

Cloud screening in AVHRR digital data over Arctic regions

SONIA C. GALLEGOS, JEFFREY D. HAWKINS

Naval Research Laboratory, Stennis Space Center, Mississippi 39529, U.S.A.

AND CHIU FU CHENG

Sverdrup Technologies, Inc., Stennis Space Center, Mississippi 39529, U.S.A.

ABSTRACT. A cloud screening method initially generated to mask cloud contaminated pixels over the ocean in visible/infrared imagery, has been revised and adapted to detect clouds over Arctic regions with encouraging results. Although the method is quite successful in eliminating very cold clouds, it underestimates low level clouds. However, this does not appear to interfere with monitoring of ice related features such as leads or the ice edge in Advanced Very High Resolution Radiometer (AVHRR) scenes. The method uses: a multiple-band approach to produce signatures not readily available in single channel data, an edge detection/dilation technique to locate features in the clouds and to join isolated edges, and a polygon identification technique to remove noise in the form of isolated pixels and separate clear regions from cloud contaminated areas. The method has been tested over a limited set of data with consistent results. Initial evaluation of the usefulness of this cloud-detection algorithm in data-fusion experiments indicate a potential in locating areas in AVHRR data which are cloud contaminated and which could yield a far superior representation of the ice features if replaced with data from a different sensor such as the Special Sensor Microwave/Imager (SSM/I).

INTRODUCTION

Cloud removal from data gathered by the polar orbiting environmental satellites over polar regions is a complicated problem. In these regions it is often difficult to discriminate between cloud-filled pixels and cloud-free pixels based exclusively on radiance differences. Visible data are available only part of the year. When available, the albedos of clouds and ice are similar and sometimes indistinguishable from one another. The thermal radiances which are highly useful to discriminate clouds from surrounding targets in lower latitudes are often plagued by tropospheric isothermal inversions, low clouds in the winter and extensive stratiform clouds in the summer.

Significant efforts to address problems in cloud identification have been made by the International Satellite Cloud Climatology Project (ISCCP). These efforts have resulted in the production of global cloud climatologies from visible and infrared acquisitions of weather satellites (Rossow and Schiffer, 1991). Recently, pattern-recognition methods have been introduced to cloud-identification routines. The addition of texture information produces more efficient and accurate algorithms than those which utilize spectral information exclusively. Ebert (1987) identified various surfaces and cloud types in AVHRR visible and infrared data sets from Arctic regions using a combination of three spectral and five textural features. Welch and others (1988) demonstrated for the first time that textural features alone

derived from a single visible channel of the Landsat Multispectral Scanner (LMS) were sufficient to classify four types of clouds. Welch and others (1990) showed that ice- and snow-covered surfaces have different texture features from clouds. This paper describes ongoing work to develop an automated technique to identify and remove clouds from AVHRR Arctic data for use in several sea-ice mapping applications. Algorithm performance is illustrated with examples, and its potential for assisting promising data fusion studies is highlighted.

METHODS

Cloud removal from satellite polar scenes is presently accomplished at the Naval Research Laboratory at Stennis Space Center (NRL-SSC) by imposing thresholds to the AVHRR data. This method is often inadequate and time-consuming. In the past year our research group developed an accurate and time efficient algorithm to remove clouds from AVHRR daytime data over oceanic surfaces (unpublished manuscript). This algorithm which will be referred throughout the text as the "over-the-water" algorithm, cannot be applied to polar regions in its original form. However, we have adapted some of the methodologies employed in its development to identify cloud targets at high latitudes. The primary aim of the "over-the-water" technique is to locate homogeneous areas of low albedo in the AVHRR

scenes. Location of these areas is accomplished using full resolution channel 1 data. Cloud-edge information is produced through a segmentation technique based on the Gray Level Co-Occurrence matrix (GLC). Areas between the cloud edges are assigned to cloudy and clear categories by a procedure referred as the polygon identification routine.

Because this work was initiated in response to the needs of a data-fusion study, emphasis was placed in removing those clouds which completely blocked the surface and in retaining as much surface information as possible. No effort was made to remove those clouds through which some surface information was still available to the observer. The cloud-masking algorithm over ice surfaces utilizes spectral data from two reflective and three infrared channels of the AVHRR. As in the "over-the-water algorithm", it uses an image segmentation technique and a polygon identification routine. Although both use similar approaches, they differ in the data utilized, computation of clear sky radiances, and inputs to the edge-detection procedure and to the area identification routine.

Initial steps of the algorithm include: (1) removal of land pixels, (2) acceptance of interactive inputs and (3) production of clear sky radiances in the scenes. The interactive inputs require the operator to enter the surface feature of interest. This study focuses on two features of naval interest. These are: (1) ice edge and (2) ice leads. Ice edge, as the name implies, is the region where the ice meets the open water. Ice leads are fractures or pass-ways through sea ice which are navigable by surface vessels (U.S. Navy, 1991). In the first path, commands direct the procedure to identify clouds blocking the view of the ice edge. In the second path, procedures are designed to mask clouds over areas from which ice-leads statistics are desired.

Generation of clear sky radiance is the same for both paths. It is calculated using two methodologies. The first method generates clear pixel information using the "over-the-water" algorithm. The second calculates clear pixel values with the Coakley and Bretherton (1982) spatial coherence techniques. The "over-the-water" algorithm is run exclusively on channel 1 data. Spatial coherence is calculated for a combination of reflective and infrared channels referred to as COM1 ((channel 1 + channel 2) - (channel 4 + channel 5)). Masson and Comiso (1991) used this combination for identification of ice types. In this study, this channel combination was chosen for no special or physical reason except to provide a means to combine reflective and thermal data in the same histogram. Pixels that have been declared "clear" by these procedures do not enter any further calculations.

For the first path (ice edge), texture features are extracted from a combination of infrared channels referred as COM2 ((channel 3 - channel 4) / (channel 3 + channel 4)). For the second path (ice leads), texture features are extracted from "degraded channel 1" data and COM2. Texture features are computed exclusively to generate cloud edges and not to produce target identification as would be the case in a classification study. The term "degraded" channel 1 data is used here to refer to 8-bit channel 1 data from which the two least significant bits have been eliminated to preserve high

albedo information. This normalized difference of channels is utilized here as a measure of cloud (water-vapor) contamination in the atmosphere. We realize, of course, that this combination may not be sufficient to identify all cloud related pixels in the scene. That was not the goal of this study. As previously indicated, this algorithm was developed to identify and remove those clouds which completely block the view of the ice features in a timely and efficient manner. Texture is calculated for 512×512 scenes using the cluster shade measure (Equation 2), based on the Gray Level Co-occurrence matrix (GLC). The routine we utilized for computing Cluster Shade values was implemented by Holyer and Peckinpaugh (1990).

Co-occurrence is a statistical approach which generates texture based on the spatial distribution of the gray levels in a scene. The (i, j) th element of the matrix is the relative frequency of occurrence of gray level i and gray level j when separated by a distance $(\Delta x, \Delta y)$ in a defined local neighborhood. Given a window of size $M \times N$ where M are the columns and N are the rows, a distance or displacement vector defined by $(\Delta x, \Delta y)$, and a range of intensity levels from 0 to $L - 1$, the GLC matrix is calculated from:

$$P(i, j | \Delta x, \Delta y) = \sum_{n=1}^{N-\Delta y} \sum_{m=1}^{M-\Delta x} A \quad (1)$$

Where

$$A = \frac{1}{(M - \Delta x)(N - \Delta y)}$$

if $f(m, n) = i$ and $f(m + \Delta x, n + \Delta y) = j$;
else $A = 0$.

The GLC matrix computations result in a second order probability matrix of size $L \times L$ from which cluster-shade measure is calculated for each local neighborhood as follows:

$$S(\Delta x, \Delta y) = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} (i + j - \mu_i - \mu_j)^2 P(i, j | \Delta x, \Delta y) \quad (2)$$

where

$$\mu_i = \sum_{i=0}^{L-1} i \sum_{j=0}^{L-1} P(i, j | \Delta x, \Delta y)$$

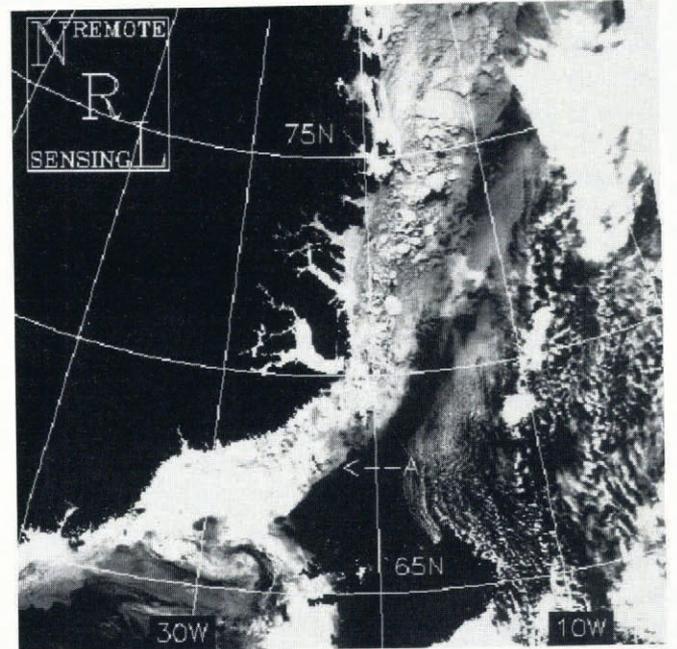
$$\mu_j = \sum_{j=0}^{L-1} j \sum_{i=0}^{L-1} P(i, j | \Delta x, \Delta y).$$

Pixel displacement is calculated in the horizontal and vertical direction for adjacent pixels separated by $(\Delta x, \Delta y) = 1$ for channel 1 and COM2. Size of the search neighborhood for the degraded channel 1 data is 8×8 and 3×3 for COM2. The size of these neighborhoods is outside the window stability limits set by Key (1990). He indicated that window sizes smaller than ten produce highly variable texture values. His studies were done using single channel 5×5 km AVHRR data. In this study, we utilize 1 km "degraded" channel 1 data in one case and a combination of channels in another. Zero crossing tests to generate

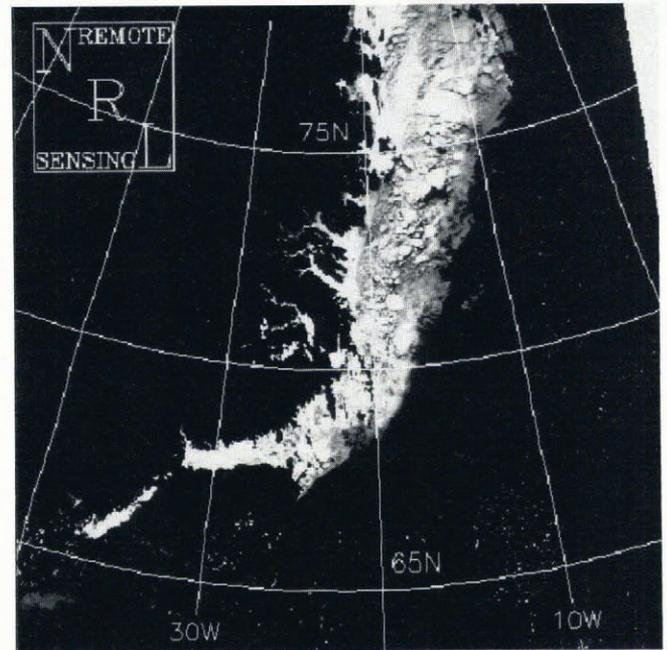
edge information from cluster shade matrices are calculated using arbitrary thresholds. An edge is generated at the site of the neighboring pixels whose cluster-shade values are opposite in sign and where both have cluster-shade intensities greater than or equal to a chosen threshold. The result of the edge-generation procedures is a binary file in which edges around various features in the scene such as: cloud edges, snow edges or ice edges are assigned a value of zero. The remaining pixels are given a value of 1.

In the case of the "degraded channel 1" data, the edges produced were not sufficient to include the full extent of the clouds. Therefore, it was necessary to resort to a dilation technique which uses arbitrary thresholds to generate additional edges around the primary edges of the scene. Dilation procedures were not necessary for the COM2 data sets. Because the edge-detection procedure was insufficient to separate clear from cloudy targets from the COM2 data, another procedure, referred as the polygon-identification routine, was produced. In this routine, a polygon is defined as a contiguous group of pixels with a value of 1, which is surrounded by a boundary of pixels with a value of zero. To separate the polygons into clear and cloud-filled areas, a procedure was used in which the mean of the pixels inside each polygon and that of the pixels which constitute its boundary were calculated from coincident COM2 data. If the mean of the boundary pixels exceeds the mean of the inside pixels, the polygon is declared a cloud-free polygon. Conversely, if the mean of the inside pixels is greater than that of its boundary, the polygon is assumed to be cloud-filled and is assimilated into the cloud mask. Pixels inside a cloud-filled polygon are set to 0. Pixels in a cloud-free polygon are set to 1.

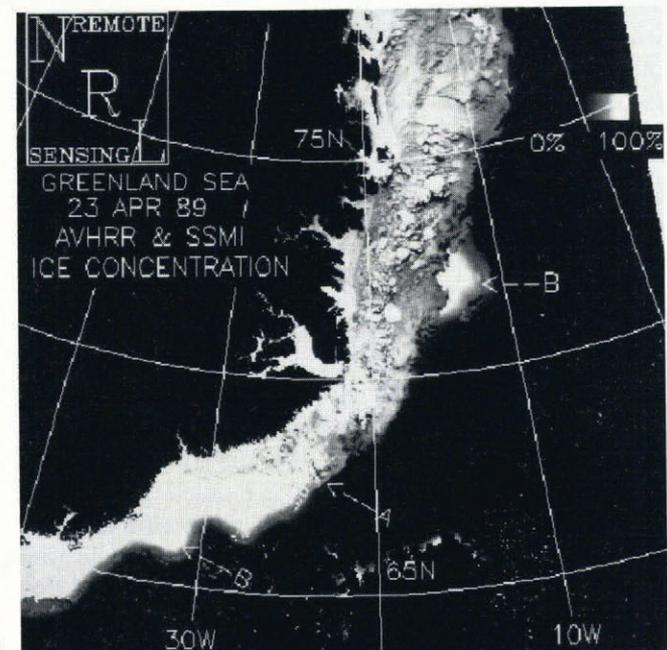
The algorithm was tested with AVHRR data sets from the Sea Ice Leads Accelerated Research Initiative (ARI) (Fetterer and Hawkins, 1991) and ice data fusion studies (Hawkins and Gineris, unpublished data). Imagery with significant sea-ice leads were processed for the time-span of April–September 1989. Ice-edge scenes were extracted from the April–July 1989 ARI data set. To demonstrate the performance of the algorithm and its potential use in data-fusion studies, scenes containing the same AVHRR data were prepared and are presented in Figures 1a to 1c and 2a to 2c. Figure 1a shows the Greenland ice edge as viewed by the AVHRR on the afternoon of 23 April 1989. In this scene, large clouds appear blocking the view of the ice edge at approximately 67° N, 30° W; 72° N, 12° W and 75° N, 03° W. Clouds observed in Figure 1a are identified



a

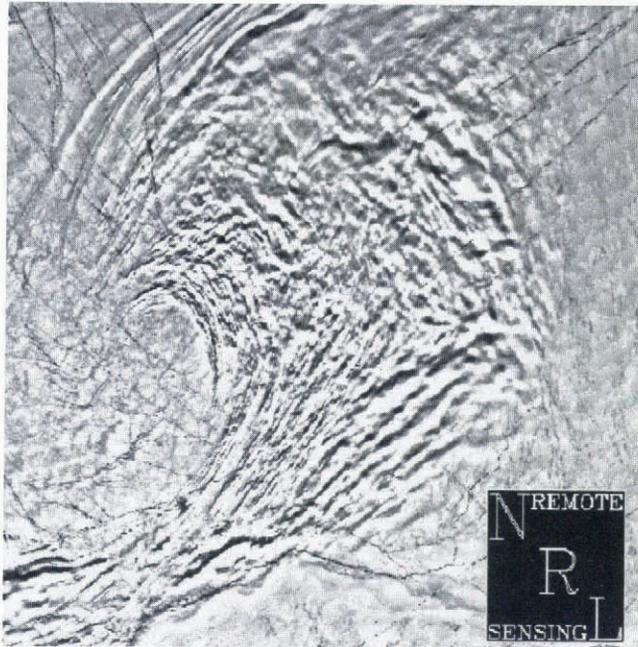


b

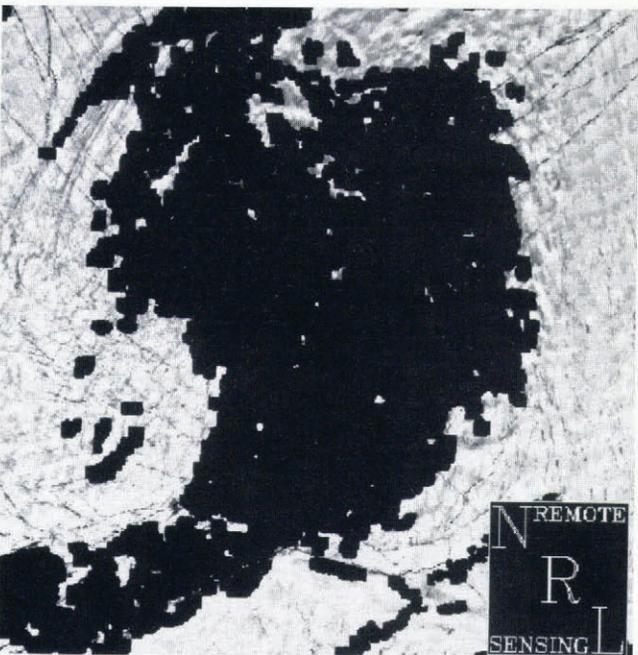


c

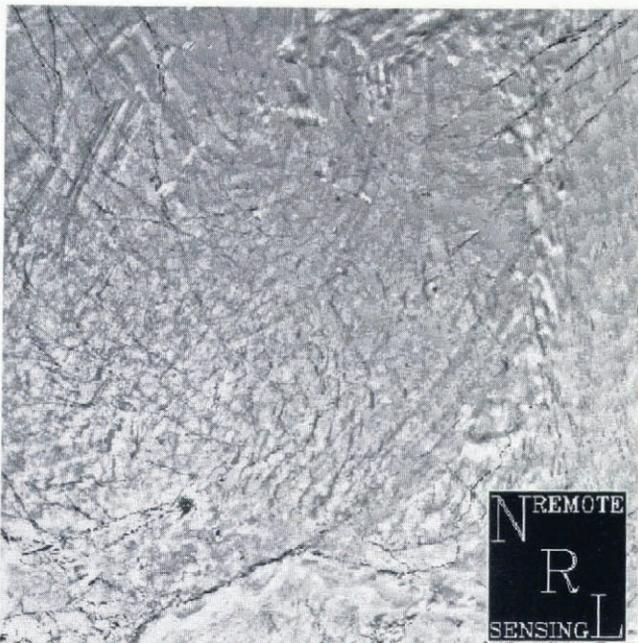
Fig. 1. A view of the East Greenland Sea ice edge by NOAA-11 AVHRR on 23 April 1989. (a) Image processed using $((Ch1 + Ch2) - (Ch3 + Ch4))$. Position of the ice edge in this image is indicated by the letter A. Clouds are observed covering large parts of the ice edge at three locations. (b) Clouds that were apparent in Figure 1a are identified by the cloud-masking algorithm and are colored black. (c) The cloud-filled pixels are replaced in this image with SSM/I data. Letters A and B show the position of the ice edge in the AVHRR and SSM/I data, respectively.



a



b



c

by the cloud algorithm and colored black in Figure 1b. Figure 1c shows a continuous ice edge along the Greenland coast. In this image, ice-edge pixels that were previously identified by the algorithm are now removed and replaced with cloud-free SSM/I data.

Pixels were replaced by a simple operation in which all the pixels which exhibited a value of zero (clouds) in the cloud mask were replaced with ice-concentration estimates derived from SSM/I data. The SSM/I data had been previously warped to fit the AVHRR spatial resolution. The ice-concentration estimates were produced from the 19 and 37 GHz channels of the SSM/I where the effect of clouds is minimal (Hollinger, 1991). Pixels in the cloud mask which had a value of zero, but for which the ice-concentration estimate was zero, remained as zero, and thus appear as clouds in the scenes.

In the second set of images, Figure 2a presents East Siberian Sea ice-lead data corresponding to 7 April 1989. A large cloud is observed blocking the ice-lead view in the central part of the image. These cloud-filled pixels appear black in Figure 2b after identification by the cloud algorithm. In Figure 2c, the cloud-filled pixels are replaced with cloud-free AVHRR data from 8 April 1989. Although this technique may prove useful for monitoring long-lived leads, it may not be too accurate for highly dynamic lead patterns which may open and close and possibly re-open in the time span between AVHRR acquisitions.

In both data sets, the cloud-masking procedures made possible the identification of the cloud-filled pixels and marked them for replacement with cloud-free pixels available from other data sets. In comparisons with the old threshold method, this new technique appears to provide a more accurate estimation of the clouds extent in the scene. In the first case (Figs 1a–c), it identified 24% more cloud pixels than the threshold method. This was mainly the result of the threshold method failure to identify various sets of clouds in the lower part of the scene. In the second case (Figs 2a–c), it removed only those pixels that upon observation appeared as clouds. Conversely, the thresholds removed lead information if set low and clear ice information if set high. We were unable to produce, without manual removal of the clouds, an accurate cloud mask using exclusively thresholds.

CONCLUSIONS AND FUTURE WORK

We have produced an algorithm that consistently identified and removed clouds in AVHRR data containing ice features required for accurate sea ice mapping

Fig. 2. A view of ice leads in the East Siberian Sea as depicted by the NOAA-11 AVHRR on 7 April 1989. (a) An extensive cloud appears covering the center part of the image. A Wallis filter has been used to enhance the cloud-filled pixels. (b) The cloud-filled pixels identified by the masking algorithm appear as black. (c) Pixels that were removed due to cloud contamination are now replaced with data from a next-day pass (8 April 1989). This composite shows that, although there is a 24 h time difference between the two data sets, ice-lead information is still intact.

efforts. The procedure was tested over limited data sets that included images from the East Greenland Sea ice edge and leads within the East Siberian Sea. Preliminary testing of this algorithm indicates a potential for development into an operational procedure. This technique failed to identify some thin cirrus and low clouds in the AVHRR data, a common problem with polar cloud-identification algorithms. However, it did not present a problem to our ice-mapping efforts because we were able to extract useful surface information through those types of clouds. We recognize that the presence of such clouds may interfere in the future with automated methods designed to extract ice-feature statistics.

The method has some limitations. Because it relies on reflective channels for cloud identification, it is restricted to the polar summer months. At present, it is not fully automated and identifies clouds based on sea-ice features of interest. Future plans for refinement of this method include the production of improved clear-sky radiance estimates, generation of additional surface-texture information and, if possible, full automation of the technique. Testing of this algorithm should be extended to a larger data set and to other regions of the Arctic and Antarctic.

ACKNOWLEDGEMENTS

This work was supported by the Chief of Naval Operations (OP-096), the Space and Naval Warfare Systems Command Satellite Applications and Technology Program, Program Element No. 0603704N, Cdr P. Ranelli, Program Manager. We thank R. Holyer and S. Peckinpaugh of NRL-SSC for providing the initial cluster-shade software upon which this algorithm was built, D. Gineris of Sverdrup Technologies Inc. for the AVHRR-SMM/I data used in the production of Figure 1a-c. and R. Grant of Sverdrup Technologies, Inc. for processing the AVHRR datasets used in testing the algorithm.

REFERENCES

Coakley, J. A., Jr, and F. P. Bretherton. 1982. Cloud cover from high-resolution scanner data: detecting and

allowing for partially filled fields of view. *J. Geophys. Res.*, **87**(C7), 4917-4932.

Ebert, E. 1987. A pattern recognition technique for distinguishing surface and cloud types in the polar regions. *J. Climate Appl. Meteorol.*, **26**(10), 1412-1427.

Fetterer, F. M. and J. D. Hawkins. 1991. An AVHRR data set for the Arctic leads ARI. Naval Oceanographic and Atmospheric Laboratory. Technical Note 118.

Hollinger, J. P. 1989. *DMSP special sensor microwave/imager calibration/validation. Final report 2*. Washington, DC, Naval Research Laboratory.

Holyer, R. J. and S. H. Peckinpaugh. 1989. Edge detection applied to satellite imagery of the oceans. *IEEE Trans. Geosci. Remote Sensing*, **27**, 46-56.

Key, J. 1990. Cloud cover analysis with Arctic Advanced Very High Resolution Radiometer data. 2. Classification with spectral and textural measures. *J. Geophys. Res.*, **95**(D6), 7661-7675.

Massom, R. and J. Comiso 1991. The detection of new sea ice and surface temperature using Advanced Very High Resolution Radiometer data and Special Sensor Microwave/Imager satellite data. *IGARSS'91. Remote sensing: global monitoring for Earth management. 1991 International Geoscience and Remote Sensing Symposium, Helsinki University of Technology, Espoo, Finland, June 3-6, 1991*, 791-795.

Rossow, W. B. and R. A. Schiffer. 1991. ISCCP cloud data products. *Bull. Am. Meteorol. Soc.*, **72**, 2-20.

U.S. Navy. 1991. *Ice observation handbook*. Washington, DC, Naval Polar Oceanographic Center.

Welch, R. M., S. K. Sengupta and D. W. Chen. 1988. Cloud field classification based upon high spatial resolution textural features. 1. Gray level co-occurrence matrix approach. *J. Geophys. Res.*, **93**(D10), 12,663-12,681.

Welch, R. M., K. S. Kuo and S. K. Sengupta. 1990. Cloud and surface textural features in polar regions. *IEEE Trans. Geosci. Remote Sensing*, **28**(4), 520-528.

The accuracy of references in the text and in this list is the responsibility of the authors, to whom queries should be addressed.