

Research Article

Scale Mixture of Gaussian Modelling of Polarimetric SAR Data

Anthony P. Doulgeris and Torbjørn Eltoft

The Department of Physics and Technology, University of Tromsø, 9037 Tromsø, Norway

Correspondence should be addressed to Anthony P. Doulgeris, anthony.p.doulgeris@uit.no

Received 1 June 2009; Accepted 28 September 2009

Academic Editor: Carlos Lopez-Martinez

Copyright © 2010 A. P. Doulgeris and T. Eltoft. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper describes a flexible non-Gaussian statistical method used to model polarimetric synthetic aperture radar (POLSAR) data. We outline the theoretical basis of the well-known *product model* as described by the class of Scale Mixture models and discuss their appropriateness for modelling radar data. The statistical distributions of several Scale mixture models are then described, including the commonly used Gaussian model, and techniques for model parameter estimation are given. Real data evaluations are made using airborne fully polarimetric SAR studies for several distinct land cover types. Generic scale mixture of Gaussian features is extracted from the model parameters and a simple clustering example presented.

1. Introduction

It is well known that POLSAR data can be non-Gaussian in nature and that various non-Gaussian models have been used to fit SAR images—firstly with single channel amplitude distributions [1–3] and later extended into the polarimetric realm where the multivariate K-distributions [4, 5] and G-distributions [6] have been successful. These polarimetric models are derived as stochastic *product models* [7, 8] of a non-Gaussian *texture* term and a multivariate Gaussian-based *speckle* term, and can be described by the class of models known as Scale Mixture of Gaussian (SMoG) models. The assumed distribution of the texture term gives rise to different product distributions and the parameters used to describe them.

In this paper we only investigate the semisymmetric zero-mean case, which is expected for scattering in the natural terrain, and the more general scale mixture model includes a skewness term to account for a dominant or coherent scatterer and a mean value vector. Extension to the non-symmetric case or expanding to a multitextural/nonscalar product will be addressed in the future. It is worth noting that these methods are general multivariate statistical techniques for covariate product model analysis and can be generally applied to single, dual, quad, and combined (stacked) dual frequency SAR images, or any type of coherent

imaging system. The significance and interpretation of the parameters, however, may be different in each case.

The scale mixture models essentially describe the probability density function giving rise to the measured complex scattering coefficients. They therefore model at the scattering vector level, that is, Single-Look Complex (SLC) data sets, which contain 4-dimensional complex values. These complex vectors represent both magnitude and phase for the four combinations of both transmitted and received signals for both horizontal and vertical polarisation. Statistical modelling is achieved by looking at a small neighbourhood of pixels around each point and the model parameters are estimated from this collection of data vectors. Parameter estimation, particularly of higher-order statistical terms, is improved by using a larger neighbourhood size, but at the expense of image resolution and the introduction of class mixture effects at the boundaries. So a compromise must be made between a small neighbourhood to avoid mixtures and blurring and a large neighbourhood to improve parameter estimation.

The model fitting procedure generates the model parameters at each image pixel location which gives rise to a new feature space description of the image and can be used for subsequent classification or image interpretation. Although many different models have been used to describe non-Gaussian data, with quite different orders of complexity

and parametric descriptions, the parameters are usually estimated from measurable sample moments. Since the parameters are simply nonlinear relations of measured moments, one can say that the moments themselves represent the rawest form, and additionally they are independent of the particular model in question. We therefore see two quite different avenues to take regarding analysis: firstly, one can choose a specific non-Gaussian model with an explicit probability density function (pdf) and use Bayesian statistical techniques to analyse the data, or alternatively, one can extract general scale mixture of Gaussian features (that are independent of any explicit model pdf) and work solely in a two-moment generic SMOG feature space. In this sense, the Gaussian-based analysis is a single moment method.

Speckle variation may be reduced by multilook averaging, either in the frequency domain during processing or in the spatial domain postimaging, and produces Multilook Complex (MLC) matrix data. Such multi-look averaging modifies the intensity distribution of the data and subsequent statistical modelling must take this into account for parameter estimation or statistical inference. The multi-looked matrix-variate distribution derived from purely Gaussian data is the complex Wishart distribution [9] and for the Scale Mixture case is the generalised Wishart distribution, for example, the K-Wishart [10]. Statistical clustering using these multi-look matrix-variate models has been demonstrated elsewhere [6, 10, 11], and here we only describe multi-look data for model parameter estimation.

The plan for this paper is to describe the modelling in Section 2, with general properties and suitability discussed in Section 3. Intercomparison and parametric feature results are shown for several data sets in Section 4, followed by our conclusions in Section 5.

We denote scalar values by either lower or upper case standard weight characters, vectors as lower case bold characters and matrices as bold uppercase characters. For simplicity, we have not distinguished between random variables and instances of random variables, as such can be ascertained through context.

2. Scale Mixture of Gaussian Scheme

The Scale Mixture of Gaussian models, also known as normal-variance mixtures [12, 13], are a statistical product model with a texture random variable times a speckle random variable. The pure speckle term has a standard complex multivariate Gaussian distribution and the texture term has any positive only scalar distribution. Since the textural random variable models the variance of the signal rather than its amplitude, it is introduced as a square root term in the data vector (described in [8]).

Mathematically, we model the vector of polarimetric scattering coefficients (\mathbf{y}) under the multidimensional SMOG scheme as

$$\mathbf{y} = \boldsymbol{\mu} + \sqrt{z}\boldsymbol{\Gamma}^{1/2} \mathbf{x}, \quad (1)$$

where $\boldsymbol{\mu}$ is the mean vector, the scale parameter z is a strictly positive random variable (scalar), $\boldsymbol{\Gamma}$ is the internal covariance structure matrix, normalised such that the

determinant $|\boldsymbol{\Gamma}| = 1$, and \mathbf{x} is a standardised, complex multivariate Gaussian variable with zero mean and identity covariance matrix, that is, $\mathbf{x} \sim \mathcal{N}_c(\mathbf{0}, \mathbf{I})$. We will hereafter assume that $\boldsymbol{\mu} = \mathbf{0}$. This assumption is well justified for natural environments (i.e., distributed targets without dominant coherent scatterers), where the complex values of \mathbf{y} are theoretically expected to be, and generally are, zero mean. Theoretically, this is the case of distributed coherent imaging where the resolution cell size and roughness are large relative to the illuminating wavelength, leading to the absolute phase variation over all scatterers in the cell being uniformly randomly distributed and the integrated in-phase and quadrature signals are therefore expected to be zero. We have chosen to normalise the covariance structure matrix instead of the scale parameter in our work, because of the analogy between the average scale, $E\{z\}$, and the radar cross section, σ , of 1-dimensional data (also described in [8]), even though this interpretation is not straight forward for multidimensional data.

This scheme describes different parametric families of distributions, depending on the scale parameter probability density function, $f_z(z)$. Given the pdf for the scale parameter, the marginal pdf for \mathbf{y} can be obtained by integrating the conditional pdf of $\mathbf{y} | z$, which is multivariate Gaussian, over the density of z . That is,

$$f_{\mathbf{y}}(\mathbf{y}) = \int_0^\infty f_{\mathbf{y}|z}(\mathbf{y} | z) f_z(z) dz. \quad (2)$$

Four scale mixture models, derived with closed form expressions in [14], are depicted in Table 1, including the Multivariate Gaussian (MG) distribution as a special case. All are heavy-tailed (sparse) and symmetric distributions, with a global shape for all dimensions, but an allowable width variation described by the covariance structure matrix. Both the MG and multivariate Laplacian (ML) distributions have fixed shapes and only vary with width parameter. The two-parameter Multivariate K-distribution (MK) and multivariate normal inverse Gaussian (MNIG) distributions describe a range of shapes as well as widths, both including the MG as a limiting shape. See Figure 1 for an example of these shapes. Both the MK and the MNIG distribution have theoretical links to the nature of distributed target scattering, as they can be derived from Brownian motion models [1, 15]. Many other models have been investigated in literature with some authors advocating the three parameter Generalised Inverse Gaussian (GIG) for the scale parameter, since it has the Gamma, Exponential, and Dirac delta distributions as subcases and is even more flexible to fit to real data. Note that having more parameters requires more complicated estimation expressions, and higher-order moment estimators are known to have higher variance. Considering the limited sample sizes used in the modelling, the benefit of such complicated modelling may not be significant. The article with the three-parameter GIG model is subsequently simplified to two parameter special cases, which proved flexible enough for modelling real data variations. We all agree that more flexibility than a purely Gaussian analysis is sometimes required.

TABLE 1: Scale mixture of Gaussian models.

z distribution	Multivariate scale mixture distribution
Constant (Dirac delta) ($z; \sigma^2$)	Gaussian, $\mathbf{MG}(\mathbf{y}; \sigma^2, \boldsymbol{\mu}, \boldsymbol{\Gamma})$
Exponential ($z; \lambda$) = $\frac{1}{\lambda} \exp\left(-\frac{z}{\lambda}\right)$	Laplacian, $\mathbf{ML}(\mathbf{y}; \lambda, \boldsymbol{\mu}, \boldsymbol{\Gamma}) = \frac{1}{\pi^d} \frac{2}{\lambda} \frac{K_{d-1}(2\sqrt{q(\mathbf{y})/\lambda})}{(\sqrt{\lambda q(\mathbf{y})})^{d-1}}$
Gamma ($z; \alpha, \mu_z$) = $\left(\frac{\alpha}{\mu_z}\right)^\alpha \frac{z^{\alpha-1}}{\Gamma(\alpha)} \exp\left(-\frac{\alpha}{\mu_z} z\right)$	K-distribution, $\mathbf{MK}(\mathbf{y}; \alpha, \mu_z, \boldsymbol{\mu}, \boldsymbol{\Gamma})$ $= \frac{2}{\pi^d} \frac{1}{\Gamma(\alpha)} \left(\frac{\alpha}{\mu_z}\right)^{(\alpha+d)/2} (q(\mathbf{y}))^{(\alpha-d)/2} K_{\alpha-d}\left(2\sqrt{\frac{\alpha}{\mu_z} q(\mathbf{y})}\right)$
Inverse Gaussian ($z; \delta, \gamma$) $= \frac{\delta}{\sqrt{2\pi}} e^{\delta\gamma} z^{-3/2} \exp\left(-\frac{1}{2}\left(\frac{\delta^2}{z} + \gamma^2 z\right)\right)$	Normal Inverse Gaussian, $\mathbf{MNIG}(\mathbf{y}; \delta, \gamma, \boldsymbol{\mu}, \boldsymbol{\Gamma})$ $= \sqrt{2\delta} e^{\delta\gamma} \left(\frac{\gamma}{\pi\sqrt{\delta^2 + 2q(\mathbf{y})}}\right)^{d+(1/2)} K_{d+(1/2)}(\gamma\sqrt{\delta^2 + 2q(\mathbf{y})})$

$q(\mathbf{y}) = (\mathbf{y} - \boldsymbol{\mu})^T \boldsymbol{\Gamma}^{-1} (\mathbf{y} - \boldsymbol{\mu})$ is the scaled squared Mahalanobis distance from the mean, with $\boldsymbol{\mu} = \mathbf{0}$ for the PolSAR case.
 $K_m(x)$ is a modified Bessel function of the second kind with order m .

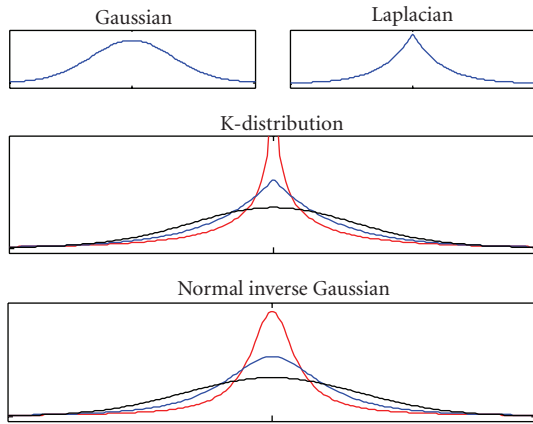


FIGURE 1: Example shapes for each model distribution, fixed width. The two parameter models, the K-distribution and the normal inverse Gaussian, can vary in shape.

Given such a general scheme as in (1), it can be readily shown that

$$\mathbf{E}\{\mathbf{y}\} = \boldsymbol{\mu} \quad (= \mathbf{0}), \quad (3)$$

$$\mathbf{E}\{(\mathbf{y} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu})^H\} = \mathbf{E}\{z\}\boldsymbol{\Gamma} \quad (= \boldsymbol{\Sigma}), \quad (4)$$

$$\mathbf{E}\left\{\left[(\mathbf{y} - \boldsymbol{\mu})^H \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu})\right]^2\right\} = \frac{\mathbf{E}\{z^2\}}{[\mathbf{E}\{z\}]^2} d(d+1), \quad (5)$$

where $(\cdot)^H$ means (Hermitian) conjugate transpose, $\mathbf{E}\{\cdot\}$ is the expectation operator, and d is the dimension (which will be 4 for PolSAR data). These equations can be used to estimate the various parameters: obtaining $\boldsymbol{\Gamma}$ and $\mathbf{E}\{z\}$ from

the sample covariance via (4) plus the normalisation $|\boldsymbol{\Gamma}| = 1$, and the second moment $\mathbf{E}\{z^2\}$ from the sample multivariate Kurtosis via (5) [16]. It is mathematically convenient to define a scale invariant measure of non-Gaussianity, the Relative Kurtosis (RK), as Mardia's multivariate kurtosis of the sample divided by $d(d+1)$. Mardia's multivariate kurtosis is scale invariant due to the $\boldsymbol{\Sigma}^{-1}$ in (5) and is relative to the complex Gaussian distribution value of $d(d+1)$. This equates to $\mathbf{E}\{z^2\}/[\mathbf{E}\{z\}]^2$ for our texture random variable, as is easily found from (5). The particular parametric form of the distribution of z is then obtained by the solution for its first and second moments given the estimates obtained for $\bar{z} = \mathbf{E}\{z\}$ and RK from (4) and (5). The solutions are shown in Table 2, but note that some smart numerical exceptions may need to be made, for example, when the sample kurtosis is less than the $d(d+1)$ due to sample variation.

In the general case, all the model parameters are free to be optimised in the fitting procedures. However, if we have some *a priori* knowledge about a parameter's value, we would expect a better model fit by actually constraining it. The most obvious constraint in our radar data is the expected zero-mean and we have further zero and pair value constraints on the covariance structure matrix. Simulated studies show that applying the constraints has a great improvement in the estimated parameters, particularly when the sample size is small, with the covariance constraints being the most significant. Figure 2 depicts an estimate of $\boldsymbol{\Gamma}_{11}$ versus sample size with and without applying either the zero-mean or Gamma matrix constraints. However, in this paper the modelling has been left free because the sample sizes are large enough that the mean is very nearly zero and the covariance constraints are approximately met anyway. The examples are all analysed with a $13 \times 13 = 169$ window size.

Estimation in the case of L-look MLC data is based upon the neighbourhood mean of the matrix-variate data, plus the variance of a mean squared Mahalanobis measure (M)

TABLE 2: Moment expressions and parameter solutions for each model.

Model	z distribution	$\mathbf{E}\{z\}$	$\mathbf{E}\{z^2\}$	Solution given \bar{z} and $\text{RK} = \frac{\bar{z}^2}{\bar{z}^2}$
MG	Constant (σ^2)	σ^2	not needed	$\hat{\sigma}^2 = \bar{z}$
ML	Exponential (λ)	λ	not needed	$\hat{\lambda} = \bar{z}$
MK	Gamma (α, μ_z)	μ_z	$\frac{(\alpha+1)}{\alpha} \mu_z^2$	$\hat{\mu}_z = \bar{z}$ $\hat{\alpha} = \frac{1}{(\bar{z}^2/\bar{z}^2) - 1}$
MNIG	Inverse Gaussian (δ, γ)	$\frac{\delta}{\gamma}$	$\left(1 + \frac{1}{\delta\gamma}\right) \frac{\delta^2}{\gamma^2}$	$\hat{\delta} = \sqrt{\frac{\bar{z}}{(\bar{z}^2/\bar{z}^2) - 1}}$ $\hat{\gamma} = \frac{\hat{\delta}}{\bar{z}}$

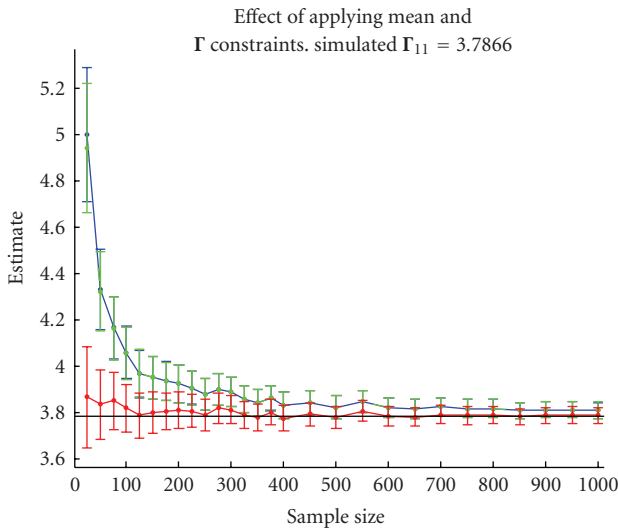


FIGURE 2: Effect of constraints on Γ_{11} estimate versus sample size: Unconstrained in blue, zero-mean constraint in green, and covariance constraints in red. Note that the green points completely overprint the blue above size 100.

which is equivalent to $\text{trace}(\Sigma^{-1}C)$. Assuming that $\mu = \mathbf{0}$ for simplicity, it is easily shown that

$$C = \frac{1}{L} \sum_{l=1}^L \mathbf{y} \mathbf{y}^H, \quad (6)$$

$$\mathbf{E}\{C\} = \mathbf{E}\{z\} \Gamma \quad (= \Sigma), \quad (7)$$

$$\begin{aligned} M &= \text{tr}(\Sigma^{-1}C) \\ &= \frac{1}{L} \sum_{l=1}^L (\mathbf{y}_l^H \Sigma^{-1} \mathbf{y}_l), \end{aligned} \quad (8)$$

$$\mathbf{E}\{M\} = d, \quad (9)$$

$$\mathbf{E}\{(M-d)^2\} = \frac{1}{L} \frac{\mathbf{E}\{z^2\}}{[\mathbf{E}\{z\}]^2} d(d+1) - \frac{1}{L} d^2. \quad (10)$$

Note that the expectation of M equals d because of the normalisation with respect to each local covariance matrix Σ in $\text{trace}(\Sigma^{-1}C)$. The parameters Γ and $\mathbf{E}\{z\}$ are obtained from the mean matrix by applying the constraint that $|\Gamma| = 1$, and RK is obtained in terms of $\text{var}(M)$ by rearranging (10). Subsequently, the texture parameters are solved for as in Table 2.

3. Properties and Suitability

All models are symmetric about the mean and although each dimension may have different relative widths, distributed by the covariance matrix Γ , they will each have a similar (global) shape governed by the scalar parameters. All models are also sparse distributions, meaning that they are more pointed in the peak and heavier tailed than the Gaussian. The MG and ML distributions have a fixed shape and the scalar parameter varies the width. The MK's and MNIG's two scalar parameters lead to a range of shapes as well as overall width. The shapes range from more pointed than Laplacian, through to rounded like the Gaussian (see Figure 1). The effect of the shape parameter on the density function is highly nonlinear with value, with the clearly visible variation occurring for small parameter values (e.g., $\alpha < 10$ for the K-distribution) and converging rapidly towards the Gaussian in shape from only moderate values (e.g., $\alpha > 15$) up to infinity. Also note that both the ML and MK distribution's pdfs can go to infinity at the mean value, whereas the MNIG always has a finite peak.

If we take our assumption of scale mixture of Gaussians modelling and our theoretical radar scattering as a vector sum with uniformly random phase, then three main properties emerge: zero-mean, semisymmetric shape, and global shape. It seemed appropriate to investigate whether the real PolSAR data showed similar general features as a validation for using such a mixture model.

Figure 3 shows three different sets of real PolSAR data distributions depicted as marginal Parzen estimates for both the real and imaginary parts of all 4 dimensions of the data vector. The intention is to observe the general features

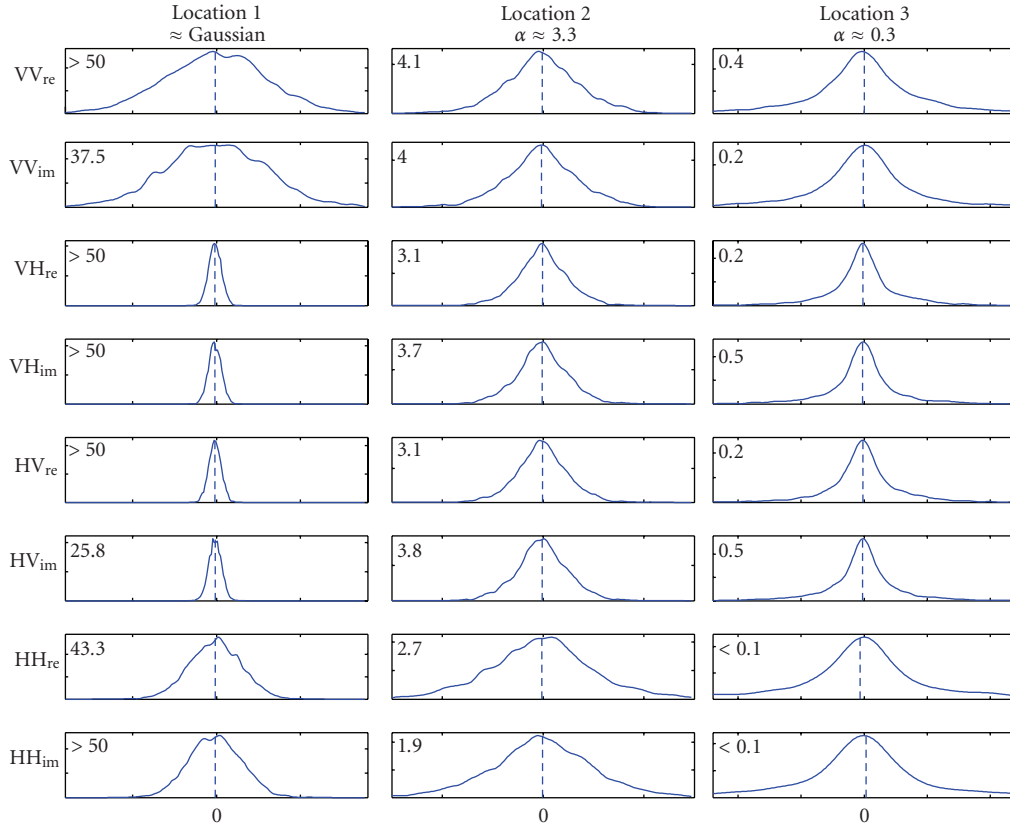


FIGURE 3: Real data observations: non-Gaussian, global symmetric shape, polarimetric width variations, zero-mean, and individual and overall shape α estimates noted. Actual locations represent water, forest, and urban areas.

of the data. Firstly, it can be seen, in all sets of data, that they do appear to be distributed with a mean of zero for each dimension. Secondly, taking each set individually (columns), it appears that the shape is consistent and symmetric down all dimensions (rows), although the shape can appear distinctly different from one sample location to another. Specifically, the first set (location 1) has a Gaussian-like roundness to each dimensional distribution, the second set (location 2) has a reasonably pointed peak with smoothly sloping sides, quite triangular in appearance, and the third set (location 3) has a marked kink in the sides and is very heavy tailed, again with each dimension showing basically the same shape. Some shape parameter values, α from the MK model, are noted within each box, and although there is some random variation for each value, probably due to estimation inaccuracy in the random sample, the shape value is consistent for each location, and clearly distinguishable from the other location values. The actual locations represent water, forest, and urban areas, respectively, and the observed progression in non-Gaussianity is expected for these target types.

It is interesting to also note that the polarimetric information becomes visible in the form of the different widths of each dimension, which can vary distinctly as in the first set, showing very little cross-polarisation scattering, or be much more evenly scaled as in the other two locations. Also note the pairwise equality in the distributions, because

the real and imaginary parts will have equal magnitudes, and the centre four dimensions being equally scaled due to reciprocity.

Clearly, the choice of semi-symmetric, zero-mean scale mixture of Gaussian models appears to be well suited for this type of PolSAR data.

4. Modelling Results

After obtaining four parametric descriptions of the data, we then compare a goodness-of-fit measure of each to determine which model fits best. Since we are comparing four different parametric descriptions to the same data set, it is sufficient to use a relative ranking measure only, and we do not require an absolute or normalised measure of fit. The log-likelihood measure is fast and efficient and simply requires summing the log of the model pdf value at each data point. The logarithmic nature of this measure also makes it sensitive to differences in the tails of the distributions and is therefore well suited for testing heavy-tailed distributions.

A “best fit” map is produced by goodness testing all four fitted models and mapping the chosen best-fitted model in different colours, white for Gaussian, green for Laplacian, red for K-distribution, and blue for normal inverse Gaussian. We also observed that although one model may be chosen as the best fit, some of the other models may be quite reasonable

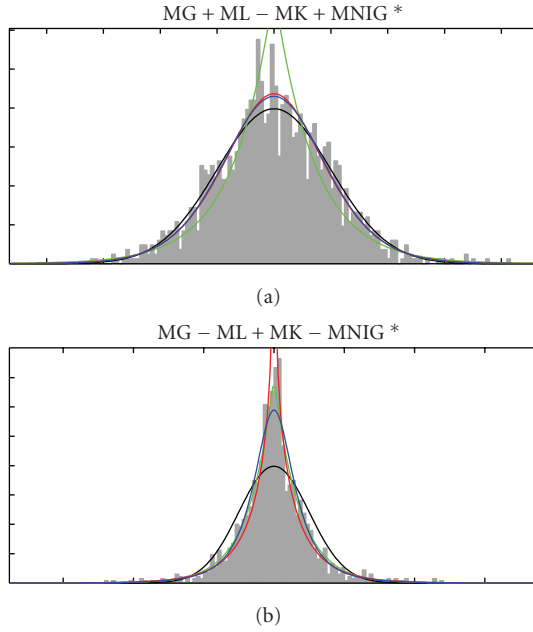


FIGURE 4: Examples of good-fit histograms showing that several models may have very similar fitting to the data, and from which we derived our 0.5% relative log-likelihood threshold. The top example shows that the MG (black), MK (red), and MNIG (blue, and the best-fit) are all acceptable fits to the data, whereas the ML (green) is not. The lower example shows that the ML and MNIG (best) are good fits, while the MG and MK are not.

fits, with visually similar fitting to the data histograms and very similar log-likelihood scores. An absolute goodness-of-fit measure for multivariate data is not a simple matter, particularly when the actual data histogram is not known, and the sample sizes relatively small. However, we found that an empirical threshold of within 0.5% of the best-fitted log-likelihood score clearly separated the bad fits from visually acceptable fits to the data histogram. Figure 4 shows two such examples where the data histogram is shown as grey bars, the MG model in black, the ML in green, the MK in red, and the MNIG in blue. The upper example shows that the MG, MK, and MNIG are all good fits, with the MNIG being the best-fitted model. The ML is clearly not a good fit to the histogram. The lower example shows that the MG is too “short and fat,” while the MK is too “tall and thin,” and they are both considered poor fits to the data. Both the ML and MNIG are good fits, with the MNIG again being the best fit of these examples. Using this threshold, we can produce a “coverage” map for each distribution that shows not only where it was the best fit but also where it was considered a “good” and poor fit too. Where each model was the best fit is depicted in red, a good fit in magenta, and a poor fit in black. Of real interest is in fact the regions where each model was considered a poor fit (black) and thus do not represent the data very well at all.

The modelling was tested on several different real PolSAR data images to compare the behaviour for quite different terrain types. The “Bleikvatnet” (a) mountain lake and forest

area and the “Okstinden” (b) mountain glacier area, both in Norway, are from airborne EMISAR flights in 1995. The “sea ice” (c) image is from an airborne CONVAIR flight in Canada in 2001. And the “Foulum” (d) agricultural and urban area is from an airborne EMISAR flight over Denmark in 1998. The best fit maps, and reference intensity maps, are shown for all four areas in Figure 5, from which the following observations may be made.

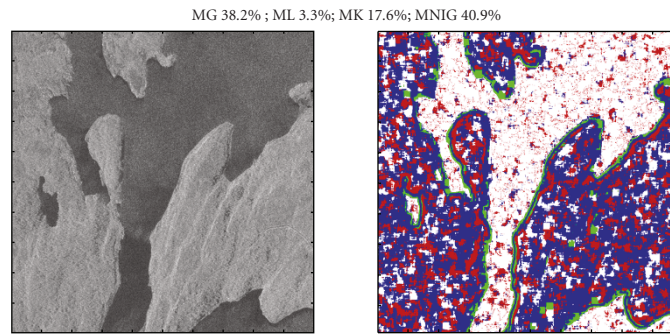
- (i) Uniform, smooth, or homogeneous areas are usually best fitted as Gaussian (white), as seen in the central lake area in (a), the large open snow areas in (b), the (presumably) snow covered old ice patches in (c), and the water inlet and several large fields in (d).
- (ii) The land in general, the visible icy crevasses, rocky outcrops, urban areas, and certainly anything with small scale details and high contrast are certainly non-Gaussian in nature and were poorly fitted by the Gaussian model.
- (iii) All types of vegetated land appear to be best described by the normal inverse Gaussian distribution, whereas the sea ice image by the K-distribution, although the difference compared to the NIG was negligible.
- (iv) The urban areas and coastlines are best fitted more often by the Laplacian; however this may be due to high contrast edge mixture effects because it appears at all water/land boundaries, around point sources like known huts within the forest, and along hedge/fence lines around fields.

The coverage maps for each separate model and for each area are shown in Figure 6. We observe the following.

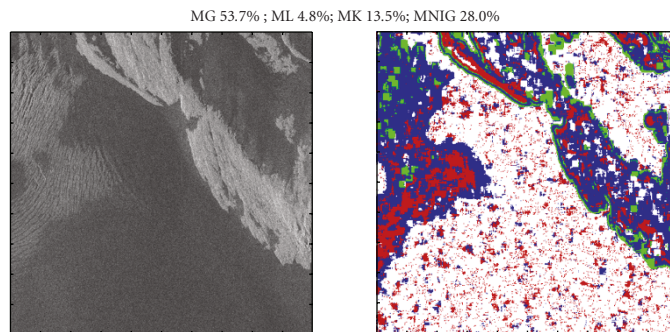
- (i) The Gaussian model is usually a poor fit for significant parts of the image area, over 20%.
- (ii) The Laplacian model is very good at detecting edges and point sources and is otherwise very poor at fitting to natural terrain types. Its seemingly good fit for urban areas is presumably because of the predominance of points and edges of mixed terrain in the urban landscape.
- (iii) In all cases the two parameters of the MK and MNIG give a shape space that finds a “good” fit for the majority of the data points (over 90%), and mostly “fail”, that is, are more poorly fitted, for the high contrast edges and point sources.
- (iv) The normal inverse Gaussian model has the greatest “good” fitted area for all images and is usually the greatest best fit also.

Our results indicate that using a single, flexible two parameter model is sufficient to capture the majority of shapes seen in real PolSAR imagery. Our results indicate that the normal inverse Gaussian model is the best choice, and the K-distribution model for sea ice analysis, although both are flexible enough for all types of data.

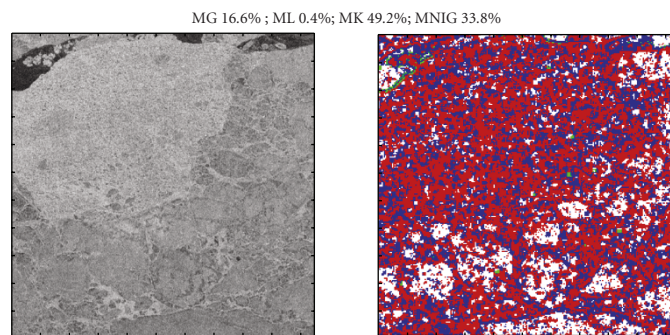
The reason that the Laplacian seems better at edges than either the MK or MNIG is probably because of the influence



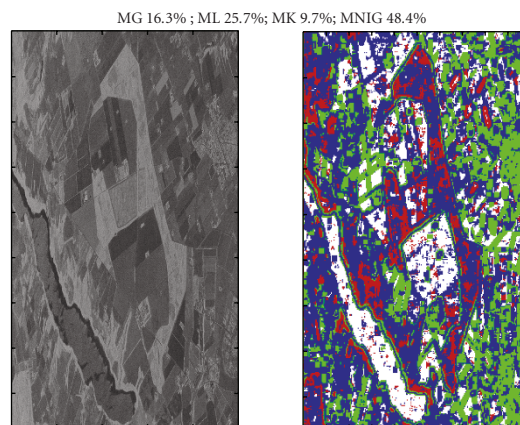
(a) Mountain lake and forest



(b) Mountain glacier: snow, ice and rock

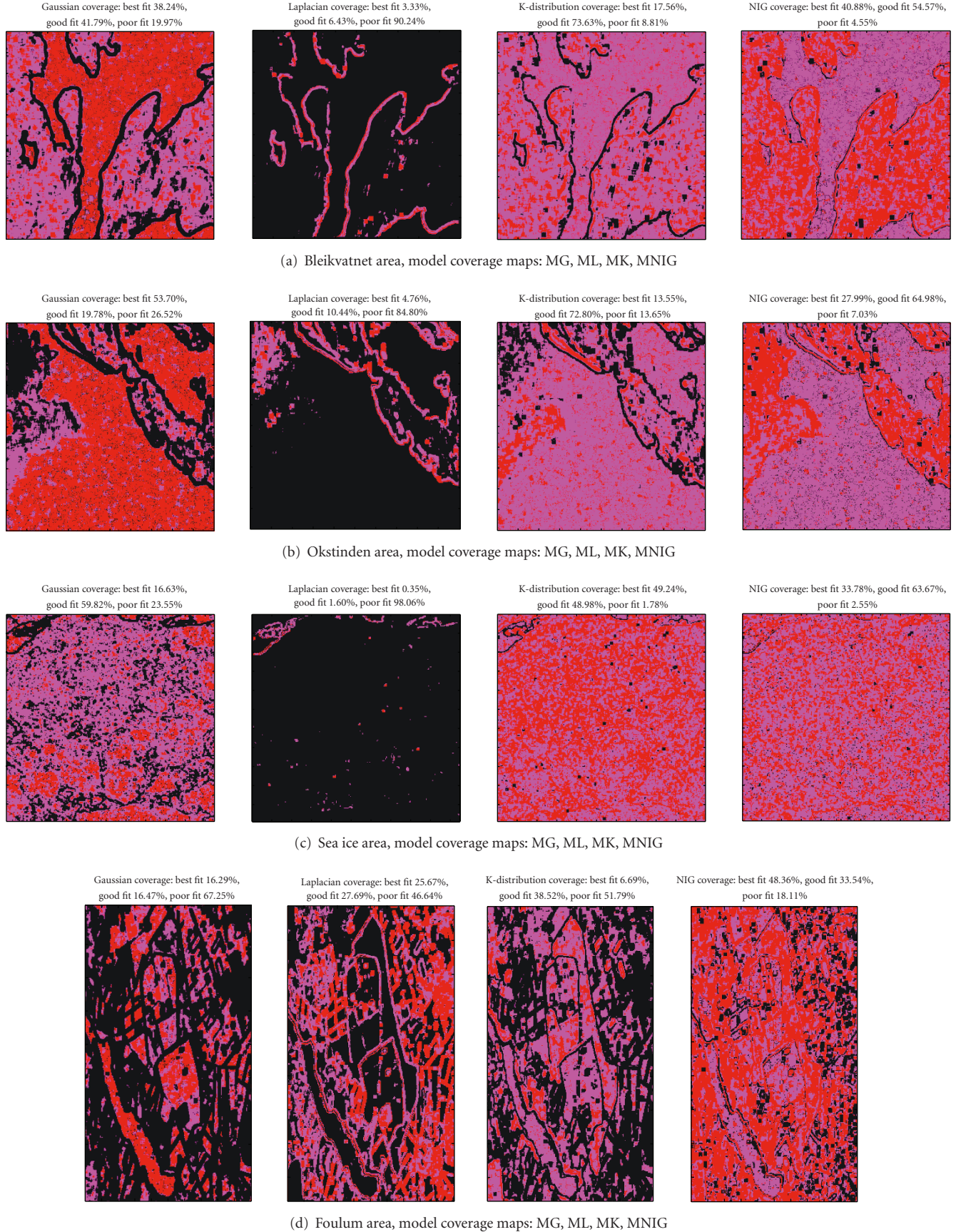


(c) Sea ice



(d) Agricultural, urban and water

FIGURE 5: Intensity and “best fit” coloured maps for the four areas: Gaussian in white, Laplacian in green, K-distribution in red, and normal inverse Gaussian in blue.



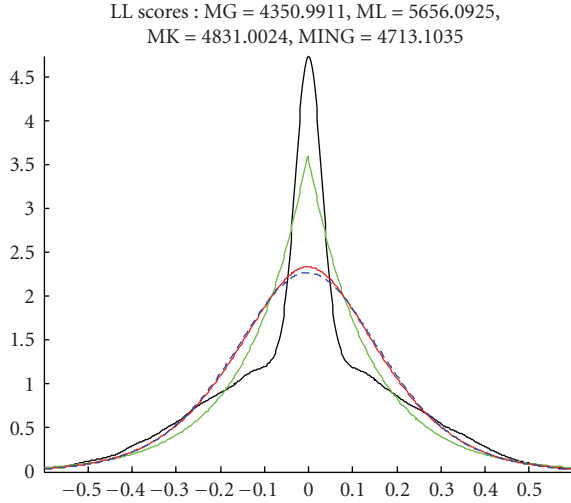


FIGURE 7: Two-part discrete mixture of Gaussians (black) fitted as Laplacian (green), K-distribution (red) and normal inverse Gaussian (blue dashed). The mixture appears to have a high non-Gaussianity measure.

of the outer shoulder in such mixed distributions. If indeed the “edge” distribution is from a very narrow and a very broad boundary mix, then it is unlikely to have the expected peak height for the apparent shoulder and tail for either the MK or MNIG and the best fit parameters seem to emphasise the tail region. The Laplacian has a fixed shape and simply fits roughly centrally over the middle “kink” and therefore leads to a higher goodness-of-fit score. This can be seen in the simple example, a mixture of two Gaussians, in Figure 7. The mixed distribution is shown in black. Observe that the Laplacian (in green) fits higher up the peak and lower in the tails, whereas the K-distribution (red) and normal inverse Gaussian (blue dashed) are almost identical and fit most tightly in the tails.

It is important to remember that only the goodness-of-fit testing of each model has been depicted in the figures so far, and not an actual image segmentation based upon the modelled parameters. The modelled parameters consist of a brightness (or total intensity) value, a non-Gaussianity (shape or texture) value, and a polarimetric matrix. The main emphasis of our method is to include the non-Gaussianity measure, which gives additional information that is otherwise ignored in a purely Gaussian-based approach. Additionally, by working with the raw non-Gaussianity measure, these features are independent of the specific scale model and can be considered a general two moment SMOG model.

The polarimetry can be interpreted in the usual manner because our matrix is simply a normalised covariance matrix. For example, a Pauli RGB colouring scheme, or the Freeman-Durden decomposition [17], can be used for display and interpretation with respect to general scattering mechanisms. Simple polarimetric features, extracted from the covariance matrix, are cross-polarisation fraction, co-pol ratio, co-pol correlation magnitude, and correlation phase. In total we

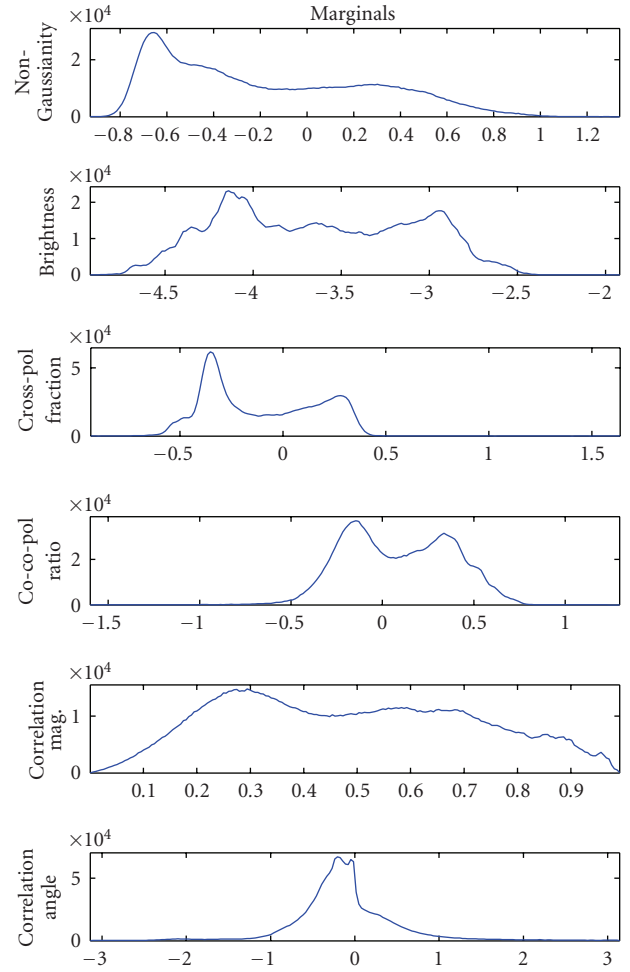


FIGURE 8: Extracted 6D feature histogram plots. Each feature shows several peaks such that a discrete mixture model may well describe the data.

have six scalar features, and we found that a logarithmic transformation of the brightness, non-Gaussianity, cross-pol fraction, and co-pol ratio improved visualisation and linearity of those features.

We demonstrate the modelling features with an airborne L-band (1.25 GHz) PolSAR data set acquired with the EMISAR instrument over the agriculture test area of Foulum, Denmark, in April 17, 1998. The resolution is 1.5 m in range and 0.75 m in azimuth. Referring to the Pauli composite image ($R = HH - VV$, $G = HV$, $B = HH + VV$) shown in Figure 12, the image contains a lake (purple-blue area stretching diagonally from left edge to bottom edge), some forest (e.g., the whitest areas in the central image), patches of cropland separated by roads, and some urban areas (greyish areas along right-hand edge). Ground truth data is described in [18]. Figure 8 depicts the six feature histograms for the Foulum image data and shows a significant amount of detail in each feature.

Figure 9 shows just the first two features as a scatter plot to demonstrate that natural compact clusters are clearly visible and a simple discrete mixture of Gaussian clustering

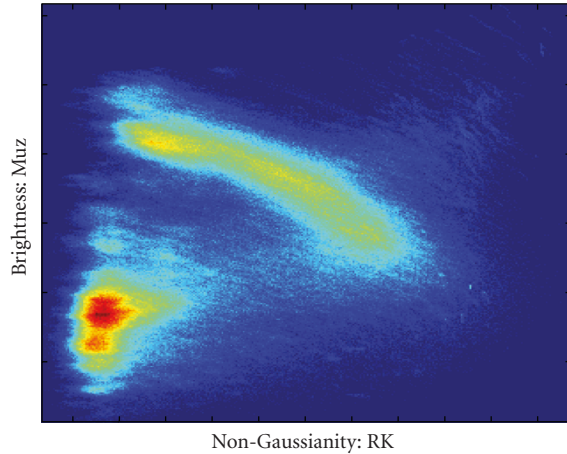


FIGURE 9: Two-feature scatter plot: non-Gaussianity versus Brightness. Note the natural clusters, but remember that there are four other dimensions too.

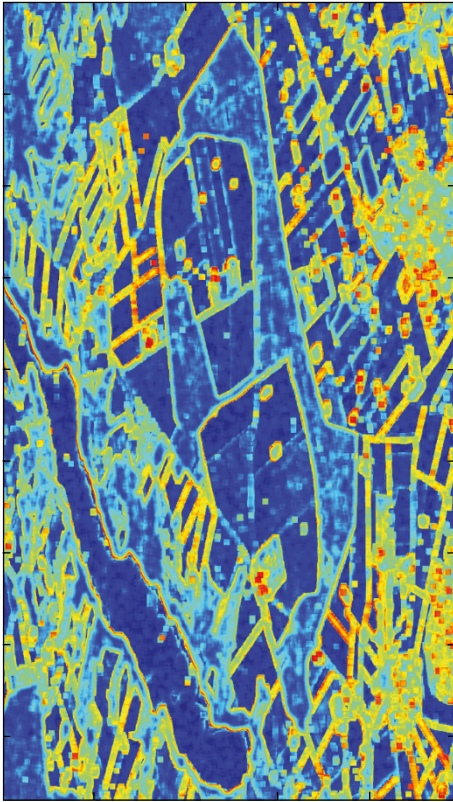


FIGURE 10: Non-Gaussianity measure (RK) image showing low values (Gaussian data) for the water and many fields, moderate values for the forest, and high values for the urban areas and mixed edges.

of the 6D feature space may be quite suitable. A stretched mixing line is also visible and probably corresponds to class boundary mixtures of bright and dark classes stretching from bright to dark via high non-Gaussianity as discussed for Figure 7.

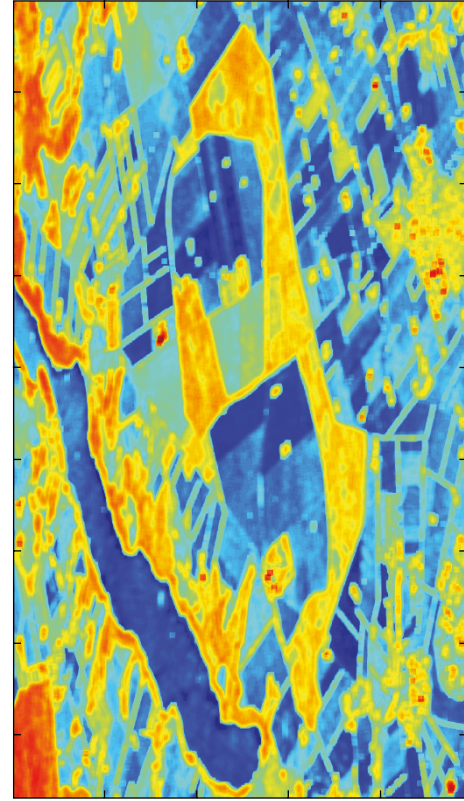


FIGURE 11: Brightness measure (μ_z) image showing low values for some fields and water, high values for forest and urban areas, and particularly bright for certain buildings that presumably faced the radar. Some range effect is also visible towards the left-hand edge of the image.

The features can also be viewed as images, to indicate where different features are significant, with a colour scale from blue for low values through yellow to red for high values. Figure 10 shows non-Gaussianity, where the water inlet and many fields have low values (dark blue), the forested areas have moderate values (light blue), and the urban areas are light blue to yellow, with bright point and edge mixtures showing the highest non-Gaussianity values in red. The brightness image is shown in Figure 11 and shows the expected range from low for smooth fields and water to bright for forest and urban areas. Some extremely bright buildings are visible, which presumably line up with the radar acquisition direction, and also note that some range effect is visible with the forest getting progressively brighter towards the left-hand side of the image, corresponding to lower incidence angles. The polarimetry is viewed as a Pauli RGB image in Figure 12, showing some distinctly different polarimetric responses.

Image segmentation of the parametric feature set derived from the modelling has a rich distinguishing power as seen in a simple preliminary segmentation of the Foulum image shown in Figure 13. Depicted is an unsupervised image segmentation created as a 16-class discrete mixture of Gaussians clustering of the 6D feature set. The segmentation looks quite good by visual inspection, but rigorous ground



FIGURE 12: Polarimetry (Γ matrix) image shown as Pauli RGB colours ($R = |HH - VV|$, $G = |HV|$, $B = |HH + VV|$).

truth testing has not been performed due to lack of data. The significance of adding the non-Gaussianity could be demonstrated by comparing segmentation images made with and without that feature; however the results can only be meaningfully compared with good ground truth testing. A quick test (not shown) did indeed show some differences; however, we have not yet investigated the details.

Besides image segmentation, the features may be useful for physical parameter extraction or physical interpretation in terms of polarimetry or backscattering brightness or texture. This has not yet been rigorously studied, but the level of detail shown in these parameters is encouraging.

5. Conclusion

The scale mixture of Gaussians models indeed seems well suited to modelling PolSAR data which show inherently heavy-tailed distributions with zero-mean and a global shape for each dimension.

We have confirmed that many terrain types are clearly non-Gaussian in nature and that a flexible two-parameter model is able to capture the full shape range of PolSAR data distributions, whereas the Gaussian model cannot. Different terrain types can show quite different distribution shapes; therefore the non-Gaussianity/shape parameter should be of benefit to subsequent image segmentation.

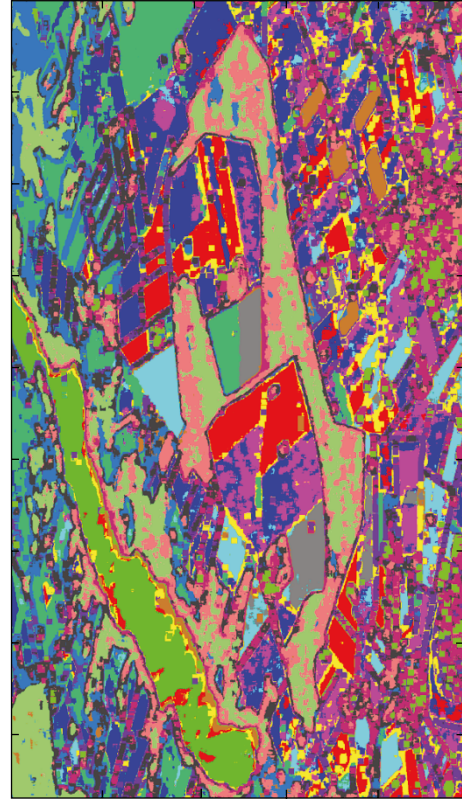


FIGURE 13: Foulum image segmentation example based upon modelled parameters. 16-class discrete mixture of multivariate Gaussian clustering on 6D feature set.

It was demonstrated that the normal inverse Gaussian distribution is the better fitting model, out of those analysed, and usually better than the more commonly used K-distribution, with the exception of over sea ice. The MNIG model captures the greater proportion of distribution shape variations and has less trouble at boundary mixtures than the MK. The normal inverse Gaussian also has strong theoretical grounds derived from Brownian motion theory. A detailed study of why it is generally superior, and why not for sea ice, has not yet been undertaken.

We also described how a generic, two-moment, scale mixture of Gaussian analysis may be performed without the need for choosing a specific model. The feature space obtained from the modelling contains non-Gaussianity, Brightness and a Polarimetric matrix. Six features were extracted, displayed, and discussed, with a final simple image segmentation as an example application.

The methods described here can be considered the foundation for our statistical analysis of PolSAR data and future work will investigate some of the observations made here as well as address several important extensions to the model that were discussed in the introduction.

Acknowledgments

The authors would like to thank Professor Henning Skriver, Dr. Jorgen Dall, and the Danish Technical University for the

Foulum data set. Thanks to Dr. Daniel Delisle, Dr. Sahebi Mahmood Reza, and the Canadian Space Agency for the sea ice data set. Both data-sets were downloaded from the European Space Agency website (<http://earth.esa.int/polsarpro/datasets.html>). Thanks to Norut Tromsø, Norway, for the Bleikvatnet and Okstinden data sets.

References

- [1] E. Jakeman and P. N. Pusey, "A model for non-Rayleigh sea echo," *IEEE Transactions on Antennas and Propagation*, vol. 24, no. 6, pp. 806–818, 1976.
- [2] E. Jakeman and R. J. A. Tough, "Generalized K distribution: a statistical model for weak scattering," *Journal of the Optical Society of America A*, vol. 4, no. 9, pp. 1764–1772, 1987.
- [3] T. Eltoft, "The Rician inverse Gaussian distribution: a new model for non-Rayleigh signal amplitude statistics," *IEEE Transactions on Image Processing*, vol. 14, no. 11, pp. 1722–1735, 2005.
- [4] S. H. Yueh, J. A. Kong, J. K. Jao, R. T. Shin, and L. M. Novak, "K-distribution and polarimetric terrain radar clutter," *Journal of Electromagnetic Waves and Applications*, vol. 3, pp. 747–768, 1989.
- [5] J. S. Lee, D. L. Schuler, R. H. Lang, and K. J. Ranson, "K-distribution for multi-look processed polarimetric SAR imagery," in *Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS '94)*, vol. 4, pp. 2179–2181, Pasadena, Calif, USA, August 1994.
- [6] 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.
- [7] A. Lopès and F. Séry, "Optimal speckle reduction for the product model in multilook polarimetric sar imagery and the wishart distribution," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 35, no. 3, pp. 632–647, 1997.
- [8] C. Oliver and S. Quegan, *Understanding Synthetic Aperture Radar Images*, SciTech, Raleigh, NC, USA, 2nd edition, 2004.
- [9] N. Goodman, "Statistical analysis based on certain multivariate complex gaussian distribution," *The Annals of Mathematical Statistics*, vol. 34, pp. 152–177, 1963.
- [10] A. P. Doulgeris, S. N. Anfinsen, and T. Eltoft, "Classification with a non-Gaussian model for PoISAR data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 10, pp. 2999–3009, 2008.
- [11] J. S. Lee, M. R. Grunes, and R. Kwok, "Classification of multi-look polarimetric SAR imagery based on complex Wishart distribution," *International Journal of Remote Sensing*, vol. 15, no. 11, pp. 2299–2311, 1994.
- [12] A. F. Andrews and C. L. Mallows, "Scale mixtures of normal distributions," *Journal of the Royal Statistical Society. Series B*, vol. 36, no. 1, pp. 99–102, 1974.
- [13] O. Barndorff-Nielsen, J. Kent, and M. Sorensen, "Normal variance-mean mixtures and z distributions," *International Statistical Review*, vol. 50, no. 2, pp. 145–159, 1982.
- [14] T. Eltoft, T. Kim, and T.-W. Lee, "Multivariate scale mixture of Gaussians models," in *Proceedings of the 6th International Conference on Independent Component Analysis and Blind Source Separation (ICA '06)*, Charleston, SC, USA, March 2006.
- [15] J. Folks and R. Chhikara, "The inverse Gaussian distribution and its application: a review," *Journal of the Royal Statistical Society. Series B*, vol. 40, pp. 263–289, 1978.
- [16] K. V. Mardia, "Measure of multivariate skewness and kurtosis with applications," *Biometrika*, vol. 57, no. 3, pp. 519–530, 1970.
- [17] A. Freeman and S. L. Durden, "A three-component scattering model for polarimetric SAR data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 36, no. 3, pp. 963–973, 1998.
- [18] H. Skriver, J. Dall, L. Ferro-Famil, et al., "Agriculture classification using POLSAR data," in *Proceedings of the 2nd International Workshop on Applications of Polarimetry and Polarimetric Interferometry (PoLinSAR '05)*, Frascati, Italy, January 2005.