

# Analyzing Lead Information from SAR Images

Michele M. Van Dyne, Costas Tsatsoulis, *IEEE, Member*, and Florence Fetterer

**Abstract**—Leads are relatively linear features in the sea ice cover, which are composed of open water or new, thin ice. Because of their composition, leads impact the ocean/air heat exchange. Automated analysis of leads from sea ice imagery may provide a means of gathering important information about the sea ice cover and its climatic influence. This paper describes: 1) a method for extracting and analyzing leads from ERS-1 synthetic aperture radar (SAR) images classified by ice type and 2) the results of using this method on images of the Beaufort Sea. The methodology consists of identifying potential lead features in the image and measuring their characteristics both before and after using a thinning or skeletonization technique on the features. The measurements obtained using this method include lead area, average width, number of leads in an area, amount of branching, and linearity of the lead. These measurements were analyzed with respect to the time of year and the latitude of the images. Results indicate that the measurements produced by the methodology are consistent with lead measurement distributions that others have found. The results of the study suggest that the methodology is appropriate to study lead characteristics on a large scale.

**Index Terms**—Beaufort Sea, image analysis, image processing, leads, lead statistics, SAR, sea ice.

## I. INTRODUCTION

GW AND Tucker [8] describe leads as cracks in the sea ice cover of the Arctic that are formed by divergent motion. This causes the ice to pull apart or shear and allows the ocean to come into direct contact with the atmosphere. Along with polynyas, leads are significant sources of heat loss to the atmosphere and during the winter constitute a major source of new ice growth and salt fluxes into the upper ocean. Zonally averaged models predict that an increase of several percent in the winter lead area would increase polar-surface air temperature by several degrees Kelvin. One estimate of lead coverage in the Arctic is that open water constitutes about 1% of the area in the winter [14] and increases to 10–20% during the summer [8]. Another estimate of area covered by leads in the central Arctic is 1–5% coverage in October, while that area reaches 5–20% in the marginal ice zone during the same time frame [20].

Leads rapidly freeze during the winter so that they may be composed of either open water or new, thin ice. Then, the identification and analysis of leads in synthetic aperture radar (SAR) images depends upon an adequate classification

of sea ice types or thicknesses. In general, winter multiyear ice (thick ice that has survived a summer's melt) has the highest backscatter and appears brightest in SAR images. This is because multiyear ice has been desalinated by the melt process and has dielectric properties, which permit volume scattering of radar energy from bubbles in the ice. First-year ice (ice in its first year that has grown to at least 30-cm thick) has lower backscatter because it is smoother than multiyear ice, and surface rather than volume scattering dominates the radar return. As a rule, new or young ice and smooth open water have the lowest backscatter near or below the noise floor of the sensor.

Classification of ice types in SAR images has been addressed using several methods. One approach uses cluster analysis and subsequent cluster labeling, which uses lookup tables for backscatter signatures of different ice types [10], [12]. This algorithm produced classified SAR ice-type images through the Alaska SAR Facility (ASF) Geophysical Processor System (GPS). The technique was validated under winter Arctic conditions, and it has been demonstrated that the classification process is very good at separating multiyear from first-year ice types [3], although the algorithm divides first-year ice into smooth and rough classes with doubtful accuracy [5]. A more serious problem is the difficulty the classification algorithm has distinguishing younger ice types. As new ice grows in leads, the backscatter of the ice generally evolves through a sequence, during which its signature can equal that of multiyear or first-year ice [18]. Also, wind-roughened water in an open lead can have a signature as bright as multiyear ice. The new ice or open water in a lead may then be misclassified as multiyear or first-year ice, depending on when in the evolution of a lead the image is acquired. For a more thorough assessment of this classification method, see Fetterer *et al.* [5].

Several methods for extracting and analyzing leads from images have been used. Key *et al.* [11] used a combination of numeric and symbolic algorithms to detect and map leads from LANDSAT data. The numeric process used was a dynamic-thresholding approach to separate potential leads from nonleads based on signature, and the symbolic approach was a rule-based system that then determined whether potential leads were in fact leads based on size, location, linearity, and orientation in relation to other objects. Lindsay and Rothrock [13] used Advanced Very High-Resolution Radiometer (AVHRR) infrared imagery to study leads in the Arctic. A threshold based on the potential open water that may occupy a pixel is used to define lead pixels; lead widths and separation distances are then obtained using a large number of transects randomly placed on the images. Fily and Rothrock [6] used

Manuscript received January 12, 1995; revised March 10, 1997. This work was supported in part by NASA Research Grant NAGW-3043 and by NSF Grant CISE-CDA-9401021.

M. M. Van Dyne and C. Tsatsoulis are with the Department of Electrical Engineering and Computer Science, University of Kansas, Lawrence, KS 66045 USA (e-mail: tsatsoulis@eecs.ukans.edu).

F. Fetterer is with the National Snow and Ice Data Center, CIRES, University of Colorado, Boulder, CO 80309-0449 USA.

Publisher Item Identifier S 0196-2892(98)00554-3.

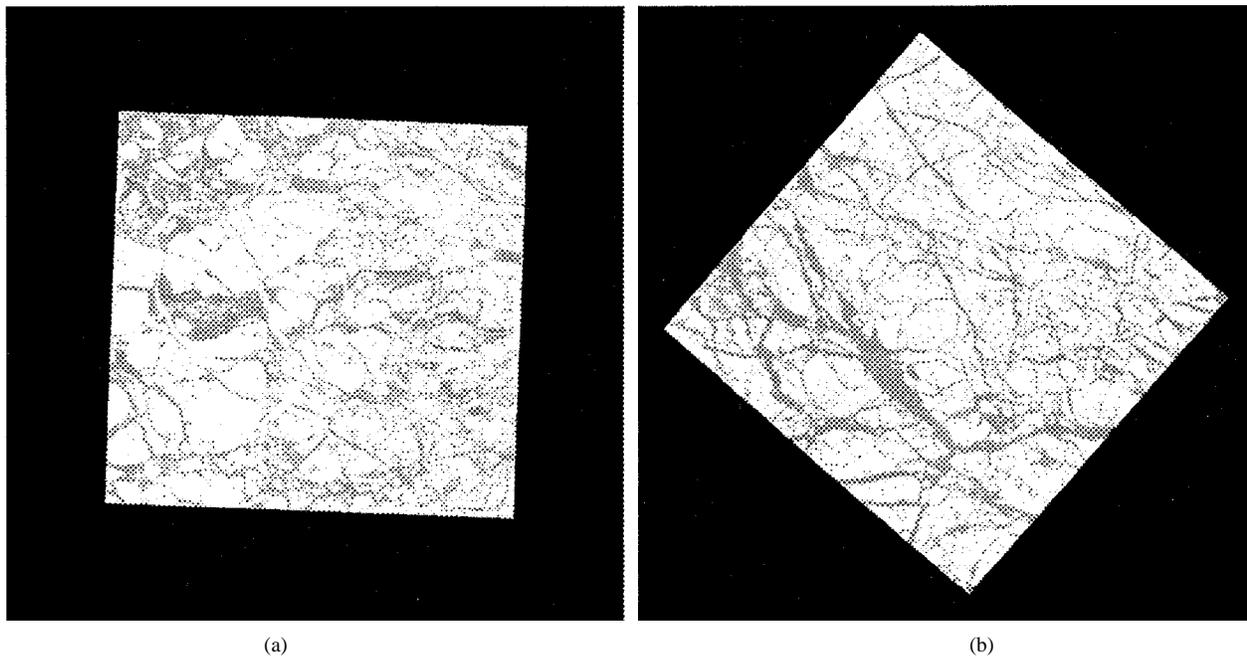


Fig. 1. Classified images. The ASF/GPS classified image product available from ASF provides a  $1536 \times 1536$ -pixel geocoded image that has been classified into the four ice types: multiyear ice, first-year rough, first-year smooth, and new ice/open water represented, respectively, by progressively darker grayscale tones in the images. (a) Image 9511 was acquired December 4, 1991, at center latitude  $73.607$ , longitude  $-139.109$ . (b) Image 8306 was acquired February 7, 1992, at center latitude  $73.291$ , longitude  $-153.771$ .

sequential SAR images to measure the opening and closing of sea ice leads. They base their identification of leads on the assumption that leads are closed linear features surrounded by thick ice, and the initial classification is done using a simple thresholding technique. Because the focus of the research was the change in lead coverage through opening and closing, additional measurements of the leads were not made. Banfield [1] used a mathematical morphology approach to provide a structural description of lead shape and to measure certain characteristics of leads in SAR imagery. Lead and nonlead pixels are determined by choosing an intensity threshold and classifying pixels above that threshold as ice and those below as leads. Banfield indicates that including additional criteria for identifying lead features, such as the linearity of the features, would improve results. Banfield discusses measurements that can be made of lead features, based on their skeletal structures, which serves as a basis for the lead measurements presented here.

Cunningham *et al.* [2] used Banfield's method [1] to measure lead-orientation characteristics in the Beaufort Sea from January to March 1992 and to relate the orientation data to the principal direction of shear within ice-motion fields. In a classified (ice-type) image, all pixels classified as nonmultiyear ice were taken as making up potential lead features, and false lead features were then filtered out based on size and excessive curvature. These authors suggest that additional work could be done based on other lead measurements available, such as lead lengths and average widths.

The images used for lead identification in this work are the ASF/GPS classified images produced from ERS-1 SAR data. In these images, pixels are classified as multiyear ice, first-year rough ice, first-year smooth ice, and new ice/open water. The pixel size is 100 m. The classified images are generated

from geocoded or earth-located images. Geocoding places the  $100\text{-km}^2$  SAR image in a fixed geographical reference frame. A total of 48 classified images of the Beaufort Sea from October 1991 to March 1992 were used in this study to test the methodology and generate lead measurements. The images come from a box bounded by  $69^\circ$  S,  $78^\circ$  N,  $-160^\circ$  W, and  $-120^\circ$  E. Fig. 1 shows two examples of classified images from the test image set.

When measuring lead characteristics in the Beaufort Sea area, we should expect, based on the observations of others [13], [17], [21], that lead size (width or area) should exhibit a roughly exponential or power-law distribution overall. Furthermore, there should be an approximately  $30^\circ$  angle for intersections between leads. This angle is a result of an angle of internal friction for sea ice of about  $15^\circ$  [4]. Here we will describe a method for identifying leads in SAR imagery that is based on Banfield's method [1] and will test its utility for obtaining lead statistics automatically by seeing if the above statements about lead-population characteristics hold true for our sample of SAR images.

## II. METHOD

The approach to lead measurements is to refine the initial ASF/GPS classification based on shape and size clues. The method used is best described in two sections: 1) the computer algorithms for automated lead identification and measurement and 2) the procedure used in analyzing the output measurements on Beaufort Sea imagery.

### A. Automated Lead Identification and Measurement

In extracting potential lead features from the ASF/GPS classified images, images are first thresholded based on ice-type classification, then each potential lead feature is distin-

guished from others by assigning it a unique number. Initial measurements are made at this point, then the leads are thinned into a “skeletonized” representation, and finally additional measurements are made of the lead. Each of these processes is described more fully in the following sections. During all processes that modify the image, the changes are displayed on the screen for visual inspection, although the display function of the program may be turned off by the user, allowing the code to run more quickly.

1) *Thresholding*: Thresholding, based on the ice-type classifications, consists of assigning the color white to all ice types of interest and black to those ice types that are not of interest for analysis. The program allows the user to choose which ice-type classifications are of interest for further analysis.

2) *Feature Identification*: After thresholding, an algorithm to label connected components or features is run on the image to distinguish different features by assigning each a unique number or “color.” (During display of the images, each unique number is drawn on the screen as a different color.) The algorithm uses an eight-neighborhood approach, in which pixels to the top, bottom, left, right, and diagonal to the current pixel are all considered connected to that pixel. The traditional approach to this process [7] is to scan the image pixel by pixel from left to right and top to bottom. When an unnumbered pixel is encountered, the order of scanning the image guarantees that its upper left, upper, upper right, and left neighbors have all been previously scanned. Then, to assign a label to the current pixel, the algorithm looks at the four neighbors it has already encountered and, if they are all blank, assigns a new number to the current pixel. If some of the neighbors already encountered have been assigned the same number and the rest are blank, that number is assigned to the current pixel. Finally, if some of the previously scanned neighbors have been assigned different numbers, the current pixel is assigned one of these numbers and then the numbers are entered into an equivalence table. This process continues until the entire image has been scanned and all nonblank pixels have been assigned a color. A second pass is then made to sort the equivalent labels and assign a single label to these classes. These new labels are assigned to the pixels in the image.

This traditional labeling process was modified in the current study so that only a single scan need be made over the full image, with numbering passes done only in areas where an unnumbered feature has been encountered. In the modified approach, when an unnumbered pixel is encountered, it is assigned a new number or label, then all of its neighboring pixels and their neighbors are then also assigned the same number using a nonrecursive “flood-fill” or region-growing process and an eight-neighbor scheme. The flood-fill process looks first at the current pixel’s right, lower left, lower, and lower right neighbors, and if they are unnumbered (but not black), they are all assigned the same label as was just assigned to the current pixel. A backward pass is then done looking at the left, upper left, upper, and upper right neighbors. The furthest pixels define a maximum rectangle about the feature, and the algorithm need never look more than one pixel beyond this boundary. As a side benefit, the process produces the coordinates of the bounding rectangle of the feature and the

area, in pixels, of the feature. The flood-fill process continues on a feature until no more unnumbered connected pixels are found. The scan of the image is then restarted at the pixel it initially encountered.

This process produces a measurement (in pixels) of the area, length, and average width of each of the identified features, the coordinates of the bounding rectangle about the feature, and a count of the number of features found in the image.

3) *Feature Measurement*: Initial measurements are made of the full feature before additional processing continues. The perimeter of the feature is determined by finding all the pixels on the periphery of that feature and is measured in terms of the number of pixels. The main length of the feature is defined as the maximum Euclidean distance between two points on the boundary of the feature. The average width of the feature is also computed. At this point, features with a small length-to-width or elongation ratio may be eliminated from further consideration. The cutoff value for the elongation ratio may be changed by the user. The program also allows the user to specify a feature size, such that features with an area under that size will not be included in further analysis.

4) *Thinning*: Next, features identified as potential leads are subjected to a skeletonization process. The result of this process is a skeleton or stick figure representation of the feature. This representation facilitates additional measurements of the feature that may not be obvious when examining the full feature. For example, the end-to-end length of the potential lead may be measured from the full feature, but using the main channel of the skeleton, or the path from end to end, gives a clearer picture of its deviation from the straight line length and, thus, a measure of the linearity of the feature. Additionally, branching off the thinned figure occurs where there are protrusions on the original figure. Using the thinned representation, lengths, and angular relationships of branches to the main channel of the potential lead can be measured more easily than can be done with the full feature.

Rather than using a mathematical morphology approach [1] to producing skeletons from leads, a thinning algorithm is used, since a drawback of the mathematical morphology approach is that connectivity of the resulting skeleton is not guaranteed. The thinning algorithm guarantees that the resultant skeleton will be maximally thin, minimally eroded, and connectivity of previously connected elements will be retained. “Maximally thin” skeletons are those that are no more than 1-pixel thick, and minimal erosion is desirable so that the endpoints of the feature are not lost during the thinning process. Finally, preservation of connectivity is desired so that the measurements made reflect the original features and not pieces of those features.

The thinning algorithm is performed by making a first pass through the image and marking pixels for deletion whenever the deletion of a particular pixel will not disconnect or erode the endpoints of the eventual skeleton. The marked pixels are then deleted, and a second pass is made using slightly different criteria than the first pass in marking pixels for deletion. The two passes are made iteratively until no more pixels are marked in either pass for deletion, that is, until the lead skeletons are of minimal width. This process also uses

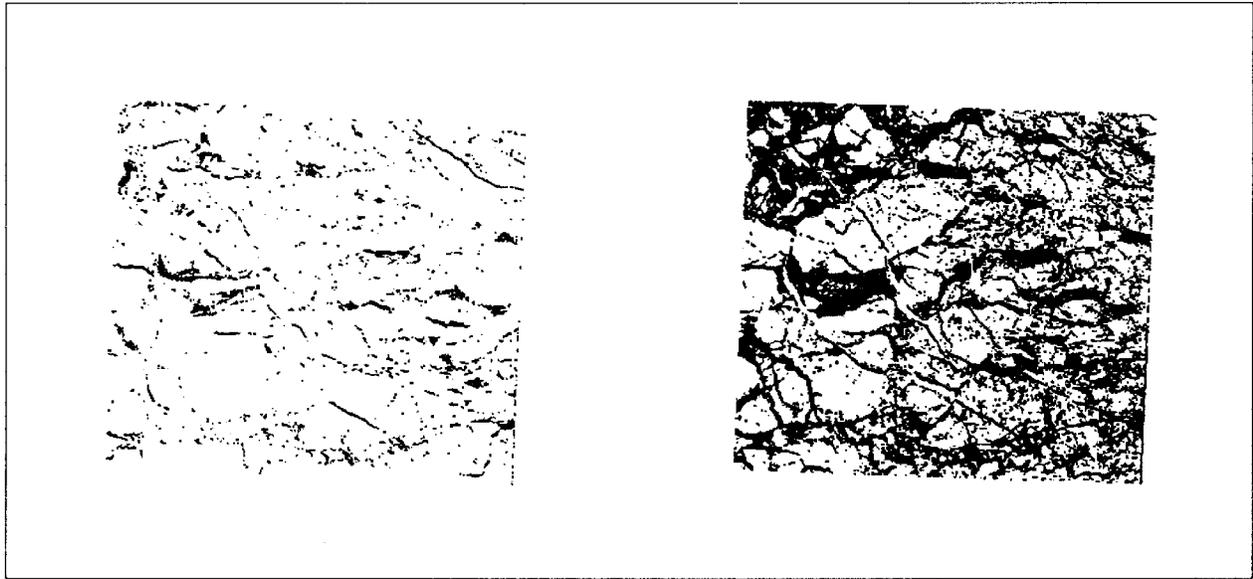


Fig. 2. Thresholded images. Thresholded representations of image 9511 of Fig. 1(a), where the ice types of interest are black and the rest of the image is white. Including three ice-types overestimates potential lead area. (a) Thresholding on the two youngest ice types (first-year smooth and new ice/open water). (b) Thresholding on the three youngest ice types (first-year rough, first-year smooth, and new ice/open water).

an eight-neighborhood approach. This algorithm is described in more detail by Gonzales and Woods [7] and Van Dyne and Tsatsoulis [19], and the procedure is defined as follows.

*Pass 1:* Flag a pixel for deletion if:

- a) the number of neighbors is between two and six;
- b) the number of transitions from zero to one when traversing through neighboring pixels is equal to one;
- c) there is at least one zero-valued pixel in the upper, right, or lower neighbors;
- d) there is at least one zero-valued pixel in the right, lower, or left neighbors.

Delete flagged pixels after the entire image has been examined.

*Pass 2:* Flag a pixel for deletion if:

- a) the number of neighbors is between two and six;
- b) the number of transitions from zero to one when traversing through neighboring pixels is equal to one;
- c) there is at least one zero-valued pixel in the upper, right, or left neighbors;
- d) there is at least one zero-valued pixel in the upper, lower, or left neighbors.

Delete flagged pixels after the entire image has been examined. Repeat until no more pixels are flagged for deletion.

Condition a) ensures that no interior points or endpoints are flagged for deletion. Condition b) prevents a pixel from being removed from a 1-pixel thick line, so that the skeleton cannot become disconnected. Conditions c) and d) in Pass 1 indicate that the current pixel is an east, south, or northwest boundary point and is not a part of the skeleton, while these conditions in Pass 2 indicate that it is a north, west, or southeast boundary point and can be deleted. The northeast and southwest corner points are found from all four c) and d) conditions.

If the original feature contains holes or islands, the thinning process will produce a skeleton that contains loops. Further-

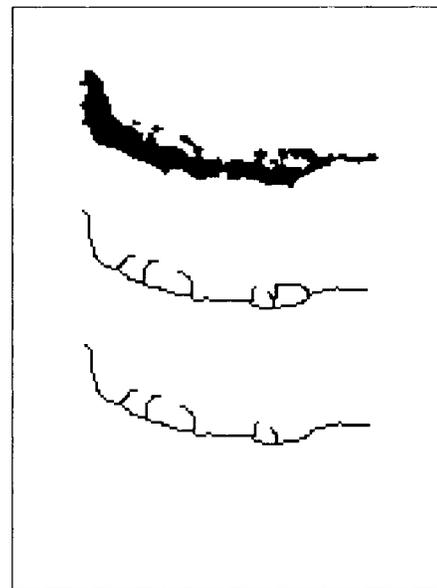


Fig. 3. Identified lead/thinned lead/thinned lead after processing branches. This is an example of one feature identified as a potential lead by the program in image 9511 of Fig. 1(a). The entire lead feature is measured then reduced to a skeletal representation by thinning. Loops that occur in the skeleton are eliminated using a graph spanning-tree traversal algorithm. Once loops are eliminated, additional measurements are taken. (a) Potential lead feature, (b) thinned feature, and (c) thinned feature after branch processing.

more, any corners in the outline of the feature will result in a branch off the skeleton. This is desirable if the corner actually represents a branch, but spurious branches may occur in some instances. These conditions are taken into consideration when making measurements from the skeleton.

5) *Skeleton Measurements:* Many of the measurements used in this work are those originally described by Banfield [1]. These include the main diagonal length, the skeletal length, the total skeletal length, linearity index, branching index, area,

TABLE I  
 DESCRIPTIVE STATISTICS. THE FIRST SIX VARIABLES DESCRIBE STATISTICS ACROSS IMAGES, USING THE 48 TEST IMAGES AS DATA POINTS. THE REMAINING VARIABLES DESCRIBE STATISTICS ACROSS IDENTIFIED LEADS WITH THE 3013 IDENTIFIED LEAD FEATURES AS DATA POINTS

Variable	Mean	Std Dev	Minimum	Maximum	Units
Non-Multiyear Ice Area	1,928.91	1,167.40	268.24	7,129.33	Square Kilometers
Non-Multiyear Ice Coverage	20%	12%	3%	72%	Percentage
New Ice / Open Water Area	642.34	503.42	7.52	2,191.42	Square Kilometers
New Ice / Open Water Coverage	7%	5%	0%	22%	Percentage
Number of Leads per Image	62.77	47.51	2.00	167.00	Nondimensional
Lead Coverage	1%	2%	0%	6%	Percentage
Lead Area	2.19	7.06	0.07	223.06	Square Kilometers
Average Width	0.32	0.24	0.07	2.79	Kilometers
Perimeter	6.31	9.10	0.60	148.90	Kilometers
Main Length	4.16	5.04	0.30	89.90	Kilometers
Skeletal Length	4.37	5.46	0.30	96.30	Kilometers
Total Length	5.32	8.04	0.30	153.30	Kilometers
Elongation Index	8.95	5.64	5.00	138.06	Nondimensional
Branching Index	0.91	0.11	0.44	1.00	Nondimensional
Linearity Index	0.97	0.05	0.85	1.00	Nondimensional
Lead Orientation	94.69	61.56	0.00	178.81	Degrees
Average Number of Branches per Lead	1.22	2.33	0.00	36.00	Nondimensional
Number of Left/Lower Branches	0.57	1.40	0.00	22.00	Nondimensional
Number of Right/Upper Branches	0.65	1.46	0.00	17.00	Nondimensional
Average Left/Lower Branch Angles	-10.38	25.15	-151.65	0.00	Degrees
Average Right/Upper Branch Angles	14.05	30.68	0.00	165.07	Degrees

average width, and main diagonal orientation. Additional measurements made include the number of branches and the average branching angles on either side of the lead. Aggregate measures are also calculated across the lead population in each image, including the overall lead coverage, average lengths, orientations, and areas.

The *main diagonal length* is defined as the distance from endpoint to endpoint of the skeleton. Endpoints are identified on the skeleton as those pixels that have one or two neighbors. If the pixel has two neighbors, these neighbors must also

be neighbors of each other for the pixel to be considered a potential endpoint. The two endpoints of the skeleton are those that result in the largest main diagonal length. In contrast to the main length measured from the full feature, which uses a Euclidean distance, the main diagonal length uses a city-block distance measurement for ease of comparison with additional length measurements, although the Euclidean distance could also be calculated. The city-block distance measurement between the two points,  $p_1$  and  $p_2$ , is defined as  $D(p_1, p_2) = |p_{1x} - p_{2x}| + |p_{1y} - p_{2y}|$ , where  $|\dots|$  indicates

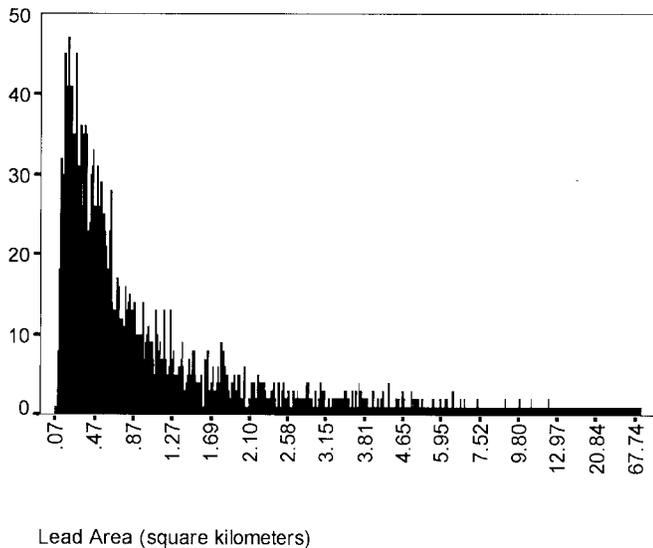


Fig. 4. Lead area. The distribution of lead area across the 48 test images, reported in square kilometers, shows a roughly exponential distribution. Each bar in the graph has a width of  $0.01 \text{ km}^2$  and extreme outliers are not included.

the absolute value and the subscripts  $x$  and  $y$  indicate the  $x$  and  $y$  coordinates of the pixel or the point within the image [7].

The *skeletal length* is determined by traversing the skeleton from endpoint to endpoint and following a minimal path, and it is either equal to or larger than the main diagonal length. Dividing the main diagonal length by the skeletal length produces a measurement of the *linearity* of the feature. The *total skeletal length* is the skeletal length plus the length of each of the branches on the skeleton and is computed by adding the lengths of the branches to the skeletal length. Dividing the skeletal length by the total skeletal length produces a *branching index* of the lead. Main diagonal *orientation* is measured by determining the angle of the main diagonal in relation to the  $x$ -axis of the image and then applying a correction so that the orientation is relative to true north (e.g., leads aligned in a northwest-southeast direction would have an orientation of  $135^\circ$ ).

The *number of branches* is determined by traversing the path of the skeletal length and finding protrusions or junctions on this main channel. Since there may be loops represented in protrusions off the main channel, the structure of a "branch" may be more of a cyclic graph at this point. A spanning tree is created from this graph with the root of the spanning tree at the junction of the branch with the main channel. Graham *et al.* [9] define a spanning tree of a graph as a subgraph of the original graph that contains all the vertices and enough of the original edges that the subgraph is connected, but contains no cycles. The endpoint of a branch is determined by finding the farthest leaf from the root of the spanning tree. Traversal of the branch is then done in the same manner as the skeletal-length traversal, that is, finding the minimum path from the junction to the endpoint, which determines the *branch length*. From the branch endpoint and its junction with the skeleton, a *branching angle* relative to the main diagonal orientation is calculated. *Average angles* are calculated by looking at all branches that are above or below the more horizontally oriented leads and branches that are to the right or left of the more vertical leads.

Within the aggregate or population measures for images, the *overall lead coverage* is the ratio of lead pixels to all pixels within the entire image. *Average lengths*, *orientation*, and *area* are all calculated by taking the averages of the individual lead measurements.

Typically, geophysicists are interested in lead breaking and intersection angles or lead width as a function of density. The additional, more esoteric descriptive measurements were derived for two reasons: first, they might prove useful for screening real leads from false ones (this proved not to be the case) and second, we were interested in seeing if such measures as length, linearity, or branching index showed regional or temporal organization. For instance, it might be expected that the leads are longer and less branched at high latitudes in winter because the pack is consolidated and relatively homogenous there.

### B. Lead Measurement Analysis

The program was first run on 120 ASF/GPS classified images taken between October 1991 and March 1992, with 20 images selected from each month. Seven of these images were subsequently removed from consideration because they included large areas of land that produced incorrect ice-type classifications. Images were then compared by latitude and longitude for each month, and those that created areas of overlap were removed so that statistical analysis would not include duplicate lead features. (Leads can appear quite static with the same lead feature identifiable in SAR imagery for a month or longer. Of course, during this time the ice in the lead may grow to first-year thickness.) After eliminating images of duplicate areas, there were 48 remaining images used for analysis. With the exception of March, which only had three images, each month had eight to eleven images.

The requirements for initial measurements of the elongation ratio were varied. Features with a length-to-width ratio of less than 5:1 were removed from consideration because, by definition, leads are relatively long and narrow. Additionally, statistics were generated on the linearity index of all the identified lead features, and the initial mean was 0.94 with a standard deviation of 0.09. Using these values, lead features with a linearity index below one standard deviation below the mean (0.85) were also eliminated from further consideration.

Experimentation was done on the inclusion or exclusion of proportionately small features in the image. The program was run on all features and then was run on only those features that had an area of greater than 100 pixels (approximately  $1 \text{ km}^2$ ). Since statistical analysis showed no difference between the inclusion or exclusion of these smaller features, their measurements were included in the full analysis of the data.

Descriptive statistics of the variables were then computed using the SPSS 6.1 software [15], [16]. The overall area covered by identified lead features versus other ice types in the image was calculated. The distribution of lead-size measurements (area, main length, and average width) and lead orientation were determined to compare with expected results. Lead coverage, area, length, and width were all analyzed, with respect to month and to latitude of the image.

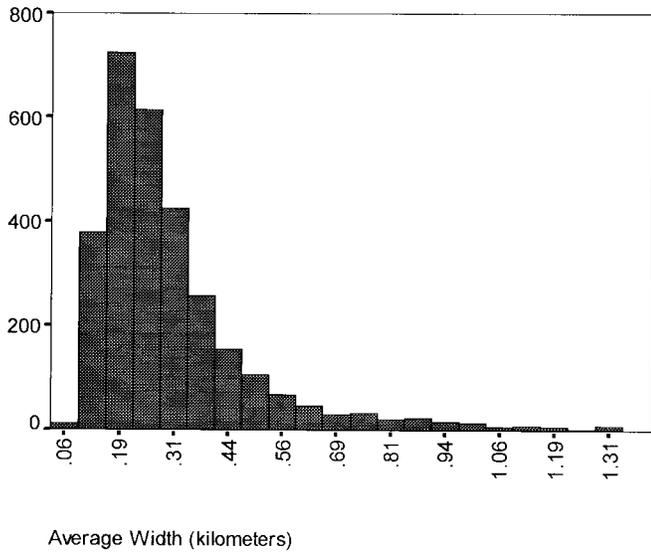


Fig. 5. Average lead width. The distribution of average lead width across the 48 test images also shows a roughly exponential distribution. Average width is shown in kilometers, and each bar in the graph has a width of 0.0625 km. Again, extreme outliers are not included.

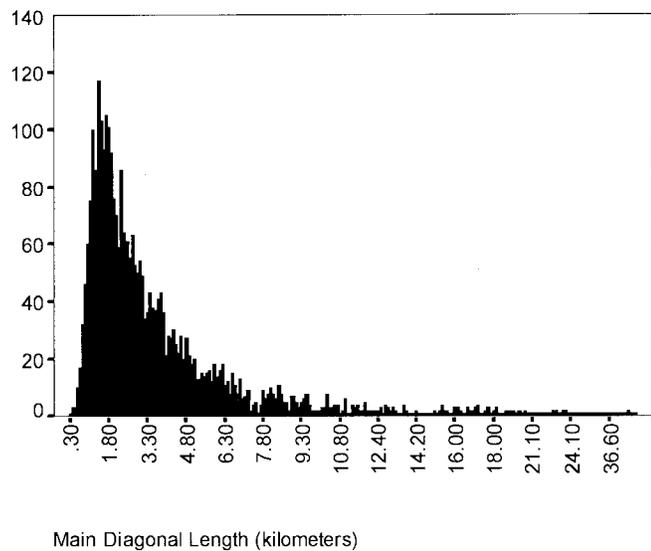


Fig. 6. Main diagonal length. The distribution of main diagonal length across the 48 test images also shows a roughly exponential distribution. Average length is reported in kilometers. Extreme outliers are not included, and each bar in the graph has a width of 0.1 km.

Average air temperatures were obtained from drifting buoys. There were eight buoys reporting valid temperatures from the Beaufort Sea during the time period the images were taken, although not all of the buoys reported data for all months. Temperatures ranged from a mean of  $-31.96^\circ$  for January to a mean of  $-6.70^\circ$  for October. With these low temperatures, the backscatter of multiyear ice is stable, and the ice-type algorithm works well for multiyear ice.

### III. RESULTS

The average processing time was 47.5 s/image on a 133-MHz Pentium machine. However, the processing time of a given image is dependent on the number and complexity

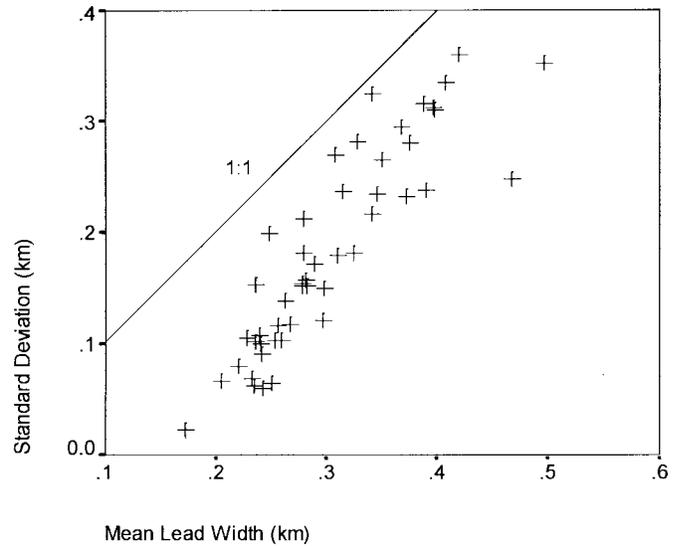


Fig. 7. Standard deviation versus mean lead width. In a true exponential distribution, the standard deviation should equal the mean of the measure. In the case of lead width, the slope is slightly greater than one, which indicates that the tail of the width distribution has more weight than a true exponential distribution.

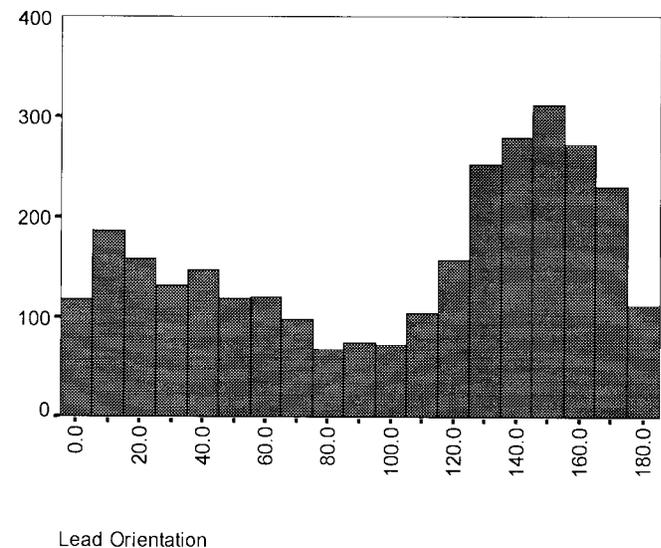


Fig. 8. Lead orientation. The distribution of the orientation of leads with respect to true north shows a bimodal distribution with peaks around 10 and 150° and a separation of about 40°. Each bar in the graph has a width of 10°.

of the lead features in the image, so there was a wide variation in processing time across images. The performance was measured with the image-display capability turned on. Additional processing-time savings can be realized by turning the display variable off.

The program was run with either three nonmultiyear ice types (first-year rough ice, first-year smooth ice, and new ice/open water) or two nonmultiyear ice types (first-year smooth ice and new ice/open water) as potential leads. Fig. 2 shows the results of both variations in thresholding to obtain potential lead pixels. The very large area identified as potential lead features, using the three youngest classification types [Fig. 2(b)], contains many pixels on the interior of floes that

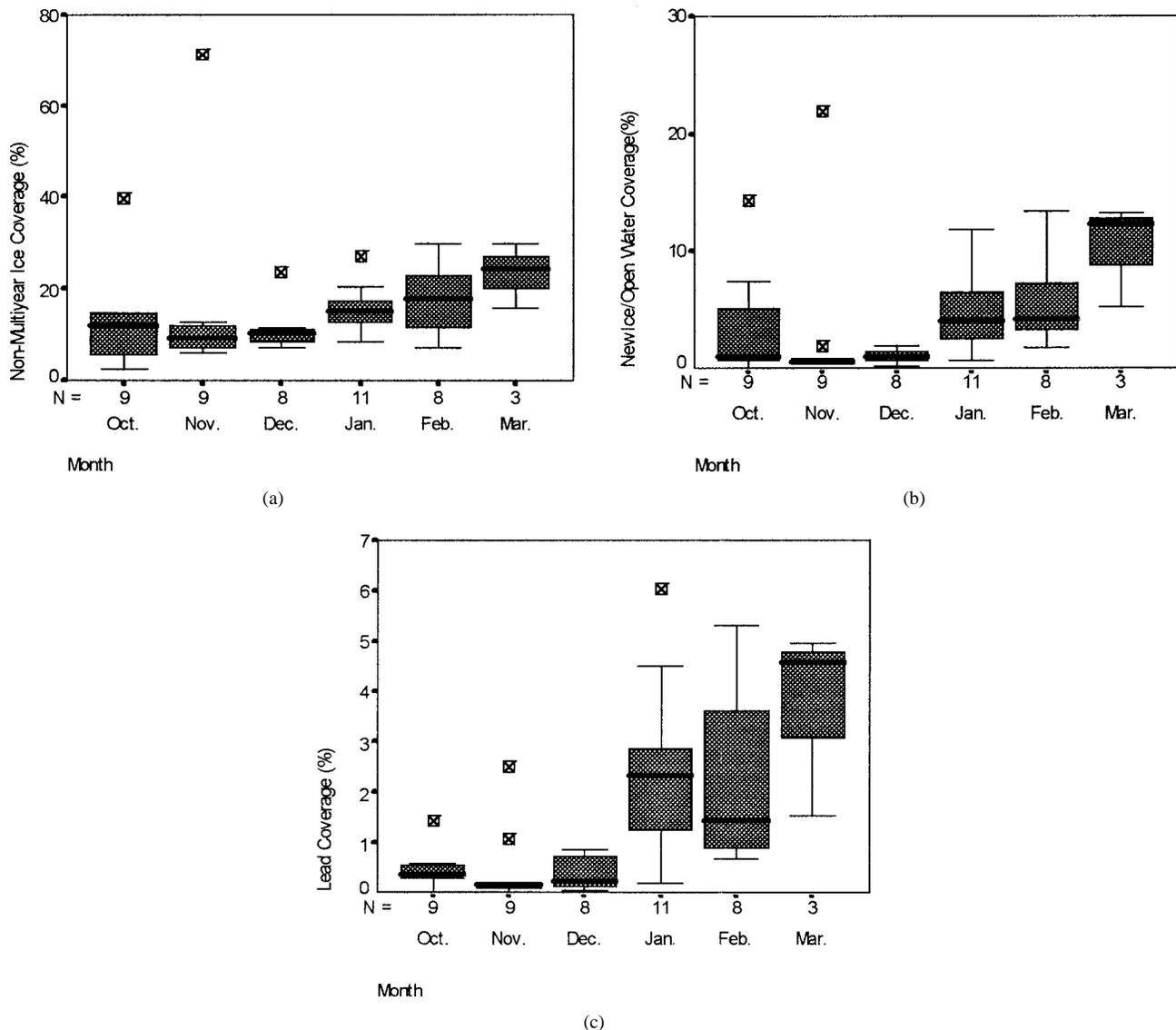


Fig. 9. Area coverage by month. Coverage is reported as the percentage of the total image area classified by ice type or lead feature versus total image area aggregated by month. The box plots show a line at the median, while the full box shows quartiles and the lines extend from the boxes to maximum and minimum values, excluding outliers. Extremes and outliers are indicated by the symbols beyond the range of the boxes and lines. The number of images included in each month is shown below the plots. (a) Nonmultiyear ice coverage by month, (b) new ice/open water coverage by month, and (c) lead coverage by month.

are clearly not lead pixels. Similar patterns of overestimation of area were found in all the images. It was determined that when using all three younger ice classifications, a tendency to join networks of lead features into single features emerged, and the use of only the two youngest ice classifications agreed more with visual inspection of the images. Subsequent analysis, therefore, uses first-year smooth ice and new ice/open water to identify potential lead features.

The thinning algorithm was found to produce the desired properties of maximally thin, minimally eroded, and connected skeletons, but it did have the side effect of producing branches where both protrusions and corners occurred in the original feature as well as producing loops around "islands" in the feature. The use of the spanning tree of a graph eliminated loops and defining branches as the longest path from the junction with the main channel of the skeleton, and an endpoint eliminated additional branching off of branches. Some spuri-

ous branching off the main channel still remains, however. Thus, the number of branches counted by the program may be somewhat higher than what would be obtained by hand analysis. This spurious branching also influences the measurement of the average branching angles. Fig. 3 shows the results of performing thinning on one of the leads in the image. Fig. 3(a) shows a feature the program identified as a lead and Fig. 3(b) is the skeleton of that lead. The skeletonized lead shows an area where there was an "island" in the original lead and a loop was created in the skeletonized version. The spanning-tree algorithm eliminates these loops, as shown in Fig. 3(c).

When run on the 48 images in the test set, the program identified 3013 features as meeting the defined elongation (length-to-width ratio) and linearity criteria of a potential lead. Table I is a summary of descriptive statistics from all the lead-feature measurements. All coverage measurements made are, with respect to the area within the image, classified as

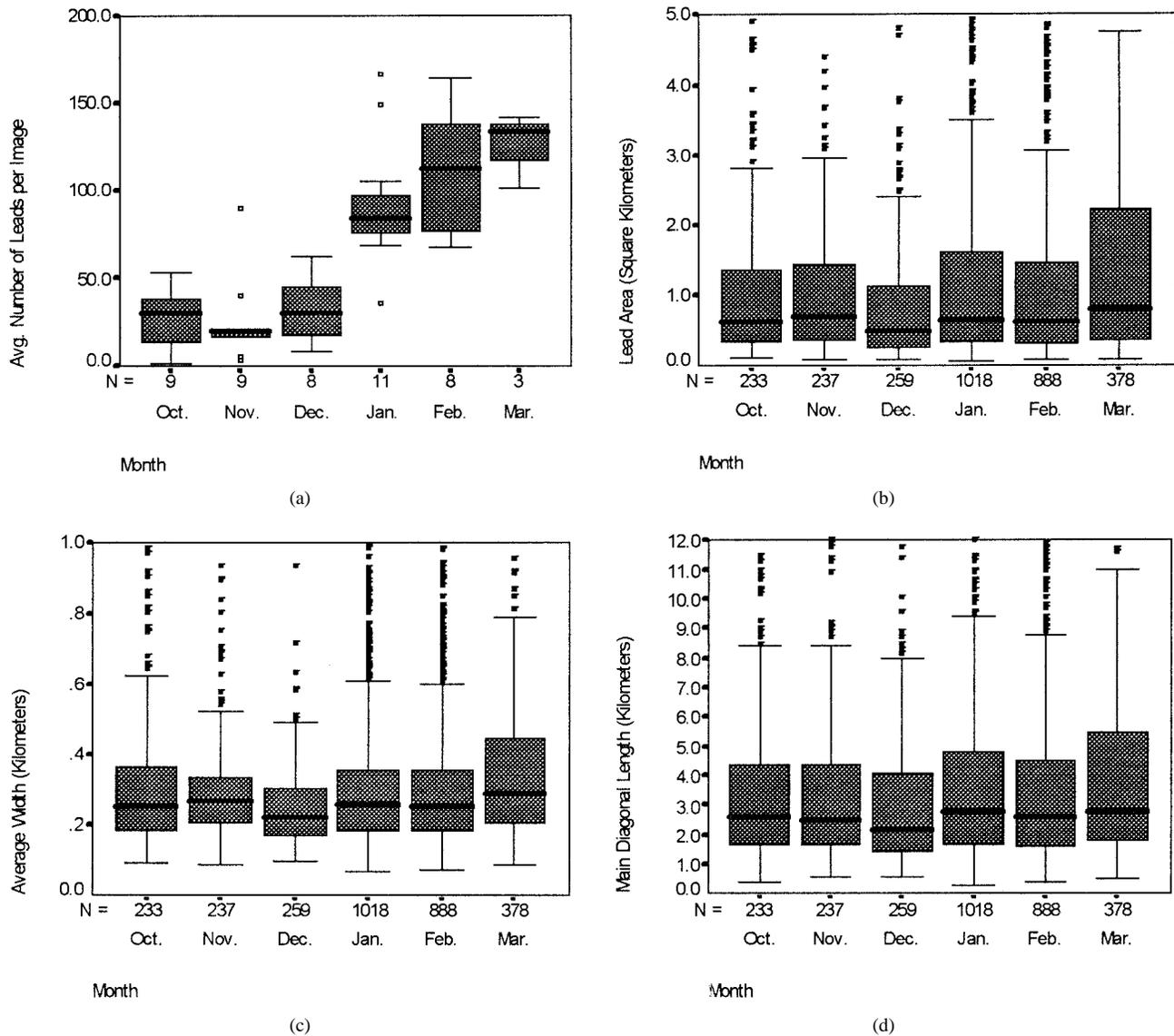


Fig. 10. Lead measurements by month. The number of leads is counted across images within a given month, while the other measurements refer to the overall lead population. Area is shown in km<sup>2</sup> and the length and width are shown in km. While the number of leads found increases in January, February, and March, their area, width, and length do not appear to vary as much. This is consistent with the trend of increasing lead coverage shown in Fig. 9(c). (a) Number of leads by month, (b) area of leads by month, (c) average width of leads by month, and (d) main diagonal length of leads by month.

ice (rather than land or blank space around the image). “Not Multiyear Ice Area” includes all pixels within the image area that are classified as not multiyear ice, that is, either new ice/open water, smooth first-year ice, or rough first-year ice. The “New Ice/Open Water Area” variable includes only pixels classified as that ice type. The “Lead Coverage” variable is a measure of area comprising lead features. Lead parameters, such as the length measures, perimeter, and average width, are reported in units of kilometers. The elongation index is the ratio of main length to average width, the branching index is the ratio of total length to skeletal length, and the linearity index is the ratio of skeletal length to main skeletal length. “Orientation” is the angular orientation of lead features, with respect to the *x*-axis of the image, corrected for its relationship to true north.

As can be seen from the descriptive statistics, the lead coverage reported by the analysis methodology, across all 48

images, averages 1%, with a minimum of 0% and a maximum of 6%. This is consistent with estimates of lead coverage in the Arctic, as reported in the literature. New ice/open water coverage was measured at an average of 7%, with a minimum of 0% and a maximum of 22%. This is somewhat higher than expected for the winter Beaufort Sea, but analysis of the classification algorithm [5] has shown that it classifies many first-year ice areas as new ice. Often these areas can be distinguished from leads by their relatively large size and lack of elongation. By using elongation, the algorithm can eliminate these nonlead features.

Analysis of the distributions of size and orientation measurements are in agreement with the hypothesized distribution. In particular, it was expected that the distributions of lead area, average width, and main diagonal length would appear as exponential distributions, and results meet this expectation (see Figs. 4–6). In Figs. 4–6, extreme outliers were eliminated

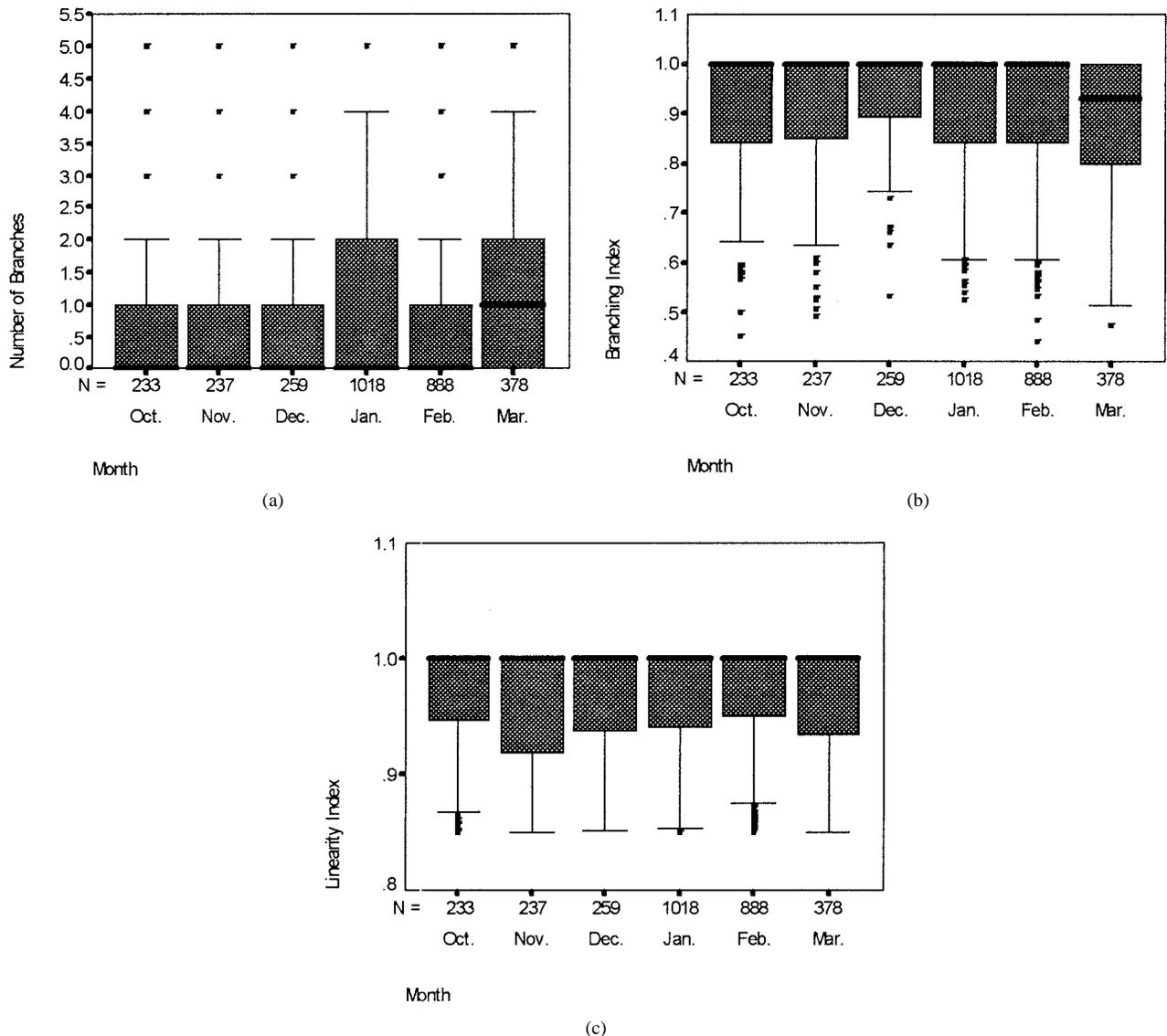


Fig. 11. Branching and linearity by month. There is little change in the number of branches, branching index, or linearity across leads by month, except for a slight trend toward more branching in March. Note that the number of branches is bounded by a minimum of zero branches, while the linearity and branching indexes are bounded by a maximum of 1.0. Linearity also has a minimum bound of 0.85 since features that are less linear are not considered leads. (a) Number of branches by month, (b) branching index by month, and (c) linearity index by month.

from the graphs for clarity in presentation. In Fig. 4, 27 leads exceeded the area of  $39.24 \text{ km}^2$  (the maximum area reported was  $223.06 \text{ km}^2$ ), while in Fig. 5, 26 identified lead features had an average width that exceeded 1.4 km (the maximum average width reported in the descriptive statistics was 2.79 km). In Fig. 6, five lead features exceeded a length of 49.6 km (maximum length was 89.9 km).

To test for a near-exponential distribution, we plotted the standard deviation versus the mean lead width for leads in each image (Fig. 7). Standard deviation should equal mean width for an exponential distribution. Lindsay and Rothrock [13] do the same for leads measured using AVHRR. It is not clear why all the points in Fig. 7 fall below the 1:1 relationship line. However, the points do follow a linear trend and, as Lindsay and Rothrock find [13, Fig. 9], the slope is slightly greater than one, which indicates that the tail of the width distribution has more weight than a true exponential distribution.

The range of widths found by Lindsay and Rothrock [13] is between about 2 and 8 km. Here, our average width is about 0.32 km (2.79 is the maximum for the entire data set; see Table I) and minimum is about 0.2 km. Lead width measurements from AVHRR are hampered by the low resolution of the instrument (1.1 km at nadir). Conversely, the SAR imagery has a pixel size of 100 m (although true resolution is about 240 m). While small leads are easy for the SAR algorithm to identify, large leads may not be. This is because wide leads often have ice at different stages of freezing within them, so that the algorithm tends not to detect a single, whole lead. Another reason for little evidence of large leads in this sample is that there are relatively few wide leads: sampling them with  $100\text{-km}^2$  images is not effective. Based on submarine upward-looking sonar data, Wadhams [21, Fig. 9] has an average figure of 100 km for the distance between 2-km wide leads in the interior pack.

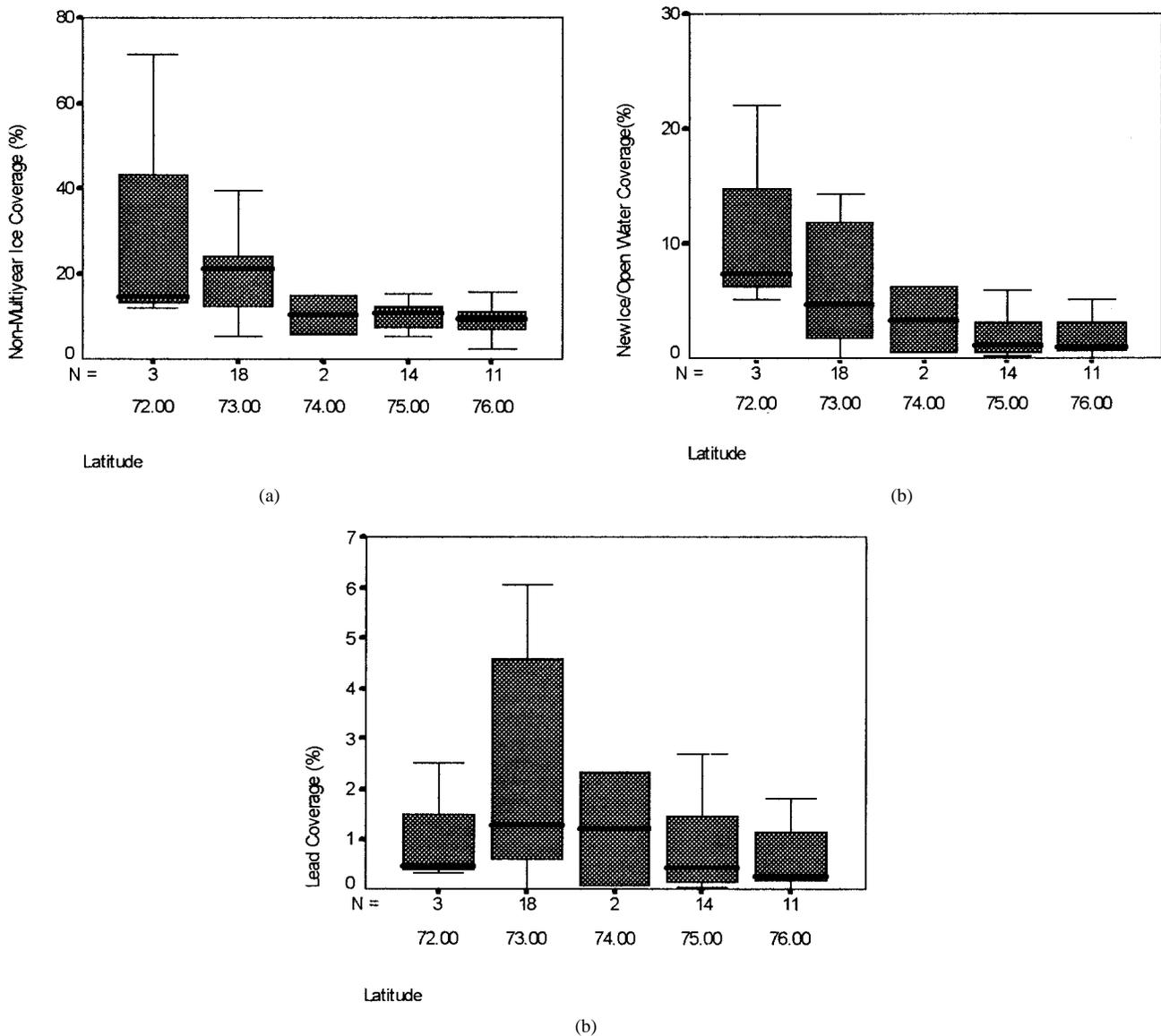


Fig. 12. Area coverage by latitude. Coverage is reported as the percentage of the total image area classified by ice type or lead feature versus total image area aggregated by latitude. The number of images included in each latitude group is shown below the plots. Nonmultiyear ice decreases at increasing latitudes, however, the areal coverage of leads remains fairly consistent. (a) Nonmultiyear ice coverage by latitude, (b) new ice/open water coverage by latitude, and (c) lead coverage by latitude.

Fig. 8 shows the distribution of orientation across all leads in all images. The angular orientation of the leads is with respect to true north and can range from 0–180°. We did not look at orientation statistics on an image-by-image basis. Overall, however, the distribution shows peaks around 10° and around 150°. Curiously, this separation of approximately 40° (between 10 and 150°) is consistent with Erlingsson’s [4] citations of observations of leads crossing at 30°±10° in the Beaufort Sea, with the fracture-pattern angle attributed to the internal angle of friction and the principal direction of stress. This is surprising since the orientation statistics are for six months, rather than a single point in time. The bimodal histogram may have two explanations. One is that wind stress has a preferred direction (from the northwest quadrant in this part of the Beaufort Sea in winter), so that fractures tend to form in the same general directions throughout the winter. Another is that ice retains a record of previous lead patterns

because the signature of leads can remain distinctive for weeks or even months. Cunningham *et al.* [2] discuss this sampling problem. In general, our results were consistent with those of Cunningham *et al.*, who find mostly north- and south-trending leads in the Beaufort Sea below a latitude of about 78° N.

Note that images were selected for this study based on availability rather than a sampling strategy for thoroughly characterizing Beaufort Sea lead characteristics. Therefore, all the derived statistics must be viewed with caution: these results are given primarily to demonstrate that the technique will be useful in a larger study.

Figs. 9–14 are box plots of measurement statistics aggregated over month and latitude range. The box plots have a line at the median, while the box shows quartiles and the lines extend from the boxes to maximum and minimum values, excluding outliers. Extremes and outliers are indicated by the symbols beyond the range of the box and lines.

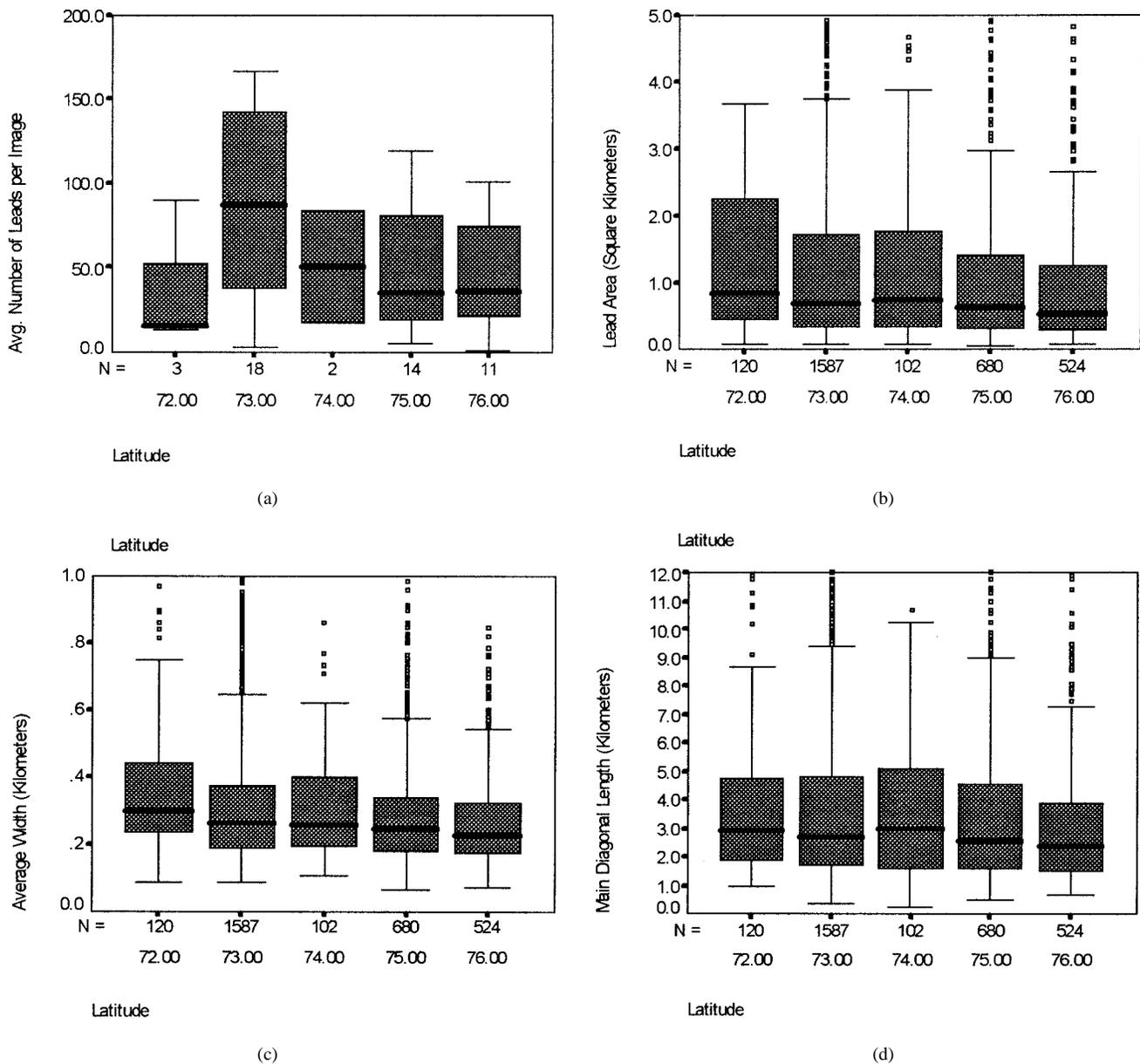


Fig. 13. Lead measurements by latitude. Area is shown in  $\text{km}^2$  and the length and width are shown in km. While the number of leads/image remains fairly constant, there is a slight downward trend in measurements of size with increasing latitude. (a) Number of leads by latitude, (b) area of leads by latitude, (c) average width of leads by latitude, and (d) main diagonal length of leads by latitude.

Fig. 9 shows the coverage of image area by nonmultiyear ice [Fig. 9(a)], by new ice/open water [Fig. 9(b)], and by identified leads [Fig. 9(c)]. There is slightly more nonmultiyear ice with time, but this could easily be a result of the selection of images. The November outlier, showing 72% of nonmultiyear ice, is fast ice near the coast of Alaska. The percentage of new ice or open water and of identified leads tracks that of nonmultiyear ice, but with greater variability. Note that lead area (coverage) generally does not exceed 5% and is a small fraction of the nonmultiyear ice area.

Interestingly, there are more identified leads in January [Fig. 10(a)], as might be expected from the increase in lead coverage shown in Fig. 9(c), but the average area of the leads [Fig. 10(b)] remains fairly constant, as do the width and length

[Fig. 10(c) and (d)]. The box plots reveal the skewed nature of what are exponential-size distributions.

Number of branches, branching index, and linearity show no organization by month [Fig. 11(a)–(c)]. As with size measurements, these measurements have skewed distributions with many more straight, lightly branched leads.

Lead measurements grouped by the latitude of the image show more organization. Nonmultiyear ice [Fig. 12(a) and (b)] decreases with latitude. (Imagery was acquired in the southern Beaufort Sea, where fast first-year ice along the Alaska shore gives way to mixed multiyear and first-year ice to the north before the mostly multiyear ice of the central pack.) The areal coverage of leads, however, remains fairly consistent [Fig. 12(c)].

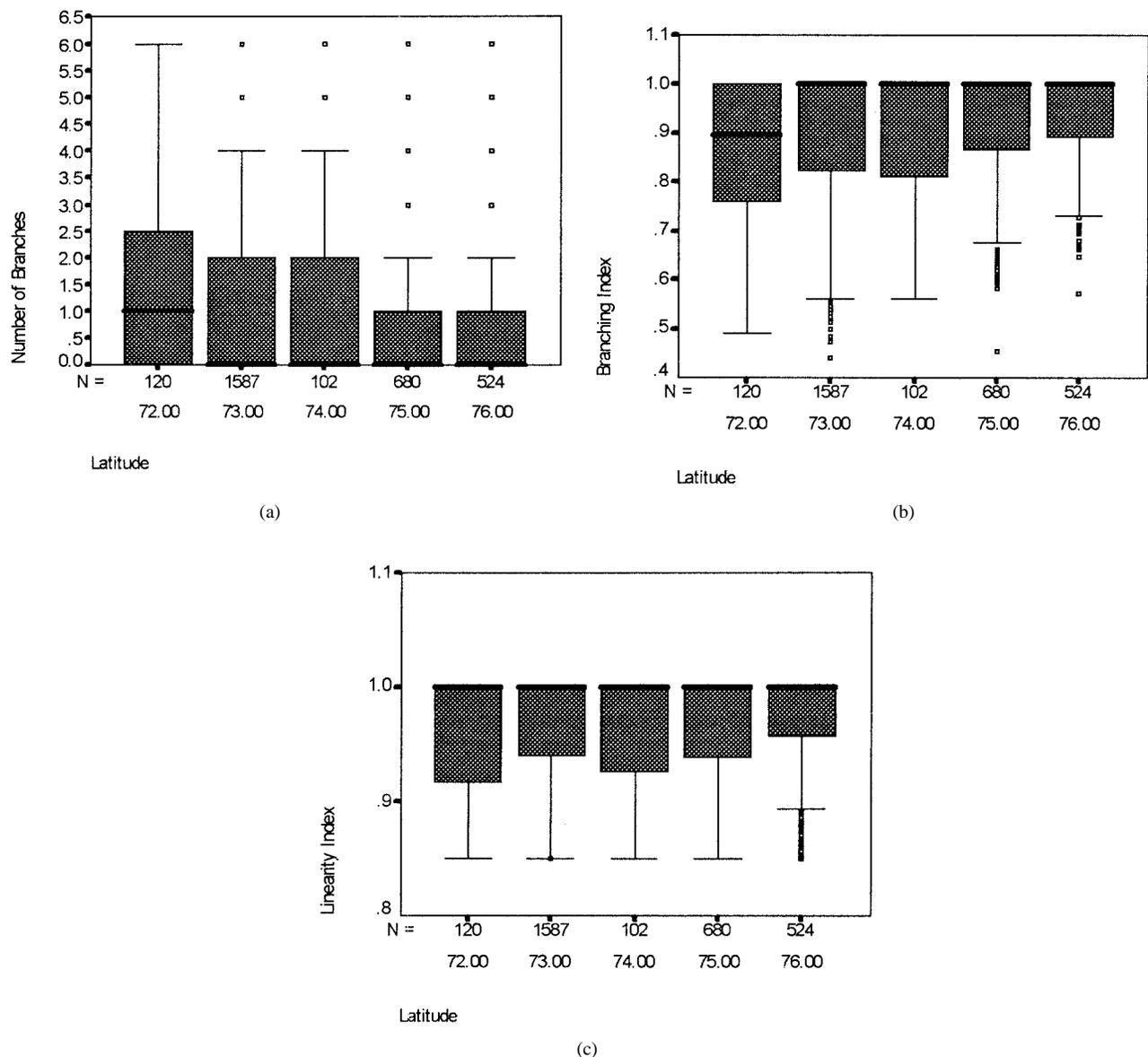


Fig. 14. Branching and linearity by latitude. The number of branches, branching, and linearity indexes all show trends that suggest straighter leads with fewer branches at higher latitudes. (a) Number of branches by latitude, (b) branching index by latitude, and (c) linearity index by latitude.

Fig. 13 shows that while the number of leads/image remained fairly constant, there is a slight downward trend in measurements of size with increasing latitude.

The number of branches, branching, and linearity indexes (Fig. 14) all have trends suggestive of straighter leads with fewer branches at higher latitudes.

#### IV. CONCLUSIONS

Correct classification of sea ice is an underlying assumption of lead identification from classified SAR imagery in this study. Any method that relies on an initial classification of ice types will be bounded in accuracy by the underlying classification system, and as the underlying classification algorithms are improved, the accuracy of the lead identification and analysis output will also improve. Some sources of error are inherent in the image-acquisition and image-processing techniques, however. Both sampling (reducing an area into a single pixel)

and quantization (requiring discrete numeric levels for each pixel) introduce error into the image itself, which affects classification accuracy. When reporting results in terms of kilometers rather than pixels, as was done in this study, it is important to remember that these are approximations of the true distances covered. Additionally, leads extending past the boundary of the image introduce measurement errors.

In the program itself, the skeletonization or thinning process used in automating lead measurement may produce some spurious branching off the main channel because of the shape of the original feature. One future direction may be to add knowledge-based techniques to the program to determine which branches should be included in the measurements and which should be removed, based on the original feature shape and knowledge about lead branching in general.

While the focus of this study was to produce an automated lead-analysis tool, the measurements that it produced point

to areas where further investigation would be interesting. Analyzing the data in conjunction with wind vector data may show some interesting relationships. In fact, retrospective analysis of wind speed and the potential for misclassifying wind-roughened water as multiyear ice might be used to improve the initial classification of the ice itself. Branching and linearity indexes appear to show trends with latitude and further investigation into these relationships may be warranted.

In order to eliminate duplication of area covered each month, more than half of the originally available images were removed from further analysis in this investigation. Another direction of future study would be to look at duplicate areas as a time series and to follow selected leads across time to analyze the ice dynamics.

In general, results indicate that the measurements produced by this methodology are consistent with known lead-measurement distributions. The results of the study suggest that the methodology may be appropriate to study lead characteristics and dynamics at a large scale. The method of identifying potential lead features may be refined over time to reflect growing understanding of the definition of what constitutes a "lead," and additional measurements may be added and verified as additional measurements are identified as useful. Using this automated method may allow us to develop a better understanding of the general dynamics of lead formation, growth, and closing. The program itself may be useful in testing parameter assumptions about leads, and the output may suggest new directions of investigation.

#### ACKNOWLEDGMENT

Special thanks go to I. Rigor at the University of Washington for making buoy temperature data available to the authors and to the Alaska SAR Facility for their assistance in image-orientation conversion.

#### REFERENCES

- [1] J. Banfield, "Skeletal modeling of ice leads," *IEEE Trans. Geosci. Remote Sensing*, vol. 30, pp. 918–923, Mar. 1992.
- [2] G. F. Cunningham, R. Kwok, and J. Banfield, "Ice lead orientation characteristics in the winter Beaufort Sea," in *Int. Geophys. Remote Sensing Symp.*, 1994, pp. 1747–1749.
- [3] G. Cunningham, R. Kwok, and B. Holt, "Preliminary results from the ASF/GPS ice classification algorithm," in *Int. Geophys. Remote Sensing Symp.*, 1992, pp. 573–575.
- [4] B. Erlingsson, "Two-dimensional deformation patterns in sea ice," *J. Glaciology*, vol. 34, no. 118, pp. 301–308, 1988.
- [5] F. M. Fetterer, D. Gineris, and R. Kwok, "Sea ice type maps from Alaska synthetic aperture radar facility imagery: An assessment," *J. Geophys. Res.*, vol. 99, no. C11, pp. 22443–22458, Nov. 1994.
- [6] M. Fily and D. A. Rothrock, "Opening and closing of sea ice leads: Digital measurements from synthetic aperture radar," *J. Geophys. Res.*, vol. 95, no. C1, pp. 789–796, Jan. 1990.
- [7] R. C. Gonzales and R. E. Woods, *Digital Image Processing*. Reading, MA: Addison-Wesley, 1992.
- [8] A. J. Gow and W. B. Tucker, "Sea ice in the polar regions," in *Polar Oceanography, Part A: Physical Science*. New York: Academic, 1990, pp. 47–121.

- [9] R. L. Graham, D. E. Knuth, and O. Patashnik, *Concrete Mathematics*, 2nd ed. Reading, MA: Addison-Wesley, 1994.
- [10] B. Holt, R. Kwok, and E. Rignot, "Ice classification algorithm development and verification for the Alaska SAR facility using aircraft imagery," in *Proc. Int. Geosci. Remote Sensing Symp.*, vol. V2, 1989, pp. 751–754.
- [11] J. Key, A. J. Schweiger, and J. A. Maslanik, "Mapping sea ice leads with a coupled numeric/symbolic system," *ACSM-ASPRS*, vol. 4, pp. 228–237, Mar. 1990.
- [12] R. Kwok, E. Rignot, B. Holt, and R. Onstott, "Identification of sea ice types in spaceborne synthetic aperture radar data," *J. Geophys. Res.*, vol. 97, no. C2, pp. 2391–2402, Feb. 1992.
- [13] R. W. Lindsay and D. A. Rothrock, "Arctic sea ice leads from advanced very high resolution radiometer images," *J. Geophys. Res.*, vol. 100, no. C3, pp. 4533–4544, Mar. 1995.
- [14] G. A. Maykut, "The ice environment," in *In Sea Ice Biota*, R. A. Horner, Ed. Boca Raton, FL: CRC, 1985, pp. 21–82.
- [15] M. J. Norusis, *SPSS for Windows Base System User's Guide*, SPSS, Inc., Chicago, IL, 1992.
- [16] ———, *SPSS for Windows Professional Statistics*, SPSS, Inc., Chicago, IL, 1992.
- [17] K. Steffen, "Fractures in Arctic winter pack ice (North Water, Northern Baffin Bay)," *Ann. Glaciology*, vol. 9, pp. 1–4, 1987.
- [18] R. Onstott, "SAR and scatterometer signatures of sea ice, in microwave remote sensing of sea ice," in *Geophysical Monograph Series*, vol. 68, F. Carsey, Ed. Washington, DC: AGU, 1992, pp. 73–104.
- [19] M. Van Dyne and C. Tsatsoulis, "Extraction and analysis of sea ice leads from SAR images," in *Proc. Int. Geosci. Remote Sensing Symp.*, vol. 2, pp. 629–631 1993.
- [20] P. Wadhams, "Sea-Ice Topography of the Arctic Ocean in the Region 70° W to 25° E," *Philos. Trans. R. Soc.*, vol. 302, pp. 45–85, 1981.
- [21] P. Wadhams, A. A. McLaren, and R. Weintraub, "Ice thickness distribution in Davis Strait in Feb. from submarine sonar profiles," *J. Geophys. Res.*, vol. 90, no. C1, pp. 1069–1077, Jan. 1985.



**Michele M. Van Dyne** received the B.S. degree in psychology in 1981 and the M.S. degree in computer science in 1985 from the University of Montana, Missoula.

She is presently pursuing the Ph.D. degree in electrical engineering at the University of Kansas, Lawrence. Her research interests include machine learning, image processing, and computer vision.



**Costas Tsatsoulis** (M'88) received the Ph.D. degree in electrical engineering in 1987 from Purdue University, West Lafayette, IN.

He is presently an Associate Professor in the Department of Electrical Engineering and Computer Science at the University of Kansas, Lawrence. His research interests include artificial intelligence, case-based reasoning, image processing, computer vision, and multiagent systems.

**Florence Fetterer**, for a photograph and biography, see this issue, p. 492.