Chapter 5

Paper 3:
Automated Non-Gaussian Clustering of Polarimetric Synthetic Aperture Radar Images

Automated Non-Gaussian Clustering of Polarimetric Synthetic Aperture Radar Images

Anthony P. Doulgeris, *Student Member, IEEE,* Stian N. Anfinsen, *Member, IEEE,* Torbjørn Eltoft, *Member, IEEE*

***Abstract*—This paper presents an automatic image segmentation method for Polarimetric SAR data. It utilises the full polarimetric information and incorporates texture by modelling with a non-Gaussian distribution for the complex scattering coefficients. The modelling is based upon the well known product model, with a Gamma distributed texture parameter, leading to the K-Wishart model for the covariance matrix. The automatic clustering is achieved through a finite mixture model estimated with a modified Expectation Maximisation algorithm. We include an additional goodness-of-fit test stage that allows for splitting and merging of clusters. This not only improves the model fit of the clusters, but also dynamically selects the appropriate number of clusters. The resulting image segmentation depicts the statistically significant clusters within the image. A key feature is that the degree of sub-sampling of the input image will affect the detail level of the clustering, revealing only the major classes, or a variable level of detail. Real world examples are shown to demonstrate the technique.**

***Index Terms*—Polarimetric Synthetic Aperture Radar, Non-Gaussian, Statistical Modelling, Clustering.**

I. Introduction

SATELLITE-BORNE Synthetic Aperture Radar (SAR) systems are well suited to environmental monitoring because of their cloud penetrating ability, day/night operation and broad coverage. Reliable automated image classification schemes are required because of the huge scale of earth monitoring and the wealth of new data scenes available from several new SAR satellite systems.

Pixel-wise analysis of SAR imagery is generally complicated due to the presence of speckle and requires that statistical modelling methods are employed. It is well known that in some circumstances radar complex scattering coefficients are non-Gaussian in distribution. For this reason, various non-Gaussian models have been proposed to represent SAR data and many of these have been extended to the polarimetric SAR case. The multivariate K-distributions [1], [2] and G-distributions [3] have been successful for modelling PolSAR data and highlight the importance of non-Gaussianity. Both these distributions are members of the so-called product model [4], which states that, under certain conditions, the backscattered signal results from the product between a Gaussian speckle noise component and the textured terrain mean reflectivity. Associated with these models are one or more so-called non-Gaussianity parameters, which account for deviation from Gaussian statistics.

The Department of Physics and Technology, University of Tromsø, Norway
e-mail: anthony.p.doulgeris@uit.no, stian.normann.anfinsen@uit.no, torbjorn.eltoft@uit.no

Manuscript submitted September 30, 2010.

Fully polarimetric SAR (PolSAR) or quadruple polarisation (quad-pol) imagery contains the full scattering matrix information which leads to the best classification results [5]. In addition, full polarimetric SAR images also allow for classification of pixels by dominant scattering mechanisms. Fully polarimetric data is not always available because the wider coverage of dual (dual-pol) or single polarisation modes are often preferred for monitoring purposes. In this paper, we show how non-Gaussianity and polarimetry can be incorporated into practical, automatic and general methods to analyse SAR image data. The methods are generic in the sense that they can be applied to single, dual or quad-pol data, of any frequency, in a common way.

Polarimetric classifications are commonly performed with the Gaussian based Wishart clustering algorithm. In this study, we use the more advanced non-Gaussian K-Wishart clustering algorithm [6], that additionally accounts for potential textural differences in the classes. The unsupervised clustering algorithm is based upon the expectation maximisation (EM) algorithm [7] for finite mixture modelling, using the K-Wishart probability density function, Eqn. (4), for the probabilities and method of moments for class parameter updates.

In addition to the estimated class parameters, we have automated otherwise *a priori* parameters for these statistical models. The *effective number of looks* (ENL) is automatically determined within the algorithm and includes a correction for texture, by using matrix log-cumulants [8]. The *appropriate number of classes* is determined by a goodness-of-fit test stage within the EM algorithm that allows for splitting or merging of classes. The result is an automatic image segmentation depicting the statistically distinct classes within the image.

A further benefit is that the algorithm is not influenced by initial conditions, as is commonly the case for clustering algorithms, and the main tuning parameter is the confidence level for the goodness-of-fit tests. The clustering simply starts out with the entire image as one class and will split and merge its way to a stable result with the goodness-of-fit tests. The fact that the goodness-of-fit tests are dependent on the number of samples, leads to a useful feature of performing a faster, less detailed analysis by sub-sampling the image, that is also less sensitive to the precision of the model fitting the data.

The automatic approach with goodness-of-fit testing was presented at EUSAR [9], but we have implemented more sophisticated parameter estimation with matrix log-cumulants [8], use matrix log-cumulant goodness-of-fit tests [10] directly on the samples, and the design has been further refined.

The resulting image segmentation, or clustering, is not a true classification, because the clusters have not been identified,

only separated. However, polarimetric decompositions or any, even limited, ground truth data may be used for subsequent automatic or assisted labelling of the found classes.

The automatic technique to find the optimum number of classes is easily adaptable to other statistical models and all that is required is that the probability density function and the first few matrix log-cumulants are known. The automatic clustering technique will be demonstrated for both simulated data and real world examples, and for the K-Wishart, standard Wishart and Relaxed-Wishart [11] models.

This paper proceeds to describes the product model and texture in section II, details the K-Wishart, Wishart and Relaxed-Wishart models and the required matrix log-cumulant theory in section III, and explains the various stages of the automatic algorithm in section IV. Results are shown in section V, followed by the discussion in VI and conclusions in VII.

II. Texture and The Product Model

The full complex scattering matrix of a polarimetric SAR using linear polarisations includes the four combinations of horizontal and vertical polarisation for both the transmitted and received signals. These four channels, HH, HV, VH and VV, are together referred to as quad-pol data, and contain the complete vector information of the scattered signal with respect to the satellite antenna system. These four complex scattering coefficients can be represented as a 4-dimensional complex vector, $\mathbf{s}$, and correspond to the single-look complex (SLC) data format.

Non-Gaussian distributed scattering coefficients have been observed for highly textured target regions imaged with high resolution. Many authors have noted better model fits to class histograms with non-Gaussian models [1], [3], [12], and although there are several specific functional descriptions, they are all based upon the well known *product model*. The product model (explained in [4]) presumes that the observed variation in SAR backscatter is composed of two independent components, one describing natural variation in the mean reflectivity, called *texture*, and another for Gaussian *speckle*. For the semi-symmetric, zero-mean case used to describe radar backscatter from natural terrain, the product model may be written as

$$\mathbf{s} = \sqrt{t}\ \mathbf{\Sigma}^{\frac{1}{2}}\mathbf{n}, \tag{1}$$

where the scale parameter t is a strictly positive scalar random variable, normalised such that its mean $\bar{t} = 1$, $\mathbf{\Sigma}$ is the polarimetric covariance matrix, and $\mathbf{n}$ is a standardised, complex multivariate Gaussian variable with zero mean and identity covariance matrix, i.e. $\mathbf{n} \sim \mathcal{N}^{\mathcal{C}}(\mathbf{0}, \mathbf{I})$. The statistical distribution of t governs the resulting non-Gaussianity of the product model. The general interpretation is that the product model describes texture modulated Gaussian speckle. The texture term is introduced here with a square root because the variable t models the variance of the signal rather than its amplitude.

Since the measured scattering coefficients have a mean value of zero, we have to evaluate the second order statistics in the covariance matrix to estimate scattering properties. The statistical modelling is therefore carried through to multi-look complex (MLC) covariance matrix data. Taking the mathematics through the sample covariance matrix expression is, in general, intractable, but suitable non-Gaussian based models were found by applying the texture product model directly to the covariance matrix data:

$$\mathbf{C} = t\ \mathbf{W}/L, \tag{2}$$

where $\mathbf{W}$ is complex Wishart distributed [13] with L looks and $\mathbf{\Sigma}$ mean covariance matrix.

A constant texture term would describe purely Gaussian vector data and Wishart covariance data and is observed only for very homogeneous and smooth areas in real SAR images. The presence of texture increases the variation beyond that of homogenous areas and produces non-Gaussian distributions for the scattering coefficients and non-Gaussian based, or generalised Wishart models for the covariance matrices [6]. Among the alternative non-Gaussian models, we choose the K-Distribution for the complex scattering coefficients, and the K-Wishart for the covariance matrix, because they appear to be very good models in many situations and they are quite well known. We shall only investigate multi-looked, MLC, data analysis here, since even SLC vector estimation is essentially performed through the covariance matrix of the vectors.

III. The model distributions

A. Finite Mixture Modelling

Our main objective is to segment the image pixels into separate clusters based upon the K-Wishart model. The statistical model for clustering the images assumes that the probability density function (PDF) for the whole data-set can be described by a finite mixture model,

$$f_{\mathbf{C}}(\mathbf{C}) = \sum_{j=1}^{m} \pi_j f_j(\mathbf{C}) \tag{3}$$

where π_j are the class *priors* such that $\sum_{j=1}^{m} \pi_j = 1$ and the $f_j(\mathbf{C})$ are the individual class model PDFs. All priors and model parameters are estimated with the iterative EM algorithm [7]. Our modifications to the EM algorithm will be explained in Sec. IV, but the basic model expressions and log-cumulant theory will be stated here.

B. Model PDFs

The K-Wishart probability density function, with L degrees of freedom, shape (non-Gaussianity) parameter α, and mean covariance matrix $\mathbf{\Sigma}$, is given by

$$\begin{aligned} f_{\mathbf{C}}(\mathbf{C}; L, \alpha, \mathbf{\Sigma}) = & \\ & \frac{2|\mathbf{C}|^{L-d}}{I(L,d)\Gamma(\alpha)|\mathbf{\Sigma}|^{L}} (L\alpha)^{\frac{\alpha+Ld}{2}} \left(\mathrm{tr}(\mathbf{\Sigma}^{-1}\mathbf{C})\right)^{\frac{\alpha-Ld}{2}} \\ & \times K_{\alpha-Ld}\left(2\sqrt{L\alpha \mathrm{tr}(\mathbf{\Sigma}^{-1}\mathbf{C})}\right) \end{aligned} \tag{4}$$

and denoted $\mathbf{C} \sim \mathrm{KW}(L, \alpha, \mathbf{\Sigma})$. Here, $\mathrm{tr}(\cdot)$ denotes the trace operation, $|\cdot|$ the determinant, $\Gamma(\cdot)$ as the standard Γ-function, d

is the vector dimension, $K_m(x)$ is a modified Bessel function of the second kind with order m and

$$I(L,d) = \pi^{\frac{d(d-1)}{2}} \prod_{i=0}^{d-1} \Gamma(L-i) \tag{5}$$

is a normalisation constant.

We have given the PDF with three parameters as $KW(L,\alpha,\mathbf{\Sigma})$, but we sometimes separate $\mathbf{\Sigma}$ into the mean backscatter intensity, μ, and a normalised internal covariance structure matrix, $\mathbf{\Gamma}$, such that $\mu = |\mathbf{\Sigma}|^{\frac{1}{d}}$ and $|\mathbf{\Gamma}| = 1$. μ can then be considered a mean *brightness* parameter and the relative polarimetry is contained in $\mathbf{\Gamma}$. Therefore, $KW(L,\alpha,\mu,\mathbf{\Gamma})$ is simply an alternative description that we have used previously in [6].

The structure of the product model, log-cumulants and the relation between the standard Wishart and the K-Wishart makes it very easy to adapt the algorithm to be an automatic clustering for the standard Wishart model. In fact, it is as simple as fixing the shape parameter, α, to infinity instead of estimating it, since the standard Wishart is an asymptotic case of the K-Wishart model. This forces the K-Wishart probability density function to be evaluated by the standard Wishart PDF for the probabilities and the log-cumulants revert to the Wishart portion only. Of course, the standard Wishart does not benefit from the flexibility of the K-Wishart and it is expected that more classes will be needed to fit any textured regions in the data.

The Wishart PDF, with L degrees of freedom and mean covariance matrix $\mathbf{\Sigma}$, is given by

$$f_{\mathbf{C}}(\mathbf{C};L,\mathbf{\Sigma}) = \frac{L^{Ld}|\mathbf{C}|^{L-d}}{I(L,d)|\mathbf{\Sigma}|^{L}} \exp\left(-L\mathrm{tr}(\mathbf{\Sigma}^{-1}\mathbf{C})\right) \tag{6}$$

which we denote as $\mathbf{C} \sim \mathrm{W}(L,\mathbf{\Sigma})$. As mentioned, the Wishart model is a limiting case of the K-Wishart, such that

$$\lim_{\alpha\to\infty} \mathrm{KW}(L,\alpha,\mathbf{\Sigma}) = \mathrm{W}(L,\mathbf{\Sigma}) \tag{7}$$

Another simple adaption, is to again fix α to infinity, but allow the individual clusters to have their own independent ENL estimates. This leads to the Relaxed-Wishart model [11] which allows for a moderate degree of texture and with very simple numerical evaluations. It uses the same PDF as the Wishart, but each class can have a different number of looks. We will show examples from both of these options as well as the K-Wishart.

C. Matrix log-cumulants

The matrix log-cumulant expressions for the K-Wishart model are given in [8] as:

$$\kappa_1 = \ln|\mathbf{\Sigma}| + \psi_d^{(0)}(L) - d\ln(L) + d\psi^{(0)}(\alpha) - d\ln(\alpha) \tag{8}$$
$$\kappa_\nu = \psi_d^{\nu-1}(L) + d^\nu \psi^{\nu-1}(\alpha) \qquad ,\nu > 1 \tag{9}$$

where $\psi_d^{(\nu)}(\cdot)$ is the multivariate polygamma function, defined as

$$\psi_d^{(\nu)}(L) = \sum_{i=0}^{d-1} \psi^{(\nu)}(L-i) \tag{10}$$

and $\psi^{(\nu)}(\cdot)$ is the ordinary polygamma function.

The standard Wishart model simply removes the texture dependent part and reduces to

$$\kappa_1 = \ln|\mathbf{\Sigma}| + \psi_d^{(0)}(L) - d\ln(L) \tag{11}$$
$$\kappa_\nu = \psi_d^{\nu-1}(L) \qquad ,\nu > 1 \tag{12}$$

D. Matrix log-cumulant Goodness-of-fit test

The goodness-of-fit testing of order p described in [10] utilises multiple sample matrix log-cumulants as a vector $\langle\boldsymbol{\kappa}\rangle = [\langle\kappa_1\rangle, \langle\kappa_2\rangle, \ldots, \langle\kappa_p\rangle]^T$ and compares them to the population matrix log-cumulants (as a vector $\boldsymbol{\kappa}$) defined by the model parameters and expressions (8) and (9). The test statistic for n samples is

$$Q_p = n(\langle\boldsymbol{\kappa}\rangle - \boldsymbol{\kappa})^T \mathbf{K}^{-1} (\langle\boldsymbol{\kappa}\rangle - \boldsymbol{\kappa}) \tag{13}$$

which has been shown to be asymptotically χ^2 distributed with p degrees of freedom. We shall be using tests of order 4 using the first four matrix log-cumulants and require population matrix log-cumulants up to order 8 to determine the log-cumulant covariance matrix $\mathbf{K}$,

$$\mathbf{K} = \begin{bmatrix} \kappa_2 & \kappa_3 & \kappa_4 & \kappa_5 \\ \kappa_3 & \kappa_4 + 2\kappa_2^2 & \kappa_5 + 6\kappa_2\kappa_3 & K_{24} \\ \kappa_4 & \kappa_5 + 6\kappa_2\kappa_3 & K_{33} & K_{34} \\ \kappa_5 & K_{42} & K_{43} & K_{44} \end{bmatrix} \tag{14}$$

where

$$K_{24} = K_{42} = \kappa_6 + 8\kappa_2\kappa_4 + 6\kappa_3^2\,, \tag{15}$$
$$K_{33} = \kappa_6 + 9\kappa_2\kappa_4 + 9\kappa_3^2 + 6\kappa_2^3\,, \tag{16}$$
$$K_{34} = K_{43} = \kappa_7 + 12\kappa_2\kappa_5 + 30\kappa_3\kappa_4 + 36\kappa_2^2\kappa_3\,, \tag{17}$$
$$K_{44} = \kappa_8 + 16\kappa_2\kappa_6 + 48\kappa_3\kappa_5 + 34\kappa_4^2 + 72\kappa_2^2\kappa_4 + 144\kappa_2\kappa_3^2 + 24\kappa_2^4\,. \tag{18}$$

The sample matrix log-cumulants are determined from the sample matrix log-moments

$$\langle\mu_\nu\rangle = \frac{1}{n}\sum_{i=1}^{n} (\log|\mathbf{C}_i|)^\nu \tag{19}$$

through the following moment to cumulant equations

$$\kappa_1 = \mu_1\,, \tag{20}$$
$$\kappa_2 = \mu_2 - \mu_1^2\,, \tag{21}$$
$$\kappa_3 = \mu_3 - 3\mu_1\mu_2 + 2\mu_1^3\,, \tag{22}$$
$$\kappa_4 = \mu_4 - 4\mu_1\mu_3 - 3\mu_2^2 + 12\mu_1^2\mu_2 - 6\mu_1^4\,, \tag{23}$$

The asymptotic distribution may not be accurate for small sample sizes and we therefore use the Monte-Carlo approach (also described in [10]) to evaluate the test p-value for sample sizes smaller than 300 pixels, although it is much slower to compute.

IV. Automated Analysis

A. Expectation Maximisation based algorithm

Our modifications to the standard EM algorithm are that we use method of multiple log-cumulants for class parameter updates and include and extra split/merge stage through goodness-of-fit testing. The basic idea has been summarised into table I, where we point out a brief comment or which estimation method is used for each parameter, while more details follow in this section.

The E-step is the standard procedure, using the K-Wishart expression (4) to calculate the posterior class probabilities for each pixel. The posterior class probabilities are used as weights for each sample for all class parameter estimates and sample log-cumulant evaluation.

In the M-step, the K-Wishart class parameter updates involve membership weighted means over the covariance matrix data points for the class mean covariance matrix $\mathbf{\Sigma}$, and a new minimum distance optimisation (solved numerically) over multiple matrix log-cumulants [8] for the texture/shape parameter α. We found better, more stable results when using several moments and by implementing the matrix log-cumulants because of their superior bias and variance properties, and known multivariate Gaussian limiting distribution derived in [8].

TABLE I
Summary of K-Wishart Clustering Algorithm

Input:	MLC matrix data	$\{\mathbf{C}_i\}_{i=1:N}$
E-stage:		
$\forall j = 1:m$	membership weights $W_{j,i}$	KW$(\mathbf{C}_i; \text{ENL}, \alpha_j, \mathbf{\Sigma}_j)$
	$\Rightarrow$ posterior weights	div. sum $W_{j,i}$ per sample
M-stage:		
$\forall j = 1:m$	ENL_j	MoLCs
	global ENL	$\text{ENL} = \text{RMS}(\text{ENL}_j)$
$\forall j = 1:m$	priors π_j	sum $W_{j,i}$
	$\mathbf{\Sigma}_j$	weighted sample mean
	α_j	MoLCs
GoF test:	· Split	$m = m + 1$
	· Merge	$m = m - 1$
	· unchanged	
Exit test:	converged or max.iterations	
Output:	· Clustered image	
	· Class parameter values	$\{\pi_j, \text{ENL}_j, \alpha_j, \mathbf{\Sigma}_j\}_{j=1:m}$

where MoLCs means method of matrix log-cumulants and RMS is root-mean-squared.

B. ENL estimation

The K-Wishart expression (4) requires the number of looks parameter L, where the looks are assumed to be independent. In reality, there is some correlation between image pixels and the actual number of correlated looks needs to be substituted with the effective number of independent looks (ENL) to fit the data. The ENL value may be known from the satellite system and the degree of multi-look smoothing applied to the image, or estimated from the data. We have incorporated ENL estimation directly into the EM-algorithm, and include a correction for texture. Estimates are obtained for each cluster from the first log-cumulant expression given the estimated cluster covariance matrix and an independent (from the log-cumulant expression point of view) estimate for the cluster texture parameter, obtained from the variance of trace$(\mathbf{\Sigma}^{-1}\mathbf{C})$ [6]. The new global ENL estimate is taken to be the root-mean-squared average of each class estimate.

C. Split and Merge

At regular iteration steps, for example every tenth iteration, we perform the goodness-of-fit testing. The testing has two phases. First, we test how well all the existing clusters fit to their models and if they fail the test they are split into two clusters. It was observed that the main type of poorly fitted clusters were those that were multi-modal and clearly made up of two or more tighter groups, which will occur when there are simply too few classes to cover the data space. The poorly fitted clusters are split based upon whether trace$(\mathbf{\Sigma}^{-1}C_i)$ is less than or greater than d, which is the expected mean value, and often immediately results in *good* clusters.

The second phase is testing for merge possibilities, and here we have to test all possible pairs of *good* clusters to see whether they are converging to the same data group. This is a common result in mixture models when too many clusters are involved, and several end up competing for the same data group and have virtually identical parameters. The samples from each pair of clusters are combined, new joint class parameters are estimated, the combined group is goodness-of-fit tested and if it passes the two are merged together. Even when only testing each *good* pair, this can involve very many tests for complicated images. Results have shown that this is necessary even when initialised from below with one cluster, because earlier stages may have over-split the image if it had not fully stabilised before goodness-of-fit testing. Alternatively, one may have initialised with very many clusters and need to down-cluster towards the optimum.

D. Goodness-of-fit test stage

We have now implemented a fully matrix-variate goodness-of-fit test that utilises the latest in methods of matrix log-cumulants directly on the class sample covariance samples [10], and the only tuning parameter is the test confidence level which has a meaningful interpretation. We have used a confidence level of 95% for the split and merge tests. The goodness-of-fit test statistic is only asymptotically χ^2 distributed and we clearly observed poor testing results for small sample sizes. Our current testing uses the fast χ^2 approximation for cluster samples sizes over 300 pixels, and uses the much slower Monte-Carlo simulation below 300.

E. Number of classes

The goodness-of-fit criterion and the number of samples in each class therefore determines how many classes are deemed statistically significant within the image. Our adaptive method is distinctly different from existing methods to determine the number of classes as it tests each cluster's goodness-of-fit to the data individually, rather than globally, and within a single clustering algorithm. In contrast, the commonly used L-method [14] first performs many complete image clusterings for different numbers of classes and then evaluates each one through the total log-likelihood function to choose an

optimum. The L-method has been applied to polarimetric SAR analysis after performing an agglomerative hierarchical clustering algorithm in conjunction with Wishart clustering in [15], but the L-method choice of optimum seemed somewhat broad and inaccurate. Our method not only finds a statistically meaningful number of clusters, but also guarantees that the resulting clusters are actually good fits to the models. It also has a very appealing statistical interpretation for the choice of the number of classes that are *supported* by the data itself to the chosen *confidence level*.

F. Initial conditions

The goodness-of-fit testing also means that the clustering algorithm is not influenced by the choice of initial conditions, because it can always be initialised as one single class and will automatically split and adapt until it is a sufficiently good fit to the data mixture model. We have not rigorously tested the convergence of the algorithm from different starting condition nor against other methods, but our algorithm appears to converge to repeatable results that are influenced by the number of samples and the confidence level of the tests. Starting from a single cluster and testing after only partial convergence can very quickly multiply to very many clusters if necessary. This method must be quicker than hierarchical methods, because we essentially only partly cluster to all the incorrect number of classes before changing size, rather than many full cluster solutions.

G. K-Wishart Numerical Problems

Bessel K: A serious problem with the K-Wishart model is that the PDF in (4) requires evaluation of the modified Bessel function of the second kind. The range of real SAR data often eventuates in numerical infinities in the Bessel-K function leading to many or all of the probabilities being incalculable. The K-Wishart expression also involves power and gamma function terms, that may also give numerical problems for large L and α. We could improve the calculable range for these terms by calculating the log-density function instead, but the Bessel-K function was still a major problem for some data values. The known asymptotic approximations did not sufficiently extend the range or did not result in a smooth transition to the approximation and could not be used for clustering probabilities. We finally avoided the Bessel-K function altogether for the troublesome data and used a numerical evaluation of the generalised Wishart integral [6, Eq. (11)].

The result is a multi-stage, sequential evaluation, where we first use the Bessel K function and evaluate the logarithm of the PDF from (4). Where that fails and α is large, i.e. $\alpha > 50(dL+1)/(d+1)$, we use the Relaxed-Wishart model instead, and for small α we use numerical integration and the adaptive Gauss-Kronrod quadrature method. If the integration fails, we use the Relaxed-Wishart model again, even though it is not so accurate for small α, but it will at least give finite results. This scheme appears to be accurate enough except for very large sample sizes, but is slow to compute if many Bessel K results are invalid.

Where the Relaxed-Wishart model is used, the class number of looks parameter is calculated from the texture parameter such that the Wishart model, $W(L_r, \mathbf{\Sigma})$, gives the same expected variance for trace$(\mathbf{\Sigma}^{-1}C)$ as the K-Wishart model, $KW(L, \alpha, \mathbf{\Sigma})$. We therefore we use

$$L_r = L/\left(1 + \frac{(d\ L + 1)}{\alpha}\right) \tag{24}$$

and limit the value to $d \geq L_r \geq L$ for valid evaluation of the Wishart pdf.

Extreme α: Another problem occurs when the texture parameter, α, is so low that the PDF has a non-zero limit as $\mathbf{C}$ approaches the null matrix. A class with such a low α has a tendency to absorb all low intensity pixels and become even more extreme. The K-Wishart expression can certainly evaluate this case, and the clusters do tend to get split up eventually, but, apart from for very low numbers of looks, this is not really representing a likely physical situation as shape due to backscatter variation from a single target media is seldom so extreme. Texture may *appear* to be extremely low because of mixtures of classes, but that is exactly the type of mixed cluster that we wish to separate. Hence, in our algorithms we have limited the texture parameter by a minimum value.

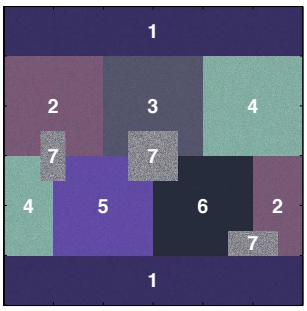
V. Results

Results will be shown for several different cases. Firstly, we will demonstrate the overall approach with automatic clustering results for simulated K-Wishart test-pattern images, where class histograms will be included for each test stage decision. Secondly, we will show results for different models on the same test pattern to show how the more restrictive models require more classes to account for the total data variation. Thirdly, we will show results for real data from the Foulum sample data-set. Here we will show the effect of sub-sampling and the resulting number of clusters, as well as the choice of model. Lastly, we will show clustering examples from other data-sets to demonstrate the generality of the algorithms.

A. Simulated examples

To validate the procedures in a clear way, we have produced a test image with clear uniformly distributed areas that are generated with 16-look, quad-pol K-Wishart distributed data. The polarimetric matrix and texture parameters were taken from a real data-set to simulate the natural range of classes expected in real images. They come from water, forest, urban and four agricultural fields from the Foulum data-set which will also be a real-world example in Sec. V-B. The image layout was designed such that each class has a boundary to every other class to help visualise the class boundary distinction.

Fig. 1 shows the simulated test pattern and some of the parameter values. The number of looks is 16, and the graininess due to texture is most visible in the most highly textured classes 4 and 7. Fig. 3 shows the clustered image which correctly found the 7 classes, after initialisation to all one class,



Looks = 16

#	type	alpha	mu$\times 10^3$
1	water	8281	0.09
2	field A	202.8	1
3	field B	91	0.6
4	forest	39	6
5	field C	275.6	0.71
6	field D	780	0.05
7	urban	1.95	4

Fig. 1. Simulated test pattern, 16-look: enhanced Pauli RGB image with class labels, texture and brightness values.

and with sub-sampling of 1/49. This degree of sub-sampling resulted in between 500 and 1500 samples per class, which is enough for the χ^2 approximation and quite acceptable class distinction.

The actual class histogram split/merge sequence is shown in Fig. 2, with the columns indicating the progressive test stages and the number of rows indicates the number of clusters at that stage. A visual display of the class goodness-of-fit is obtained by plotting the one dimensional trace$(\mathbf{\Sigma}^{-1}\mathbf{C})$ as histograms for each class. The corresponding class density functions are evaluated as $KW(L\,d, \alpha, d)$ for each of the K-Wishart clusters. The scale is normalised for each class so inter-comparison is not directly possible, but the goodness-of-fit of each class is readily apparent.

Notice how the single initialisation cluster is clearly a poor fit to the model curve (in red) and that the number of clusters actually rose to 8 before merging back to 7. The final ENL value was estimated to be 16.17 within the algorithm, but was much lower during the earlier mixed cluster stages.

Fig. 4 shows the result for the standard-Wishart and the Relaxed-Wishart models on the same K-Wishart simulated test image, also with 1/49 sub-sampling. As expected, the less flexible distributions found more than the 7 true classes, the standard-Wishart found 16 and the Relaxed-Wishart found 13 clusters. Note that the degree of class fragmentation coincides very well with ordering of the non-Gaussianity parameter. The Non-Gaussianity increases in the order: 1, 6, 5, 2, 3, 4, and 7. In both the standard-Wishart and Relaxed-Wishart, the most fragmented classes are 7, 4 and 3. Note also that the moderate flexibility of the Relaxed-Wishart model managed to cleanly cluster class 5, whereas the standard-Wishart did not. We would also comment that the very speckled appearance of the extra clusters and the fact that there are multiple clusters evenly mixed within whole true class regions, but none crossing or blending different true clusters, supports the case that they are model mismatch. The restricted models simply need more clusters to describe the large variability of the highly textured classes.

B. Foulum example

The automatic segmentation will be demonstrated on a small and reasonably simple section of the image such that a visual validation is possible. The image is an agricultural area from an airborne EMISAR, L-band, quad-pol SAR flight over Foulum Denmark in 1998 (also analysed and described in [6]).

Figure 5 shows an enhanced Pauli RGB image (top) and the subsequent 'fine', 'medium' and 'coarse' K-Wishart segmentations due to different degrees of sub-sampling. The sub-sampling was 1/4, 1/16, and 1/49, respectively.

Visual comparison with the Pauli image confirms that the major fields are cleanly segmented and distinguished, and that the different classes within the forest (the large whitish areas in the Pauli) clearly match the brighter or darker regions within the forest. Overall this gives a visually satisfactory segmentation.

All of the three degrees of detail show the main fields within the image. The coarsest image (bottom plot) separates the purple and reddish fields from the centre of the Pauli image (into the pink and yellow classes, respectively). However, it shows a mixed field in the lower left corner as a cross between the two (pink and yellow), and a mixed (pink and dark blue) for the smaller dark field in the upper centre-left of the Pauli. The Pauli colours of these fields do fall somewhere between the other colours and it is therefore quite understandable why they are mixed when we have so few classes. The medium and fine detail images have clarified these cases and given each their own class, without changing the main coarse detail classes. The medium and fine images show other examples of extra detail, particularly many edge classes along the tree lined roads and near the buildings, the bright white and pink dots in the Pauli, and in the forest. There is also a sub-class forming in the reddish field of the Pauli, which has become a class of its own in the fine detail image (the olive green colour within the pink). This class seems to correspond to a richer coloured region within the reddish class and probably corresponds to some actual variation in moisture or vegetation growth that is statistically distinct when given enough data points to cluster. Also note that the pink spots within the forest (black class) are a separate sub-class, because there are quite a few of those pixels. These are certainly believable sub-class choices after all the more obvious classes have been separated, however we do not have any ground truth data to confirm what they are. The most we can claim is that the algorithm gives a visually pleasing clustering of the images, finding the clearly different polarimetric regions before sub-classing the smaller details.

Fig. 6 shows the result for the standard-Wishart and the Relaxed-Wishart models on the same K-Wishart simulated test image, with 1/4 sub-sampling. These less flexible models found 44 and 30 clusters, respectively, and although some fields are still reasonably smoothly separated, the more non-Gaussian forest area has become very many clusters. The sub-clusters within the fields are certainly representing real differences in the data intensities and may be useful in some applications, but the number of clusters can become so large as to be quite difficult to interpret. Image segmentation applications are usually trying to simplify the description of the image content, and hence may prefer the K-Wishart result.

In our previous work [6], we compared fixed number of classes clustering results of the K-Wishart against the Wishart model, but this was not a fair comparison of the two models.

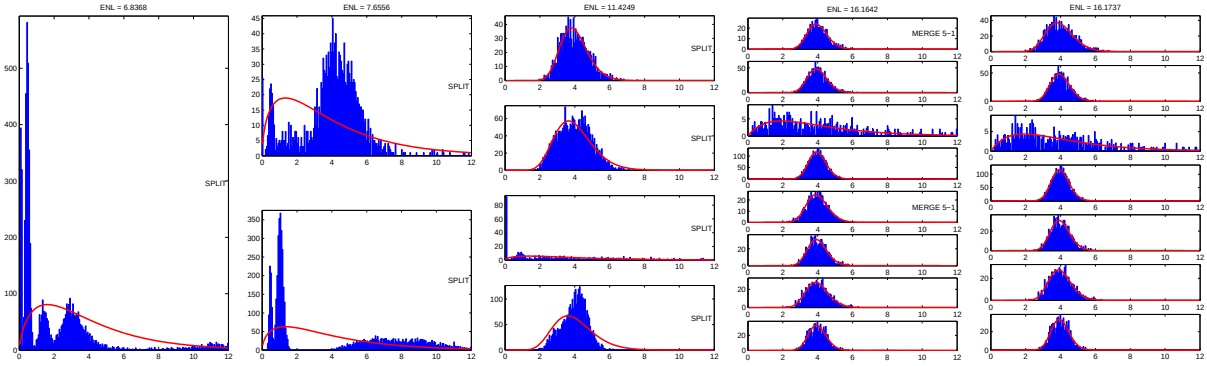


Fig. 2. Class histogram sequence, from initial single class to final 7 class automatic cluster result, sub-sampling to 1/49 samples.

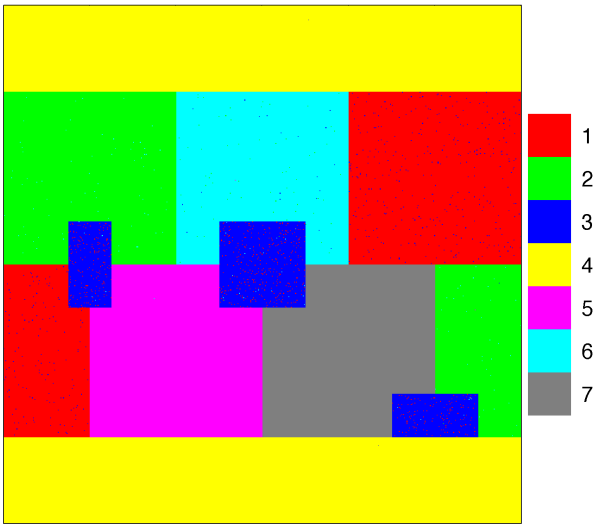


Fig. 3. Automatic K-Wishart clustering result of simulated K-Wishart test pattern. 7 classes correctly found, sub-sampling to 1/49 samples.

The new results shown here in Fig. 6 really show the true behaviour of the Wishart model, and therefore the full difference compared to the K-Wishart model.

C. Kongsvegen glacier example

This scene is from an Envisat, C-band, dual-pol ASAR image over Kongsvegen glacier, Svalbard in 2005 (previously analysed in [16]). The dual-pol data was VV and VH polarisations, and so the clustering algorithm was working with 2×2 matrix MLC images.

Figure 7 shows the results of a 'fine' (1/4 sub-sampling) and a 'coarse' (1/16 sub-sampling) segmentation of the dual-pol glacier data. The mountains have been masked out for convenience, and again a careful comparison with the dual-pol quasi-Pauli image (top) indicates that it has successfully found the major brightness and polarimetric classes within the image. More sub-sampling (less samples) results in fewer classes, as expected, yet maintains the major clusters. The dual-pol

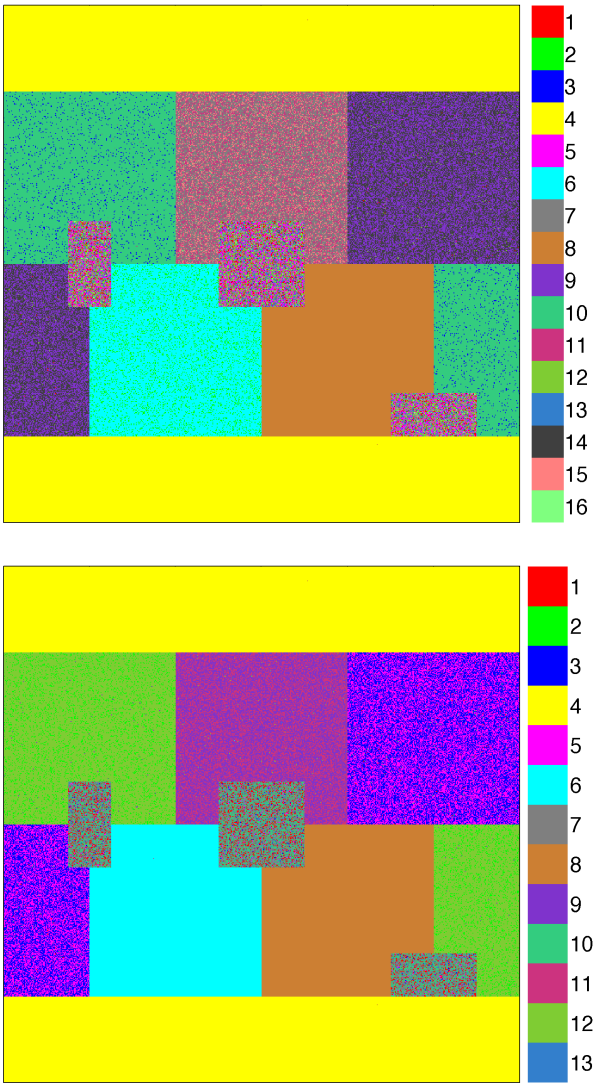


Fig. 4. Automatic standard-Wishart (top) and Relaxed-Wishart (bottom) clustering results for simulated K-Wishart test pattern, sub-sampling to 1/49 samples. 16 and 13 classes found, respectively.

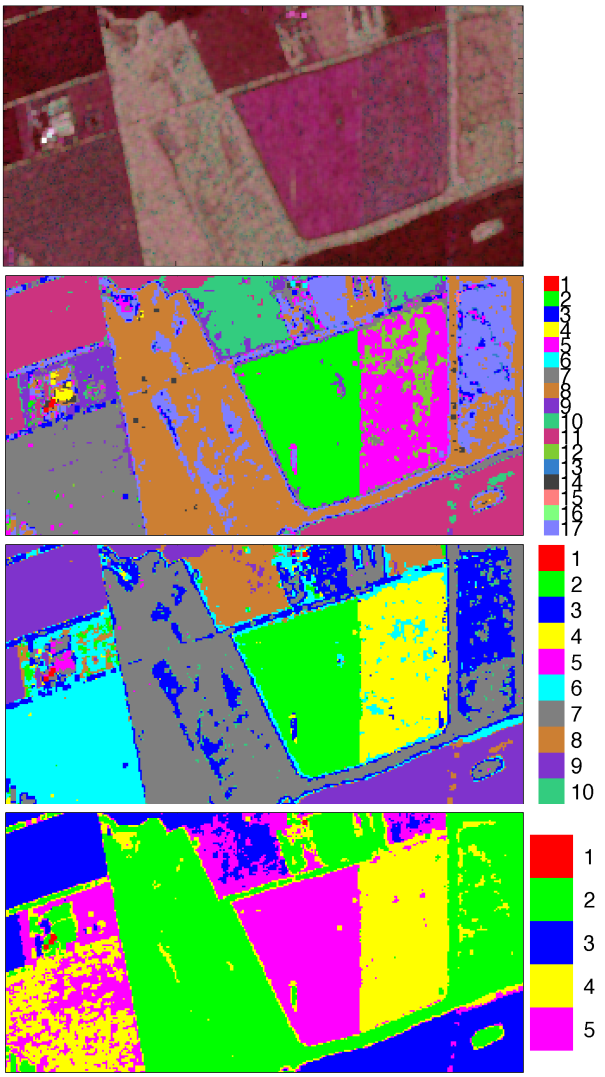


Fig. 5. Enhanced Pauli RGB image (top) and Auto-Segmentation of Foulum dataset, 18 looks, to 'fine', 'medium', and 'coarse' (bottom) sub-sampling.

glacier data seems to have more trouble finding clean, solid clusters and results in quite a bit of overlap between some classes. However, this cluster confusion appears to be visible in the Pauli image and is therefore part of the actual data.

D. Flevoland example

This AIRSAR, L-band, quad-pol image was taken over Flevoland, The Netherlands, in August 1989, and has been saved in the 3-dimensional reduced data format. The clustering therefore worked in 3×3 matrix space.

Figure 8 shows the results of a 'fine' (1/25 sub-sampling) and a 'coarse' (1/100 sub-sampling) segmentation of the airborne agricultural scene. Note that the mottled appearance of most fields come through in the detailed image, but has been partly smoothed over in the coarse image, without significant

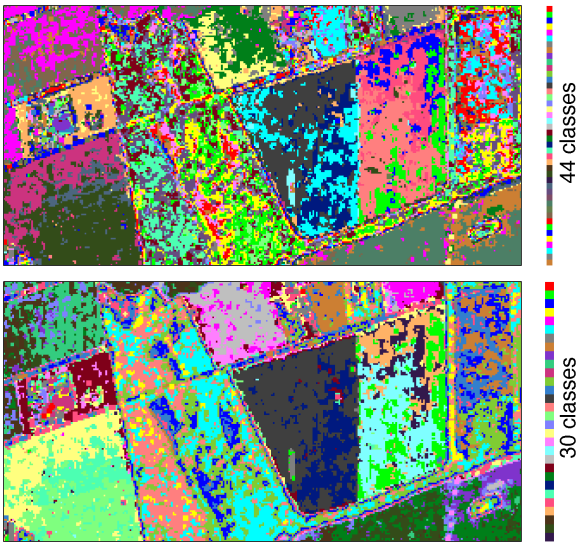


Fig. 6. Foulum dataset, 18 looks, automatic standard-Wishart (top) and Relaxed-Wishart (bottom) clustering results, sub-sampling to 1/4 samples. 44 and 30 classes found, respectively.

blurring of the boundaries. Visual inspection shows that the main different coloured fields have been reasonably well separated, and even the small bright objects (probably houses and wind turbines) are well clustered even on the coarse image, based on very few pixels, because they are so distinctly different from the other pixels.

E. Sea Ice example

This scene is a CONVAIR, C-band, quad-pol image over sea ice from Canada in March 2001.

Figure 9 shows the results of a 'fine' (1/25 sub-sampling) and a 'coarse' (1/64 sub-sampling) segmentation of the airborne quad-pol sea ice data. Again, using the Pauli as a visual guide, we can observe that the main classes are well distinguished even though the image has quite complicated structures throughout the ice fields.

VI. Discussion

The examples given in the results section were selected to show how the approach works and that it works for several different imaging types: both different sensor configurations and different ground terrain types. We would now like to comment upon some of the key observations and features.

We have demonstrated that the automatic algorithm finds the correct number of clusters and the correct ENL estimate for simulated data. We have, of course, observed this in many more situations than shown here. It is, however, more difficult to discuss the correct number of clusters for real data, but the examples shown appear to be finding an appropriate number of classes according to our visual assessment of the Pauli image. The statistical interpretation of the number of clusters is appealing because the statistical confidence level gives meaning to the threshold choice.

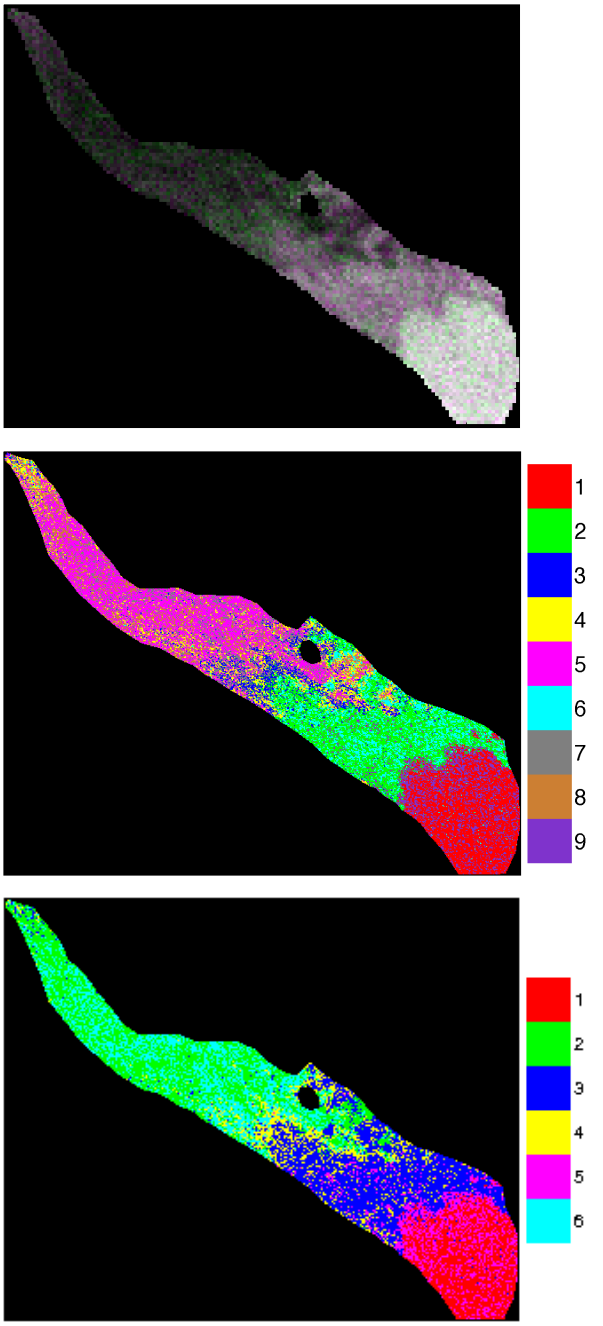


Fig. 7. Quasi-Pauli (dual-pol) RGB image (top) and Auto-Segmentation of Kongsvegen dataset, 40 looks, for high (middle) and low (bottom) sample sizes.

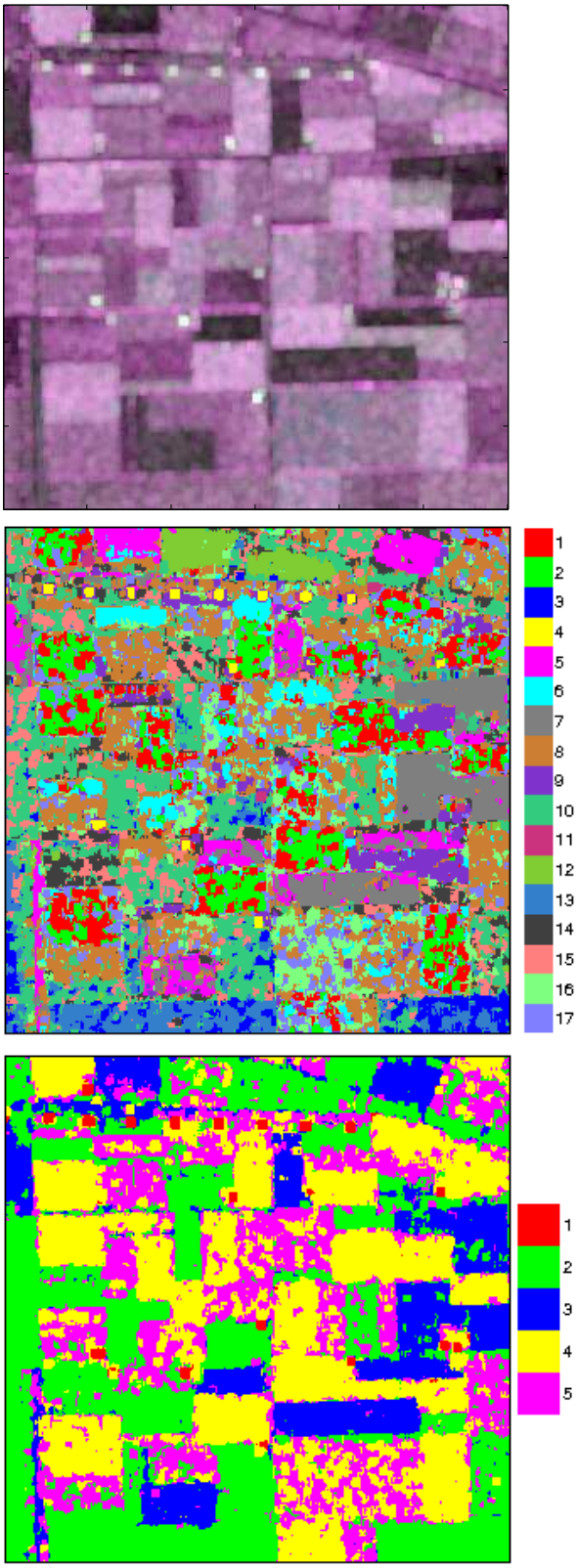


Fig. 8. Enhanced Pauli RGB image (top) and Auto-Segmentation of Flevoland dataset, 25 looks, for high (middle) and low (bottom) sample sizes.

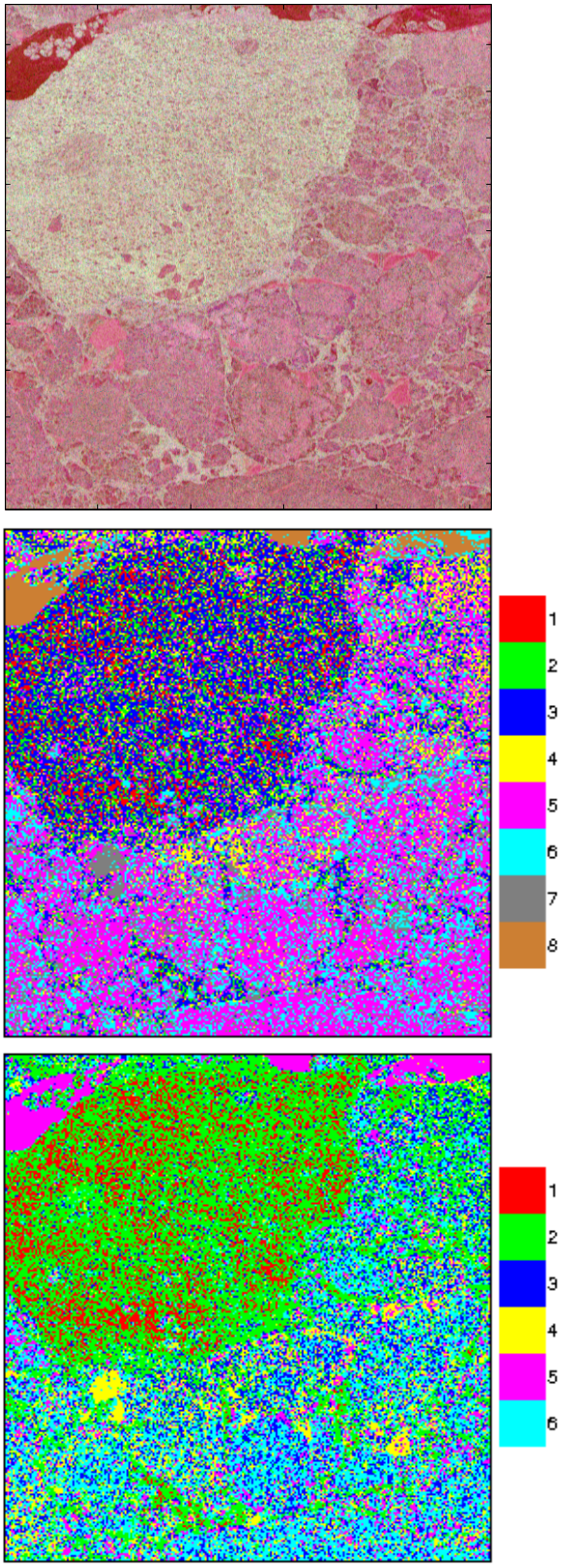


Fig. 9. Enhanced Pauli RGB image (top) and Auto-Segmentation of Canadian sea ice dataset, 20 looks, for high (middle) and low (bottom) sample sizes.

The statistics also governs the detail level effect of reducing the sample size through sub-sampling, and again, in a statistically meaningful manner. Any finite sample from a statistical distribution will exhibit some natural variation in estimated parameters that will decrease with increasing sample size. The reduction in variance allows smaller details to be detected in large sample sizes, and conversely, the large variance for small sample sizes makes the tests less sensitive to small deviations. These small deviations may originate from small sub-classes of a major class, but may also be caused by inaccurate parameter estimation or an inexact statistical model to describe the real world data. The ability to decide the degree of sensitivity, on an application to application basis, by simply sub-sampling the image is a very useful feature.

The comparison to the standard Wishart model results highlights the flexibility of the K-Wishart model to incorporate more intensity variation than the standard Wishart model. Note that this increased intensity variation will still be for a common polarimetric matrix parameter, under the model assumptions. If these brighter or darker pixels had significantly different polarimetry, then they would have been split into extra classes. This means that the extra classes that the standard Wishart model finds, are most likely just brightness variation, and still represent the same polarimetric media and thereby the same thematic class. The class histograms of the final K-Wishart clusters were also observed to be generally good, smooth and continuous fits to the data,which strengthens the choice of the non-Gaussian modelling.

We are aware of applications where seeing this extra information of brightness level may be useful, for example in physical parameter estimation, but we would suggest that a clustering algorithm is then not the most suitable approach to take. For physical parameter inversion, we would suggest a local modelling approach and then interpret the local model parameter values directly, rather than clustering to global class mean representatives. It may still be useful to consider non-Gaussian modelling for this case, as it is additional information on top of the polarimetric matrix values [2].

The ENL estimation appears to be quite robust, even though it starts out with very low values in the early, mixed cluster situations. Once the splitting has obtained nearly all good clusters, then the estimate converges to believable values for the systems and level of multi-looking in question. It is quite important for the goodness-of-fit testing that the ENL estimate is not too low for the least textured classes, as the ENL limits how narrow the PDF can fit to the data. Since we observed that even after correcting for α, the ENL estimate from mixed, multi-modal clusters was still too low, we have used the root-mean-squared average to emphasise the higher values and reduce the influence of this bias in the global estimate. The algorithm always obtained quite precise results for a range of simulated data-sets.

We did observe an occasional problem of cyclically splitting and re-merging the same clusters over and over again, due to some small parameter drift during the intervening iterations. This seems to occur most often for very complicated images, where either there is a continuous mixing transition between classes (and therefore a lot of overlap with neighbouring

transition classes), or where the K-Wishart model (or its approximated numerical evaluation) may not adequately describe the actual data cluster variation. We have therefore incorporated a ramping confidence level to help break such cyclic behaviour, where the split-confidence level is ramped from 95% to 99.999% and the merge is ramped from 95% to 85% over several test stages. This progressive reduction of the sensitivity of the tests breaks the cyclic stale-mate situations and appears to give acceptable results. The ramping only starts after several split merge stages to allow the main classes to be divided, and continues for several tests after the ramp stops to stabilise the clustering. This does not appear to affect other, cleaner, situations where the algorithm converge to stable clusters.

VII. Conclusions

We have demonstrated an automatic segmentation method that incorporates both polarimetry and non-Gaussian analysis and demonstrated good results for both simulated and several real-world data-sets. We have used the non-Gaussian based K-Wishart distribution in a mixture model algorithm for the multi-look covariance matrix data, and have also compared to the standard Wishart and the Relaxed-Wishart models. The automatic techniques are readily adaptable to other models and only requires the probability density function and the first few matrix log-cumulants.

The main new work presented here is the addition of a goodness-of-fit test stage that allows for splitting and merging of clusters. This not only improves the fit of the model to the data, but also results in automatically determining the number of clusters. The goodness-of-fit tests can be readily visualised with the class histograms and seems appropriate by visual inspection. The algorithm requires no initialisation, and determines the number of statistically distinct clusters supported by the actual data-set.

The dependence on sample sizes to statistically supporting the number of classes is expected and desirable, since it gives the opportunity to do either 'fine' or 'coarse' analyses and potentially save processing time. We have presented examples of both coarse and fine segmentations for several images and observe that the main clusters are identified even with coarse sub-sampling, and that the more samples given, the more small details are identified.

The standard Wishart algorithm, is more restrictive in intensity variation and results in many more clusters begin detected. The Relaxed-Wishart algorithm, with a variable number of looks parameter, can only adapt to a moderate degree of texture and its results are intermediate between the Wishart and K-Wishart models.

The K-Wishart probability density function expression is numerically troublesome due to the modified bessel function term, and although we have implemented methods to overcome this, they are slow to compute and we feel that the effect of the approximations is evident for large sample size clustering. We would like to investigate other equally flexible models that are numerically better behaved.

We shall continue to investigate the class drifting that occurs in some situations to determine its cause. One suspicion is that some classes that are too small to be supported on their own are skewing the parameter estimation for their nearest supported cluster. It may be practical to introduce a broad 'residual' class whose prior probability is limited to some small fraction of the total pixels to absorb these outlying but influential pixels. Another possibility is that the sample mean estimator for the covariance matrix is not suitable to the more logarithmic scaling of radar data, and we shall investigate if log-cumulant type methods can be used to improve the covariance matrix parameter estimation.

Acknowledgment

The authors would like to thank: Professor Henning Skriver, Dr. Jorgen Dall, and the Danish Technical University for the Foulum data set; NASA/JPL-Caltech for the Flevoland data set; Dr. Daniel Delisle, Dr. Sahebi Mahmod Reza, and the Canadian Space Agency for the sea ice data set. These data-sets were downloaded from the Euro- pean Space Agency website (http://earth.esa.int/polsarpro/ datasets.html). Thanks to Norut Tromsø, Norway, for the Kongsvegen data set.

References

[1] J. S. Lee, D. L. Schuler, R. H. Lang, and K. J. Ranson, "K-distribution for multi-look processed polarimetric sar imagery," in *IEEE Int. Geosci. Remote Sensing Symp.*, 1994, pp. 2179–2181.

[2] A. P. Doulgeris and T. Eltoft, "Scale mixture of gaussian modelling of polarimetric sar data," *EURASIP Journal on Advances in Signal Processing*, vol. 2010, no. 874592, p. 12, 2010.

[3] C. C. Freitas, A. C. Frery, and A. H. Correia, "The polarimetric G distribution for SAR data analysis," *Environmetrics*, vol. 16, pp. 13–31, 2005.

[4] C. Oliver and S. Quegan, *Understanding Synthetic Aperture Radar Images*, 2nd ed. Raleigh, USA: SciTech Publishing, 2004.

[5] A. Doulgeris, S. Anfinsen, Y. Larsen, K. Langley, and T. Eltoft, "Evaluation of polarimetric configurations for glacier classification," in *International POLinSAR Workshop*, Frascati, Italy, January 26-30 2009.

[6] A. Doulgeris, S. Anfinsen, and T. Eltoft, "Classification with a non-gaussian model for polsar data," *IEEE Trans. Geoscience and Remote Sensing*, vol. 46, no. 10, pp. 2999–3009, Oct. 2008.

[7] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via teh EM algorithm," *Journal of the Royal Statistical Society. Series B*, vol. 39, no. 1, pp. 1–38, 1977.

[8] S. N. Anfinsen and T. Eltoft, "Application of the Matrix-Variate Mellin Tranform to Analysis of Polarimetric Radar Images," *IEEE Trans. Geoscience and Remote Sensing*, Submitted 2010.

[9] A. P. Doulgeris and T. Eltoft, "Automated Non-Gaussian clustering of polarimetric SAR." in *8th European Conference on Synthetic Aperture Radar (EUSAR)*, Aachen, Germany, June 7-10 2010.

[10] S. N. Anfinsen, A. P. Doulgeris, and T. Eltoft, "Goodness-of-Fit Tests for Multilook Polarimetric Radar Data Based on the Mellin Transform," *IEEE Trans. Geoscience and Remote Sensing*, Submitted 2010.

[11] S. Anfinsen, T. Eltoft, and A. Doulgeris, "A relaxed Wishart model for polarimetric SAR data," in *Proc. 4th Int. Workshop on Science and Applications of SAR Polarimetry and Polarmietric Interferomentery (POLinSAR2009), Frascati, Italy, 26-30 January*, ser. ESA SP-668, April 2009, p. 8 pp.

[12] E. Jakeman and P. N. Pusey, "A model for non-Rayleigh sea echo," *IEEE Trans. Antennas Propagat.*, vol. 24, 6, pp. 806–814, Nov. 1976.

[13] J. Lee, M. Grunes, and R. Kwok, "Classification of multi-look polarimetric SAR imagery based on the complex Wishart distribution," in *Int. J. Remote Sensing*, vol. 15, 11 1994.

[14] S. Salvador and P. Chan, "Determining the number of clusters/segments in hierarchical clustering/segmentation algorithms," in *Tools with Artificial Intelligence, 2004. ICTAI 2004. 16th IEEE International Conference on*, nov. 2004, pp. 576 – 584.

[15] F. Cao, W. Hong, Y. Wu, and E. Pottier, "An unsupervised segmentation with an adaptive number of clusters using the SPAN/H/α/A space and the complex wishart clustering for fully polarimetric SAR data analysis," *IEEE Trans. Geoscience and Remote Sensing*, vol. 45, no. 11, pp. 3454–3467, November 2007.

[16] A. P. Doulgeris, K. Langley, and T. Eltoft, "Analysis and classification of high arctic glaciers with ASAR data," in *IEEE International Geoscience and Remote Sensing Symposium*, Boston, Massachusetts, USA, July 6-11, 2008.

Anthony P. Doulgeris (S'06) received the B.Sc. degree in physics from The Australian National University, Canberra, in 1988, and the M.Sc. degree in physics from the University of Tromsø, Tromsø, Norway, in 2006. He is currently working toward the Ph.D. degree in physics in the Department of Physics and Technology, University of Troms, investigating remote sensing with polarimetric synthetic aperture radar and multidimensional statistical modelling.

Stian Normann Anfinsen (S'06) received the Cand. Mag. (B.Sc.), Cand. Scient. (M.Sc.) and Ph.D. degrees in physics from the University of Tromsø, Tromsø, Norway in 1997, 2000 and 2010, respectively. He received the M.Sc. in Communications, Control and Digital Signal Processing from the University of Strathclyde, Glasgow, U.K. in 1998. He was with satellite ground station system provider Kongsberg Spacetec in Tromsøfrom 2001 to 2005, before he moved to the University of Tromsø, Department of Physics and Technology, to conduct research on statistical analysis and pattern recognition applied to polarimetric radar images. He is currently a postdoctoral research fellow at the same department, and is also affiliated with the Barents Remote Sensing School. His current research focuses on statistical modelling, parameter estimation, classification, change detection and target detection in synthetic aperture radar images.

Torbjørn Eltoft (A'92M'98) Cand.Real. (M.Sc.) and Dr.Sci. degrees from the University of Tromsø, Tromsø, Norway, in 1981 and 1984, respectively. Since 1988, he has been with the University of Tromsø, where he is currently a Professor in electrical engineering in the Department of Physics and Technology. He is also an Adjunct Professor with Norut AS, Tromsø. His current research interests include multidimensional signal and image analysis with application in remote sensing, statistical models, neural networks, and machine learning. Dr. Eltoft is a recipient of the 2000 Outstanding Paper Award in Neural Networks from the IEEE Neural Networks Council and an Honorable Mention in the 2003 Pattern Recognition Journal Best Paper Award.