Scale Mixture of Gaussians Modelling of Polarimetric SAR Data

Anthony P. Doulgeris and Torbjørn Eltoft

University of Tromsø
Institute of Physics
9037 Tromsø
NORWAY

E-mail: adoulgeris@ieee.org and torbjorn.eltoft@phys.uit.no

ABSTRACT

This paper discusses a multivariate, non-Gaussian parametric modelling technique to analyse polarimetric SAR data. We investigate a simple class of multivariate non-Gaussian distributions, the 'Scale mixture of Gaussians', and assess its "Goodness-of-fit" to the radar data. Four models are analysed and various characteristics of the models are interpreted, together with practical considerations with regard to parameter estimation. We observe that SAR data is often not Gaussian in distribution, being more highly peaked at zero and falling off more slowly than the Gaussian. It is shown that a single 'flexible' model is sufficient to capture the statistics of the SAR data, leading to a feature set of the modelled parameters. Image classification is then studied by means of the modelled data and compared with an existing land cover map.

1. INTRODUCTION

A considerable amount of research has been devoted to the study of the statistics of single polarisation SAR images. The Non-Gaussian nature of these signals is often modelled using non-Rayleigh amplitude models (e.g. the K model [1] and the RiIG model [2]). Polarimetric Synthetic Aperture Radar data is multivariate, with a complex $2 \times 2$ scattering matrix representing the returned amplitude and phase of the wave reflected from the target area. Modelling non-Gaussian polarimetric data has previously been studied using a similar multivariate K distribution [3].

Symmetric distributions that are peaked at zero and asymptotically fall off slower than the Gaussian are called sparse distributions. Multivariate sparse distributions have frequently been represented using 'mixture of Gaussians' models, in which the non-Gaussian distribution is modelled as a sum of several independent Gaussian distributions (each with their own mean and covariance matrix). More recently, multivariate data has been modelled with a multivariate extension to the class of distributions known as 'scale mixture of Gaussians' models [4], in which case the non-Gaussian distribution is constructed as a continuously scaled mixture of a normalised Gaussian distribution (with just one mean and normalised covariance structure). The scaling parameter itself is considered a positive random variable whose own density distribution function governs the characteristics of the mixture model distribution function. The marginal distribution of the mixture can be calculated by integrating the scaled Gaussian function over the distribution of the scale parameter, and several closed form solutions have been derived [5].

This paper investigates three such models in addition to the multivariate Gaussian: the multivariate Laplacian, multivariate K and multivariate Normal Inverse Gaussian. Modelling is accomplished using maximum likelihood estimation and method of moments to obtain the parametric description of the mixture model. The common generation method as a scale mixture of Gaussians leads to a common method of determining the mean and covariance structure parameters. The scalar parameters for each model are determined from combinations of one or more moment estimates related back to the scale parameter's distribution function. The parametric estimation routines will always find some parameters that fit each model to the given data. Goodness-of-fit testing is subsequently used to find which of the four fitted models best describes the data set. In this work we have chosen to use the log-likelihood measure, primarily due to its speed and simplicity.

Our main objective of the model fitting is to produce a new feature space to describe the data based upon the statistical model parameters. This feature space can then be used to classify the images. As an example we have applied this method to airborne polarimetric SAR data from a mountainous area in Norway. We show some initial results of simple classification based upon the parameter features.

2. THE MODELS

The multidimensional extension of the 'scale mixture of Gaussians' model is expressed as

\[ Y = \mu + \sqrt{Z} \Gamma^{1/2} X, \]  

where $\mu$ is the mean vector, $Z$ is the (positive only) scalar scale parameter, $\Gamma$ is the internal covariance structure matrix, normalised such that $\det \Gamma = 1$, and $X$ is a standardised Gaussian variable with zero mean and identity covariance matrix, i.e. $X \sim \mathcal{N}(0, I)$. 

1-4244-0413-4/06/$20.00 ©2006 IEEE 18 
NORSIG 2006
Observe that since \( Z \) and \( X \) are independent, we obtain the expectations
\[
\text{mean } \mathbf{Y} = \mathbb{E}\{\mathbf{Y}\} = \mathbf{\mu} \quad \text{and} \quad \text{cov. } \mathbf{Y} = \mathbb{E}\{(\mathbf{Y} - \mathbf{\mu})(\mathbf{Y} - \mathbf{\mu})^T\} = \mathbb{E}\{Z\} \quad \Gamma. \quad (3)
\]

Hence, for all of the models, we can estimate the parameter \( \mathbf{\mu} \) directly from the sample mean, and \( \Gamma \) from the normalised sample covariance matrix. It is also useful to interpret the scale \( Z \) as a global scale because the \( \Gamma \) matrix contains further relative scaling for each dimension. The multivariate Gaussian distribution can be considered a special case of this scale mixture model when the scale parameter \( Z \) is a constant.

Given a probability density function (pdf) for the scale parameter, \( f_Z(z) \), the marginal pdf for \( \mathbf{Y} \) can be obtained by averaging the pdf of \( \mathbf{Y} \mid Z \) (7) over the density of \( Z \)
\[
f_Y(y) = \int_0^\infty f_{Y|Z}(y|z) f_Z(z) \, dz 
\]
\[
= \int_0^\infty \frac{1}{(2\pi)^{\frac{d}{2}}\lambda} \exp\left( -\frac{1}{2\lambda} q(y) \right) f_Z(z) \, dz, 
\]
where for brevity we have defined
\[
q(y) = (y - \mathbf{\mu})^T \Gamma^{-1}(y - \mathbf{\mu}). 
\quad (5)
\]
The three distributions derived in this manner are named as multivariate extensions to existing one dimensional distributions and retain the general characteristics of their namesakes.

### 2.1. Multivariate Laplacian

The multivariate Laplacian, \( \text{ML}(\lambda, \mathbf{\mu}, \Gamma) \), is derived with \( Z \) from an exponential distribution
\[
f_Z(z) = \frac{1}{\lambda} \exp\left( -\frac{z}{\lambda} \right),
\quad (6)
\]
and the \( d \) dimensional marginal pdf of \( \mathbf{Y} \), using (4), becomes
\[
f_Y(y) = \frac{1}{(2\pi)^{\frac{d}{2}}\lambda} \frac{2}{\lambda} K_{\frac{d}{2}-1}\left(\sqrt{\frac{d}{q(y)}}\right)
\quad \left(\sqrt{\frac{d}{q(y)}}\right)^{\frac{d}{2}-1}, 
\quad (7)
\]
where \( K_m(x) \) denotes the modified Bessel function of the second kind and order \( m \), evaluated at \( x \).

### 2.2. Multivariate K

The multivariate K, \( \text{MK}(\alpha, \lambda, \mathbf{\mu}, \Gamma) \), is formed when \( Z \) is from a Gamma distribution
\[
f_Z(z) = \frac{\lambda^{\alpha+1} z^\alpha}{\Gamma(\alpha+1)} \exp(-\lambda z),
\quad (8)
\]
and the general \( d \) dimensional pdf of \( \mathbf{Y} \) is
\[
f_Y(y) = 
\frac{2}{(2\pi)^\frac{d}{2}\Gamma(\alpha+1)} \left(\sqrt{\frac{q(y)}{2\lambda}}\right)^{\alpha+1-\frac{d}{2}} K_{\alpha+1-\frac{d}{2}}\left(\sqrt{2\lambda q(y)}\right). 
\quad (9)
\]

### 2.3. Multivariate Normal Inverse Gaussian

When \( Z \) has an inverse Gaussian distribution,
\[
f_Z(z) = \frac{\delta}{\sqrt{2\pi\delta^2 + q(y)}} \exp\left( -\frac{1}{2} \left( \frac{\delta^2}{z} + \gamma^2 z \right) \right), 
\quad (10)
\]
the mixture model produces the multivariate Normal Inverse Gaussian, \( \text{MNIG}(\delta, \gamma, \mathbf{\mu}, \Gamma) \) with marginal density function
\[
f_Y(y) = 
\frac{2\delta e^{\gamma\sqrt{\frac{\gamma}{2\pi(\delta^2 + q(y))}}}}{\Gamma(\frac{d}{2})} K_{\frac{d}{2}-1}\left(\gamma\sqrt{\delta^2 + q(y)}\right). 
\quad (11)
\]

### 2.4. Properties

All models are symmetric about the mean and although each dimension may have different relative widths, distributed by the covariance matrix \( \Gamma \), they will each have a similar shape governed by the scalar parameters. The MG and ML distributions have a fixed shape, the rounded Gaussian and the pointed Laplacian. 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 the rounded Gaussian (see figure 1). The shape parameter does not vary linearly in value with most of the variation occurring near zero and converging asymptotically towards the Gaussian curve. 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.

The mixture of Gaussians method was also reviewed, because many of the non-Gaussian shapes may be created by a mixture of several independent Gaussian distributions. However, there is no clear rule to determine how many Gaussians to use or why, and the estimation would be poor with the reasonably small sample sizes used (we use \( 169 \times 6-1 \) samples). The scale mixture of Gaussians is a much more compact description, with only 3 or 4 parameters, that still captures the overall shape of the distribution.

### 3. ESTIMATION AND TESTING

In section 2, we explained how the mean \( \mathbf{\mu} \) and covariance structure matrix \( \Gamma \) can be estimated for each model. The other parameters can be estimated from one or two moment estimates for \( Z \), related back to each chosen distribution. Initially, the EM iterative approach was used (as in [5]), however the long processing time and occasional numerical limitations proved impractical to use over a large image.
At the expense of precision we have chosen to estimate the first moment of $Z$ from the determinant of the covariance matrix of $Y$, and the second moment of $Z$ from a simple fourth order moment in $Y$. The performance of both procedures is analysed using simulated data sets (figure 2). We accepted the compromise given the 15 fold speed increase. Note that there appears to be some bias for small sample sizes, but both methods are reasonably unbiased from about 180 samples onwards, for 6-D data (like our SAR data).

In the general case, all the model parameters are free to be optimised in the fitting procedures. However, it is normal for a random sample to vary slightly from the ideal distribution, purely due to it being a random sample. Therefore, if there is some reasonable basis for a parameter to be fixed then we would expect a better model fit by actually constraining it. The most obvious constraint is the zero-mean constraint. Additionally certain model parameters may have restricted range, for instance, the MK-dist. alpha value must be greater than -1, and lambda greater than zero. In our radar data we have further zero and pair value constraints on the covariance structure matrix. Examples in this paper are calculated without applying any constraints, however the mean was found to be very close to zero in all cases, and similarly the covariance terms expected to be zero were all very small.

In the case of the multivariate Laplacian, the first moment of $Z$ from (6) can be shown to be $\mathcal{E}\{Z\} = \lambda$. Combined with the observation of (3), we can estimate $\lambda$ directly from the determinant of the sample covariance thus

$$\hat{\lambda} = \mathcal{E}\{Z\} = \left(\text{det} \left( \text{cov}(Y) \right) \right)^{\frac{1}{d}}$$  \hspace{1cm} (12)

where $d$ is the data dimension.

The multivariate K and NIG distributions have two parameters each, and require two moment estimates to solve for two parameters. The simplest second moment relating to the scale mixture of Gaussians derives from the fact that the kurtosis of the standard Gaussian variable $X$ is $3\gamma$. A little manipulation reveals that by averaging over each dimension and all the data we can estimate the second moment thus

$$\mathcal{E}\{Z^2\} = \frac{1}{3dN} \sum_{k=1}^{d} \sum_{i=1}^{N} [\Gamma^{-\frac{1}{2}}(y_i - \mu)]_k^2$$  \hspace{1cm} (13)

where $\Gamma$ and $\mu$ are calculated from the sample mean and normalised covariance.

For the K-distribution with $Z$ Gamma distributed, (8), we have

$$\mathcal{E}\{Z\} = \frac{\alpha + 1}{\lambda}$$  \hspace{1cm} (14)

$$\mathcal{E}\{Z^2\} = \frac{(\alpha + 1)(\alpha + 2)}{\lambda^2}$$  \hspace{1cm} (15)

which can be solved to find $\alpha$ and $\lambda$.

The inverse Gaussian distribution, (10), has moments

$$\mathcal{E}\{Z\} = \frac{\delta}{\gamma}$$  \hspace{1cm} (16)

$$\mathcal{E}\{Z^2\} = \left(1 + \frac{1}{\delta^2}\right) \frac{\delta^2}{\gamma^2}$$  \hspace{1cm} (17)

and we can therefore solve for $\delta$ and $\gamma$.

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, 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 sparse distributions. Other
measure based upon the integrated squared differences of the pdfs were reviewed but found to be too cumbersome to compute because the integrals over Bessel functions did not have analytical solutions.

A 'best fit' map was produced by goodness testing all four fitted models and mapping the chosen best-fitted model in a different colour (figure 3). Of interest is the observation that although one model may be chosen as the best fit, some of the other models may be quite reasonable fits, with similar log-likelihood scores. The two parameters of the MK and MNIG give a shape space that finds a very close fit for data varying from Laplacian-like to Gaussian-like. For example, the MK and MNIG distributions, respectively, have 98% and 99% coverage as best or within 0.5% log-likelihood score, while the coverage for Gaussian was only 63%. This means that either of the MK or MNIG models alone may be used in place of any of the other models depicted here, and we are assured of fitting the data reasonably well. Even though the fit may not be quite as good, it is probably within the random sampling variation inherent in the data, and the shape parameter will probably represent the data fit well enough for classification purposes. Initial results have only looked at the MK distribution because of numerical difficulties encountered while testing the MNIG distribution. The decision to use just one model to fit the entire image, means that the classification stage becomes a single feature-set classifier, instead of a potential multiple pass or branched classifier with, no doubt, rather complex output merging problems.

4. IMAGE SEGMENTATION

We have chosen a 1000 × 1000 test area from a single look composite complex image. The region contains both water and land and comes from an EMISAR flight in July 1995 over Bleikvatnet, Norway. Starting with the 4 complex elements of the scattering matrix, we have taken the two diagonal terms and one of the cross terms (due to the reciprocity theorem), and split the complex numbers into their real and imaginary parts. Therefore, our data-set becomes a six component real vector, with constraints of certain equal magnitude pairs and many zero elements in the 6 × 6 covariance structure. The 6 dimensional representation was then analysed in 13 × 13 neighbourhoods (≈ 20 meter squares) using the multivariate K distribution only. The resulting parameter feature set was then classified with a simple mixture of Gaussian clustering, trained over 1% of the data, and then the whole image was classified with a Bayesian style classifier. The results are compared to a classification image based on Landsat STM that was supplied by Norut IT, Tromsø.

5. CONCLUSIONS

In summary, we have shown that a single flexible non-Gaussian model may be used to create a new parameter feature space from a particular class of multidimensional data. The scalar parameters may be interpreted as global shape and width terms, and the vector mean and covariance structure matrix describe spatial position and relative internal orientations. This technique was then shown to produce a smooth classification image with realistic looking regions compared to conventional classification. Further work is required to optimise and determine potential advantages of using this method.

REFERENCES

[1] E. Jakeman and P. N. Pusey, A model for non-Rayleigh sea echo. IEEE Trans. Antennas Propagat., 24, 6:806–814, November 1976.
[2] T. Elliott. The Rician inverse Gaussian distribution: A new model for non-Rayleigh signal amplitude statistics. IEEE Trans. Image Process., 14:1722–1735, November 2005.
[3] S. H. Yueh, J. A. Kong, J. K. Jao, R. T. Shin, and L. M. Novak. K-Distribution and polarimetric terrain radar clutter. J. Electro. Waves Appli., 3:747–768, 1989.
[4] A. F. Andrews and C. L. Mallows. Scale mixtures of normal distributions. Journal of the Royal Statistical Society, Series B, 36:99–102, no. 1 1974.
[5] T. Elliott, T. Kim, and T-W. Lee. Multivariate scale mixture of Gaussians models. Proceedings of ICA 2006, Charleston, SC, USA, March 2006.