Joint distributions for correlated radar images

B. Storvik
Norwegian Computing Center (NR)
Email: Baard.Storvik@nr.no

G. Storvik
University of Oslo
Email: geirs@math.uio.no

R. Fjørtoft
Norwegian Computing Center (NR)
Email: Roger.Fjortoft@nr.no

***Abstract*— For a given ground cover class, there is no straightforward way of expressing the joint distribution of a set of correlated radar images represented in amplitude or intensity. In this article we propose a general transformation method that permits incorporation of inter-image covariance while keeping a good fit to the marginal distributions. The approach is here studied for Gamma marginals, and the results of tests on a multi-temporal series of ERS-1 multi-look images are presented.**

I. Introduction

With the growing number and diversity of earth observation satellites, the coverage of the earth in space, time and the electromagnetic spectrum is increasing fast. This creates a demand for image analysis methods that can handle multi-sensor, multi-scale and multi-temporal data sets covering a certain region. We have developed a new statistical model for classification of such compound data sets [1], which is currently being validated.

The observed pixel values of a given ground cover class in multi-spectral optical images are well modeled by a multivariate Gaussian distribution, and the same model can be used for the joint distribution of a set of overlapping multi-spectral images. However, for detected radar images (amplitude or intensity) neither marginal nor joint distributions are Gaussian.

Let us consider a set of radar images acquired over a given area. The images will generally appear somewhat different, e.g. because of:

- different acquisition geometry
- different acquisition dates (multi-temporal)
- different wavelengths (multi-frequency)
- different polarization combinations (polarimetric)

Despite the differences, the values of overlapping pixels in the different images will in many cases be correlated. This correlation can easily be taken into account for single-look complex (SLC) images, where a multivariate complex Gaussian distribution is well suited. However, for detected radar images it is far more complicated to express a joint distribution incorporating dependence between the images.

Assuming fully developed speckle [2] and ignoring spatial correlations, the intensity $\bar{I}$ of a pixel in a multi-look radar image is Gamma distributed

$$f_{\bar{I}}(x;L,R) = \frac{1}{\Gamma(L)}\left(\frac{L}{R}\right)^{L}\exp(-\frac{Lx}{R})x^{L-1} \qquad (1)$$

where $x \geq 0$ is a realization of $\bar{I}$, $R = E[\bar{I}]$ is the local radar reflectivity, and $L = R^2/Var[\bar{I}]$ is the equivalent number of independent looks (ENIL) of the image. If the radar reflectivity of a given class has texture, it is frequently assumed to be Gamma distributed as well, in which case the observed intensities of the class are K distributed [3].

Some multivariate Gamma distributions are presented in [4]. However, there are restrictions on the dependence structure that make these multivariate distributions unsuited for our application.

In this paper we propose to use meta-Gaussian distributions to model dependence between detected radar images. This approach is very general and does not imply strong restrictions on the dependence structure, as opposed to the multivariate Gamma distributions in [4]. The meta-Gaussian approach can be used to combine virtually any kind of marginal distributions (e.g. Gaussian, Gamma and K distributions) into multivariate distributions. It can therefore be a useful tool for joint analysis of multi-temporal, multi-frequency and polarimetric radar data represented in amplitude or intensity, and for combinations of radar data and optical data.

II. Meta-Gaussian Distribution

The basic idea of meta-Gaussian distributions is to transform the marginal values so that they become Gaussian, measure the correlation on the Gaussian scale, and transform them back again.

Let $\mathbf{X} = (X_1, ..., X_N)$ be a stochastic vector with marginal density g_j for the jth component X_j of $\mathbf{X}$. (In our setting X_j is the value of a given pixel in image number j out of N overlapping images.) Let furthermore G_j be the cumulative distribution function corresponding to g_j and Φ the cumulative distribution function for the standard normal distribution. General probability theory then says that

$$Y_j = \Phi^{-1}(G_j(X_j)) \qquad (2)$$

is a standard normally distributed variable. The meta-Gaussian approach is to model the dependence between the different components of $\mathbf{X}$ through the dependence between the components of $\mathbf{Y} = (Y_1, ..., Y_N)$. In particular, it is assumed that $\mathbf{Y}$ is a multivariate Gaussian distributed vector with expectation vector $\mathbf{0}$ and covariance matrix $\mathbf{\Sigma}$. In order to keep each Y_k standard normal, we require the diagonal elements of $\mathbf{\Sigma}$ to be equal to 1. Inverting (2), we obtain

$$X_j = G_j^{-1}(\Phi(Y_j)). \qquad (3)$$

Further, by using standard results from probability theory on transformations, the multivariate density of $\mathbf{X}$ is

$$f(\mathbf{x};\gamma) = |\mathbf{\Sigma}|^{-1/2} \exp\{-\frac{1}{2}\mathbf{y}(\mathbf{x};\gamma)^T(\mathbf{\Sigma}^{-1} - \mathbf{I})\mathbf{y}(\mathbf{x};\gamma)\} \times \prod_{j=1}^{p} g_j(x_j;\gamma) \quad (4)$$

where $\boldsymbol{\gamma}_j$ are the parameters of the marginal distribution g_j, $\mathbf{y}(\mathbf{x};\boldsymbol{\gamma}) = (y_1(x_1;\boldsymbol{\gamma}_1), \ldots, y_N(x_N;\boldsymbol{\gamma}_N))^T$ and $y_j(x_j;\boldsymbol{\gamma}_j) = G_j^{-1}(\Phi(x_j);\boldsymbol{\gamma}_j)$.

It should be noted that for $\mathbf{\Sigma} = \mathbf{I}$, the distribution reduces to a product of independent marginals, making the interpretation of $\mathbf{\Sigma}$ similar to the correlation matrix for multivariate Gaussian distributions. No assumptions are here made about g_j, except that the inverse of the cumulative distribution G_j must exist.

In practice, g_j will usually be chosen from a parametric family of distributions. If all g_j are Gaussian, the density (4) reduces to a multivariate Gaussian distribution. If all g_j are lognormal, we obtain the ordinary multivariate lognormal distribution. For g_j being Gamma distributions, we obtain a multivariate Gamma distribution. If some g_j are Gaussian and some are Gamma, a multivariate distribution combining Gaussian marginals with Gamma marginals is obtained. Such combinations permit joint analysis of optical and radar images.

In this article we concentrate on Gamma marginals and multivariate Gamma distributions obtained through the meta-Gaussian approach.

III. Classification

Using the framework introduced in the previous section, we may for each class $k \in \{1, ..., K\}$ define a multivariate density $f_k(\mathbf{x})$ describing the distribution of a vector of observations $\mathbf{x}$ from class k. Define z_i to be the class of pixel i and $\mathbf{x}_i$ to be the observed values in pixel i. Neglecting contextual dependence, the Bayes classification rule is

$$\hat{z}_i = \underset{k}{\operatorname{argmax}} \{\pi_k f_k(\mathbf{x})\}. \quad (5)$$

Contextual classification methods can also be applied in the ordinary way. Assume e.g. a Potts model

$$p(\mathbf{z}) \propto e^{\sum_i \alpha_{z_i} + \beta \sum_{i \sim j} I(z_i = z_j)}$$

where $I(\cdot)$ is the indicator function and $i \sim j$ means that i and j are neighbors in a graph. Making the usual assumption of conditional independence of observations given classes, the posterior distribution for $\mathbf{z}$ is given by

$$p(\mathbf{z}|\mathbf{x}) \propto p(z) \prod_i f_{z_i}(\mathbf{x}_i). \quad (6)$$

Maximum a posteriori (MAP) estimates of $\mathbf{z}$ can be obtained by global maximization of (6). Such a maximization is recognized as a difficult problem and therefore approximative algorithms such as the iterative conditional modes (ICM) [5] are usually applied. An efficient algorithm for obtaining global maxima has been presented in [6].

TABLE I

Ground Truth

Class label	Class name	Number of pixels in training set	Number of pixels in test set
1	forest	2559	11985
3	orchard	48	66
4	hard wheat	2985	8195
5	soft wheat	2264	5782
6	maize	2876	10598
7	sunflower	2384	5479
8	barley	141	161
9	oilseed rape	2749	7012
10	peas	623	1573
11	clover	488	793
14	prairie	722	1899
17	bare soil	1162	2993
20	road	404	923
21	water	537	1990
24	urban area	1581	4008

IV. Estimation

In order to apply the classification rules discussed in the previous section, the parameters involved needs to be estimated. Based on a training set with known classes, maximum likelihood (ML) estimation can in principle be performed. Such estimates are, however, computationally costly to obtain, mainly because of the constraints on the covariance matrix $\mathbf{\Sigma}$ (all diagonal elements needs to be equal to one, and in addition, the matrix needs to be positive definite). We have therefore also considered a simpler approach, where L_j and R_j, $j = 1, \ldots, N$, first are estimated marginally based on data from the corresponding component only. Estimates of $\mathbf{\Sigma}$ are then obtained by maximizing the likelihood with the estimated L_j and R_j inserted.

Based on theory on estimation functions [7], it can be shown that the estimates obtained are asymptotically consistent and normally distributed. The asymptotic variances for these estimates will differ from the ML estimates, but in our experience the efficiency loss is small.

V. Results

The pixelwise Bayes classification rule (5) has been used to examine whether the use of meta-Gaussian distributions significantly improves the classification accuracy compared to marginal Gamma distributions that are assumed to be independent. It should be stressed that the focus is not on achieving the highest possible classification accuracy, but on revealing differences between the two approaches.

The data set considered here consists of a multi-temporal series of 6 ERS-1 images of Bourges, France. The images were acquired with monthly intervals during the summer season 1993, and 4-look amplitude images were generated from the original SLC images. The training set consists of vectors of amplitude observations from 21 523 pixels where the ground truth (class label) is known. The test data set contains 63 457 pixels. Table I contains the name, label value and number of pixels in training set and test set of each of the 15 classes.

The training set is used to estimate the parameters of the models and to construct the classification rule. The test set is used to find the probabilities of correct classification on the basis of the classification rule.

TABLE II

CONFUSION MATRIX FOR IML METHOD

$C \setminus \widehat{C}$	1	3	4	5	6	7	8	9	10	11	14	17	20	21	24
1	0.472	0.128	0.007	0.003	0.021	0.058	0.034	0.067	0.021	0.075	0.030	0.063	0.002	0.002	0.018
3	0.076	0.394	0.015	0.061	0.061	0.030	0.045	0.030	0.106	0.045	0.106	0.000	0.030	0.000	0.000
4	0.006	0.050	0.431	0.265	0.002	0.001	0.048	0.002	0.017	0.019	0.109	0.002	0.042	0.005	0.002
5	0.007	0.048	0.249	0.384	0.003	0.001	0.048	0.004	0.017	0.023	0.106	0.002	0.095	0.010	0.003
6	0.130	0.088	0.018	0.005	0.184	0.053	0.052	0.102	0.048	0.100	0.033	0.111	0.004	0.001	0.072
7	0.080	0.069	0.011	0.005	0.028	0.334	0.043	0.057	0.120	0.158	0.019	0.048	0.002	0.001	0.023
8	0.062	0.087	0.056	0.037	0.000	0.012	0.447	0.050	0.012	0.075	0.118	0.012	0.019	0.012	0.000
9	0.070	0.055	0.009	0.009	0.042	0.036	0.099	0.428	0.080	0.083	0.051	0.023	0.007	0.003	0.004
10	0.010	0.051	0.009	0.010	0.006	0.116	0.025	0.074	0.528	0.076	0.053	0.022	0.013	0.000	0.005
11	0.076	0.098	0.050	0.019	0.015	0.086	0.097	0.103	0.086	0.295	0.058	0.009	0.003	0.001	0.004
14	0.007	0.096	0.091	0.197	0.001	0.006	0.064	0.028	0.046	0.023	0.351	0.005	0.077	0.008	0.000
17	0.158	0.098	0.014	0.007	0.058	0.065	0.038	0.079	0.069	0.073	0.054	0.241	0.007	0.002	0.038
20	0.017	0.039	0.015	0.135	0.003	0.010	0.026	0.011	0.016	0.014	0.131	0.008	0.556	0.013	0.005
21	0.002	0.004	0.005	0.020	0.000	0.001	0.005	0.003	0.001	0.001	0.008	0.000	0.121	0.828	0.003
24	0.278	0.059	0.008	0.008	0.015	0.044	0.024	0.023	0.012	0.039	0.030	0.034	0.005	0.004	0.415

TABLE III

CONFUSION MATRIX FOR MEF METHOD

$C \setminus \widehat{C}$	1	3	4	5	6	7	8	9	10	11	14	17	20	21	24
1	0.471	0.114	0.007	0.003	0.024	0.059	0.033	0.063	0.021	0.077	0.030	0.069	0.003	0.003	0.023
3	0.076	0.394	0.030	0.030	0.030	0.045	0.045	0.015	0.091	0.061	0.106	0.030	0.030	0.000	0.015
4	0.005	0.059	0.453	0.250	0.006	0.002	0.039	0.002	0.016	0.018	0.107	0.001	0.036	0.007	0.001
5	0.006	0.056	0.281	0.366	0.004	0.003	0.043	0.004	0.018	0.021	0.108	0.002	0.076	0.012	0.001
6	0.126	0.087	0.017	0.009	0.239	0.050	0.045	0.105	0.045	0.104	0.030	0.106	0.004	0.002	0.032
7	0.073	0.066	0.009	0.011	0.035	0.345	0.041	0.060	0.126	0.151	0.020	0.046	0.003	0.001	0.013
8	0.050	0.118	0.037	0.075	0.000	0.012	0.435	0.043	0.012	0.062	0.099	0.012	0.019	0.025	0.000
9	0.066	0.062	0.009	0.009	0.038	0.034	0.085	0.438	0.078	0.093	0.047	0.024	0.009	0.005	0.003
10	0.007	0.055	0.010	0.008	0.009	0.120	0.024	0.076	0.526	0.078	0.051	0.022	0.013	0.001	0.003
11	0.082	0.087	0.043	0.030	0.014	0.076	0.088	0.105	0.091	0.310	0.057	0.008	0.006	0.000	0.004
14	0.008	0.113	0.092	0.162	0.002	0.007	0.060	0.028	0.044	0.025	0.345	0.005	0.099	0.008	0.001
17	0.141	0.100	0.014	0.012	0.061	0.059	0.033	0.083	0.073	0.077	0.049	0.273	0.006	0.002	0.016
20	0.020	0.054	0.016	0.150	0.002	0.010	0.027	0.011	0.023	0.012	0.113	0.009	0.531	0.013	0.011
21	0.003	0.012	0.007	0.022	0.000	0.000	0.003	0.002	0.002	0.001	0.007	0.000	0.096	0.841	0.008
24	0.283	0.052	0.009	0.007	0.015	0.049	0.022	0.022	0.014	0.044	0.032	0.041	0.007	0.005	0.396

We compare two approaches. One consists in assuming that all components are independent with Gamma marginals. ML is used to estimate the parameters involved in this case. We will denote this method by independent maximum likelihood (IML). The other approach is the meta-Gaussian with Gamma marginals. For this model, both ML estimation and the use of estimation functions are considered. These methods are denoted by MML and MEF, respectively.

For both models, the marginal distributions (1) are described by parameters $\boldsymbol{\gamma}_j = (L_j, R_j)$. For the meta-Gaussian model, the dependence is described through the correlation matrices $\mathbf{\Sigma}$ on the Gaussian scale (one for each class).

The overall portion of correctly classified pixels in the test set for the three methods were 0.387 (IML), 0.400 (MML) and 0.397 (MEF), i.e., the differences are very small. Tables II (IML) and III (MEF) shows the confusion matrices. The results for MML were similar to those of MEF. We would expect that high correlations within a class would give less confusion with other classes when taking the covariances into account (MEF) than when assuming independence (IML). This is mostly the case, but there are exceptions.

To further investigate the impact of the magnitude of the inter-image correlation, we performed classification into a reduced number of classes, corresponding to those having the strongest correlation between components, which were the ones with labels 6, 8, 17, 20, and 24. In this case, the overall portion of correctly classified pixels for the three classification rules were 0.411 (IML), 0.475 (MEF), and 0.458 (MML), i.e., a significant improvement is obtained by incorporating covariance through meta-Gaussian distributions.

VI. CONCLUSION

We propose a general transformation method that permits incorporation of inter-image covariance while keeping a good fit to the non-Gaussian marginal distributions of radar images.

Tests on a multi-temporal series of 4-look ERS-1 images indicate that the advantage of taking inter-image covariance into account increases with its strength. The proposed method should therefore be tested on data sets with stronger inter-image covariance. Partially polarimetric and multi-frequency radar images are of particular interest.

Using a contextual classification rule such as (6), rather than the pixelwise classification rule (5) used here, would give much higher classification accuracy.

REFERENCES

[1] G. Storvik, R. Fjørtoft, and A. H. S. Solberg, "Integration of multisource image data at different resolutions and time points – a mathematical framework for EOtools," Norwegian Computing Center (NR), Oslo, Norway, Internal report SAMBA/25/2002, August 2002.

[2] J. W. Goodman, "Statistical properties of laser speckle patterns," in *Laser Speckle and Related Phenomena*, 2nd ed., J. C. Dainty, Ed. New York: Springer-Verlag, 1984.

[3] E. Jakeman, "On the statistics of K distributed noise," *Journal of Physics A: Mathematical and General*, vol. 13, pp. 31–48, 1980.

[4] S. Kotz, N. Balakrishnan, and N. L. Johnson, *Continous Multivariate Distributions*, 2nd ed. New York: John Wiley Interscience Publication, 2000, vol. 1: Models an Applications.

[5] J. Besag, "On the statistical analysis of dirty pictures," *Journal of the Royal Statistical Society, Series B*, vol. 48, no. 3, pp. 259–302, 1986.

[6] G. Storvik and G. Dahl, "Lagrangian based methods for finding MAP solutions for MRF models," *IEEE Trans. Image Processing*, vol. 9, no. 3, pp. 469–479, 2000.

[7] K. Y. Liang and S. L. Zeger, "Inference based on estimating functions in the presence of nuisance parameters (with comments)," *Statistical Science*, vol. 55, pp. 158–173, 1995.