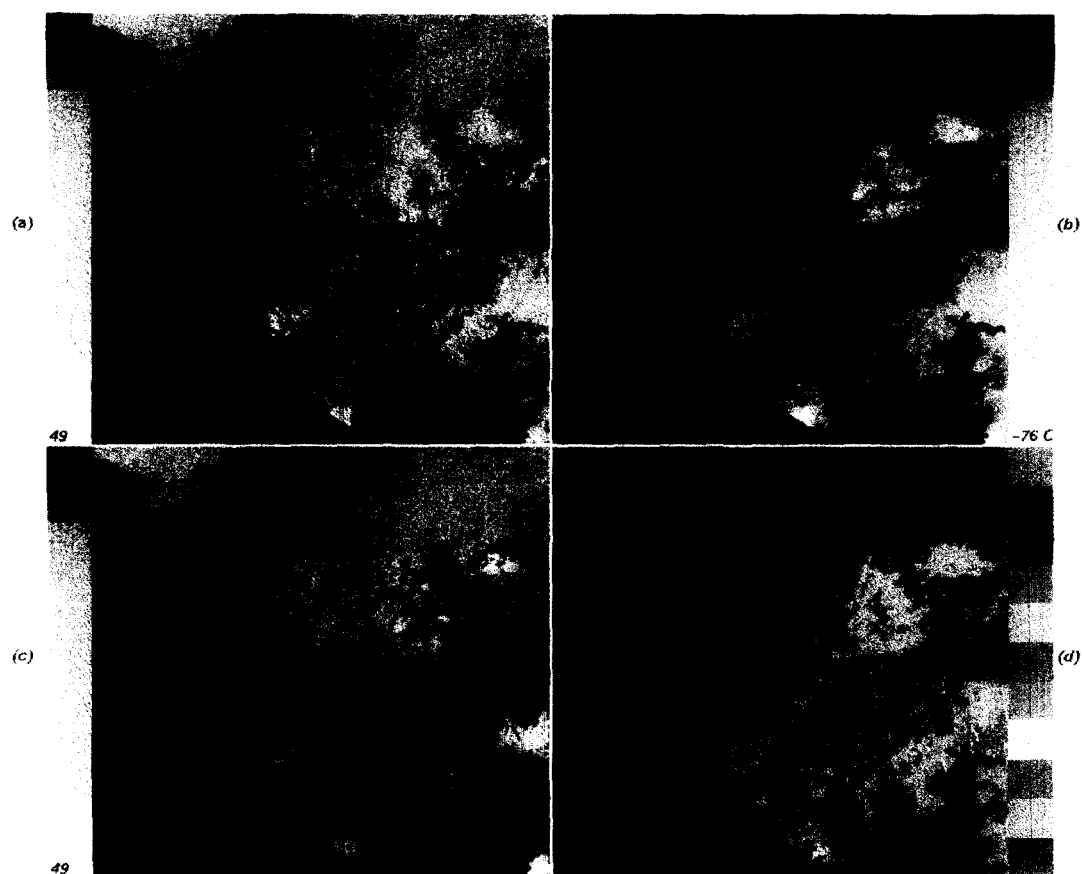


Figure 15. Analogous to Figure 3, except over a region of Pakistan and India. Ambiguous pixels are in pink. (Image Reference #AR3980.)

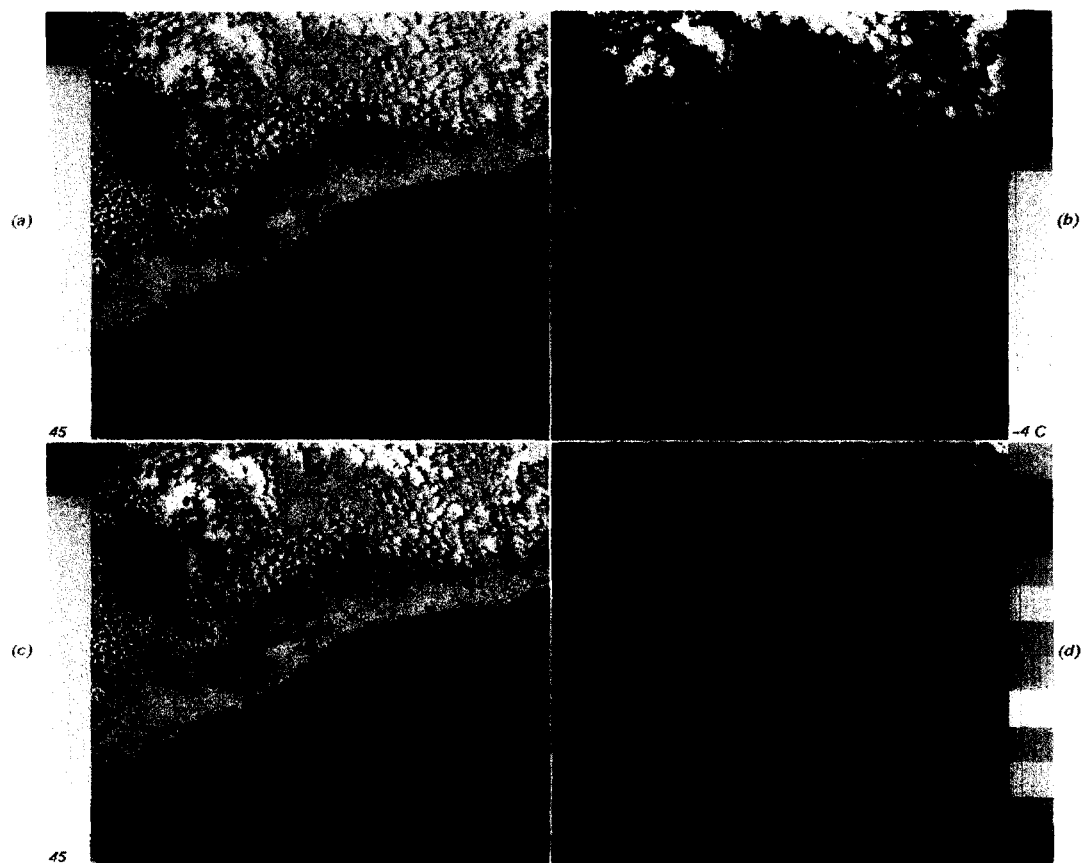


Figure 16. Analogous to Figure 3, except over a region of Mozambique, Africa. (Image Reference #AR3982.)

mask (Fig. 15c) contains many uncertain areas (colored pink). These pixels probably are cloud but cannot be classified as unambiguously as the pixels colored yellow in Figure 15c.

Image 14 (Fig. 16) shows a region of Mozambique taken on 9 October 1984. Again, cloud structure is fairly complex and diverse (Fig. 16a, b). Twelve clusters (Fig. 16d) were required to segment this scene. These clouds were not as warm and moist as those over the Pakistan and India scene (Fig. 15), and hence the labeler unambiguously assigned cloudy pixels as cloud (pixels marked yellow in Fig. 16c).

DISCUSSION

The ASMC Algorithm

Accurate cloud detection in satellite scenes over land is a difficult task complicated by spatially and temporally varying surface reflectivities and emissivities (Figs. 3–16). For these reasons neither image-specific static thresholds nor climatologically based thresholds are likely to produce robust cloud detection in land scenes. The problem often is further compounded by sensor-related problems, which in the case of the AVHRR instrument manifest themselves as highly variable noise in channel 3 mid-infrared data and atmospheric water vapor attenuation of the channel-4 infrared temperature data.

The AVHRR Split-and-Merge Clustering algorithm for cloud detection over land provides a way to minimize these difficulties. The algorithm consists of two steps: 1) a split-and-merge clustering of the input data, which segments the scene into its natural groupings; and 2) a

cluster-labelling procedure that uses scene-specific, joint three-dimensional, adaptive thresholds to label the clusters as either cloud or cloud-free land. The input data consist of calibrated AVHRR channel-2 albedo, calibrated AVHRR channel-4 temperature, and a calibrated temperature difference (AVHRR channel 3 minus AVHRR channel 4). Results to date indicate that raw channel-3 data generally can be used; that is, the ASMC algorithm generally is not affected by moderate levels of sensor noise in channel 3 (see below). In general, the ASMC algorithm performs high-quality cloud detection for an arbitrary scene (Figs. 3–16). Results from different geographical regions (Fig. 3–16, Table 1) and for different seasons (Table 1) show that the ASMC algorithm performs well independent of either geographical region or season. A quantitative comparison of scene statistics prior to and after cloud screening with the ASMC algorithm confirms the effectiveness of the approach (Tables 2, 3, 4, and 5). Performance of the algorithm is such that a complex 512×512 scene can be cloud screened in about 1 minute on a HP 9000/735 workstation with 32 MB of RAM.

The better spectral characteristics and improved thermal spatial resolution of AVHRR data compared to GOES data contribute significantly to the improved segmentation and labelling achieved by the ASMC algorithm. In addition to the binary clear/cloud classification defined by the sign of the distance of a given cluster center to the decision plane, which defines the cloud-free polyhedron (solid black triangle in Fig. 1b), the labelling scheme also can make use of the actual distance of the cluster center from the decision plane and use it to attach a reliability estimate to the resulting

Table 5. Percentages of Cloudy Pixels Found by Various Algorithms

Image #	Location	ASMC Percentage of Cloud Pixels	ASMC Percentage of Clear Pixels	ASMC Percentage of Ambiguous Pixels	Percentage of Water Pixels	Percentage of Cloud Pixels by Texture Method	Percentage of Cloud Pixels by Threshold Method	Percentage of Cloud Pixels by CLAVR Method
1 (AR7733)	Baja	7.26	26.44	0.00	66.30	N/A	N/A	N/A
2 (AR2589)	Gobi	29.01	70.98	0.00	0.01	N/A	N/A	N/A
3 (AR2695)	N. Africa	33.23	66.77	0.00	0.00	N/A	N/A	N/A
4 (AR3981)	Somalia	12.33	47.49	0.00	40.19	N/A	N/A	N/A
5 (AR8987)	Crt. Lakes	31.36	46.82	0.00	21.81	N/A	N/A	N/A
6 (AR5510)	Rockies	32.50	66.81	0.00	0.70	N/A	N/A	N/A
7 (AR2750) ^a	W. Europe	34.94	23.12	0.00	41.93	N/A	55.84	N/A
8 (AR2688)	Italy	12.31	23.84	0.00	63.85	N/A	N/A	N/A
9 (AR3982)	S. Africa	61.64	38.34	0.00	0.02	N/A	N/A	N/A
10 (AR4046)	G. Britain	29.10	18.69	10.42	41.79	N/A	N/A	N/A
11 (AR2377) ^a	S. Americ	8.68	90.79	0.00	0.52	20.28	N/A	5.83
12 (AR2293)	S. America	43.19	56.26	0.00	0.53	N/A	N/A	N/A
13 (AR3980–2)	Pakistan	2.05	53.88	8.69	35.38	N/A	N/A	N/A
14 (AR3982–2)	Mozamb.	6.02	46.90	0.00	47.09	N/A	N/A	N/A
15 (AR7948) ^a	Oregon	4.61	93.76	0.00	1.62	N/A	N/A	30.96

Water pixels are given for information only. Such pixels are excluded from ASMC procedure.

^a Image numbers 7, 11, and 15 also have comparisons with other methods.

label (Fig. 1b). In fact, additional planes (parallel to the primary decision surface) can be added to the decision rules (Eq. 10) to define zones of ambiguity (Eq. 11). Pixels assigned to these ambiguous zones have neither robust cloudy nor cloud-free signatures. Generally, such pixels occur near cloud-land boundaries in which either on board amplifier hysteresis effects or subpixel cloud effects (relative position of clouds to the sensor's integrating aperture) compromise the pixel's signature. Alternately, wet versus dry sand in a desert region can significantly alter expected reflectances and emissivities and give rise to ambiguous pixel signatures. Attenuation of radiances due to atmospheric water vapor and the presence of snow cover also can lead to an ambiguous pixel assignment. Generally, however, the number of ambiguous clusters in a scene is either zero or very small (Table 3). Use of information contained in the ambiguous assignment ultimately is a scientific decision predicated on programmatic needs (e.g., recovering as much clear land as possible, ensuring that clear land labels are assigned with a high degree of certainty, or identifying regions of possible subpixel cloud contamination). The labelling scheme of the ASMC algorithm easily accommodates such needs and allows for automated relabelling of ambiguous clusters if desired (Table 3).

The ability of the ASMC decision polyhedron to more accurately deal with complex decision surfaces compared to that of static thresholds again is shown in Figure 17. For reference, Figure 17a shows the channel-2 albedo of the image of western Europe (Image 7). Figure 17b is analogous to Figure 17a, except that only the gray pixels (cloud over land) in Figures 2j, k, and l, which either have boundaries adjacent to or are very near black pixels (cloud-free land) are overlaid in yellow in Figure 17b. Close examination of this figure shows that these pixels almost always occur near cloud-land boundaries in the scene or within small, thin clouds. The spectral properties of such pixels generally are not those of cloud-free land because the actual pixel values are a complex function of the position of the land and cloud boundaries within a given pixel relative to the on-board integrating aperture function of the AVHRR sensor. For most land applications, it is desirable that such pixels be excluded from the land analysis. The ASMC algorithm provides for their exclusion.

Experimental Variants of the ASMC Algorithm

Several variants of the input data used to cluster and then label clusters under the ASMC scheme are possible because of the multispectral, high spatial resolution nature of AVHRR data. A summary of the results obtained using these variants is given.

A four-dimensional decision space can be constructed using channel-2 albedo, channel-4 temperature, the gradient of albedo and the gradient of tempera-

ture. Results from 19 test images (not shown) indicate that a less-robust clustering is achieved compared to that produced by the ASMC process defined earlier. Albedo and temperature data simply do not contain enough information to properly cluster and classify the data. Gradient information added little useful information as can be seen by comparing the standard ASMC results based on α , T , and δ segmentation and labelling for a scene over the Amazon basin (Fig. 14) with those computed using the ASMC process with α , T and Δ where Δ is the spatial gradient of albedo (Fig. 18). Clearly the resulting cloud mask produced using reflectance gradient as an input to ASMC is inferior to that produced when the $T_3 - T_4$ difference is used. Thus, the AVHRR case is different from the GOES case because GOES thermal data have only 8 km spatial resolution (often resampled at 4 km) compared to 1 km for AVHRR thermal data. Gradient of reflectance computed on a spatial scale smaller than 8 or 4 km, however, does prove helpful in the detection of clouds in GOES scenes over land, especially in the absence of AVHRR channel-3 information (Simpson and Gobat, 1995a).

Another four-dimensional decision space can be formed using albedo, temperature, and the two thermal differences $T_3 - T_4$ and $T_4 - T_5$. This combination of variables produces good segmentation of the images using the ASMC clustering scheme. Labelling, based on the ASMC adaptive-thresholding scheme, however, is very sensitive to small changes in any of these parameters (i.e., the slope of the four-dimensional decision plane becomes too steep in one direction). Also, the $T_4 - T_5$ difference is more susceptible to water vapor contamination than the $T_3 - T_4$ difference and therefore can add ambiguity to the labelling.

Variant #3 consisted of albedo, temperature, and the signed $T_3 - T_4$ difference as inputs to the ASMC process. Again, good segmentation of the image resulted. The two-sided nature (both positive and negative difference values) of the distribution of the signed $T_3 - T_4$ differences, however, is inconsistent with the simple model that the polyhedron decision surface tries to represent. Thus, the fact that the cloud-free land pixels are bunched between very large positive differences and very large negative differences is inconsistent with the labeler's expectation that high differences represent cloud and low differences represent land. This difficulty can be overcome by using the absolute value of the temperature difference; however, this procedure effectively folds the negative side of the distribution onto the positive side. This is inappropriate because different information is contained in the sign of the difference. Clouds tend to have large positive difference in the $T_3 - T_4$ distribution, whereas negative differences tend to be associated with noise in the AVHRR channel-3 data. The actual variant chosen for the ASMC algorithm ($T_3 - T_4$ subject to the constraint that negative differ-

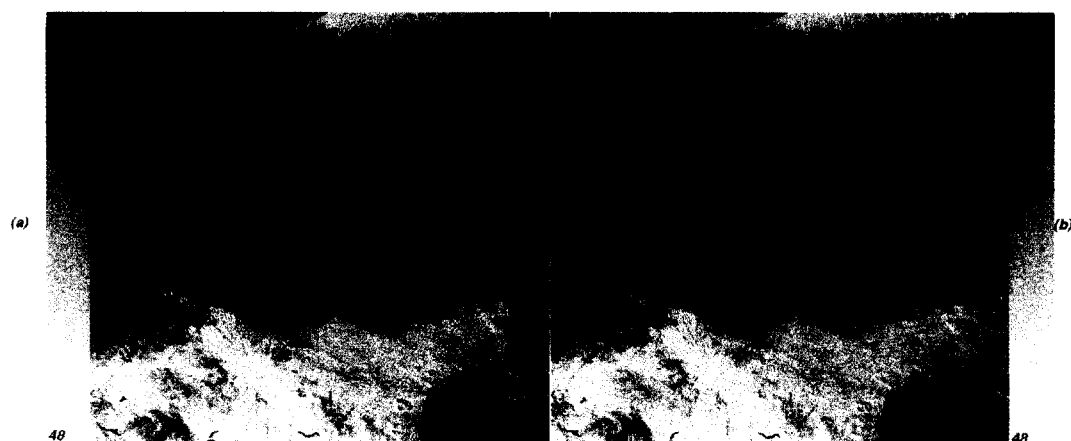


Figure 17. a) Channel-2 albedo for Image 7 (western Europe), and b) identical to (a) except pixels in yellow correspond to cloud pixels (gray), which form the boundary between clear land and cloud pixels for this image based on ASMC results (i.e., Figures 2j, k, and l). (Image Reference #AR2750.)

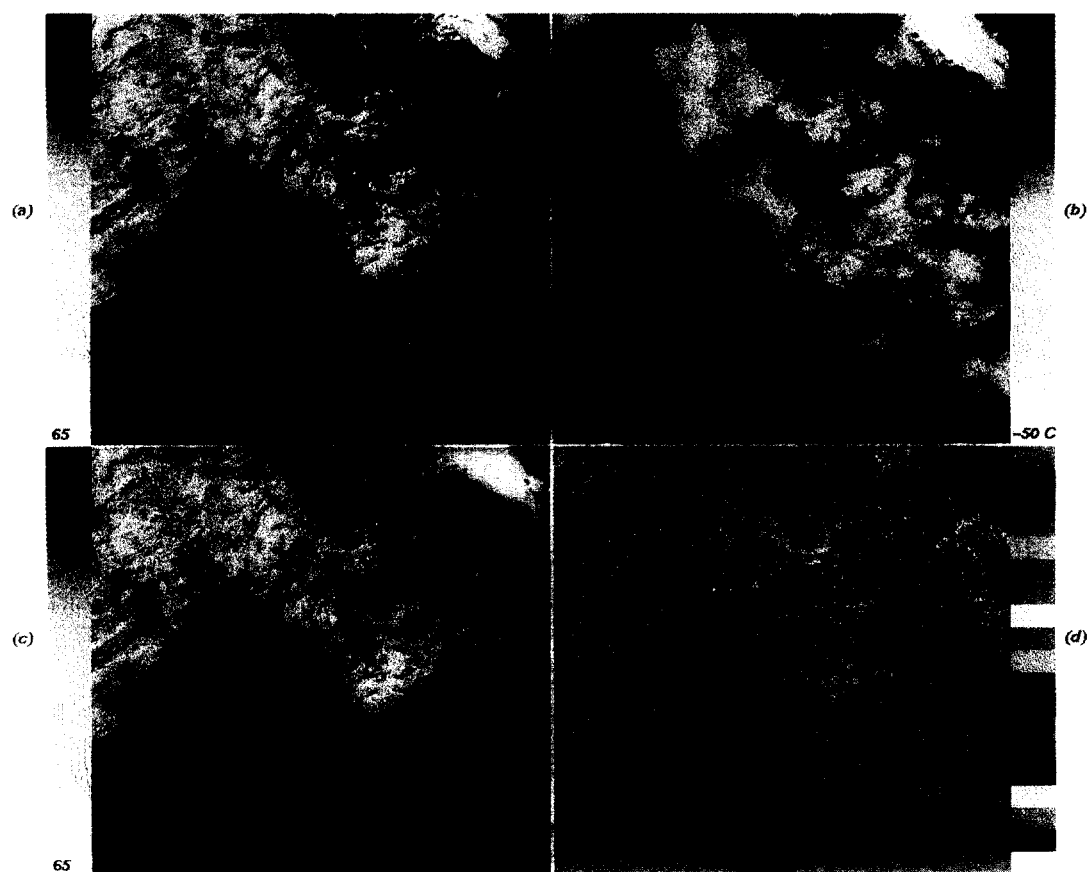


Figure 18. Identical to Figure 14, except the ASMC procedure was given inputs of channel-2 albedo, channel-4 temperature, and the spatial gradient of albedo. The resulting cloud mask (Fig. 18c) is inferior to that produced by the standard ASMC input, which utilizes the $T_3 - T_4$ difference instead of spatial gradient of albedo (Fig. 14c). (Image Reference #AR2293.)

ences are mapped to 0) retains the information contained in the sign of the temperature difference for labelling.

The last variant tested modified the chosen ASMC inputs by normalizing the albedo with the cosine of the solar zenith angle. This change did little to improve either the segmentation or labelling capabilities of the ASMC algorithm. Thus, for computational efficiency it was dropped from further consideration.

Comparison with Other Methods

A variety of methods have appeared in the literature to deal with the difficult problem of cloud detection in AVHRR data over land. Here six of these methods are reviewed briefly to facilitate a comparison (Table 5) between results from these other approaches and those obtained with the ASMC algorithm. The six techniques are: 1) cloud detection using albedo and temperature threshold, 2) texture model, 3) histogram models and morphological operations, 4) NOAA's CLAVR-I algorithm, 5) the spatial coherence method of Coakley and Bretherton (1982), and 6) the procedure of Saunders and Kriebel (1988a, b).

Albedo and temperature thresholds have been used either separately or in combination to detect clouds in AVHRR scenes over land. This approach also may include other criteria in the cloud detection (e.g., satellite viewing angle). Often, an initial coarse level cloud screening is performed and then a finer scale cloud detection follows. Three criteria for inclusion in the coarse screening traditionally have been invoked. First, the pixel selected must not be ocean. This criterion is met easily using GIS information and polygon fill operations (e.g., Simpson, 1992). Second, all pixels must satisfy an initial albedo threshold. For example, CLAVR-I uses a threshold of 44% for its coarse reflectance test. Stowe et al. (1991) say this value is a good first test based on climatology. Third, all pixels must satisfy gross thermal tests. For example, CLAVR-I uses $T \geq 21^\circ\text{C}$ to flag pixels as land, and $T < -24^\circ$ to flag pixels as cloud in its coarse temperature tests. Fine-scale cloud detection then can be performed in a variety of ways using different statistical criteria (Gutman et al., 1987). Figure 19a shows the albedo distribution of Image 7 and Figure 19b shows the corresponding ASMC cloud mask in yellow. Figure 19c shows a cloud mask (yellow) produced using joint albedo and temperature thresholding criteria. To produce this latter mask a gross thermal threshold criterion was used to build a cloud-free set of land pixels. This set was used to develop a statistical model of cloud-free albedo thresholds that incorporated variation in solar zenith angle. This statistical model of albedo was used to produce the final cloud mask. The difference between this mask and the ASMC mask (Fig. 19d) shows how conservative the joint albedo

and temperature mask is relative to the ASMC result. Purple indicates both methods classified the pixel as cloud-free, brown indicates both methods classified the pixel as cloud contaminated, green ASMC assigns pixels as cloud-free land, whereas the joint bispectral thresholds assign pixels as cloud, and orange is the converse of green. Statistics for this comparison are given in Table 5. This example illustrates that static thresholds do not have the discriminatory capabilities of adaptive thresholds, an ability clearly needed in the highly variable terrestrial environment. This result also is consistent with the general formulation of the labelling problem in Figures 1 and 2 and the corresponding discussion in the section entitled "Complementary Roles of Clustering and Labelling Procedures."

Textural models (e.g., the contrast measure) can be computed to determine a statistically robust measure of spatial homogeneity in a land scene. For this purpose, either band 2 or band 4 of the AVHRR data often is used. The assumption is that the texture information is contained in the overall or average spatial relationship that the gray tones have to one another. For computational efficiency the Gray Level Difference Vector approach (Welch et al., 1989), as opposed to the traditional co-occurrence matrix approach (Haralick et al., 1973) often is used to compute texture models. Gallegos et al. (1993), for example, use a texture model and spectral information to detect clouds in AVHRR data for the simpler oceanic case only. Their algorithm, however, is not intended for use over land.

The contrast texture measure used in this article was computed by digitizing the albedo information onto the interval [0, 15], computing a contrast measure of this digitized data using a 4×4 kernel, histogram equalizing the resulting contrast image, and applying a static contrast texture threshold across the histogram-equalized result. Note that a static threshold value of the texture parameter (e.g., contrast) must be specified by the user in order to separate cloud-free land pixels from cloudy pixels in the contrast texture model. The value of the static contrast texture threshold used in this example was 0.5, which means that pixels in the final histogram-equalized contrast image with values greater than 127 (0.5×255) will be classified as clouds. The classification is extremely sensitive to the choice of the chosen threshold, and in general, a threshold suitable for one image will produce less desirable results in a different image.

Figure 20a is over a part of the Amazon basin and is identical to that of Image 11 (Fig. 13a). The corresponding ASMC cloud mask and the texture-derived cloud mask are shown in Figures 20b and c, respectively. Panel d shows the difference mask; the color code for this panel is identical to that defined for Figure 19d. Detailed statistics for this comparison are given in Table 5. The texture model could not resolve



Figure 19. a) Channel-2 albedo for Image 7, and b) the corresponding ASMC mask shown in yellow, c) cloud mask formed using joint bispectral albedo and thermal criteria, also shown in yellow, and d) the difference between panel (b) and (c). In panel d, brown indicates that both methods assign the pixel as cloud, purple indicates both methods assign the pixel as cloud-free land, green ASMC assigns pixels as cloud-free land, whereas the joint bispectral albedo and thermal criteria assign them as cloud; orange is converse of green. (Image Reference #AR2750.)

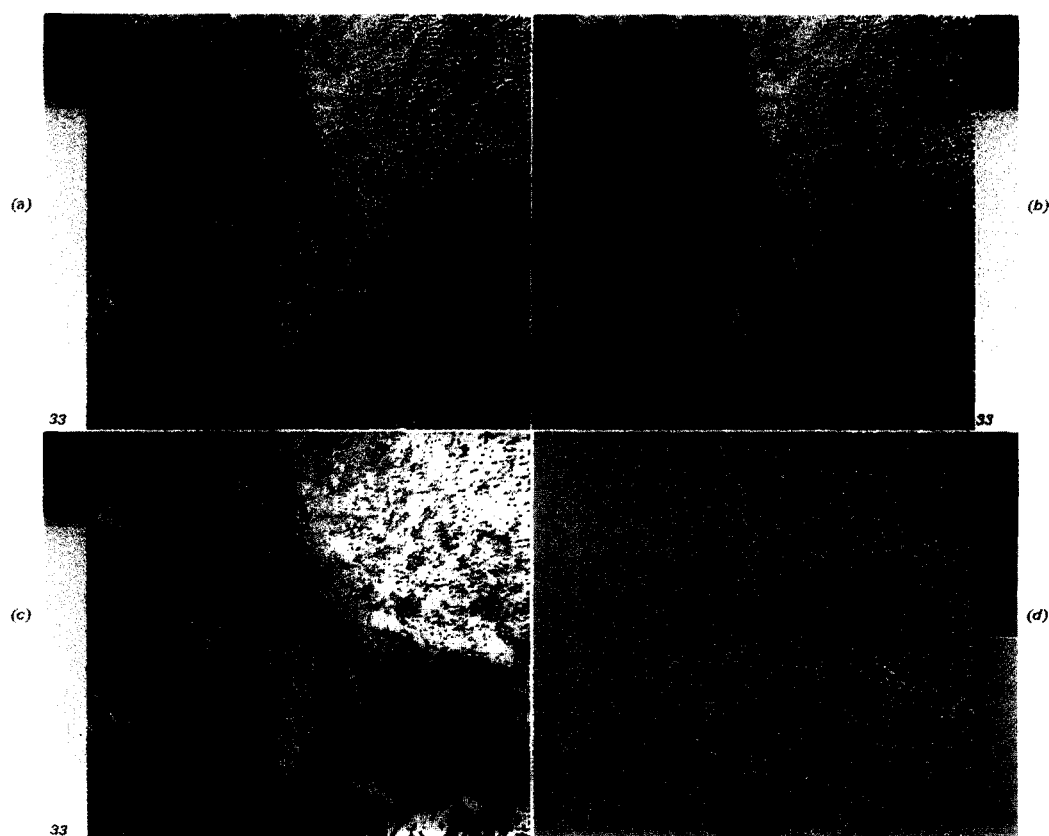


Figure 20. a) Channel 2 albedo for Image 11 of the Amazon basin (same as Fig. 13a), and b) the corresponding ASMC mask. c) The cloud mask produced by the texture model, and d) the difference of panels b and c with the color codes defined as for Figure 19d. (Image Reference #AR2377.)

the small-scale cloud structure as well as the ASMC procedure. Moreover, the texture model erroneously detected parts of the Amazon river as cloud. Texture models can be viewed as the statistical analog of spatial gradients. Thus, the results from the texture model (Fig. 20) are consistent with those obtained with the ASMC procedure when spatial gradient is used instead of the $T_3 - T_4$ temperature difference (Fig. 18); namely that spatial gradient is less useful than $T_3 - T_4$ differences in detecting cloud over land in AVHRR data.

Histograms of the initial results of cloud detection over land often are examined for residual clouds. Both albedo and thermal data can be used in this context. For albedo, the threshold is defined as the albedo coordinate of the first bin from the right of the histogram distribution, (i.e., brighter than) to have a frequency greater than some value, (e.g., 0.1%). For temperature, the threshold is defined by the temperature coordinate of the first bin from the left of the histogram (i.e., colder than), which has a frequency greater than some value (e.g., 0.1%). In other words, such probability checks attempt to chop off the "cloudy" tails of any remaining misclassified data from the initial cloud detection. Morphological dilation also can be used to screen out residual clouds and/or cloud shadows at cloud boundaries. Such techniques usually are applied as postprocessing steps to a cloud mask produced using more traditional cloud-detection techniques.

NOAA's CLAVR-I (clouds from AVHRR) algorithm for cloud detection uses a battery of threshold and restoral tests and all five channels of AVHRR data to discriminate clear pixels from cloudy pixels (Stowe et al., 1991). In general, the algorithm is intended for use over land and ocean, during night and day, and for any region of the earth's surface. The details of the algorithm, however, vary from case to case—particularly for land versus ocean and day versus night, and to a lesser degree for polar versus nonpolar data. The CLAVR-I algorithm for daytime land data, for example, subjects 2×2 arrays of pixels to a series of cloud and restoral tests. For simplicity and ease of comparison with the ASMC results, we will limit our discussion to the CLAVR-I algorithm for use with nonpolar, daytime, land data. Furthermore, we only consider CLAVR-I's clear and cloudy classifications (i.e., we do not account for CLAVR-I's mixed classification) because the final ASMC classification, which is being compared with CLAVR-I results, only produced either clear or cloudy clusters.

Figures 21a and b show the channel-2 albedo and channel-4 thermal structure, respectively, of a complex region of Oregon. The right part of the image contains dry desert regions (Columbia Plateau) of central Oregon. The lower left hand corner of the image contains lakes and bright sand dune areas (Great Sandy Desert, Harney Basin). The cloud-free nature of the highly re-

flective regions of this scene was confirmed by matching shapes in the image with corresponding features in high-resolution maps. Part of the Oregon Cascades are shown in the image as is part of the Willamette Valley. Figure 21c shows the cloud mask produced by the ASMC algorithm (yellow) and Figure 21d (salmon) shows the cloud mask produced by the CLAVR-I algorithm. The difference between these two masks is shown in Figure 21e, where purple indicates both algorithms assigned the pixel as cloud-free land, brown indicates both algorithms assigned the pixel as cloud, light green indicates CLAVR-I assigned pixel as cloud, whereas ASMC assigned pixel as cloud-free land and orange indicates the converse. A comparison between Figures 21c and 21d shows that CLAVR-I erroneously detected cloudy pixels in large segments of the image (statistics are given in Table 5). Our analysis indicates that the thermal uniformity test (TUT) of CLAVR-I contributed most significantly to the misclassification (i.e., 30.65% of the total 30.96% cloud cover detected by CLAVR-I for this image was associated with the TUT test, 0.01% with the RGCT test, 0.26% with the RUT test and 0.05% with the FMFT test). The ASMC algorithm avoids the use of all spatial uniformity tests. This is a distinct advantage for images with very high spatial variability over small space scales (Fig. 21).

Figure 22 shows an analogous example for an image over the Amazon basin. The comparison (including statistical results in Table 5) between the two algorithms is much better in this case but CLAVR-I failed (Fig. 22d) to properly identify several obvious small-scale cloud structures that were properly identified by the ASMC procedure (Fig. 22c). Again, the TUT test in the CLAVR-I algorithm contributed most to the final CLAVR-I classification (i.e., 4.75% of the total 5.87% cloud cover detected by CLAVR-I for this image was associated with the TUT test, 1.08% with the RUT test, and 0.04% with the RGCT test).

Two other methods of cloud detection [maximum likelihood, spatial coherence (e.g., Coakley and Bretherton, 1982) often are used. Maximum likelihood techniques assume that the image data are normally distributed. Our examination of arbitrary image data, however, shows that, in general, complex scenes over land have data that are not normally distributed [e.g., often tear-shaped (Fig. 2) distributed]. Therefore, we do not recommend use of this approach for cloud detection in satellite scenes over land.

Gallaudet and Simpson (1991) did an extensive comparison between their Principal Component Transformation Split-and-Merge Clustering (PCTSMC) algorithm and the spatial coherence algorithm of Coakley and Bretherton (1982) to detect clouds in AVHRR data over the ocean (see figs. 3, 10–15 and table 6 in Gallaudet and Simpson, 1991). The spatial-coherence method performed poorly relative to the PCTSMC algo-

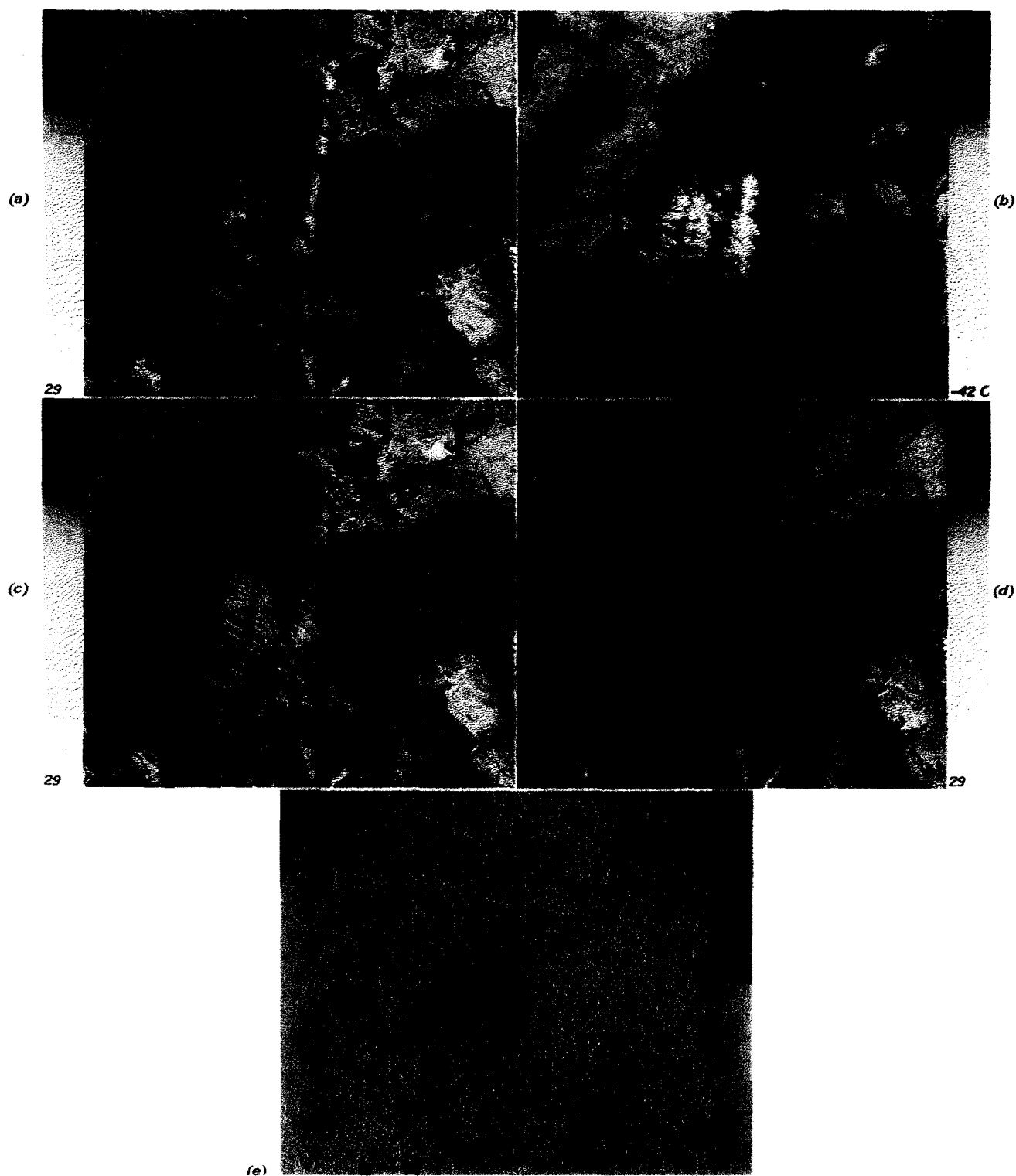


Figure 21. a) AVHRR channel-2 albedo and b) channel-4 temperature for a highly spatially variable region of Oregon (see text for description). c) The cloud mask produced by the ASMC algorithm (yellow), and d) the cloud mask produced by the CLAVR-I algorithm (salmon). Panel e is the difference of these masks with purple indicating both algorithms assigned pixel as clear land, dark reddish-brown indicates both algorithms assigned the pixel as cloudy, light green CLAVR-I assigned pixel as cloud and ASMC assigned it as clear, and orange is the converse of light green. (Image Reference #AR7948.)

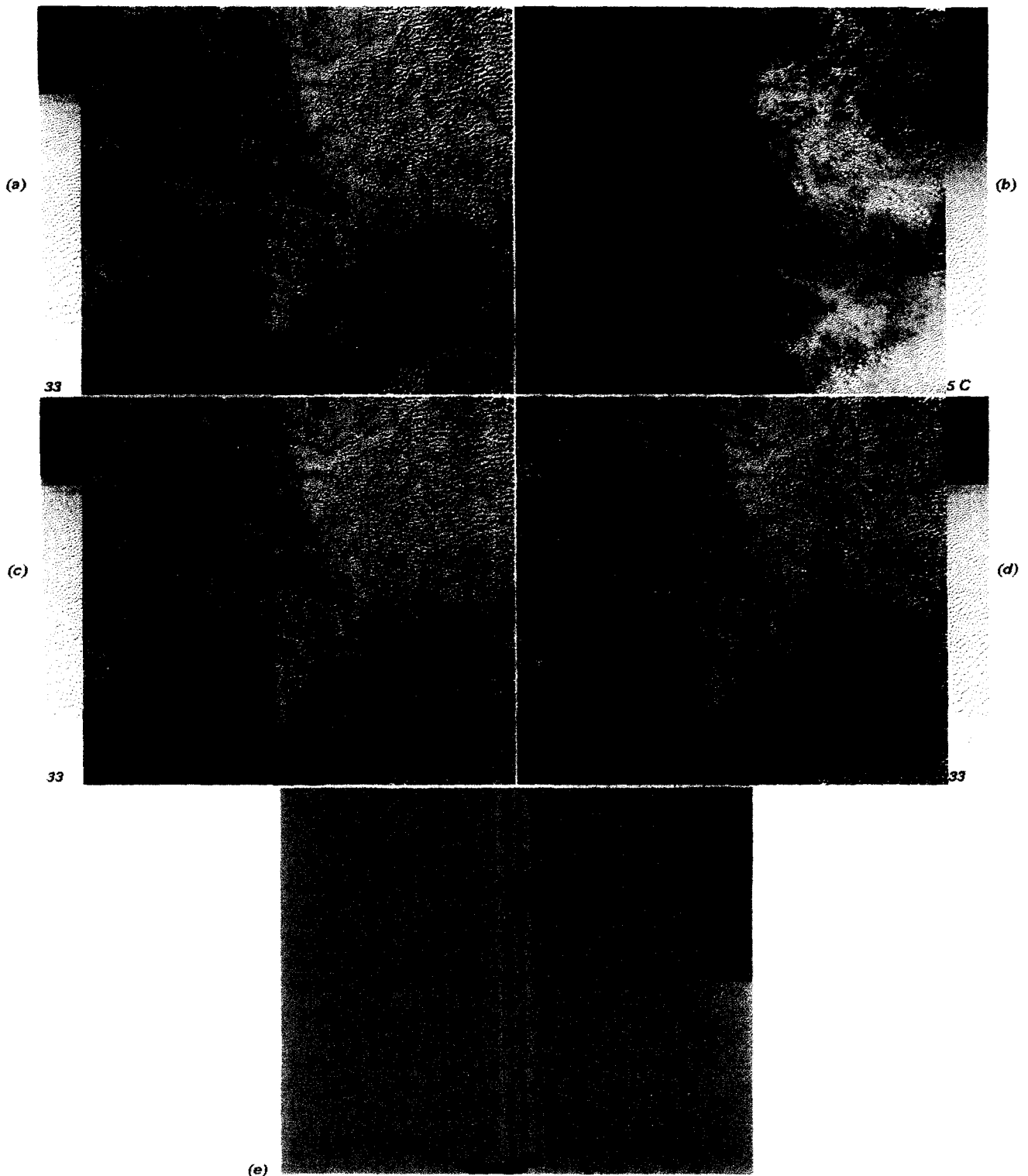


Figure 22. Analogous to Figure 21, except for a region of the Amazon basin. (Image Reference #AR2377.)

rithm for each of six images evaluated because the spatial-coherence method: 1) assumes that clouds congregate in layers that persist over large areas, which is often not the case; and 2) assumes that the thermal emission from cloud-free and completely cloud-covered regions exhibit a degree of spatial uniformity usually

unachieved by partially cloud covered regions. The Galaudet and Simpson (1991) study also showed that the spatial coherence method incorrectly masked cloud-free areas of high oceanic thermal gradient as cloud. Oceanic thermal gradients generally are much smaller and less variable than land surface thermal gradients. Therefore,

this latter failure mode of the spatial coherence method is likely to be even more severe when applied to an arbitrary terrestrial scene. Moreover, the spatial coherence method uses an 8×8 pixel local standard deviation array for its computation. This gives rise to false cloud detection, for example, at cloud-free land-ocean boundaries (see Gallaudet and Simpson, 1991, figure 15). This problem would be more severe in a terrestrial scene with rivers and lakes. For these reasons, and because the ASMC clustering procedure is a slight adaptation of the original PCTSMC clustering procedure, we have not explicitly included additional ASMC and spatial coherence comparisons in this work.

Gallaudet and Simpson (1991, Figures 12, 13, and 14) provide an extensive comparison between their PCTSMC algorithm and results obtained with that of Saunders and Kriebel (1988a, b) for the case of cloud detection over the ocean. In general, the Saunders and Kriebel algorithm incorrectly interpreted cloud-free regions of oceanic thermal gradient structure as cloud. Again, because the thermal gradient structure over the land generally is much larger and more variable than that over the ocean, this failure mode of the Saunders and Kriebel algorithm is likely to be exacerbated with an arbitrary land scene. Moreover, the Saunders and Kriebel algorithm must be tuned to individual images for optimal results, whereas the ASMC algorithm requires no image-specific tuning. Rather, the ASMC procedure automatically adapts itself to each image. For these reasons we do not recommend use of the Saunders and Kriebel (1988a, b) approach for cloud detection in an arbitrary land scene.

Noise Issues

Three types of noise can affect the performance of any algorithm designed to accurately detect clouds in AVHRR scenes over land. Each is discussed briefly and a preprocessing procedure to eliminate the problem also is identified.

Dropouts in satellite images usually are associated with loss of synchronization between the satellite and ground-receiving station. Generally they result in anomalous values being inserted into a satellite image. There is no information in such sections of an image and they will corrupt the cloud-detection process. Such lines simply should be deleted from the image prior to analysis.

Occasionally randomized speckled noise may appear in satellite data. Such noise must be removed from the data prior to use of the ASMC algorithm. A selective median filter that only operates in the regions of the speckled noise works well for this application (Simpson and Gobat, 1995a).

The AVHRR channel-3 data often are contaminated with highly variable sensor noise. The ASMC procedure is unaffected by moderate levels of channel-3 noise.

Occasionally, however, the peak-to-peak amplitude of the channel-3 noise is too large and poor segmentation results. Use of Wiener filtering techniques (Simpson and Yhann, 1994) will produce a restored channel-3 signal compatible with the ASMC process.

Limitations

Algorithmic Limitations

The ASMC algorithm is a physically motivated, multispectral, multistep statistical procedure to accurately and automatically detect clouds in an arbitrary AVHRR scene over land. In principle, its results may be sensitive to two image specific values: 1) the size of the scene, and 2) the percentage of cloud-free pixels in the scene or equivalently the percentage of cloudy pixels in the scene. All the scenes used in this work were either 512×512 or 1024×1024 pixels² in size (Table 1). Most of these scenes, however, included some water pixels (oceans, lakes, rivers), which occupied between 66.30%–0.01% of the scene (Table 5). Water pixels are excluded as inputs to the ASMC analysis. Therefore, based on data in Tables 1–5, one can conclude that the ASMC procedure will perform well on scenes 256×256 pixels², provided all the data are nonwater pixels. In fact, this estimate of minimum image size is conservative because, as will be discussed, it is the statistical distribution of the data rather than the number of data points that ultimately sets the theoretical limits on applicability of the ASMC procedure. Moreover, although we do not recommend it, the ASMC procedure has been used successfully on scenes as small as 50×50 pixels².

The cloudy versus cloud-free percentage of the effective image (i.e., image minus excluded water and coastline pixels) also is an issue to consider. Data (Table 5) shows that the ASMC procedure performs well over a very broad range of percentage composition of analyzed image (e.g., percentage of pixels that are clear, cloudy, or ambiguous). Image 13 (Fig. 15), for example, only has 2.05% cloud pixels, 52.88% clear pixels, and 8.69% ambiguous (remaining pixels are water pixels), whereas Image 15 (Fig. 21) has 4.61% cloud, 93.76% clear, and 1.62% water pixels. The ASMC procedure produced good cloud masks over land for both of these images. Theoretically, however, the ASMC algorithm will fail to converge to a valid classification in the limits of either 100% clear or 100% cloud. To examine this potential failure mode, we studied a 512×512 pixel² cloud-free image in which only 8 pixels had cloud-like properties. This image was located using the SSABLE browse system (Simpson and Harkins, 1993) and few images in SSABLE's 9,000-image catalog satisfied these near cloud-free requirements. The ASMC algorithm masked the image correctly (results not shown). Additional experiments were run on this image by successively deleting one additional pixel with cloud-like properties per

each new experiment until only 1 cloud-like pixel remained in the data. This corresponds to about a 0.0001% cloud-contaminated image. Again, the ASMC procedure produced a valid classification. When all 8 cloudy pixels were removed from the image, the procedure failed to converge; a result consistent with the theoretical limit. Other experiments were run on additional near 100% cloudy or 100% cloud-free images and similar results were obtained.

All the experiments support the conclusion that the ASMC procedure will perform properly, even for relatively small images, provided there is a minimum number of cloud-like and cloud-free pixels in the image whose properties are sufficiently different (in a statistical sense), such that the clustering procedure has enough statistical dimensionality to segment the image. It should be emphasized that the limitations previously discussed place no practical constraint on the routine operational use of the ASMC procedure. In fact, we are currently using the ASMC procedure in a fully operational mode to compute a new satellite-based cloud climatology.

Data Limitations

The ASMC algorithm was designed for improved cloud detection in AVHRR scenes over land. The algorithm is not intended to segment snow and/or ice from clouds. A separate study using a fuzzy-logic classifier is under way to deal with snow/ice versus cloud segmentation.

One of the input variables to the ASMC procedure is the albedo in channel 2. Conditions of either poor illumination and/or poor viewing geometry can degrade the quality of the channel-2 albedo (e.g., Simpson and Humphrey, 1990, appendix). Therefore, the ASMC algorithm should not be used under low sun-angle conditions. Typically, such conditions occur within 1–2 hours of sunrise and sunset.

Some scenes will contain a mixture of land and ocean or large lakes. Because the ranges of reflectance and emissivity over the ocean generally are much smaller than over land, these scenes are best handled by land masking the scene using polygon fill and morphological transformations (Simpson, 1992). The land part of the scene then can be cloud masked with the ASMC algorithm and the oceanic part with the LDTNLR algorithm (Simpson and Humphrey, 1990) or the PCTSMC algorithm (Gallaudet and Simpson, 1991). Alternately, two different approaches for daytime and nighttime cloud detection in AVHRR data using neural networks (Yhann and Simpson, 1995) can be used for the oceanic parts of the scene. The union of the two masks produces the best cloud detection for such composite images.

Cloud Shadows

Cloud shadows also can negatively affect terrestrial applications. Our approach, however, is to treat cloud-

shadow issues separately from cloud-detection issues because: 1) atmospheric scientists interested in clouds over land generally are interested in the cloud pixels only, and 2) pixels affected by cloud shadow are best identified after rigorous cloud detection has been performed. We are developing a fully automated cloud-shadow detection procedure that uses geometric optics, details of orbit geometry, and morphological transformations to detect cloud shadow on a pixel-by-pixel basis in AVHRR scenes (Simpson and Stitt, unpublished data). A complete mathematical description of this new method is beyond the scope of the present work. Preliminary results of this new approach for a few selected scenes, however, are shown in Figure 23. Note, cloud shadow can produce an anomalous low temperature and albedo as detected by the satellite. Therefore, if cloud shadows are not removed directly, then some type of postprocessing should be used to detect these pixels.

Applications to MODIS-N Data

NASA's Earth Observing System (EOS) is scheduled to launch a 25-channel moderate resolution imaging spectrometer (MODIS-N) on its first EOS satellite later this decade. A 1.64 μm channel will be available for snow/cloud differentiation; a 2.130 μm channel will be used for cloud penetration observations; a 1.131 μm channel should prove useful for cloud water absorption; numerous visible wavelength channels, as well as polarization measurements, also are planned (EOS, 1989). The clustering procedure of the ASMC algorithm already accommodates arbitrary and mixed inputs. Moreover, the MODIS-N suite of measurements provides an abundance (compared to AVHRR) of new physical criteria for labelling the segments produced by the first step of the ASMC procedure. Thus, use of the ASMC procedure with MODIS-N data only should require: 1) selection of inputs to the clustering procedure, and 2) modification of the labelling procedure to take maximum advantage of MODIS-N's enhanced measurement capabilities.

CONCLUSIONS

The AVHRR Split-and-Merge Clustering algorithm provides a robust way to detect clouds in AVHRR data over a wide variety of land forms. Joint three-dimensional adaptive labelling, coupled with channel-2 albedo, channel-4 temperature, and the channel-3 minus channel 4 temperature difference, provides a mechanism for specifying a zone of ambiguous pixels in the labelling procedure. Generally, such pixels result from instrumentation problems at cloud-land boundaries, subpixel effects, atmospheric water vapor attenuation of infrared radiation or poor satellite geometry (poor illumination, poor viewing angles). Results obtained with the ASMC procedure compared favorably with those obtained

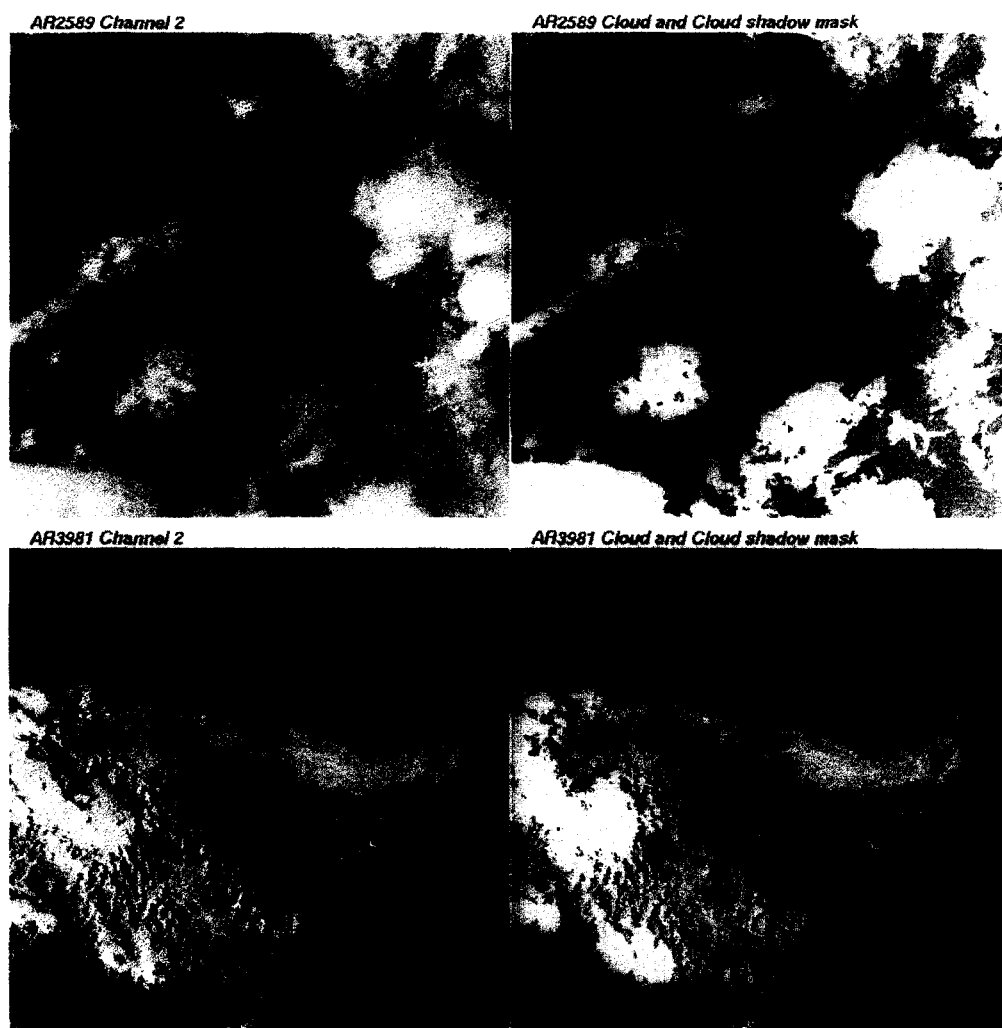


Figure 23. Preliminary results of a separate cloud shadow detection process for: a) Image 2 (Fig. 4a); b) cloud mask (ASMC) and cloud shadow pixels; c) Image 4 (Fig. 6a), and d) corresponding cloud shadow mask. Pixels identified by the ASMC cloud-detection process are shown in yellow. Cloud shadow pixels identified by the prototype cloud shadow algorithm (Simpson and Stitt, 1995) are shown in green.

when using other standard cloud detection techniques. The procedure is suitable for use in tropical, temperate, or desert environments and is not affected by seasonal variations in these environments. The ASMC algorithm is not intended to segment clouds from snow and therefore should not be used in polar regions. Moreover, the ASMC algorithm should not be used under conditions of low solar illumination.

This work was funded in part by the Marine Life Research Group of the Scripps Institution of Oceanography, by the National Sea Grant College Program, National Oceanic and Atmospheric Administration, U.S. Department of Commerce, under grant number NOAA NA89AA-D-SG138, project number R/OE-27, through the California Sea Grant College, and by the California State Resource Agency. The views expressed herein are those of the authors and do not necessarily reflect the views of NOAA or any of its subagencies. The U.S. Government is

authorized to reproduce and distribute this work for government purposes. A NASA Atmospheric Sciences project and a NOAA Operational Measurements Global Change project also provided partial funding. Special thanks to Drs. J. Dodge and A. Gruber for their continued support. M. Wilke, S. Cowley, and V. Batchelder typed the manuscript. P. Fani and A. Schmidt assisted with computer-figure generation and G. Taper assisted with final figure preparation.

REFERENCES

- Choudhury, B. J. (1990), Monitoring arid lands using AVHRR-observed visible reflectance and SMMR-37 GHz polarization difference, *Int. J. Remote Sens.* 11:1949-1956.
- Coakley, J. A., Jr., and Bretherton, F. P. (1982), Cloud cover from high resolution scanner data: Detecting and allowing for partially filled fields of view, *J. Geophys. Res.* 87:4917-4932.

- Earth Observing System (1989), From patterns to processes: The strategy of the Earth Observing System. *EOS Science Steering Committee Report*, Volume II and Appendices a-h. NASA 140 pp.
- Gallaudet, T. C., and Simpson, J. J. (1991), Automated cloud screening of AVHRR imagery using split-and-merge clustering, *Remote Sens. Environ.* 38:77–121.
- Gallegos, S. C., Hawkins, J. D., and Cheng, C. F. (1993), A new automated method of cloud masking for Advanced Very High Resolution Radiometer full-resolution data over the ocean, *J. Geophys. Res.* 98:8505–8516.
- Gutman, G. G. (1991), Vegetation indices from AVHRR: An update and future prospects, *Remote Sens. Environ.* 35:121–136.
- Gutman, G. G. (1992), Satellite daytime image classification for global studies of earth's surface parameters from polar orbiters, *Int. J. Remote Sens.* 13:209–234.
- Gutman, G., Tarpley, D., and Ohring, G. (1987), Cloud screening for land surface characteristics in a reduced resolution satellite data set, *Int. J. Remote Sens.* 8:859–870.
- Haralick, R. M., Shanmugan, K., and Dinstein, I. (1973), Texture features for image classification, *IEEE Trans. Syst. Man. Cybern.* SMC-3:610–621.
- Justice, C. O., Townshend, J. R. G., Holben, B. N., and Tucker, C. J. (1985), Analysis and phenology of global vegetation using meteorological satellite data, *Int. J. Remote Sens.* 6:1271–1318.
- Kaufman, Y. J., and Sendra, Y. (1988), Algorithm for automatic atmospheric corrections to visible and near-IR satellite imagery, *Int. J. Remote Sens.* 9:1357–1381.
- Lauritson, L., Nelson, G. G., and Porto, R. W. (1979), Data extraction and calibration of TIROS-N / NOAA A-G radiometers, *NOAA Tech. Memo. NESS 107*, U.S. Department of Commerce, Washington D.C.
- NOAA (1985), *Polar Orbiter Data Users Guide*, U.S. Dept of Commerce, NOAA, National Environmental Satellite, National Climatic Data Center, Satellite Data Service Division, Washington DC.
- Ohring, G. A., Gruber, A., and Ellingson, R. (1985), Satellite determination of the relationship between total longwave radiative flux and infrared window radiances, *J. Clim. Appl. Meteor.* 23:416–425.
- Pairman, D., and Kittler, J. (1986), Cluster algorithms for use with images of clouds, *Int. J. Remote Sens.* 7:855–866.
- Rand McNally (1977), *The International Atlas*, Section on World Scene Climate Regions pp. 306–307, New York, p. 533.
- Rossow, W. B., Moshier, F., Kinsella, E. et al. (1985), ISCCP cloud algorithm intercomparison. *J. Clim. Appl. Meteor.* 24:877–903.
- Saunders, R. W., and Kriebel, K. T. (1988a), An improved method for detecting clear sky and cloudy radiances from AVHRR data, *Int. J. Remote Sens.* 9:123–150.
- Saunders, R. W., and Kriebel, K. T. (1988b), Errata (i.e., An improved method for detecting clear sky and cloudy radiances from AVHRR data. *Int. J. Remote Sens.* 9:123–150, *Int. J. Remote Sens.* 9:1393–1394.
- Sellers, P. J., Rasool, S. I., and Bolle, H. J. (1990), A review of satellite data algorithms for studies of the land surface. *Bull. Am. Meteor. Soc.* 71:1429–1447.
- Simpson, J. J. (1992), Image masking using polygon fill and morphological operations, *Remote Sens. Environ.* 40:161–183.
- Simpson, J. J., and Humphrey, C. (1990), An automated cloud screening algorithm for daytime Advanced Very High Resolution Radiometer imagery, *J. Geophys. Res.* 95:13,459–13,481.
- Simpson, J. J., and Harkins, D. N. (1993), The Scripps satellite archive and browse for localized environments (SSABLE) system. *IEEE Trans. Geosci. Remote Sens.* 31:515–525.
- Simpson, J. J., and Yhann, S. R. (1994), Reduction of noise in AVHRR channel 3 data with minimum distortion, *IEEE Trans. Geo. Remote Sens.* 32:315–328.
- Simpson, J. J., and Gobat, J. I. (1995a), Improved cloud detection in GOES scenes over land, *Remote Sens. Environ.* 52:36–54.
- Simpson, J. J., and Gobat, J. I. (1995b), Improved cloud detection in GOES scenes over oceans, *Remote Sens. Environ.* 52:79–94.
- Simpson, J. J., and Keller, R. H. (1995), An improved fuzzy logic segmentation of sea ice, clouds and ocean in remotely sensed arctic imagery. *Remote Sens. Environ.* 54:290–312.
- Singh, S. M. (1988), Simulation of solar zenith angle effect on global vegetation index (GVI) data, *Int. J. Remote Sens.* 9:237–248.
- Spanner, M. A., Pierce, L. L., Running, S. W., and Peterson, D. L. (1990), The seasonality of AVHRR data of temperate coniferous forests: Relationships with leaf area index, *Remote Sens. Environ.* 33:97–112.
- Stowe, L. L., McClain, E. P., Carey, R., et al. (1991), Global distribution of cloud cover derived from NOAA / AVHRR operational satellite data, *Adv. Space Res.* 11:51–54.
- Tarpley, J. D., Schneider, S. R., and Money, R. L. (1984), Global vegetation indices from the NOAA-7 meteorological satellite, *J. Clim. Appl. Meteorol.* 23:491–493.
- Taylor, B. F., Dini, P. W., and Kidson, J. W. (1985), Determination of seasonal and interannual variation in New Zealand pasture growth from NOAA-7 data, *Remote Sens. Environ.* 18:177–192.
- Welch, R. M., Navar, M. S., and Sengupta, S. K. (1989), The effect of spatial resolution upon texture-based cloud field classification, *J. Geophys. Res.* 94:14767–14781.
- Yhann, S. R., and Simpson, J. J. (1995), Application of neural networks to AVHRR cloud segmentation, *IEEE Trans. Geo. Remote Sens.* 33:590–604.