An Automatic $\mathcal{U}$-distribution and Markov Random Field Segmentation Algorithm for PolSAR Images.

Anthony P. Doulgeris, *Member, IEEE*

Abstract—We recently presented a novel unsupervised, non-Gaussian and contextual clustering algorithm for segmentation of polarimetric SAR images [1]. This represents one of the most advanced PolSAR unsupervised statistical segmentation algorithms and uses the doubly flexible, two parameter, $\mathcal{U}$-distribution model for the PolSAR statistics and includes a Markov Random Field approach for contextual smoothing. A goodness-of-fit testing stage adds a statistically rigorous approach to determine the significant number of classes. The fully automatic, algorithm was demonstrated with good results for both simulated and real datasets. This paper discusses a re-thinking of the overall strategy and leads to some simplifications. The primary issue was that the Markov random field optimisation depends on the number of classes and did not behave well under the split-and-merge environment. We explain the reasons behind a separation of the cluster evaluation from the contextual smoothing as well as a modified rationale for the adaptive number of classes. Both aspects have simplified the overall algorithm whilst maintaining good visual results.

Index Terms—Polarimetric Synthetic Aperture Radar, Non-Gaussian, Statistical Modelling, Clustering, Number of Classes.

I. INTRODUCTION

We have developed and demonstrated an advanced automatic clustering algorithm that combines a flexible non-Gaussian class model for multi-look complex (MLC) covariance matrix data, a Markov random field (MRF) for contextual smoothing, and goodness-of-fit testing to optimise the segmentation and determine an appropriate number of classes [1]. The main features of this state-of-the-art approach are summarised in Sec. II. This article will subsequently discuss recent simplifications to the approach and demonstrate the new algorithm.

Satellite-borne polarimetric synthetic aperture radar (PolSAR) systems have many benefits, but analysis is hindered by complicated non-Gaussian statistical methods. PolSAR data models are generally derived from the product model [2], which states that the backscattered signal results from the product between a Gaussian speckle noise component and the textured terrain backscatter. For multi-looked PolSAR matrix data, the scaled Wishart distribution [3], [4], $\mathcal{W}_d$, is the simplest model to analyse but contains no texture parameter, based upon purely Gaussian speckle. The $\mathcal{K}_d$ (or K-Wishart) distribution [5], [6] and the $\mathcal{G}_d^0$-distribution [7], [8] are more flexible, with one texture parameter, and are successful models

for many PolSAR classes. The two parameter Kummer-U distribution has been used to model PolSAR vector data [9], with promising contiguous segmentation results and demonstrated that the two parameter model is even more flexible and able to fit real data classes. The multivariate extension of the Kummer-U distribution for MLC matrix data, hereafter simply called the $\mathcal{U}_d$-distribution, that was recently introduced [1], [10], and used for real analysis in [11], gives improved results because of its flexibility to model more varied textures and because it includes the $\mathcal{W}_d$, $\mathcal{K}_d$ and $\mathcal{G}_d^0$ models as asymptotic cases.

As with many of these product models, the probability density functions (PDFs) are complicated. Many variations lead to different cases of the hypergeometric function that are complicated to evaluate, often needing numerically integration, and complicated with respect to maximising parameters in the order term of the functions. Hence, maximum likelihood parameter estimators are not usually available with closed-form solutions. A practical solution is to estimate the model parameters with the method of matrix log-cumulants [4], because they have relatively simple numerical expressions and possess lower bias and variance compared to single channel (marginal) estimates or moment methods for product based distributions.

Unsupervised clustering is achieved with the expectation maximisation algorithm (EM-algorithm) [12], which is an iterative method that repeatedly estimates the class posterior probabilities based on the current parameters and then updates the class parameters based upon the estimated probabilities. It requires an initial state, the number of classes, PDF expressions for the class models, and update expressions for the parameters. The $\mathcal{U}_d$-distribution shall be used for the model and the method of matrix log-cumulants for parameter estimation. The initial state and the number of classes are addressed with an automatic strategy by consistently starting as one class and adaptively splitting classes until a statistical criterion is satisfied.

The number of classes is determined from the additional information contained in the goodness-of-fit of the data to the estimated model [13]–[15]. There are several different ways to implement this general framework, and our new strategy is significantly improved and simplified.

Contextual smoothing is desired to improve the accuracy and robustness of the image segmentation. It is achieved in the clustering algorithm with an MRF approach that integrates the $\mathcal{U}_d$-distribution for the PolSAR data statistics conditioned to each image cluster and a Potts model for the spatial context. This extends our previous work with the K-Wishart distribution [16], however, the parameter of the MRF model

are now estimated with a mean-field like method [17]. The inclusion of the MRF is not expected to compromise the goodness-of-fit testing stage, because the MRF only affects the local priors and the underlying model remains a mixture of $\mathcal{U}_d$-distributions.

The previous strategy applied the MRF based local prior probabilities and optimised the MRF global smoothing parameter at every iteration, but this causes serious problems in relation to the adaptive split-and-merge mixture modelling. We now propose to entirely separate the EM-algorithm mixture modelling from the MRF smoothing process in the new strategy.

The proposed improvements still combine all the benefits of a flexible non-Gaussian model for the covariance matrix data classes, an MRF for contextual smoothing, and goodness-of-fit testing to optimise the segmentation and determine an appropriate number of classes.

The previous, state-of-the-art algorithm is described in Section II, the main characteristics of the new strategy are introduced in Section III, are demonstrated for both simulated and real data-sets in Section IV, and, finally, the main conclusions are given in Section V.

II. STATE-OF-THE-ART ALGORITHM

The scope of this algorithm is to analyse MLC data images, where the data is an image of covariance matrices, $\mathbf{C}$. We assume the scalar product model is valid and that the MLC data is formed by a simple box-car multi-look average from the single-look complex scattering coefficients. The box-car is used so that we can assume a global number of looks (degree of smoothing) and "simpler" statistical models, as opposed to advanced dynamic speckle smoothing methods that produce a variable number of looks that complicate the modelling. The number of looks, L is in practise substituted with an equivalent number of looks (ENL) due to pixel correlations. The ENL is either estimated in a pre-processing step or optimised during the iterations by a minimum distance method using log-cumulant expressions given all the current class model parameters simultaneously.

Our main objective is to segment the image pixels into separate clusters based upon the $\mathcal{U}_d$-distribution model. The statistical approach for clustering the images uses the iterative EM-algorithm with a few modifications, as has previously been described in detail in [13], [14], [16]. The extension proposed in [1], was that each class is modelled with the $\mathcal{U}_d$-distribution PDF and that context has been incorporated with an MRF technique based upon the Potts model. The key features are detailed here.

Its main drawback seems to be computation time, but this can be partly alleviated with a sub-sampling approach, as in [14], that still finds the major classes of interest but sacrifices smaller sub/side classes for reduced computation time.

A. Non-Gaussian Modelling: the $\mathcal{U}_d$-distribution

Bombrun et al. [10] have shown the potential of the $\mathcal{U}_d$ PDF, with texture parameters α and λ, to model both *extremely heterogeneous*, moderately *heterogeneous* and *homogeneous*

clutter. It encompasses the other models as special cases, such that it reverts to the $\mathcal{K}_d$ as $\lambda \to \infty$, the $\mathcal{G}_d^0$ as $\alpha \to \infty$, and the $\mathcal{W}_d^{\mathbf{C}}$ as both $\alpha, \lambda \to \infty$. Therefore, this one model supersedes many previous modelling algorithms. Nevertheless, these special cases may be implemented to improve the numerical evaluation in different parameter ranges. Table I lists the probability density functions and the matrix log-cumulant expressions for the matrix variate $\mathcal{W}_d$, $\mathcal{K}_d$, $\mathcal{G}_d^0$, and $\mathcal{U}_d$ distributions.

B. Parameter Estimation

Texture parameter estimation is achieved with the method of matrix log-cumulants (MoMLC) because closed-form maximum likelihood solutions don't exist. MoMLC has become the popular estimation choice, because it is fast to compute and achieves the most accurate results among the practical alternatives [4]. Although the log-cumulants are also not in closed-form, they are faster to compute than the probability density functions. A simple gradient search algorithm in the multivariate log-cumulant domain is used to optimise the two texture parameters, α and λ, for each class, from membership weighted log-cumulant estimates, and then the ENL is similarly optimised, but over all classes simultaneously.

The model covariance matrix $\mathbf{\Sigma}$ is determined with the sample covariance matrix estimator, which becomes a membership weighted mean over the $\mathbf{C}$ matrix data samples in the fuzzy clustering sense. Although some authors, e.g. [18], have recently promoted the fixed-point estimator for the covariance matrix under the product model, it has a significant bias in low texture regions and only becomes beneficial in quite high texture areas. This is clearly demonstrated in [19]. The fixed-point estimator was tested in early versions of the algorithm, but showed no significant benefit, even for high texture cases, given the large sample sizes and that it is slower to compute.

C. Goodness-of-fit testing

The EM-algorithm was modified to incorporate a split-and-merge stage, at regular intervals within the main iterations. The stage introduces an hypothesis test based upon the goodness-of-fit of the current cluster model to the observed data. Given that the assumed model distribution is appropriate, then a poor fitting data-set must represent a mixture of clusters and we therefore split that cluster in two and continue the EM-algorithm clustering. Conversely, a merge test checks whether two clusters are converging to the same data group, and was found to be necessary due to the chance of over-splitting by testing before full convergence. The true population model will never be known, so we test against the estimated model instead and just assume that the variation around the true model and the variation around the estimated model are similar. Our end objective is to determine whether the current number of classes and model parameters *could* explain the data-set to the given confidence level.

The hypothesis test may be implemented in several ways and we have used two distinct methods with acceptable results. Firstly, we have used a distance measure in a multiple log-cumulant domain (well described in [14], [20]), since we

TABLE I
PDFs AND LOG-CUMULANT EQUATIONS FOR THE COVARIANCE MATRIX DISTRIBUTIONS UNDER THE PRODUCT MODEL [4].

Family	$f_{\mathbf{C}}(\mathbf{C})$ of covariance matrix $\mathbf{C}$	MoMLC equations
$\mathcal{W}_d^{\mathbf{C}}(L, \boldsymbol{\Sigma})$	$\frac{L^{Ld}|\mathbf{C}|^{L-d}}{\Gamma_d(L)|\boldsymbol{\Sigma}|^L} \exp\left(-L\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{C})\right)$	$\kappa_1\{\mathbf{C}\} = \ln|\boldsymbol{\Sigma}| + \psi_d^0(L) - d\ln L$ $\kappa_{\nu>1}\{\mathbf{C}\} = \psi_d^{\nu-1}(L)$
$\mathcal{K}_d(L, \boldsymbol{\Sigma}, \alpha)$	$\frac{2|\mathbf{C}|^{L-d}}{\Gamma_d(L)\Gamma(\alpha)|\boldsymbol{\Sigma}|^L}(L\alpha)^{\frac{\alpha+Ld}{2}}\left(\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{C})\right)^{\frac{\alpha-Ld}{2}}$ $\times K_{\alpha-Ld}\left(2\sqrt{L\alpha\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{C})}\right)$	$\kappa_1\{\mathbf{C}\} = \ln|\boldsymbol{\Sigma}| + \psi_d^0(L) + d\left(\psi^0(\alpha) - \ln(\alpha L)\right)$ $\kappa_{\nu>1}\{\mathbf{C}\} = \psi_d^{\nu-1}(L) + d^\nu\psi^{\nu-1}(\alpha)$
$\mathcal{G}_d^0(L, \boldsymbol{\Sigma}, \lambda)$	$\frac{L^{Ld}|\mathbf{C}|^{L-d}}{\Gamma_d(L)|\boldsymbol{\Sigma}|^L}\frac{\Gamma(Ld+\lambda)(\lambda-1)^\lambda}{\Gamma(\lambda)}$ $\left(L\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{C}) + \lambda - 1\right)^{-\lambda-Ld}$	$\kappa_1\{\mathbf{C}\} = \ln|\boldsymbol{\Sigma}| + \psi_d^0(L) + d\left(\ln(\frac{\lambda-1}{L}) - \psi^0(\lambda)\right)$ $\kappa_{\nu>1}\{\mathbf{C}\} = \psi_d^{\nu-1}(L) + (-d)^\nu\psi^{\nu-1}(\lambda)$
$\mathcal{U}_d(L, \boldsymbol{\Sigma}, \alpha, \lambda)$	$\frac{L^{Ld}}{\Gamma_d(L)}\frac{|\mathbf{C}|^{L-d}}{|\boldsymbol{\Sigma}|^L}\frac{\Gamma(\alpha+\lambda)\Gamma(Ld+\lambda)}{\Gamma(\alpha)\Gamma(\lambda)}\left(\frac{\alpha}{\lambda-1}\right)^{Ld}$ $\times U\left(Ld+\lambda, Ld-\alpha+1, L\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{C})(\frac{\alpha}{\lambda-1})\right)$	$\kappa_1\{\mathbf{C}\} = \ln|\boldsymbol{\Sigma}| + \psi_d^0(L) + d\left(\psi^0(\alpha) - \psi^0(\lambda) + \ln(\frac{\lambda-1}{\alpha L})\right)$ $\kappa_{\nu>1}\{\mathbf{C}\} = \psi_d^{\nu-1}(L) + d^\nu\left(\psi^{\nu-1}(\alpha) + (-1)^{\nu-1}\psi^{\nu-1}(\lambda)\right)$

were already using log-cumulants throughout the algorithm. This was initially implemented with the log-cumulants for the compacted measure with sampling distribution,

$$\mathrm{trace}(\boldsymbol{\Sigma}^{-1}\mathbf{C}) \sim \mathcal{U}_d(Ld, d, \alpha, \lambda). \tag{1}$$

This sampling distribution was determined though logical construction from the properties of the $\mathcal{U}_d$ distribution under linear combinations, and has been empirically verified through large sample simulations. Although it may only be asymptotically correct, it is sufficiently accurate for the algorithm's model fitting test. We used the first 3 to 5 log-cumulants in a multi-variate distance test, and either assumed the asymptotic Chi-squared model in [20], or used a Monte-Carlo technique for low sample sizes. Secondly, we have used a simple histogram test based on the multinomial distribution and Pearson's Chi-squared test. We use an irregular, equiprobable partition of the data and compare directly to the PDF. The normalised total squared error is asymptotically Chi-squared distributed as detailed in standard text books, e.g. [21]. This test is quite generic and may be easily implemented for any model with numerical inversion directly from the PDF expressions that are used in the EM-algorithm.

We recommend the Pearson's test, because the log-cumulant distance test shows some loss of sensitivity at both high texture and low sample sizes that probably reflects that the distance measure is only asymptotically Chi-squared.

D. Markov Random Fields

Markov Random Field modelling is a contextual smoothing technique which gives more weight to the class memberships of spatially neighbouring classes. The class label image is modelled as an MRF together with an isotropic second-order neighbourhood system, defining the eight surrounding pixels as the neighbourhood for each site. The class label MRF easily combines with a finite mixture model's spectral clustering, i.e.,

based on the pixel covariance matrix distributions, by replacing the global class prior probabilities with spatially varying local prior probabilities determined from the local neighbourhoods. We introduce the MRF for the class labels, $\mathcal{L}$ with sites $\mathcal{S}$, as a global, isotropic Gibbs distribution, as in [22], with the energy function being proportional to the local neighbourhood mean probability for each class, that is, we implement a mean-field like method [17] and we optimise the global MRF smoothing parameter β with the pseudo-likelihood approach of [23].

Therefore, the k class mixture model for the matrix-variate data at the ith location, $\mathbf{C}^{(i)}$, may be summarised as:

$$P_{\mathbf{C}}(\mathbf{C}^{(i)}) = \sum_{j=1}^k \mathcal{U}_d(\mathbf{C}^{(i)}; L, \boldsymbol{\Sigma}_j, \alpha_j, \lambda_j)\pi_j^{(i)}(\beta, \mathcal{L}) \tag{2}$$

where the local priors for each class, $\pi_j^{(i)}$, are derived from the mean of the current class memberships (posterior probabilities), $\overline{m}_j^{(i)}$, for each class, j, over the neighbourhood of site i, thus

$$\pi_j^{(i)}(\beta, \mathcal{L}) = \frac{\exp\left(\beta\overline{m}_j^{(i)}\right)}{\sum_{l=1}^k \exp\left(\beta\overline{m}_l^{(i)}\right)} \tag{3}$$

The MRF spatial smoothing parameter $\beta > 0$, which is a measure of correlation between neighbouring pixels, is found with the pseudo-likelihood method [23], at each MRF iteration, by maximising

$$\beta = \arg\max_\beta \sum_{i\in\mathcal{S}}\sum_{j=1}^k P(\mathcal{L}^{(i)} = j|\mathbf{C}^{(i)}, \beta, \mathcal{L})\log\pi_j^{(i)}(\beta, \mathcal{L}) \tag{4}$$

III. IMPROVEMENTS

The modifications discussed here keep all the advantages of the state-of-the-art algorithm whilst simplifying the overall

procedure. They cover three main issues: 1) the MRF optimisation under the split-and-merge strategy; 2) computation speed; and 3) test sensitivity. These points are discussed here and demonstrated in the Results section, IV.

As mentioned in the introduction, we previously combined the mixture modelling and MRF smoothing at each iteration of the EM-algorithm and we used both split and merge operations, because we did not wait until full convergence to test the model fit. Both of these choices have had serious consequences in relation to the mixture modelling.

Firstly, the early iterations of the EM-algorithm are a poor representation of the final image and will cause the MRF smoothing parameter, β, optimisation to be severely inappropriate. The β parameter value is also dependent on the number of clusters and hence the splitting and merging would make β inappropriate at the next iteration. The previous β parameter is generally too strong when increasing the number of classes and allowed multiple clusters to persist even though they were competing for one density cluster of samples.

Secondly, testing the model fit before full convergence sometimes split classes based on poor parameter estimates and often ended up with more classes than necessary. This was "fixed" by also testing for competing clusters and merging any that are good-fits to a single combined model. This required estimating the parameters for the merged data, which is computationally slow for the flexible $\mathcal{U}$-distribution model, and testing all pair-wise combinations, which quickly becomes very many individual tests as the number of classes increases. In combination with the effect of overly strong β parameter smoothing, this added a huge computational burden to the algorithm. To avoid these problems, we re-thought the whole process and realised that the mixture modelling can be entirely separated from the MRF smoothing and that the split-and-merge procedure may be faster when simplified to a sequential split-only procedure with full EM-algorithm convergence.

The contextual smoothing only affects the prior probabilities, replacing the global priors with locally derived prior probabilities based upon the neighbourhood, and although the prior probabilities are changing locally, the actual pixel values never change. Therefore, the independent collection of samples, without spatial relations, is still correctly modelled with global priors and a maximum a posteriori mixture model. Thus, the entire adaptive mixture modelling to find the number classes and model parameters can be accomplished without the complications of the MRF. A foreseeable drawback is that very small classes may be lost in the total sample clustering (either not represented in the sub-sampling, or being swamped by larger classes) that may be more distinct on a local level. The option of sub-sampling to improve the processing speed, also at the expense of class distinction, is greatly simplified without the contextual smoothing with its dependence on the spatial neighbours.

We also observed that under certain situations the combined multi-variate tests (i.e., compacted to a 1-dimensional measure using (1)) were not always detecting differences where the classes were only different in some polarimetric channels and were similar in others. This can be understood by the well-fitting dimensions diluting the sensitivity of the poor-fitting ones. We now suggest testing each of the d dimensions individually with a confidence level d times less sensitive to retain the same overall false alarm rate (essentially the naive Bonferroni correction, from standard textbooks). The compaction effect must be non-linear with respect to the confidence level thresholds, because even this d times reduced sensitivity gives better results than the compacted test. This may not be ideal, and does not consider correlation effects (to which we are exploring an whitened version of the testing), but is simple to implement and appears to work. This also simplifies the test, because we can use the class model PDF directly in the Pearson's goodness-of-fit test, albeit in 1-dimension, and don't need to derive the statistical model of the compaction. In addition, this directs us to a simple splitting mechanism, since we know exactly which dimension has the worst fitting model, and we can split the cluster down the mean or median value in that dimension and re-calculate the parameters. The actual splitting implementation is not critical because the next EM-algorithm stage will adapt the parameters and it only needs to give two distinct groups. These could be determined by random assignment, however, we found that splitting in the middle often produces excellent separation at the next iteration.

A. New Strategy

The new recommended strategy, summarised in Fig. 1, is to separate the clustering from the MRF smoothing, fully converge the EM-algorithm at each number of classes, goodness-of-fit test each dimension individually and for each class, and only split the worst class if above the confidence level threshold. Once the automatic and adaptive algorithm has found the number of classes and their class parameters, then the new MRF contextual smoothing stage only requires the maximum likelihood probability value for each class and each pixel, calculated once, but needs to iteratively adapt and spread the MRF field values, care of the Markov property. That is, the class parameters and pixel-wise probability values are fixed, and only the mean-field posterior weights and smoothing parameter are iteratively updated. Starting with the global priors achieves a smooth transition to the MRF local priors.

The full convergence and split-only approach is a much simpler solution and still finds the minimum number of classes that are able to describe the data-set to the given confidence level.

B. Processing Speed

Initially, we gained a great reduction in processing time when we separated the clustering from the MRF, because we could take advantage of sub-sampling in the clustering phase and still get full smoothing. We also believed that the split-and-merge strategy was slower than the split-only, because of excessive merge testing, but we now realise that this was caused by other issues. Particularly, slower parameter estimation and inaccurate log-cumulant tests for extreme texture cases, such as mixtures, that was causing cyclic behaviour (with the same clusters repeatedly splitting and then re-merging). Such deadlock conditions were previously solved
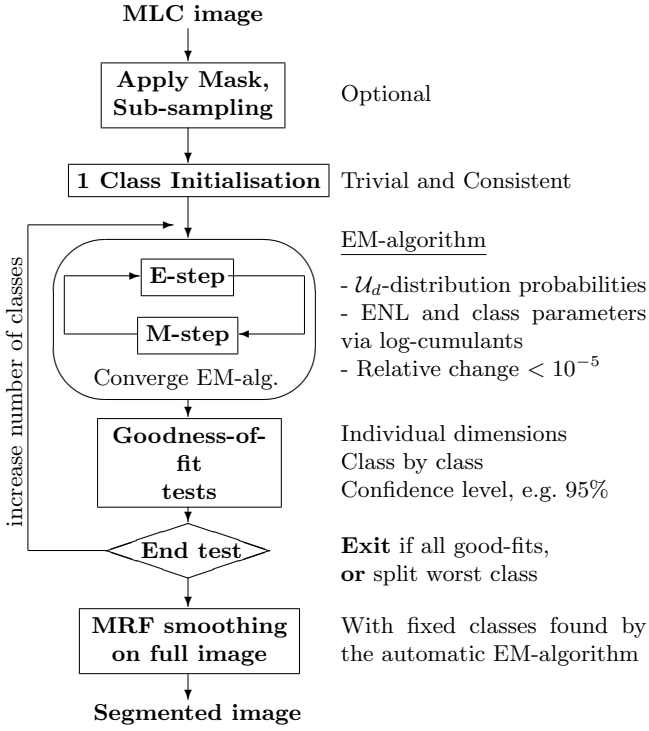
Fig. 1. New Strategy: workflow with comments.

incorrect degree of freedom will still work consistently, but with a slightly different sensitivity or false alarm rate than the chosen confidence level. We will be using the simpler Pearson's test hereafter.

Table II indicates the chosen confidence level versus the observed false alarm rate for the Pearson's test as a demonstration of the goodness-of-fit testing obtaining the specified confidence levels. The result is for simulated data $N = 1000$, $d = 4$, 9-looks, $\alpha = 14.8, \lambda = 52.8$ (approx. texture of a forest), and Σ included complex off-diagonal terms for the co-polar correlation. A wide range of parameter values and dimensions were tested and all showed this same representative behaviour.

TABLE II
CONFIDENCE LEVEL VERSUS MEASURED FALSE ALARM RATE. PEARSON'S TEST WITH 10 BINS AND 8 DEGREES OF FREEDOM, 1000 REPEATS.

Confidence level %	Measured failures %
90.0	9.655
95.0	4.700
99.0	0.915
99.9	0.105

empirically by ramping the confidence level sensitivity as the iterations proceeded, but it is much better to avoid this situation altogether.

After fixing these problems, and with equivalent code for all aspects except the split-and-merge strategy, i.e. the same parameter estimation and the same Pearson's testing on the compacted measure, then the timing was roughly equivalent between the split-and-merge and split-only strategies and depends mostly on the image complexity. Specifically, how many classes are involved and how well separated or overlapped the clusters are. We did, however, observe subtle differences in the segmented results depending on the strategy. We find the sequential, split-only approach slightly more sensitive and has a cleaner interpretation with less concern about the β optimisation.

IV. RESULTS

We demonstrate the characteristics specifically related to the new algorithm with both simulated data, for validation, and with real Pol-SAR images.

A. Confidence Level Check

We found that the Pearson's test with 10 to 20 bins gave very good sensitivity using a chi-squared with the number of bins minus two degrees of freedom. If using the log-cumulant based test, we found that the degrees of freedom is a little more complicated and depends on the number of cumulants used in both the testing and in the estimation (constraints), as well as the dimensions. However, using an

B. Test Individual Dimensions

By chance, one of our test images for Wishart data classes had two clusters that were difficult to distinguish and at higher sub-sampling factors were considered a good single class. The polarimetric colours look distinctly different and we therefore tried to improve the sensitivity of the testing by looking at individual dimensions. Fig. 2 shows the segmented images and the class histogram for both the compacted scalar test, i.e., the trace statistic in (1), and the individual dimension tests. Note how the pink class in the "Compacted" image is correctly separated into the pink and grey classes in the "Individual" image. The "Compacted" histogram was considered a good fit at 99%, whilst the "Individual" histogram shows significant variation in some dimensions that were subsequently split by the algorithm. Note that the individual dimension test uses a confidence level that is d times less sensitive, but still manages to be better than the single compacted test statistic. Since this is with the same number of samples, the individual dimension test is considered more sensitive (See also Fig. 4). The compacted test did manage to separate the classes when using a lower sub-sampling factor, because the greater number of samples were better able to distinguish the mixture at the given confidence level.

C. Confidence level and sub-sampling effects

The number of clusters that are considered significant depends on the chosen confidence level and the actual separation between the data points. This may be used to simplify the image clustering result by decreasing the sensitivity, for example changing from 95% to 99% or higher, and will likely find fewer classes (and fewer false alarms) since each class test would be more tolerant of random variation. Extremely poor fits will still be separated and hence the major class
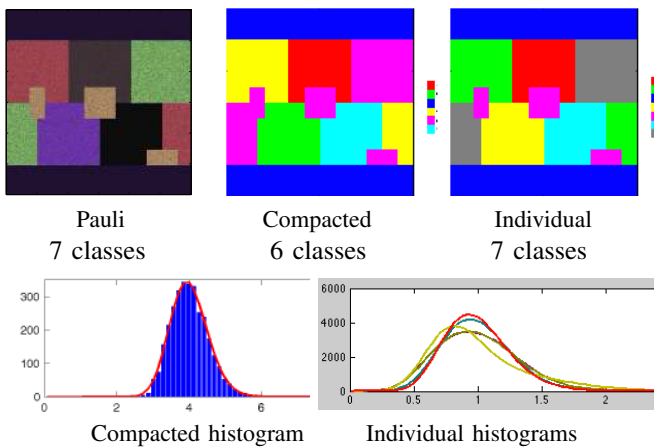
Fig. 2. Compacted test versus individual dimension tests. Note how the pink class in the "Compacted" image is correctly separated into the pink and grey classes in the "Individual" image. The "Compacted" histogram was considered a good fit (to the model density, red line) at 99%, whilst the "Individual" histogram shows significant variation in some dimensions (coloured lines, model in red), even when reduced to 99.75% confidence.

divisions are still detected, but less distinct clusters (i.e., sub-class divisions) are grouped together. A simple example of this effect may be seen in the first two segmentation images, (b) and (c), of Fig. 3, where 14 classes were found at 95% and 11 classes at 99.99%. The image is taken from the EMISAR sample scene over Foulum, Denmark, 1998, L-band, multi-looked by 18, 150×300 pixels. Image (a) shows the Pauli polarimetric decomposition for reference, where different colours represent different polarimetric properties, and depicts agricultural fields, forest and buildings.

An alternative method to simplify the segmentation is by increasing the sub-sampling factor, which reduces the number of samples being tested, and hence lowers the sensitivity of the test. Fewer samples also dramatically speeds up the algorithm, because far fewer computations are needed during the slow iterative clustering stage. This effect may be seen in the top and bottom segmentations, (b) and (d), of Fig. 3 and seems to have a more dramatic effect than changing the confidence level. Both images use 95% confidence, but sub-sampling by 25 obtained 14 classes, and sub-sampling by 100 found 8 classes. The main class distinction between forest (whiter regions in Pauli) and fields (darker, redder regions in Pauli), and the most different fields (i.e., different colouring) are consistent in both of these two approaches to simplifying the image segmentation.

Table III gives an indication of the processing times for the Foulum data sub-set. The intention is to see the benefit of sub-sampling compared to the full resolution analysis. The tests were run on a 2.66GHz Intel Core i7 Apple laptop with 4GB memory using MATLAB [24], and have not been fully optimised. The old approach used to take hours to days, depending on the size of the scene, but there have been too many other coding improvements for a fair comparison of the underlying strategy.
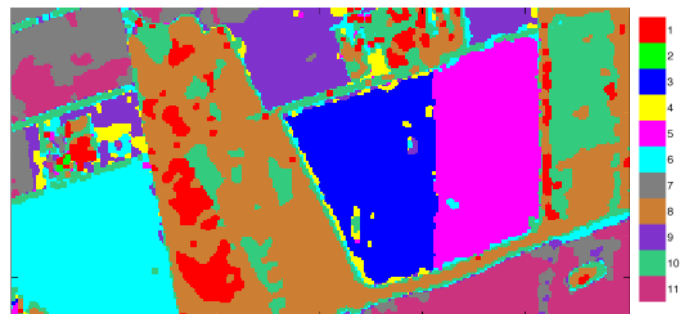
Note that the huge increase in both time and number of classes for sub-sampling level 1 probably indicates a problem
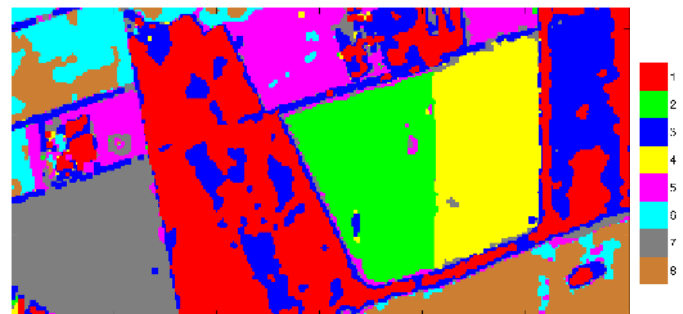

(a) Pauli image for reference.
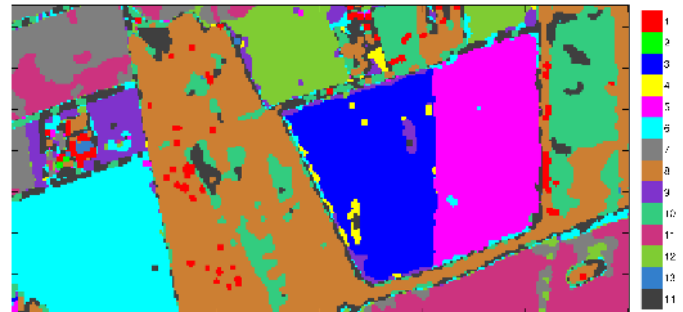
(b) 95% confidence, sub-sampling by 25, 14-classes.

(c) 99.99% confidence, sub-sampling by 25, 11-classes.


(d) 95% confidence, sub-sampling by 100, 8-classes.

Fig. 3. Lowering the sensitivity finds fewer classes, but retains the major class divisions. EMISAR image from Foulum, Denmark, 1998.

TABLE III
SUB-SAMPLING TESTS FOR REAL "FOULUM" IMAGE EXTRACT. NEW
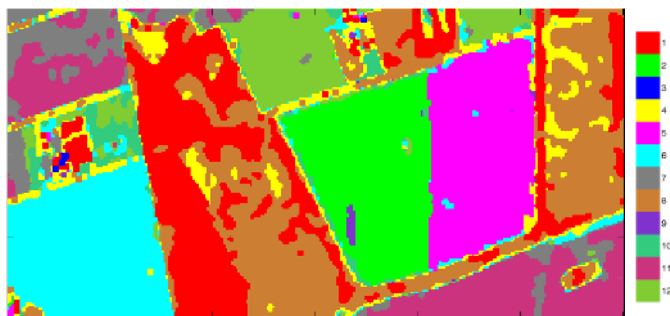SPLIT-ONLY STRATEGY, INDIVIDUAL DIMENSION TESTS, 99%.

sub-sampling	pixels	classes	time (including final MRF)
1	45000	> 100	over 1 day
4	11250	29	3 hr. 18 min. 17 sec.
16	2850	14	36 min. 32 sec.
49	946	12	32 min. 50 sec.
100	450	7	12 min. 22 sec.
400	120	5	3 min. 21 sec.
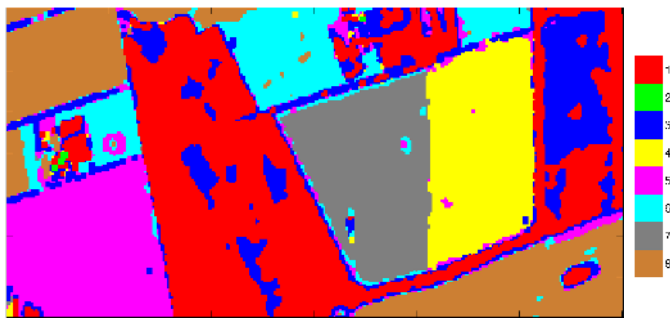
(a) Old strategy, split-and-merge, internal MRF, compacted tests.
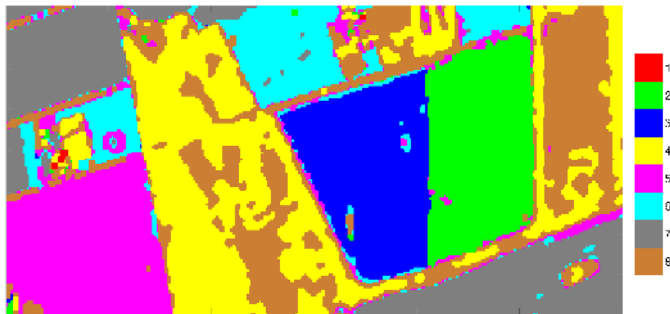
(b) New strategy, split-only, final MRF, compacted tests.

(c) New strategy, split-only, final MRF, individual dimension tests.

Fig. 4. Results for different strategies. Main class boundaries are similar, and all are smooth. The individual dimension test is more sensitive and distinguishes more classes. EMISAR image from Foulum, Denmark, 1998.
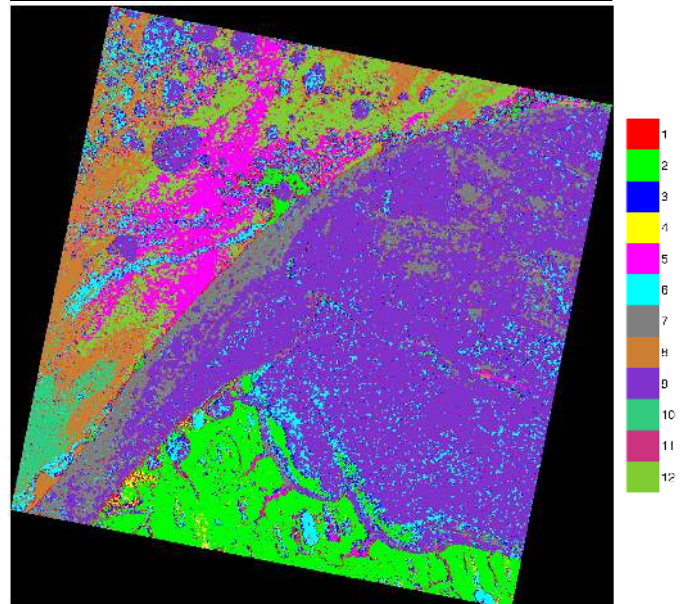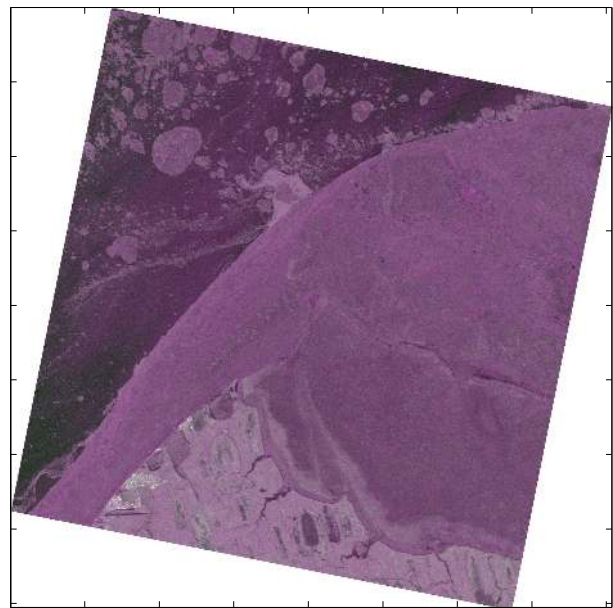

Fig. 5. Sea ice around Point Barrow, Alaska. Pauli image (top) and automatic segmentation into 12 classes (bottom). Radarsat-2, C-band, June 2011, 16-look averaging, sub-sampling by 256, 99% confidence level.

and was therefore cancelled before it finished. The goodness-of-fit tests become very sensitive when the sample sizes are very large. Imprecise parameter estimation, particularly the ENL parameter, or an incorrect model distribution could both make the test fail under high sensitivity, even though they are acceptable for a lower number of samples. Thus, there may be extra benefits of using the sub-sampling, or setting a maximum number of samples to test per class, but an optimum level is probably application and data dependent.

D. Different Strategies

Some differences between the old split-and-merge approach and the new split-only approach are shown in Fig. 4. All are from the Foulum data-set, 18-look averaging, and sub-sampled by 49. The segmentation with the old, split-and-merge strategy and compacted tests found 8 major classes in (a), and is very similar to the new split-only strategy results under the same test conditions in (b). Segmentation (c) shows the result for the new strategy when testing individual dimensions, which are

more sensitive, and found 12 classes. In all cases, the major class boundaries are similar, the regions are smooth and solid, and they can be seen to correspond with different coloured regions in the Pauli image in Fig. 3 (a). Note that the boundary between sparse forest and the dense forest is different in Fig. 4 (a) (red & blue classes) than in (b) (yellow & brown classes), and must reflect the influence of the MRF strategy, because all other aspects of the algorithm were the same. However, given that both scenes have a constrained number of classes due to sub-sampling, it is difficult to state which is best.

E. More Real Data Examples

We include two very different real image examples just to demonstrate the flexible modelling and simplified segmented

images that may be generally used as an initial step for further studies.

Fig. 5 shows an area of sea ice around Point Barrow, Alaska, from a Radarsat-2, C-band, 16-look scene from 2011. It clearly separates the darkest regions of open water (several classes due to varying wind conditions) from the ice floes, the land-fast ice and the land. The research interest here is in the few classes in the land-fast ice region, which is of interest to the local community who travel and hunt on the ice. This segmentation may be used for further analysis of the polarimetric properties to identify the type of ice in the regions.

Fig. 6 shows San Francisco city from the Radarsat-2 sample scene, 25-looks, from 2008. Although 23 classes is a little complicated, the different coloured segments do appear to correlate with different coloured regions in the Pauli image and has quite fine distinction at this sub-sampling level. Note that urban areas have the most varied textures, which are satisfactorily clustered with this flexible model. Besides capturing these visible regions well, this segmentation also demonstrates that the rotated buildings and the forested vegetation are distinguished in the polarimetric data. Urban analysts may subsequently explore the specific properties or features of the polarimetry that best separate these regions.

V. Conclusions

We have proposed an improved strategy for our advanced non-Gaussian clustering algorithm which maintains the benefits of the flexible $\mathcal{U}_d$-distribution for multi-look covariance matrix data classes, Markov random fields for contextual smoothing, and goodness-of-fit testing to optimise the number of clusters. The new strategy simplified the logic of the algorithm, making it more robust, improves the sensitivity, and often leads to faster results.

This paper demonstrated that the goodness-of-fit testing is statistically correct by measuring the expected failure rate, gave an example to demonstrate the improved sensitivity by testing each dimension separately, demonstrated how changing the sensitivity and sub-sampling affects the number of classes and can be used to simplify the segmentation result, gave an example of the subtle differences that the choice of strategy may produce, and included further examples for real PolSAR images to demonstrate the generality of the algorithm for different scene types and that the optimised MRF smoothing level seems appropriate.

Visual inspection indicates that it achieves good results that appear valid for real data images and is fully automatic. Subsequent interpretation and identification of the segments may be accomplished with ground truth information, or by exploring the polarimetric class parameters.
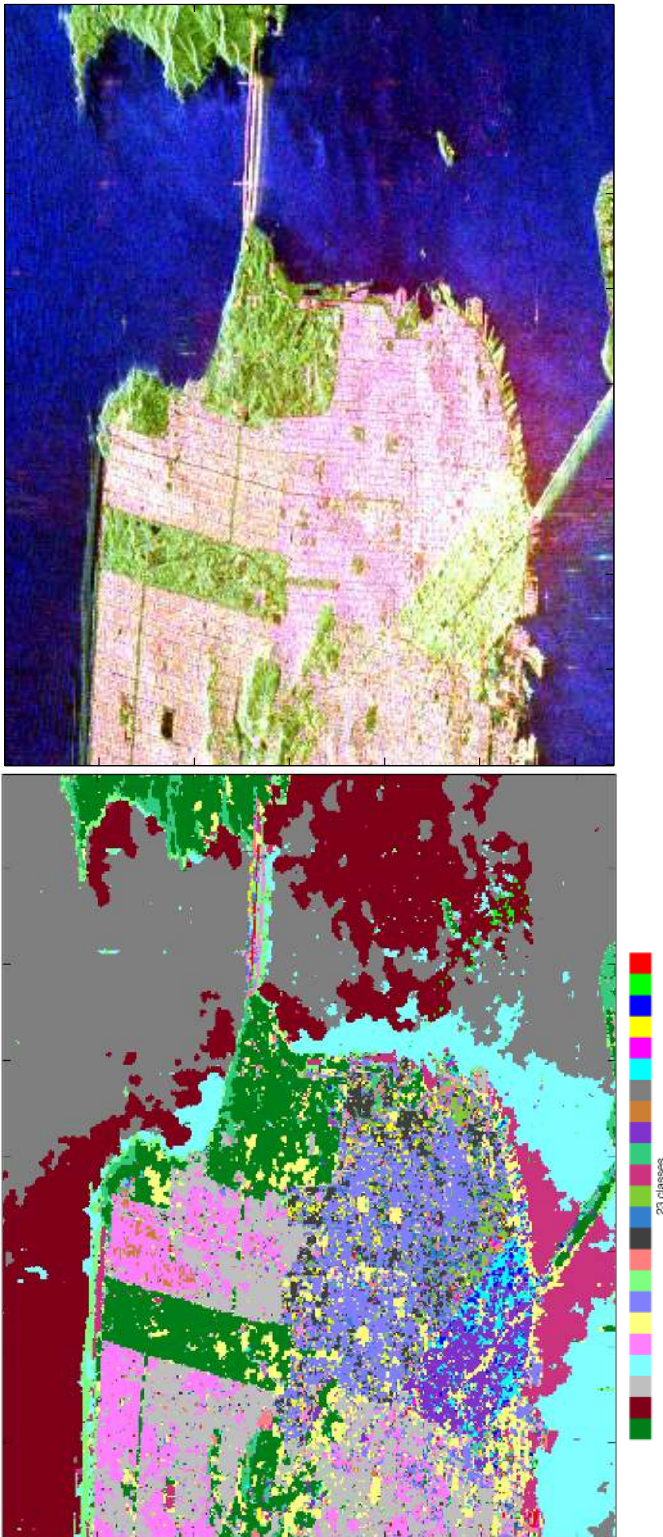
Fig. 6. San Francisco city. Pauli image (top) and automatic segmentation into 23 classes (bottom). Radarsat-2, C-band, June 2008, 25-look averaging, sub-sampling by 25, 99% confidence level.

REFERENCES

[1] A. P. Doulgeris, V. Akbari, and T. Eltoft, "Automatic PolSAR segmentation with the U-distribution and Markov random fields." in *9th European Conference on Synthetic Aperture Radar (EUSAR2012)*, Nuremberg, Germany, April 23-26 2012.

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

[3] N. Goodman, "Statistical analysis based on certain multivariate complex gaussian distribution," in *Ann. Math. Statist.*, vol. 34, 1963, pp. 152–177.

[4] 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*, vol. 49, no. 6, pp. 2281–2295, June 2011.

[5] 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.*, vol. 4, Pasedena, USA, August 1994, pp. 2179–2181.

[6] A. Doulgeris and T. Eltoft, "Scale mixture of Gaussians modelling of polarimetric SAR data," in *International POLinSAR Workshop (POLinSAR2007)*, Frascati, Italy, January 22-26 2007.

[7] C. C. Freitas, A. C. Frery, and A. H. Correia, "The polarimetric G distribution for SAR data analysis," *Environmetrics*, vol. 16, no. 1, pp. 13–31, 2005.

[8] M. Horta, N. Mascarenhas, A. Frery, and A. Levada, "Clustering of fully polarimetric sar data using finite gp0 mixture model and sem algorithm," in *Systems, Signals and Image Processing, 2008. IWSSIP 2008. 15th International Conference on*, June 2008, pp. 81 –84.

[9] L. Bombrun, G. Vasile, M. Gay, and F. Totir, "Hierarchical segmentation of polarimetric sar images using heterogeneous clutter models," *Geoscience and Remote Sensing, IEEE Transactions on*, pp. 1 –12, 2010.

[10] L. Bombrun and J.-M. Beaulieu, "Fisher distribution for texture modeling of polarimetric sar data," *Geoscience and Remote Sensing Letters, IEEE*, vol. 5, no. 3, pp. 512 –516, Jul. 2008.

[11] V. Akbari, A. Doulgeris, and T. Eltoft, "Monitoring glacier changes using multitemporal multipolarization sar images," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. PP, no. 99, pp. 1–13, 2013 (IEEE early access).

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

[13] A. P. Doulgeris and T. Eltoft, "Automated Non-Gaussian clustering of polarimetric SAR." in *8th European Conference on Synthetic Aperture Radar (EUSAR2010)*, Aachen, Germany, June 7-10 2010.

[14] A. P. Doulgeris, S. N. Anfinsen, and T. Eltoft, "Automated non-Gaussian clustering of polarimetric synthetic aperture radar images," *IEEE Trans. Geoscience and Remote Sensing*, vol. 49, no. 10, pp. 3665–3676, 2011.

[15] A. P. Doulgeris, "A simple approach to choosing the number of classes in finite mixture modelling," *IEEE Signal Processing Letters*, p. pp. 4, 2013 (submitted).

[16] V. Akbari, A. Doulgeris, G. Moser, T. Eltoft, S. Anfinsen, and S. Serpico, "A textural–contextual model for unsupervised segmentation of multi-polarization synthetic aperture radar images," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 51, no. 4, pp. 2442–2453, 2013.

[17] G. Celeux, F. Forbes, and N. Peyrand, "Em procedures using mean field-like approximations for markov model-based image segmentation," *Pattern Recognition*, vol. 36, no. 1, pp. 131–144, 2003.

[18] P. Formont, F. Pascal, G. Vasile, J. Ovarlez, and L. Ferro-Famil, "Statistical classification for heterogeneous polarimetric sar images," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 5, no. 3, pp. 567–576, June 2011.

[19] D. Tao, S. Anfinsen, and C. Brekke, "A comparative study of sea clutter covariance matrix estimators," *Geoscience and Remote Sensing Letters, IEEE*, vol. 11, no. 5, pp. 1010–1014, May 2014.

[20] 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*, vol. 49, no. 8, August 2011, 18 pp.

[21] R. Larsen and M. Marx, *An Introduction to Mathematical Statistics and Its Applications*, 5th ed. Prentice-Hall, Incorporated, 2011.

[22] S. Li, *Markov random field modeling in image analysis*. Springer, 2009.

[23] J. Besag, "Efficiency of pseudo-likelihood estimation for simple gaussian fields," *Biometrika*, vol. 64, pp. 616– 618, 1977.

[24] (2013). [Online]. Available: mathworks.com

Anthony P. Doulgeris (S06M12) received the B.Sc. degree in physics from The Australian National University, Canberra, Australia, in 1988, the M.Sc. degree and the Ph.D. degree in physics from the Department of Physics and Technology, University of Tromsø, Tromsø, Norway, in 2006, and 2011, respectively.

He joined the Department of Physics and Technology at the University of Tromsø- The Arctic University of Norway, in 2007 and is currently an associate professor in applied Earth observation. His research interests focus on investigating remote sensing, pattern recognition and multidimensional statistical modelling, in particular with polarimetric synthetic aperture radar images.