

# An analysis of co-occurrence texture statistics as a function of grey level quantization

David A. Clausi

---

**Abstract.** In this paper, the effect of grey level quantization on the ability of co-occurrence probability statistics to classify natural textures is studied. Generally, as a function of increasing grey levels, many of the statistics demonstrate a decrease in classification ability while a few maintain constant classification accuracy. None of the individual statistics show increasing classification accuracy throughout all grey levels. Correlation analysis is used to rationalize a preferred subset of statistics. The preferred statistics set (contrast, correlation, and entropy) is demonstrated to be an improvement over using single statistics or using the entire set of statistics. If the feature space dimension only allows for a single statistic, one of contrast, dissimilarity, inverse difference normalized, or inverse difference moment normalized, is recommended. Testing that compares (using all orientations separately), the average of all orientations and look direction averaging, when determining the co-occurrence features, indicates that the look direction or all orientations is preferred. The Fisher linear discriminant method is used for all classification testing. The Fisher criterion is used as a separability index to provide insight into the classification results. Testing is performed on Brodatz imagery as well as two separate SAR sea-ice data sets.

**Résumé.** Dans cet article, on étudie l'effet de la numérisation des niveaux de gris sur la capacité des statistiques probabilistiques à effectuer une classification des textures naturelles. En général, plusieurs types de statistiques montrent une diminution de la capacité à effectuer une classification lorsque les valeurs de niveaux de gris sont plus élevées et peu de types de statistiques conservent une précision de classification constante. Aucune statistique ne présentent une augmentation de la précision des statistiques de classification parmi tous les niveaux de gris. On utilise l'analyse de la corrélation pour suggérer les meilleurs sous-ensembles de statistiques. Il appert que les meilleurs ensembles de statistiques (contraste, corrélation et entropie) constituent une amélioration par rapport aux statistiques simples ou à l'utilisation de tous les ensembles de statistiques. Si l'espace des éléments ne permet qu'un seul type de statistiques, le contraste, la dissimilarité, la différence inverse normalisée ou la différence de moment inverse normalisé sont recommandés. Lors de la détermination des éléments de cooccurrence, les tests qui comparent les diverses orientations de façon séparée, la moyenne de toutes les orientations et la moyenne des directions de visée indiquent qu'il est préférable d'utiliser la direction de visée ou toutes les orientations. La méthode des discriminants linéaires de Fisher est utilisée pour toutes les vérifications de classification. Le critère de Fisher est utilisé comme index de séparabilité pour aider à comprendre les résultats de classification. Les tests ont été effectués sur des images de Brodatz ainsi que sur deux images RSO de glaces de mer.

[Traduit par la Rédaction]

## Introduction

The interpretation of synthetic aperture radar (SAR) images is an important ongoing research field. Scientists' efforts are directed towards improving calibration of radar platforms, participating in field validation programmes to correlate ground measurements with remotely sensed data, and predictive modelling of SAR backscatter, as well as developing computer-aided algorithms for the identification of pertinent features. This paper focuses on a subset of the last topic by studying particular abilities of co-occurrence probabilities to perform spatial texture analysis using SAR sea-ice imagery.

Sea-ice information, important for assisting ship navigation in ice-infested waters and climate change monitoring in Polar Regions, should be produced in a timely fashion (Barber et al., 1992; Carsey, 1989). Given the abundant SAR digital imagery that must be analyzed, it makes sense to provide computer-assisted interpretation techniques, a requirement that has been recognized by a number of agencies (Agnew et al., 1999; Brown et al., 1999; Partington and Bertoia, 1997). The development of reliable, robust methods for the consistent

classification of SAR sea-ice data has been evasive, even though considerable efforts have been attempted (Barber et al., 1993; Barber and LeDrew, 1991; Holmes et al., 1984; Nystuen and Garcia, 1992; Shanmugan et al., 1981; Shokr, 1991; Soh and Tsatsoulis, 1999). Since SAR sea-ice imagery contains spatially dependent class characteristics, texture extraction methods have been commonly used to generate feature information for sea-ice classes. The most common texture feature extraction method for remotely sensed data is the application of co-occurrence probabilities.

This paper advances the research field by considering the ability of co-occurrence statistics to classify across the full range of available grey level quantizations. This is important, since users usually set the image's grey level quantization arbitrarily without considering that a different quantization might produce improved results. In fact, a popular commercial

---

Received 19 June 2001. Accepted 25 September 2001.

**D.A. Clausi.** Department of Systems Design Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: dclausi@engmail.uwaterloo.ca).

remote sensing image analysis package (PCI's EASI/PACE<sup>®</sup>) determines grey level quantization, preventing the user from providing a more sound choice to potentially improve their results. By investigating the behaviour of the co-occurrence statistics across the full range of grey level quantizations, a choice of grey level quantization and co-occurrence statistics can be made. The author is not aware of any other published research that examines co-occurrence probabilities in this manner.

This paper is organized as follows. In the next section, the application of grey level co-occurrence texture features within the context of SAR sea-ice classification is described. This leads to the "Methods" section that clearly identifies the research questions to be answered, indicates the information content techniques that will be used, and describes the image data sets. Testing results and discussion, with respect to each research question, follows.

## Grey level co-occurrence texture features

### Generation of co-occurrence texture features

The co-occurrence probabilities provide a second-order method for generating texture features (Haralick et al., 1973). These probabilities represent the conditional joint probabilities of all pair wise combinations of grey levels in the spatial window of interest given two parameters: interpixel distance ( $\delta$ ) and orientation ( $\theta$ ). The probability measure can be defined as:

$$\Pr(x) = \{C_{ij} \mid (\delta, \theta)\}$$

where  $C_{ij}$  (the co-occurrence probability between grey levels  $i$  and  $j$ ) is defined as:

$$C_{ij} = \frac{P_{ij}}{\sum_{i,j=1}^G P_{ij}}$$

where  $P_{ij}$  represents the number of occurrences of grey levels  $i$  and  $j$  within the given window, given a certain  $(\delta, \theta)$  pair; and  $G$  is the quantized number of grey levels. The sum in the denominator thus represents the total number of grey level pairs  $(i, j)$  within the window.

Statistics (**Table 1**) were applied to the co-occurrence probabilities to generate the texture features. Although many statistics exist (Haralick et al., 1973), eight grey level shift invariant statistics that are commonly applied are used in this study. Only statistics that are grey level shift invariant should be considered when applied to SAR imagery so that classification is not a function of tone. Descriptions of such statistics with respect to texture characteristics have been the subject of several studies (Barber and LeDrew, 1991; Nystuen and Garcia, 1992; Baraldi and Parmiggianni, 1995).

Historically, the probabilities are stored inefficiently in a sparse matrix referred to as the grey level co-occurrence matrix

**Table 1.** Grey level shift-invariant co-occurrence texture statistics.

Maximum probability (MAX)	$\max \{C_{ij}\}$ for all $(i, j)$
Uniformity (UNI)	$\sum C_{ij}^2$
Entropy (ENT)	$\sum C_{ij} \log C_{ij}$
Dissimilarity (DIS)	$\sum C_{ij}  i - j $
Contrast (CON)	$\sum C_{ij} (i - j)^2$
Inverse difference (INV)	$\sum \frac{C_{ij}}{1 +  i - j }$
Inverse difference moment (IDM)	$\sum \frac{C_{ij}}{1 + (i - j)^2}$
Correlation (COR)	$\sum \frac{(i - \mu_x)(j - \mu_y)C_{ij}}{\sigma_x \sigma_y}$

**Note:** Short forms used in the text indicated in upper case within round brackets. All summations are over all  $(i, j)$  pairs.

or GLCM. Since the matrix is dimensioned to  $G$ , the fewer the number of grey levels the faster the computation when the statistics are applied. More efficient storage of co-occurrence probabilities is implemented by using a grey level co-occurrence linked list (GLCLL) (Clausi and Jernigan, 1998) or grey level co-occurrence hybrid structure (GLCHS) (Clausi and Zhao, 2001). To gain computational speed faster than the GLCM approach these methods avoid storing zero probabilities for grey level pairs.

### Interpretation of SAR sea-ice imagery using co-occurrence features

Scientists have investigated different aspects of co-occurrence texture features with respect to SAR sea-ice image interpretation (**Table 2**). These investigations are performed to produce recommendations on how to set necessary parameters including:  $G$ ,  $\delta$ ,  $\theta$ , statistics (number and type), and window size. In this paper, the role of  $G$  and  $\theta$  are tested, while window size is not discussed since it is accepted that larger window sizes will, in theory, provide more accurate classifications. Window size is a critical parameter for image segmentation, where large windows have a higher potential of overlapping two or more classes, but this aspect is not considered here for the classification problem.

### Grey level quantization

Typically, only one  $G$  is selected for a given study as indicated in **Table 2**, although it is possible to produce features based on multiple grey level quantizations. Smaller values for  $G$  accelerate the calculation of the co-occurrence texture features and reduce noise, however, this is offset by a reduction in information. Soh and Tsatsoulis (1999) comment that, "We assume that the information gain in noise-effect reduction does not compensate the loss of information as a result of quantization". It is expected that coarser quantization would reduce both classification accuracy and feature space

**Table 2.** Summary of selected papers that consider the application of co-occurrence texture features for the interpretation of SAR sea-ice imagery.

Author(s)	$\delta$	$\theta$	G	Co-occurrence statistics used
Barber and LeDrew (1991)	1*, 5, 10	0*, 45, 90	16	CON, COR, DIS, ENT, UNI
Holmes et al. (1984)	2	Average	8	CON, ENT
Nystuen and Garcia (1992)	4*-10	Variety	N.P.	CON, COR, ENT, IDM, UNI, cluster prominence
Shokr (1991)	1, 2*, 3	Average	16, 32	CON, ENT, IDM, UNI, MAX
Baraldi and Parmiggiani (1995)	1	0	N.P.	CON, COR, ENT, IDM, UNI, variance
Soh and Tsatsoulis (1999)	1, 2, ...N/2	Average	64	COR, ENT, IDM, UNI, autocorrelation, contrast

\*Indicates a preferred choice; N.P. indicates that this information was not provided in the paper.

separability of the classes. Conversely, finer quantization is expected to improve both accuracy and separability.

Soh and Tsatsoulis (1999) investigated feature values produced by individual statistics as a function of quantization level. Shokr (1991) also considered the effect of quantization on co-occurrence feature values. In this paper, a different approach is taken. Here, the effects of grey level quantization on the classification ability and feature space separability of the co-occurrence statistics are studied. Uniform quantization is used since it has been demonstrated to produce preferred results (Clausi, 2001).

**Table 2** shows that a significant degree of quantization has been used in other studies. Original images have 256 grey levels and the quantization levels are 8, 16, or 32. Sometimes, the quantization level is not reported. Soh and Tsatsoulis (1999) indicate that 64 grey levels would be a better choice to improve the recognition rate. Note that EASI/PACE<sup>®</sup> automatically quantizes the image data to 16 grey levels prior to capturing co-occurrence texture features.

### Co-occurrence statistics

Many of the statistics produce correlated texture features (Barber and LeDrew, 1991; Shokr, 1991; Baraldi and Parmiggiani, 1995). From a pattern recognition standpoint, it is preferred to minimize redundancy and the feature space dimensionality in a feature set so as to improve classification accuracy (Duda et al., 2001; Hughes, 1968). Barber and LeDrew (1991) recommend the use of only three statistics. Fewer statistics lead to fewer computations. Most of the studies listed in **Table 2** utilize grey level shift invariant texture features with all using entropy (ENT). Note that the statistics listed are the ones that the scientists used for the final studies, although they might have considered many others at the onset of their papers.

### Orientation ( $\theta$ )

Several of the studies listed in **Table 2** average the four inter-pixel orientations since the authors assume that texture measures are insensitive to the direction of the sensor and/or recognize that sea-ice formations rotate and position themselves in all possible orientations. However, Barber and LeDrew (1991) determined that the orientation of the look direction produces results that have greater statistical significance. In this paper, the set of all  $\theta$ 's for nearest

neighbours ( $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ ,  $135^\circ$ ) is used in all cases except in research Question #4 where all orientations, look direction averaging, and overall average are compared.

### Displacement ( $\delta$ )

Barber and LeDrew (1991) statistically demonstrate that  $\delta = 1$  produces a significantly superior classification when compared to  $\delta = 5$  and  $\delta = 9$ . Holmes et al. (1984) indicate that "Experimentation with the distance parameter led to our selection of  $\delta = 2$  as the most appropriate", but do not give details of the experimental methodology or results. Shokr (1991) experimentally compares  $\delta = \{1, 2, 3\}$  and concludes that  $\delta = 2$  is appropriate. Nystuen and Garcia (1992) use a variety of distances from one to ten and conclude that the features are consistent for distances greater than four. Soh and Tsatsoulis (1999) advocate using additional displacement values for classifying SAR sea-ice imagery. In this paper,  $\delta$  will be fixed at one to allow relative comparisons to be made with respect to the parameters under consideration.

## Methods

### Objectives

Based on the background information provided above, the following research questions are posed.

(1) What is the classification performance of each of the individual statistics (**Table 1**) across the full range of quantized grey levels for each data set? Do the statistics produce higher classification accuracies with increasing G?

(2) What is the correlation relationship of the statistics as a function of G?

(3) Given the responses to questions (1) and (2), can a preferred set of statistics be advocated for consistent texture recognition? How does this improved set compare with a brute force approach of simply selecting all statistics?

(4) Is there any improvement to the classification accuracy across the range of grey levels if look angles or the average of the angles are considered?

### Information content

As recommended by Kurvonen and Hallikainen (1999) the information content of the texture measures can be evaluated with test classification and a separability index. Test

classification considers the absolute accuracy of the texture measures with respect to the true classification. A separability index determines a relative estimate for information content of the texture measure under scrutiny, to provide a basis for comparing different methods.

### Classification

A common technique for the classification of feature vectors is the use of the maximum likelihood (ML) classifier. This technique is not used here because the restricted number of samples per class creates a class feature space representation that can be sparse. The Fisher linear discriminant (FLD) (Fisher, 1950; Duda et al., 2001) is less sensitive to the number of features compared to ML because FLD uses a pooled covariance matrix compared to the individual class covariance matrix employed by ML. Also, the FLD is convenient since it is non-parametric. Support for using the linear discriminant over the ML technique is motivated by Tom and Miller (1984). Further details of the FLD implementation used in this paper is found in Clausi (2001).

Discriminants are created using training data. The effectiveness of the discriminants is determined by applying them to separate test data. Classification of the test data produces error matrices. Kappa ( $\kappa$ ) coefficients and associated confidence intervals ( $\sigma$ ) are used to evaluate each error matrix (Bishop et al., 1975). When two error matrices are compared, the following test statistic can be used to determine a significance value (using a significance level of 5%):

$$Z \sim \frac{\kappa_1 - \kappa_2}{\sqrt{\sigma_1^2 + \sigma_2^2}}$$

where the  $\kappa$  coefficient is defined as

$$\kappa = \frac{T \sum_{i=1}^M X_{ij} - \sum_{i=1}^M X_{i+} X_{+i}}{T^2 - \sum_{i=1}^M X_{i+} X_{+i}}$$

$M$  is the number of classes;  $x_{ij}$  is the  $i$ th value on the error matrix diagonal;  $x_{i+}$  and  $x_{+i}$  are the marginal sums of rows and columns, respectively; and  $T$  is the total number of samples.

Classifying test data not used to generate the discriminants is advocated since the true test of a classification system is its ability to identify “unknown” data. Discriminants can be designed to have a high classification rate for training data, yet the same discriminants can have poor classification accuracy on test data. A proper means for checking the reliability of a classification system is to use separate training and test feature vectors (Randen and Husoy, 1999). Even though this approach is recognized to be preferred, recent published sea-ice recognition research mix the training and test sets for some of

their results (Soh and Tsatsoulis, 1999; Nystuen and Garcia, 1992), which can lead to misleading interpretations.

### Separability index

The Fisher criterion (Duda et al., 2001) is used as a measure of the separability of two classes in the feature space. This criterion calculates a ratio of the between class separability and the within class variation. Larger Fisher criterion values demonstrate improved separation of two classes.

### Comparison of lines on plots

For the classification results, plots are created with  $G$  as the independent variable and  $\kappa$  as the dependent variable. Different criteria can be used to determine if a certain line on the plot is an improvement over another line. Ideally, a set of co-occurrence parameters would produce classifications that have a strong consistent measurement over all quantized grey levels. This is preferred so that classifications are not a function of  $G$ , allowing users to select a quantization value that will produce consistent results. In the testing, it was noted that lines sometimes had significant variance, and this variance could vary as a function of the statistics used or as a function of  $G$ . Thus, if there are two lines that are generating consistent results, one may not be a statistically significant improvement over the other. Maximum and minimum  $\kappa$  values could be considered, however, such values would identify specific quantization levels which may not be consistent between data sets, so this consideration is not used in this paper. Slope, variance, and statistical significance are used in this paper to determine if one line is an improvement on another.

If the slope of the line is close to zero, then the choice of parameters is not a function of  $G$ , which is preferable. Thus, even though two lines could have the same mean, the line that is most consistent over the grey level range is a preferable choice. High variance leads to classifications that are suspect and unpredictable. Selections of  $G$  in ranges with relatively higher  $\kappa$  variance are discouraged. The average  $\kappa$  and standard deviation for each quantization level across a range of  $G$  can be determined for a particular line that has a constant mean, i.e., approximately zero slope. These values can be used to test statistical significance between a pair of lines using the  $Z$ -statistic. A significance level of 5% is used in this paper.

### Image data sets

To address the robustness of the parameter selections, three texture image sources are studied: Brodatz images, SAR aerial imagery from the Labrador Ice Margin Experiment (LIMEX), and SAR satellite imagery from the North Water (NOW) Polynya project.

### Brodatz data set

Brodatz imagery (Brodatz, 1966) is the most common test imagery used in the generic texture interpretation research literature. This imagery provides a variety of classes and allows

comparison with other research. Also, the training and test image samples are assured to contain one class only; the same cannot hold absolutely true for the SAR image samples, although efforts are made to accomplish this. Eight different Brodatz textures are used: cloth (D19), cork (D4), cotton (D77), grass (D9), paper (D57), pigskin (D92), stone (D2), and wood (D68). These textures were chosen since they have a noticeable, but not necessarily regular, textural pattern and several of the textures are similar in nature, somewhat mimicking textures found in SAR sea-ice imagery. A key distinction between Brodatz imagery and SAR imagery is the absence of speckle noise in Brodatz imagery. Obviously, speckle-free images (given all other parameters and data equal) would produce superior texture classifications. The cotton texture, used as a control, is the only one that has a well-defined repeating pattern. Training samples are selected by dividing the top left quadrant of a 256×256 image into sixty-four 16×16 sub-images. Sixty-four test samples are represented by the 16×16 sub-images from the bottom right quadrant. Due to space limitations and the availability of this imagery, these images are not presented in this paper.

#### LIMEX data set

The C-band HH polarization image (**Figure 1**) (Hirose et al., 1993) was obtained using the Canada Centre for Remote Sensing (CCRS) SAR carried on a Convair-580 platform for LIMEX during the springtime of 1989 (the air temperature was 8°C) (Tang and Manore, 1992). The image is dimensioned to 2000×2000 pixels with 100 m resolution and was acquired in nadir mode. A sub-scene that covers between 50 and 70 degrees incidence angle was selected. Further details on the processing characteristics are described by Ikeda and Tang (1992). This image contains brash ice, open water (with the wind blowing left-to-right across the image), and first year ice. One hundred 8×8 samples of each class were selected for training and sixty-four 8×8 samples of each class were selected for testing.

#### NOW data set

Detailed information concerning the RADARSAT-1 SAR imagery is contained in Mundy and Barber (2001). In brief, ScanSAR data were acquired from the Canadian Ice Services (CIS) with 150 m nominal resolution and 100 m pixel spacing as part of the International NOW Polynya project. Nine classes are included: nilas/new-ice, grey-ice, grey white-floes, medium first-year ice-floes, rough open-water, calm open-water, smooth first-year ice, rubble landfast-ice, and multiyear ice. These ice types are illustrated in **Figure 2**. SAR sea-ice classification studies usually consider fewer ice classes. Two sets of image samples were created. The “validated” data set contains those regions with field observations (during the NOW programme) that were co-registered in the images. It was these samples that were used for testing in this paper. The “inspected” samples are based on regions that were in the NOW vicinity, but not directly observed during the field programme. These class samples were selected by inspection of the images,

to have similar visual characteristics as the validated classes. It was these samples that were used for training in this paper. Thirty-two non-overlapping samples of each class within each set were selected. The window size of each sample was 16×16.

## Results and discussion

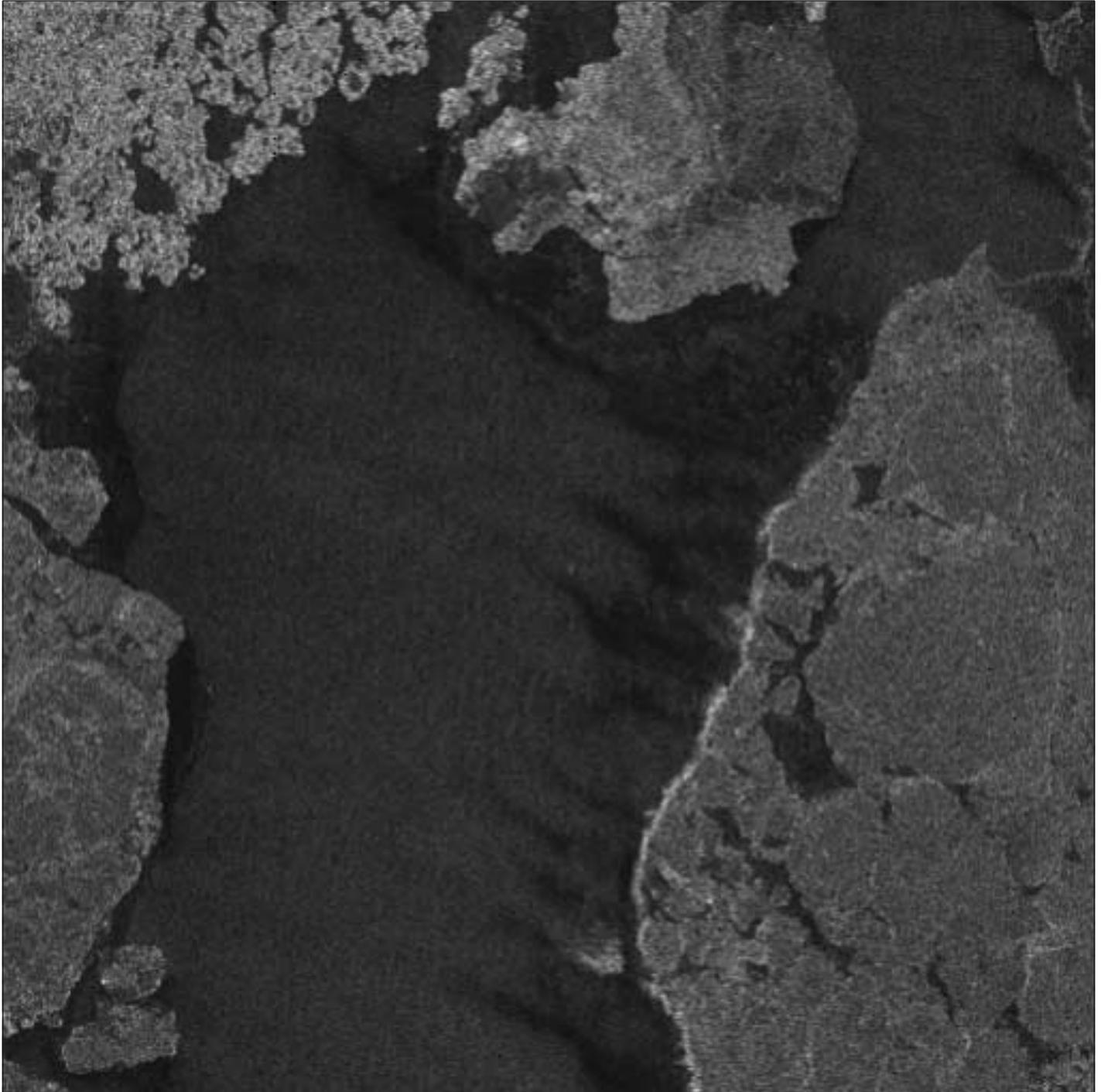
Each of the research questions will now be addressed in sequence.

- (1) What is the classification performance of each of the individual statistics (**Table 1**) across the full range of quantized grey levels? Do the statistics produce higher classification accuracies with increasing  $G$ ?

**Figure 3** contains three plots, one for each data set. Each plot shows the  $\kappa$  test classification results for each statistic alone (using  $\theta = 0, 45, 90, 135$  degrees and  $\delta = 1$  pixel, for a total of four features). For comparison purposes, the y-axis is always scaled between 0 and 1. A total of sixty-three quantization values are used (8, 12, 16, ..., 256). The classification accuracies for the NOW data are much lower than the classification accuracies of the LIMEX data. This difference is probably due to LIMEX data having fewer classes (three versus nine) and a finer resolution (100 m versus 150 m).

With higher  $G$ , it would be expected that, more information is provided and improved classification would be achieved for each statistic. Similarly, it is expected that a lower  $G$  would have less information and poorer classification. However, this is not always the case and for many statistics, the opposite occurs, that is, poorer classification with higher  $G$ . A number of these data trends are explained.

- The dissimilarity (DIS) and contrast (CON) statistics (**Table 1**) produce consistently strong classifications across all data sets. For the LIMEX and NOW data sets, coarse quantization ( $G$  roughly less than 24 grey levels) results in lower classification accuracy. For  $G$  approximately greater than 24 grey levels, results are quite consistent with minimal variability. Thus, above a certain threshold, classification using DIS and CON is independent of the value of  $G$ . DIS produces slightly higher results for the Brodatz data set and CON produces slightly higher results for the two SAR data sets. However, for each data set, there is no statistical difference between the DIS and CON results.
- The statistics entropy (ENT), uniformity (UNI), and maximum probability (MAX) (**Table 1**) all have a strong decrease in classification accuracy with increasing  $G$ . A decrease in classification accuracy with increasing  $G$  was not expected. However, these three statistics are based on homogeneity, that is, quality of disbursement of the co-occurring probabilities throughout the GLCM, in contrast to the smoothness statistics (CON, DIS, inverse difference (INV), inverse difference moment (IDM), which use a weighted distance from the main diagonal of the GLCM, (i.e., location). At full dynamic range, few (if any) grey

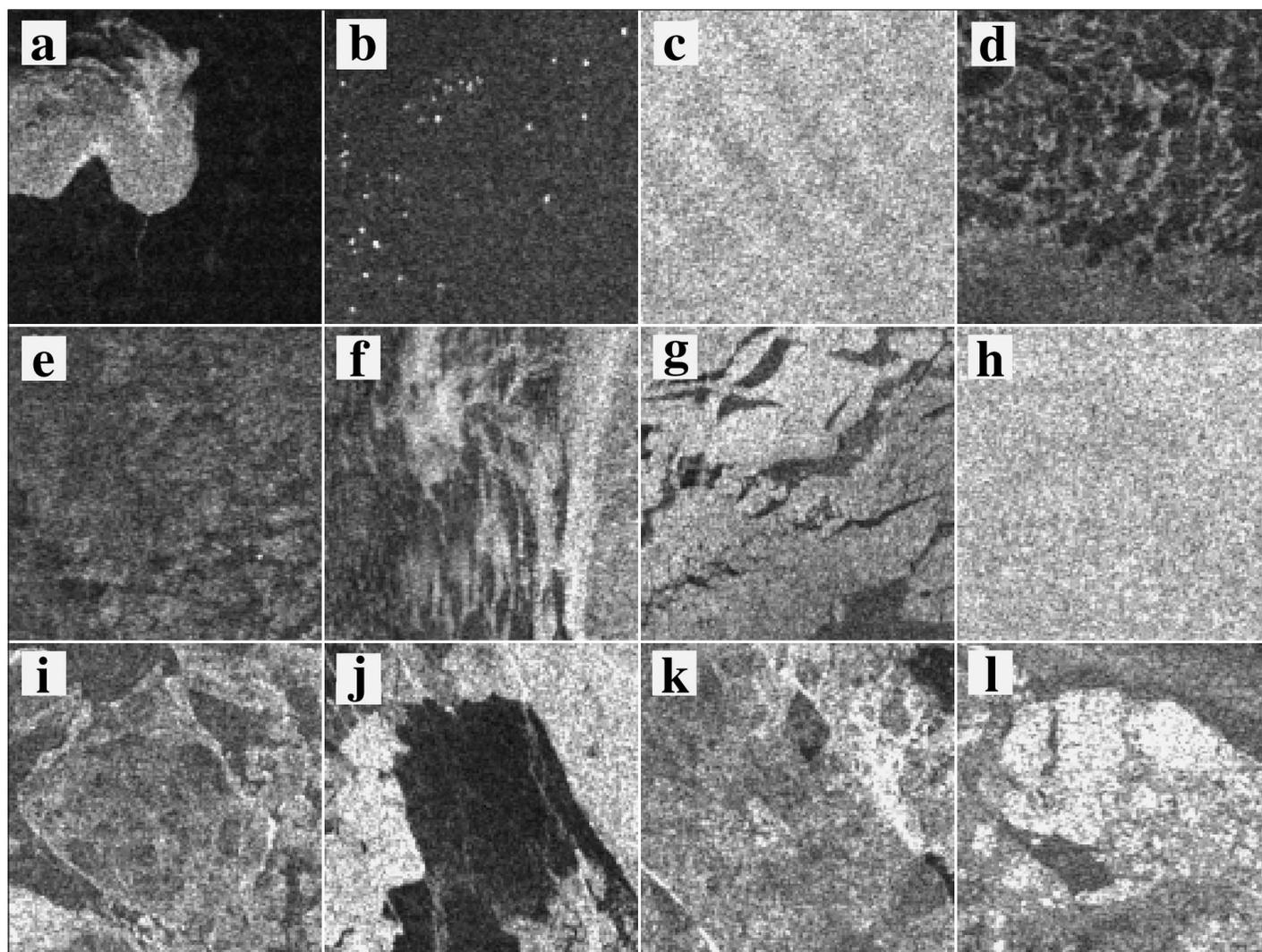


**Figure 1.** Aerial SAR image (Hirose et al., 1993) containing brash ice, open water, and first year ice.

level pairs are repeated. Classification using ENT, UNI, or MAX at full dynamic range is poor, since the probabilities tend to a near maximum state of entropy causing similar texture feature values to be generated for each class. Thus, quantization is a necessity to enhance the ability of the statistics ENT, UNI, and MAX, and to move the system away from the full entropy state. Historically, quantization has been used to reduce the computational load; however,

it has not been recognized that the classification ability of these statistics were simultaneously enhanced.

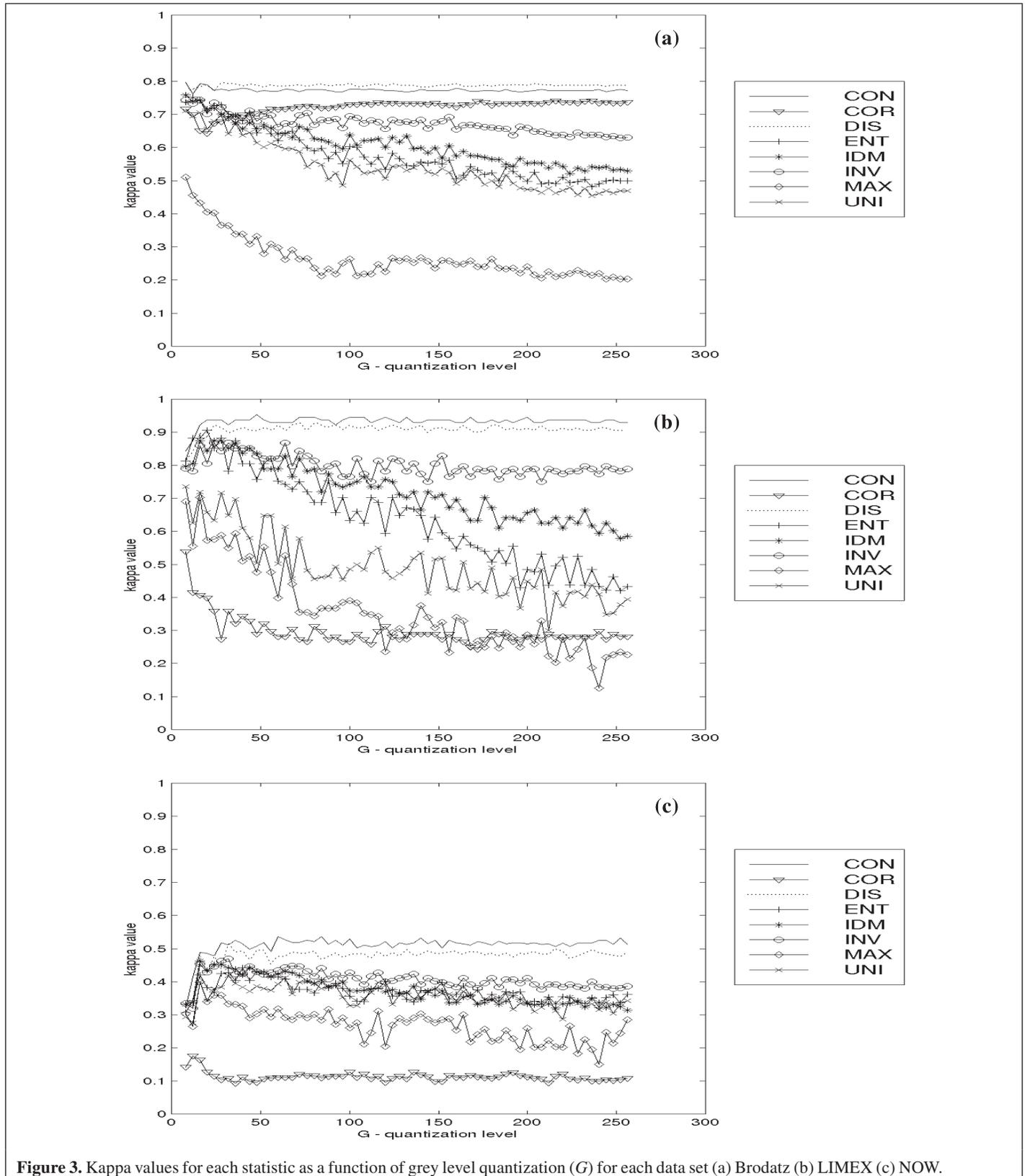
- The statistics INV and IDM have decreasing classification accuracy with increasing  $G$ . This improvement in the classification rate with increasing  $G$  was expected. These statistics are based on the sum of  $C_{ij}$  weighted by a numerical series ( $\{1, 1/2, 1/3, 1/4, \dots\}$  for INV and  $\{1, 1/2, 1/5, 1/10, \dots\}$  for IDM). For a smooth texture ( $i \approx j$ ) these



**Figure 2.** RADARSAT subimages of selected ice types identified during NOW program (Mundy and Barber, to appear) (with permission). Bold indicates those ice classes included in this study. Omitted ice types from this study had insufficient spatial coverage to provide samples for classification testing. (a) Bright signature represents brash ice, which consists of very small ice floes broken up through wave and wind disturbances. Dark homogeneous signature of **calm open water** surrounds brash ice. (b) Icebergs are apparent as distinct white dots surrounded by slightly roughened open water. (c) Rough open water showing a very low frequency wave pattern (approximately 2 km wavelength) across diagonal. (d) Wavy texture signature covering majority of window represents **nilas/new ice** with tear fractures and extensive finger rafting. Toward lower edge of window, signature becomes a more constant grey tone, which represents transition from nilas to **grey ice**. (e) Grey ice with minimal rafting (brighter signature) present throughout. (f) Bright linear signatures are strips of small pancake ice, less than 0.5 m in diameter, created during wind event. Darker signature represents open water as well as frazil and grease ice formation which smooths surface. (g) The bright signature represents newly formed frost flowers on top of nilas/grey ice cover. Darker signatures are small leads with a new ice cover. (h) Broken up through wind events, this window is covered with small (<25 m in length) angular **grey-white floes** with deformed edges where floes converge with each other. (i) Large **medium thick first-year ice** floe nearly covers whole window. Well-defined ridge circles floe where it has converged with other floes. Floe is conglomerate of smaller angular floes such as those in window H. Consequently, ridging within large floe was also present. (j) Middle of window is covered with **smooth thick first year ice** represented by a very dark homogeneous signature. Surrounding brighter signature represents **rubbled first year ice**. (k) Two parallel rubble ice shear zones amongst landfast first year ice are present in this window, represented by very bright linear features oriented at diagonals across the window. Very thick ridging had occurred along these zones. (l) The brighter signature represents a **multiyear ice** floe. Dark breaks can be observed within and around bright signature. These breaks represent leads in and around multiyear ice floe where new ice types were being formed.

statistics will likely sum  $C_{ij}$  values to approximately one, while coarse textures will likely sum to approximately zero. The outcome is a sparser cluster for a smooth (predictable) texture relative to a coarse (noisier) texture. This outcome is opposite to the expected effect of these statistics. Namely, texture features that use grey level

differences should generate relatively tighter clusters for smooth textures and relatively sparser clusters for coarse textures, leading to textures that are obviously smooth being confused with other classes. That the IDM statistic has a quadratic term accentuates this confusion and, as a



**Figure 3.** Kappa values for each statistic as a function of grey level quantization ( $G$ ) for each data set (a) Brodatz (b) LIMEX (c) NOW.

result, its classification accuracies are poorer relative to INV, especially with increasing  $G$ .

To improve the classification ability of these two statistics, the difference between  $i$  and  $j$  can be normalized by the

number of grey levels  $G$ . The modified statistics are indicated by:

Inverse difference normalized (INN)

$$\sum_{i,j=1}^G \frac{C_{ij}}{1 + |i - j|^2 / G^2}$$

Inverse difference moment normalized (IDN)

$$\sum_{i,j=1}^G \frac{C_{ij}}{1 + (i - j)^2 / G^2}$$

These normalized statistics were used to classify the NOW data set as a function of  $G$  (Figure 4). Clearly, the normalized statistics results are a close match to the CON results. Since the same type of result occurred for the Brodatz and LIMEX data sets, only the NOW data set is displayed. It will be shown in the results for Question #2 that the features produced by these normalized statistics (INN and IDN) are highly correlated with the features produced by CON and DIS.

- The statistic COR has different classification ability for Brodatz versus the SAR imagery. COR is the third best statistic after CON and DIS for classifying the Brodatz data set. However, COR does not classify the SAR data as well as the Brodatz data. Brodatz texture contains more structure than the SAR textures, which is an appealing characteristic for the COR statistic.

Figure 5 illustrates the use of the Fisher criterion to compare classification abilities of the different statistics. The value on the y-axis represents the average Fisher criterion over all pair wise classes generated when the discriminant rules were created using the given statistic. These results support those obtained in Figure 2 for the  $\kappa$  values. Once again, coarse quantization decreases the separability of the classes in the

feature space. This decrease is especially noted for IDM, INV, and ENT in the SAR data sets. As in the classification case CON, DIS, and COR produce consistent results for quantization levels greater than about  $G = 24$ . The statistics INV, IDM, ENT, MAX, and UNI all have decreasing average Fisher criteria with greater  $G$ . With increasing information, the quality decreases for many of the statistics' features.

The threshold value of  $G = 24$  is smaller than the  $G = 64$  advocated by Soh and Tsatsoulis (1999); however, they agree that large  $G$  values ( $G = 128$  or  $256$ ) are unnecessary. Both thresholds are demonstrated to be acceptable and are larger than the values typically used in other co-occurrence texture studies. Setting  $G = 16$ , as in EASI/PACE®, can produce unreliable classification results. The reason for EASI/PACE's low quantization level is probably based on maintaining minimal computational demands. Current efforts to upgrade this code with a faster algorithm (Clausi and Zhao, 2001) are underway.

It is possible that with higher  $G$ , discriminants are better tuned to the training data and are less able to classify the testing data. This is another potential explanation for the decrease in classification accuracy with increasing  $G$  observed in Figure 2. However, Figure 5 illustrates that this is not the case. If the discriminants overfitted the training data (Duda et al., 2001), then there should be an increase in the average FLD with increasing  $G$ , which is not seen.

- (2) What is the correlation relationship of the statistics as a function of  $G$ ?

Figure 6 illustrates how the average correlation of all features, averaged over all orientations, statistics, and classes, varies as a function of  $G$  for each data set. The NOW data set shows the highest correlations, followed by LIMEX, and then Brodatz. Lower average correlations of all features have the potential of greater feature diversity and were hoped to lead to improved feature sets. However, according to Figures 3 and 5, none of the statistics show classification improvement for

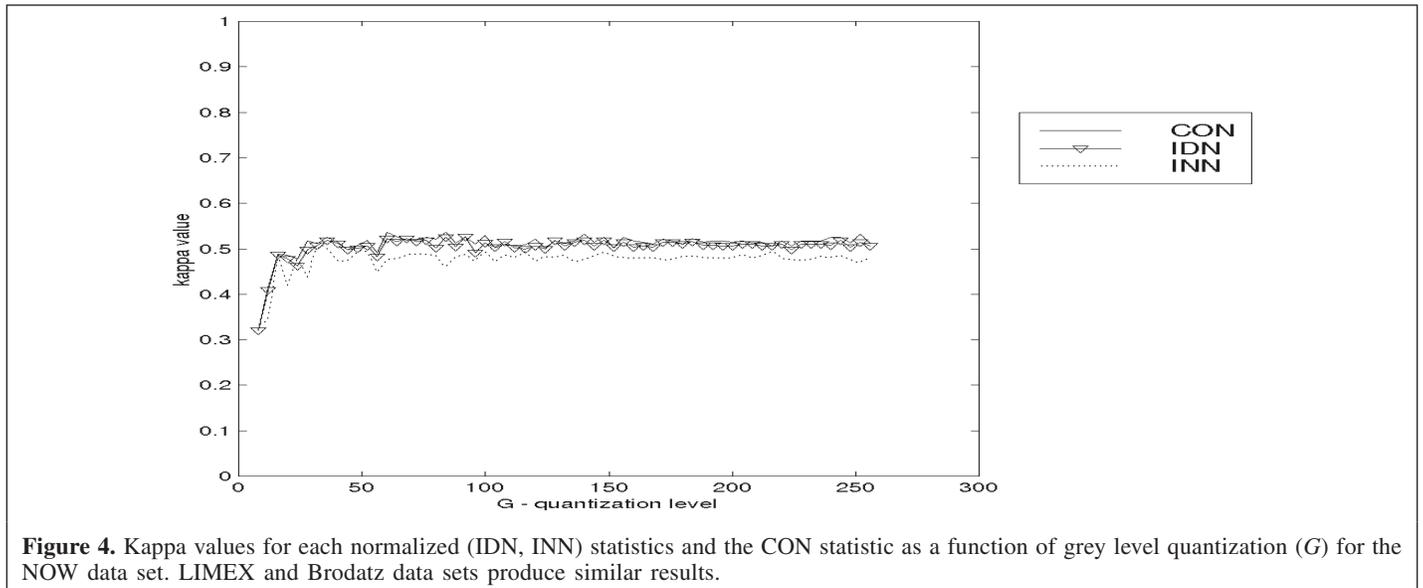
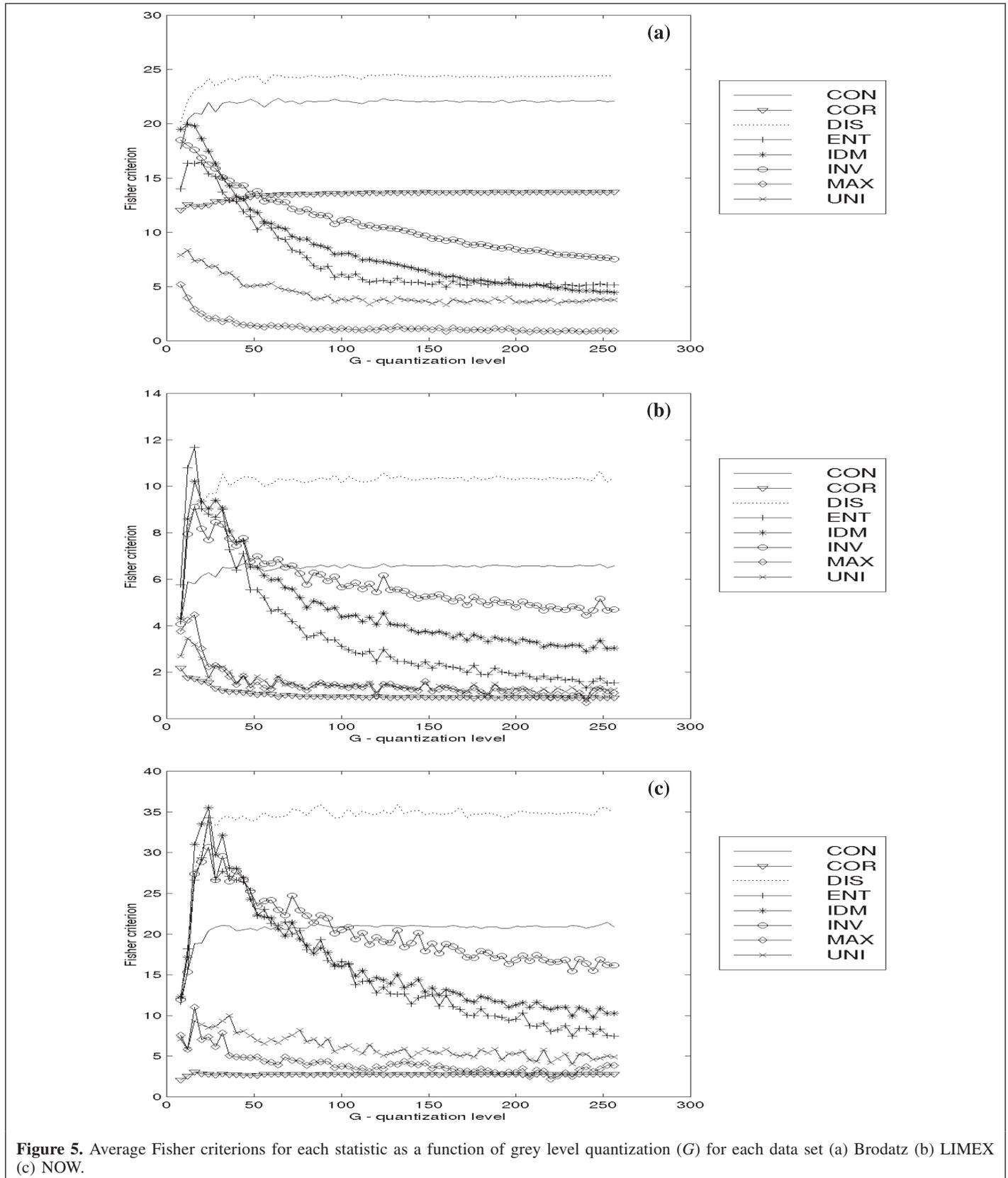


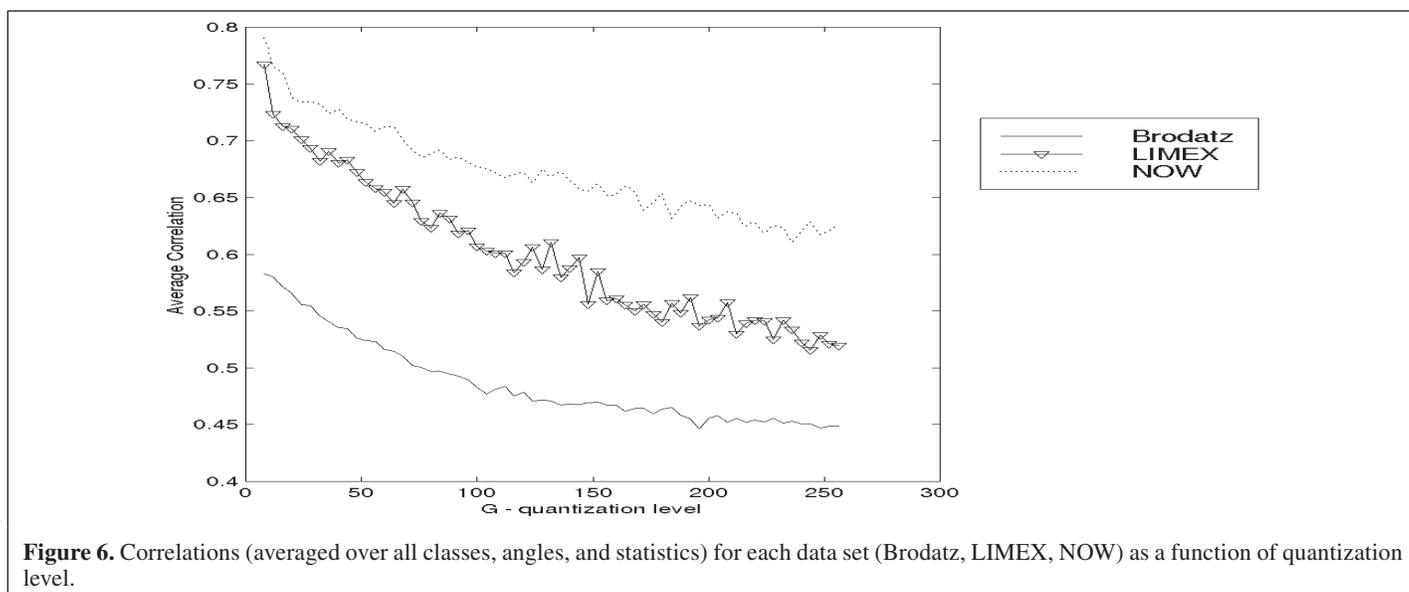
Figure 4. Kappa values for each normalized (IDN, INN) statistics and the CON statistic as a function of grey level quantization ( $G$ ) for the NOW data set. LIMEX and Brodatz data sets produce similar results.



**Figure 5.** Average Fisher criteria for each statistic as a function of grey level quantization ( $G$ ) for each data set (a) Brodatz (b) LIMEX (c) NOW.

increasing  $G$ . That the Brodatz data set would have the lowest correlations is not unexpected given the relative distinctiveness of the imagery. That this reduced correlation does not generate

higher classification rates compared to the LIMEX data set is probably due to the higher number of classes in the Brodatz imagery (eight) compared to LIMEX (three).



**Figure 6.** Correlations (averaged over all classes, angles, and statistics) for each data set (Brodatz, LIMEX, NOW) as a function of quantization level.

**Table 3** represents pair wise correlations for a fixed value of  $G = 64$  for each data set. Since orientational sensitivity is definitely important for the Brodatz data set, its correlations (Table 3a) for each statistic are only for  $\theta = 0$  degrees. For the SAR data sets (Table 3b and 3c), the average correlations over all values of  $\theta$  is reported. These tables show a number of important trends. The statistics CON, DIS, IDN, and INN are all highly correlated; absolute range of 0.930 to 1.000 across all data sets. The statistics IDM and INV are also well correlated with this group; absolute range of 0.723 to 0.998 across all data sets. This suggests that only one of these six statistics is required to be calculated for any given data set. The statistics ENT, MAX, and UNI also have strong correlations; absolute range of 0.619 to 0.966 across all data sets. The statistic COR is surprisingly uncorrelated with the rest of the statistics. Although results are only presented for  $G = 64$ , comparable results were produced for each value of  $G$  considered.

(3) Given the responses to questions (1) and (2), can a preferred set of statistics be advocated for consistent texture recognition? How does this improved set compare with a brute force approach of simply selecting all statistics?

The responses to Questions #1 and #2 leads to a natural grouping of the statistics: smoothness statistics, homogeneity statistics, and COR. Each of these will now be discussed.

Moments of the non-zero probability entries about the GLCM diagonal measure the degree of smoothness of the texture. The statistics CON, DIS, IDN, INN, IDM, and INV measure textural smoothness. Since all of these statistics produce correlated features, only one smoothness statistic needed to be selected. Since IDM and INV produce poorer results with finer quantization, they were removed from consideration, which parallels the recommendation of Baraldi and Parmiggiani (1995) who dismissed the use of IDM since it did not provide any additional information beyond what other favoured statistics provided. Due to strong correlation between

IDM and INV, INV must be removed from consideration. Thus, only one of CON, DIS, IDN, and INN needed to be selected. Since CON produced the highest results for the SAR data sets, it was selected to be used in this paper.

Homogeneity statistics (ENT, MAX, and UNI) measure the uniformity of the non-zero entries in the GLCM. If the grey levels in a particular window tend to be homogeneous, then only a few grey level pairs represent the texture. Non-homogeneity generates many different pairs of grey levels. Since these homogeneity statistics tend to be well correlated, only one needs to be selected. Due to the poor classification ability of the MAX statistic, plus the fact that it does not add any new information to the classification process, it was not considered an important statistic. Soh and Tsatsoulis (1999) excluded MAX from consideration since they felt it was volatile. Since ENT consistently produces the best results of any of the homogeneity statistics, this statistic was used as part of the preferred statistic set.

The COR statistic is an independent measure, that is, not correlated with any of the other statistics, that provides appropriate features for the classification process. It was included as one of the selected statistics in this paper. Thus, to generate a complete feature set, one homogeneity statistic (ENT), one smoothness statistic (CON), and COR need to be selected. Choosing only three statistics agrees with recommendations by Barber and LeDrew (1991) and parallels those of Baraldi and Parmiggiani (1995) who felt that CON and UNI were the two most important statistics (they also felt that UNI and ENT were quite similar in behaviour). If computational time restrictions and limited feature space dimensionality only permit the use of two statistics, then CON and ENT are recommended by the author.

Does the application of CON, ENT, and COR improve the classification accuracy over using individual statistics? **Figure 7** displays classification results for each data set given five different cases. CON, ENT, and COR are considered

**Table 3.** Representative example (G = 64 grey levels) of inter-statistic correlations for each data set.

(a)	CON	DIS	IDN	INN	IDM	INV	ENT	MAX	UNI	COR
CON	1	0.973	-1.000	-0.964	-0.727	-0.777	0.689	-0.367	-0.591	-0.295
DIS	—	1	-0.976	-0.999	-0.846	-0.885	0.755	-0.436	-0.674	-0.298
IDN	—	—	1	0.967	0.732	0.782	-0.692	0.37	0.595	0.296
INN	—	—	—	1	0.863	0.900	-0.763	0.446	0.684	0.297
IDM	—	—	—	—	1	0.995	-0.749	0.568	0.734	0.247
INV	—	—	—	—	—	1	-0.766	0.568	0.743	0.259
ENT	—	—	—	—	—	—	1	-0.619	-0.956	0.090
MAX	—	—	—	—	—	—	—	1	0.747	-0.010
UNI	—	—	—	—	—	—	—	—	1	-0.077
COR	—	—	—	—	—	—	—	—	—	1

(b)	CON	DIS	IDN	INN	IDM	INV	ENT	MAX	UNI	COR
CON	1	0.945	-1.000	-0.930	-0.723	-0.765	0.746	-0.432	-0.612	0.100
DIS	—	1	-0.951	-0.999	-0.881	-0.913	0.870	-0.558	-0.759	0.072
IDN	—	—	1	0.937	0.734	0.776	-0.755	0.438	0.622	-0.097
INN	—	—	—	1	0.897	0.927	-0.879	0.570	0.773	-0.067
IDM	—	—	—	—	1	0.995	-0.896	0.689	0.858	-0.023
INV	—	—	—	—	—	1	-0.908	0.691	0.862	-0.033
ENT	—	—	—	—	—	—	1	-0.701	-0.941	0.267
MAX	—	—	—	—	—	—	—	1	0.813	-0.097
UNI	—	—	—	—	—	—	—	—	1	-0.229
COR	—	—	—	—	—	—	—	—	—	1

(c)	CON	DIS	IDN	INN	IDM	INV	ENT	MAX	UNI	COR
CON	1	0.976	-1.000	-0.965	-0.815	-0.831	0.870	-0.529	-0.784	0.035
DIS	—	1	-0.979	-0.999	-0.909	-0.920	0.930	-0.611	-0.868	0.011
IDN	—	—	1	0.968	0.821	0.836	-0.874	0.534	0.789	-0.033
INN	—	—	—	1	0.926	0.935	-0.939	0.626	0.883	-0.007
IDM	—	—	—	—	1	0.998	-0.931	0.732	0.934	0.034
INV	—	—	—	—	—	1	-0.936	0.728	0.934	0.032
ENT	—	—	—	—	—	—	1	-0.705	-0.966	0.189
MAX	—	—	—	—	—	—	—	1	0.795	-0.037
UNI	—	—	—	—	—	—	—	—	1	-0.141
COR	—	—	—	—	—	—	—	—	—	1

**Note:** For (a) Brodatz, correlations are represented by  $\theta = 0$  degrees. For (b) LIMEX and (c) NOW, correlations are averages over all four angles. For each data set, correlations are also averaged over all classes.

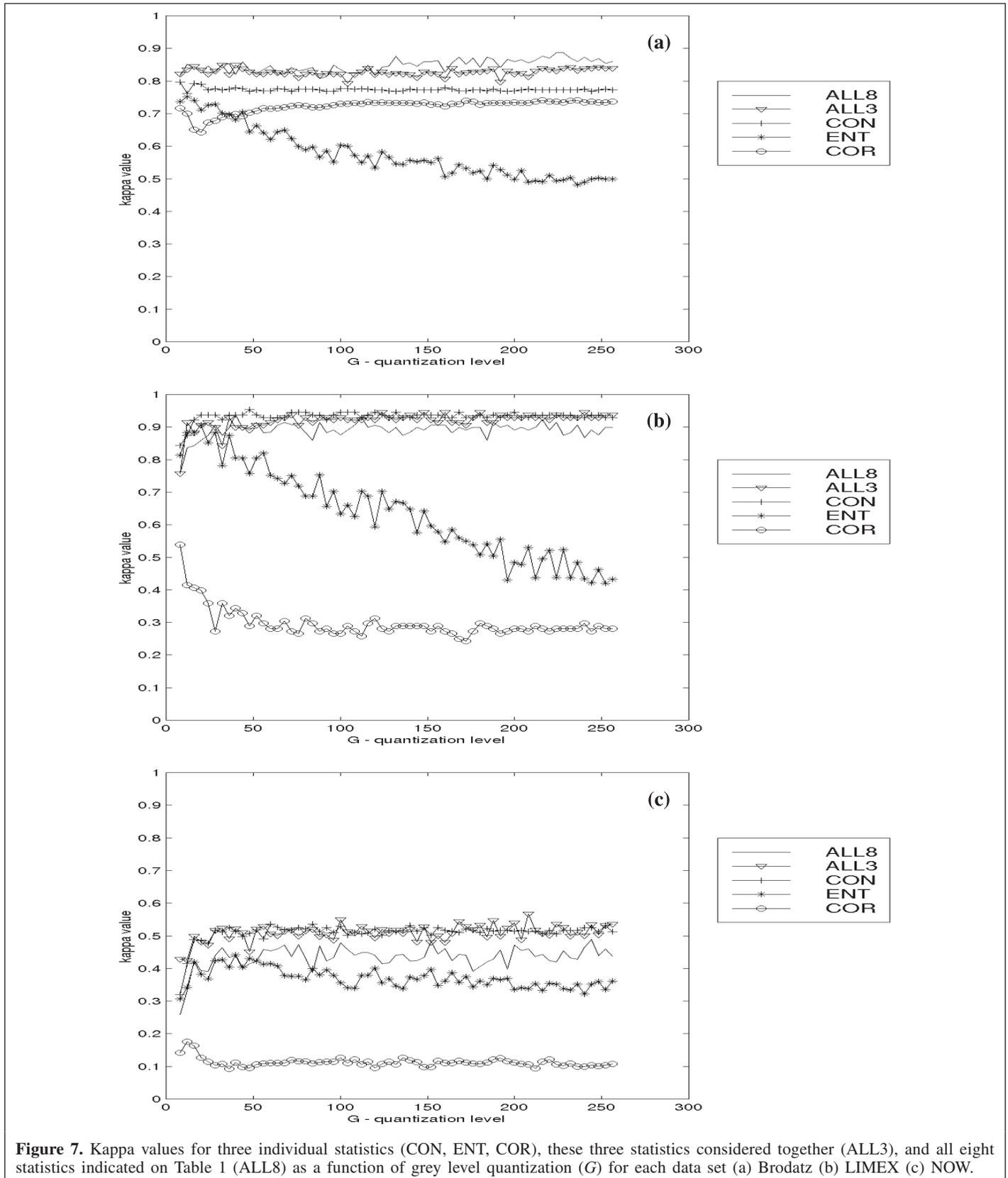
independently (same results as **Figure 2**) and as a group of three (ALL3). In addition, the brute force case of using all eight statistics (**Table 1**) at once is also provided (ALL8).

ALL8 in the Brodatz set (**Figure 7a**) produces results comparable to ALL3 (no statistical significance). Each of ALL8 and ALL3 are statistically more significant than CON alone. So, ALL8 and ALL3 are improvements compared to using any single statistic. However, ALL3 is the preferred choice due to computational reasons and because it has relatively lower variance (0.011 compared to 0.023 of ALL8). In the LIMEX results (**Figure 7b**), there is no statistical significance between ALL8, ALL3, and CON. In this case, CON would be the preferred choice because its variance (0.006) is much lower than that of ALL8 (0.015) or ALL3 (0.010) and its computational requirements are lower. For the NOW data set (**Figure 7c**), ALL3 and CON are a statistical significant improvement over ALL8. CON is the preferred

measure since it has lower variance compared to ALL3 (0.008 compared to 0.019, respectively).

That ALL8 would lead to a decrease in effectiveness for the NOW data set was not unexpected. Using eight statistics with four orientations creates a 32-dimensional feature space. Each class in the NOW data set has thirty-two samples. Sixty-four samples (thirty-two samples from each class) are used to create the pooled covariance matrix required for the FLD, so two samples per feature space dimension (spd) are provided by the NOW data set. These two samples are probably insufficient information to model covariance behaviour. Both LIMEX and Brodatz have better representation (6.2 and 4 spd respectively). Thus, the decrease in effectiveness of the ALL8 data set is probably due to ineffective covariance representation.

Other papers may be suffering from the same drawback. For example, Soh and Tsatsoulis (1999) use six samples to represent a covariance matrix in a 10-dimensional feature space (0.6 spd) to perform Bayes classification, for three of their



**Figure 7.** Kappa values for three individual statistics (CON, ENT, COR), these three statistics considered together (ALL3), and all eight statistics indicated on Table 1 (ALL8) as a function of grey level quantization ( $G$ ) for each data set (a) Brodatz (b) LIMEX (c) NOW.

seven classes. Nystuen and Garcia (1992) use values as low as one spd using a Mahalanobis classifier. Barber and LeDrew (1991) recognize this potential shortcoming in their testing and

have some tests where class covariance matrices are estimated with spd values below one when using a Mahalanobis classifier.

Overall, this discussion demonstrates that a brute force approach of selecting all statistics (ALL8) does not improve the classification ability relative to choosing only CON, ENT, and COR. For the SAR data alone, CON seems to be the best choice. However, with structured textures (as in the Brodatz case) ALL3 is the preferred choice. Since it may not be known *a priori* that the SAR sea-ice textures are highly structured (especially in an unsupervised application), it is safer to create a classification system using ALL3 statistics, given that a sufficient number of samples per class are available to properly estimate class covariance. Using this approach, a robust texture feature set will be created.

To support the classification results, plots of the Fisher criterion measured using the training data are provided in **Figure 8**. Comparatively, the individual Fisher criterion lines and the classification lines follow similar patterns as a function of  $G$  for each texture feature set. Surprisingly, ALL8 generates dramatically higher Fisher criterion values compared to ALL3, ENT, COR, and CON even though ALL8 does not demonstrate superior classification ability (**Figure 7**). ALL8 is demonstrating that the decision rules for classification are overfitted (Duda et al., 2001), that is, ALL8 generates features that are fine-tuned for the training data set but cannot classify the test set properly. If the classification of training data is only considered, results can be biased. In this paper, if the training results alone were considered, ALL8 would be the overwhelming choice for statistic selection for each data set. However, it is quite clear from the classification of the test data that this is not the preferred statistic set. Similarly, looking only at the training data would indicate that ALL3 would be a preferred choice compared to the CON statistic, however, this trend is not followed in the classification of the test data.

- (4) Is there any improvement to the classification accuracy across the range of grey levels if look angles or the average of the angles is considered?

Three test scenarios are considered in this paper. Earlier testing considered feature extraction using separate orientations (SEP), for example 0, 45, 90, 135 degrees. Since Barber and LeDrew (1991) concluded that the look direction produces statistically significant results, a feature set that accounts for the look direction (LOOK) can be created. In this paper, the following are considered independent: the look direction ( $\theta = 0$  degrees), normal to the look direction ( $\theta = 90$  degrees), and 45 degrees relative to the look direction ( $\theta =$  average of 45 and 135 degrees). The third scenario is averaging over all four orientations (AVG).

Since both CON and ALL3 produced preferred results in the above-mentioned testing of Question 3, they will both be considered. **Figure 9** depicts the results for LIMEX and NOW using the ALL3 set of statistics and the three orientation considerations while **Figure 9** illustrates LIMEX and NOW results using the CON statistic. The Brodatz data set is not considered since its textures are noticeably directionally sensitive.

When ALL3 statistics are used with both LIMEX and NOW data (**Figures 9a** and **9b**) there is no statistically significant difference in any of the results. Using CON alone with the LIMEX data, there is no statistical significance between the three plots (**Figure 10**). However, when CON is applied to the NOW data, the AVG features generate classification accuracies that are statistically lower compared to LOOK and SEP. These results tend to support the use of ALL3 statistics and either LOOK or SEP orientation selections.

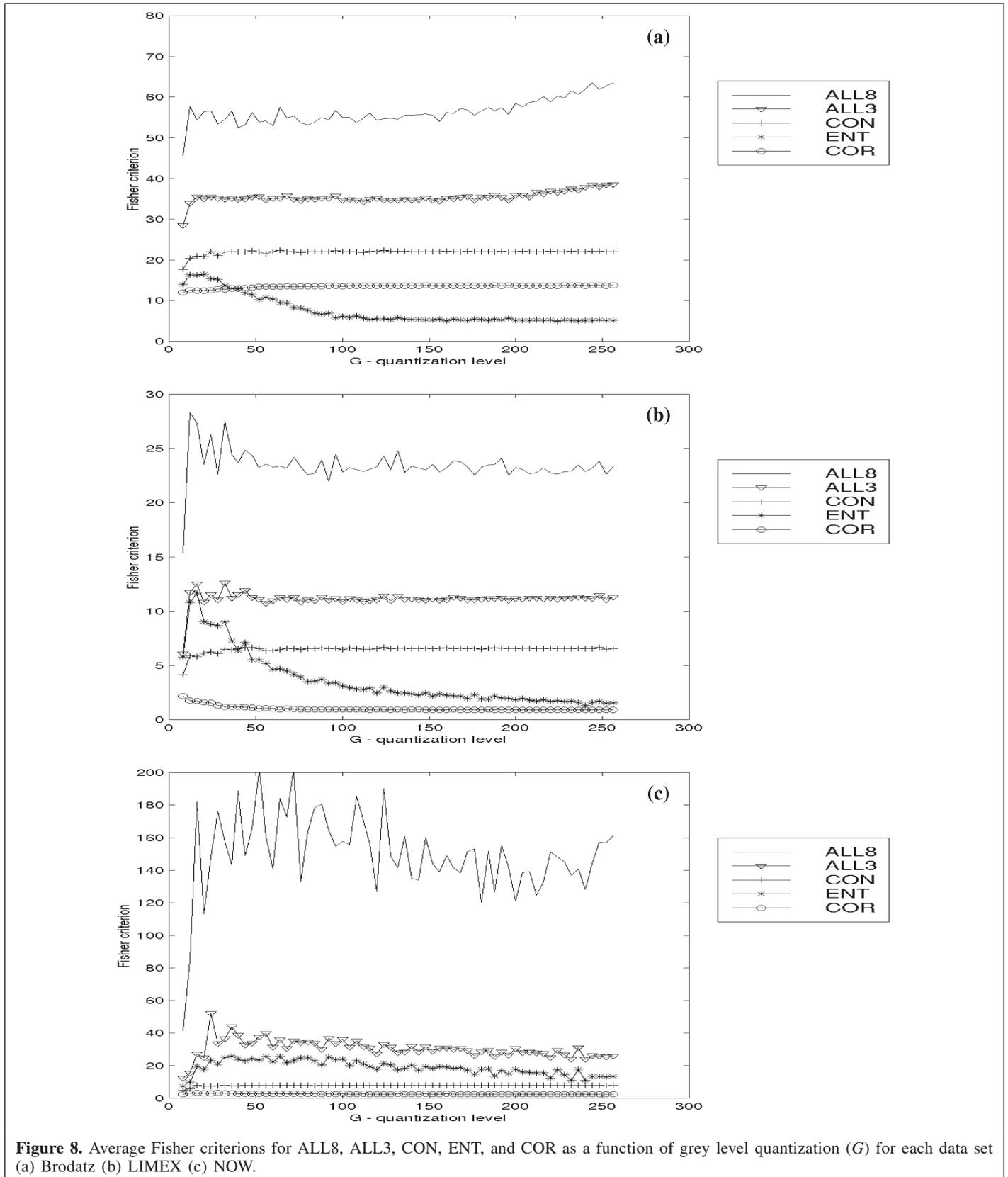
## Conclusions

This paper makes a number of important observations with respect to the classification of SAR sea-ice imagery using co-occurrence texture features.

Many of the individual statistics have peak classification accuracy at a relatively coarse quantization level, as well as a decrease in classification accuracy with increasing  $G$ . Arbitrary selections of  $G$  for these statistics can produce misleading results for the classification of textured imagery. It is recommended that the COR, ENT, IDM, INV, MAX, or UNI statistics not be used as a sole statistic for any classification study. One of CON, DIS, IDN, or INN is recommended in situations where the number of samples per class are insufficient to support a high dimensional feature space, and only one statistic can be selected.

A preferable choice of statistics is the combined use of three fairly independent statistics: CON, ENT, and COR. This statistic set demonstrates consistent classification accuracy over the range of grey level quantizations and consistently produces preferred classification results compared to individual statistics and all eight statistics. The use of all eight statistics is perilous, since, given selected orientations and displacements, a high dimensional feature space may be produced and there may be insufficient class samples to accurately represent the class covariance matrices. Also, generating texture features given eight statistics requires additional computations.

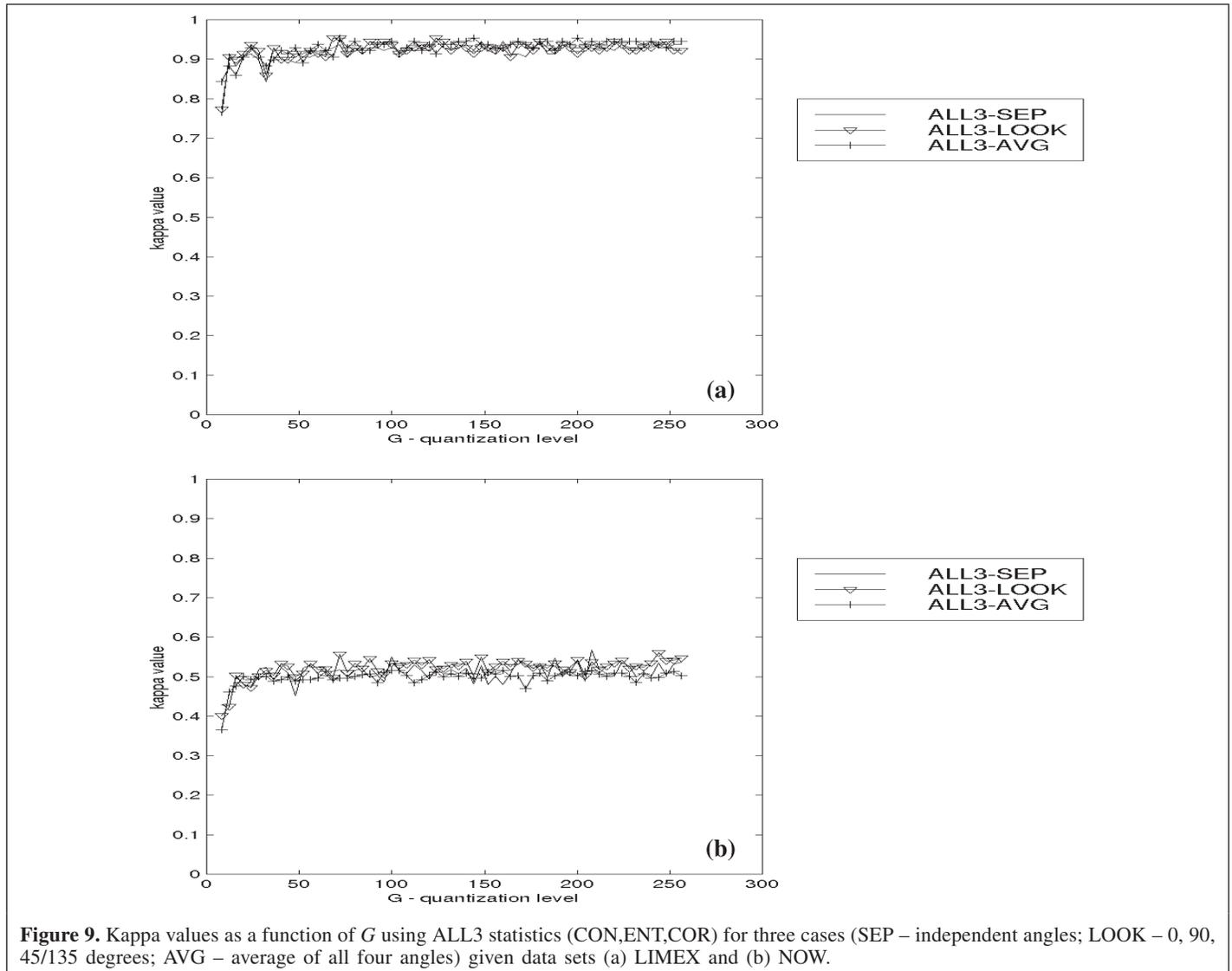
As indicated in **Table 2**, various authors consider statistics that measure the same characteristic based on a GLCM – a weighted value of  $|i - j|$  (CON, DIS, IDM, and INV as well as INN and IDN introduced in this paper). IDM and INV are excluded from consideration since they do not contributing any additional information to the classification process (Baraldi and Parmiggiani, 1995) and because their classification rates are more sensitive to  $G$  than the other four statistics. CON, DIS, INN, and IDN all generate similar classification patterns as a function of  $G$ , probably because they produce features that are extremely well correlated. These statistics are subtly different, so the calculated feature values are different. As a result, each statistic has different absolute classification success with respect to a certain data set. For example, CON may generate higher classification rates than DIS for one data set while the opposite might occur for another data set. Regardless, the features produced by CON and DIS will produce results that are expected to be statistically the same. As a result, this author highly recommends that only one of these four statistics ever be



**Figure 8.** Average Fisher criteria for ALL8, ALL3, CON, ENT, and COR as a function of grey level quantization ( $G$ ) for each data set (a) Brodatz (b) LIMEX (c) NOW.

considered in any given study. There really is no need to determine which statistic is producing the best absolute

classification since each is expected to perform statistically at about the same level.



**Figure 9.** Kappa values as a function of  $G$  using ALL3 statistics (CON,ENT,COR) for three cases (SEP – independent angles; LOOK – 0, 90, 45/135 degrees; AVG – average of all four angles) given data sets (a) LIMEX and (b) NOW.

The same type of argument holds for ENT and UNI. These statistics measure the same type of information found in a GLCM – the homogeneity of the probability values – and they produce features that are highly correlated. There is no need to include both of these statistics in a given study. In fact, inclusion of redundant features unnecessarily increases the dimensionality of the feature space and are not expected to improve the classification quality.

In terms of selecting appropriate orientations, the results in this paper seem to support Barber and LeDrew’s (1991) observation that independent orientations can produce significantly improved results. While the motivation presented by others for using the average of the orientations is sound, the evidence presented in this paper seems to encourage the use of independent orientations.

In the analysis of the three data sets, using the preferred statistic set {CON, ENT, COR}, any value of  $G$  greater than twenty-four is advocated. However, large values of  $G$  (greater than 64) are deemed unnecessary since they do not improve the classification accuracy and are computationally costly. Setting

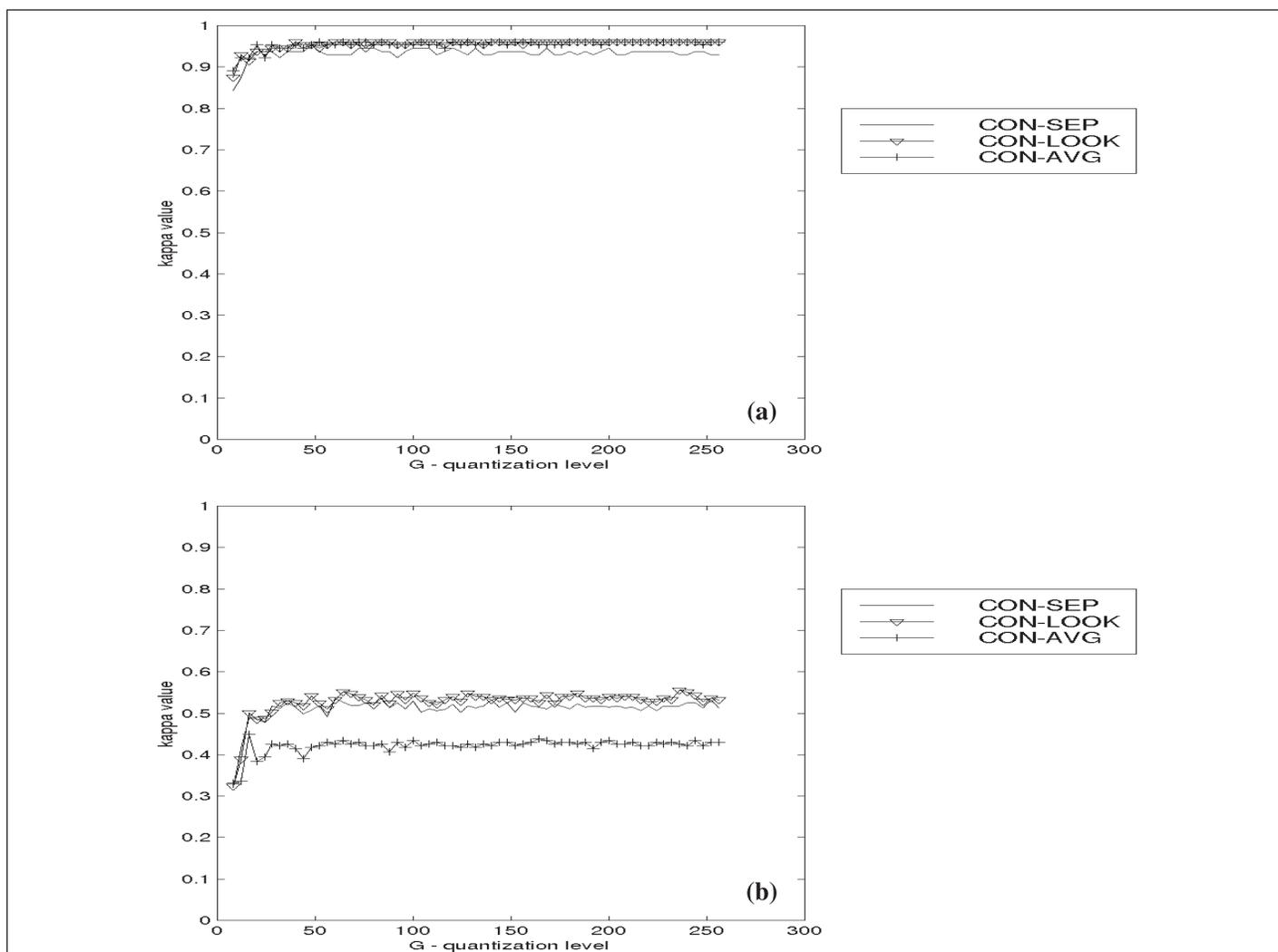
$G$  to a value under twenty-four can produce unreliable classification results.

### Acknowledgements

Thanks are extended to C.J. Mundy and D.A. Barber (University of Manitoba, Manitoba, Canada) who generously collected and supplied the necessary data sets for this study. Referees who reviewed this paper are thanked for their thoughtful comments that improved the quality of the final product. CRYSYS (Cryosphere System in Canada, <<http://www.crysys.uwaterloo.ca/>>) and GEOIDE (Geomatics for Informed Decisions, <<http://www.geoide.ulaval.ca/>>) are thanked for their support.

### References

Agnew, T., Brown, R., Flato, G., Melling, H., and Ramsay, B. 1999. Canadian Contributions to GCOS – Sea Ice. Atmospheric Environment Service (AES) – Environment Canada, Toronto, Ont.



**Figure 10.** Kappa values as a function of  $G$  using the CON statistic for three cases (SEP – independent angles; LOOK – 0, 90, 45/135 degrees; AVG – average of all four angles) given data sets (a) LIMEX and (b) NOW.

- Baraldi, A., and Parmiggiani, F. 1995. An investigation of the textural characteristic associated with gray level cooccurrence matrix statistical parameters. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 33, No. 2, pp. 293–304.
- Barber, D.G., and LeDrew, E.F. 1991. SAR sea ice discrimination using texture statistics: a multivariate approach. *Photogrammetric Engineering and Remote Sensing*, Vol. 57, No. 4, pp. 385–395.
- Barber, D.G., Manore, M.J., Agnew, T.A., Welch, H., Soulis, E.D., and LeDrew, E.F. 1992. Science issues relating to marine aspects of the cryosphere: implications for remote sensing. *Canadian Journal of Remote Sensing*, Vol. 18, No. 1, pp. 46–55.
- Barber, D.G., Shokr, M.E., Fernandes, R.A., Soulis, E.D., and Flett, D.G. 1993. A comparison of second-order classifiers for SAR sea ice discrimination. *Photogrammetric Engineering and Remote Sensing*, Vol. 59, No. 9, pp. 1397–1408.
- Bishop, T., Fienberg, S., and Holland, P. 1975. *Discrete Multivariate Analysis – Theory and Practice*. Cambridge, Mass.: MIT Press.
- Brodatz, P. 1966. *Textures: A Photographic Album for Artists and Designers*. New York: Dover.
- Brown, R., Agnew, T., Burgess, M., Cogley, G., Demuth, M., Duguay, C., Flato, G., Goodison, B., Koerner, R., Melling, H., Prowse, T., Ramsay, B., Sharp, M., Smith, S., and Walker, A. 1999. Development of a Canadian GCOS workplan for the cryosphere. Atmospheric Environment Service (AES), Environment Canada, Toronto, Ont.
- Carsey, F. 1989. Review and status of remote sensing of sea ice. *IEEE Journal of Oceanic Engineering*, Vol. 14, No. 2, pp. 127–138.
- Clausi, D.A. 2001. Comparison and fusion of co-occurrence, Gabor and MRF texture features for classification of SAR sea ice imagery. *Atmosphere-Ocean*, Vol. 39, No. 4, pp. 183–194.
- Clausi, D.A., and Jernigan, M.E. 1998. A fast method to determine co-occurrence texture features. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 36, No. 1, pp. 298–300.
- Clausi, D.A., and Zhao, Y. 2001. Rapid co-occurrence texture feature extraction using hybrid data structure. *Computers and Geosciences*. In press.
- Duda, R.O., Hart, P.E., and Stork, D.G. 2001. *Pattern Classification*, 2nd ed. New York: John Wiley & Sons Inc., 664p.

- Fisher, R.A. 1950. The use of multiple measurements in taxonomic problems. In *Contributions to Mathematical Sciences*. New York: John Wiley & Sons Inc.
- Haralick, R.M., Shanmugan, K., and Dinstein, I. 1973. Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 3, No. 6, pp. 610–621.
- Hirose, T., Heacock, T., and Duncan, R. 1993. A report on the evaluation of ice classification algorithms. Prepared by Noetix Research Inc. for Environment Canada Ice Branch and Ice Applications Group, Canada Centre for Remote Sensing.
- Holmes, Q.A., Nuesch, D.R., and Shuchman, R.A. 1984. Textural analysis and real-time classification of sea-ice types using digital SAR data. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 22, No. 2, pp. 113–120.
- Hughes, G.F. 1968. On the mean accuracy of statistical pattern recognition recognizers. *IEEE Transactions on Information Theory*, Vol. 14, pp. 55–63.
- Ikeda, M., and Tang, C.L. 1992. Detection of the Labrador current using ice-floe movement in synthetic aperture radar imagery and ice beacon trajectories. *Atmosphere-Ocean*, Vol. 30, No. 2, pp. 223–245.
- Kurvonen, L., and Hallikainen, M.T. 1999. Textural information of multitemporal ERS-1 and JERS-1 SAR images with applications to land and forest type classification in Boreal zone. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 37, No. 2, pp. 680–689.
- Mundy, C.J., and Barber, D.A. 2001. On the relationships between spatial patterns of sea ice type and the mechanisms which create and maintain the north water (NOW) polynia. *Atmosphere-Ocean*, Vol. 39, No. 4, pp. 327–341.
- Nystuen, J.A., and Garcia Jr., F.W. 1992. Sea ice classification using SAR backscatter statistics. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 30, No. 3, pp. 502–509.
- Partington, K.C., and Bertoia, C. 1997. Science Plan — Version 6.0. Washington, D.C.: National Ice Center.
- Randen, T., and Husoy, J.H. 1999. Filtering for texture classification: A comparative study. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 21, No. 4, pp. 291–310.
- Shanmugan, K.S., Narayanan, V., Frost, V.S., Stiles, J.A., and Holtzman, J.C. 1981. Textural features for radar image analysis. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 19, No. 3, pp. 153–156.
- Shokr, M.E. 1991. Evaluation of second-order texture parameters for sea ice classification from radar images. *Journal of Geophysical Research*, Vol. 96, No. C6, pp. 10 625 – 10 640.
- Soh, L.-K., and Tsatsoulis, C. 1999. Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 37, No. 2, pp. 780–795.
- Tang, C.L., and Manore, M. (Editors). 1992. LIMEX Special Issue. *Atmosphere-Ocean*, Vol. 30, No. 2.
- Tom, C.H., and Miller, L.D. 1984. An automated land-use mapping comparison of the Bayesian maximum likelihood and linear discriminant analysis algorithms. *Photogrammetric Engineering and Remote Sensing*, Vol. 50, No. 2, pp. 193–207.