JOINT DISTRIBUTIONS FOR MULTI-TEMPORAL SERIES OF RADAR IMAGES

B. STORVIK
Norwegian Computing Center
Email: Baard.Storvik@nr.no

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

R. FJØRTOFT
Norwegian Computing Center
Email: Roger.Fjortoft@nr.no

For a given ground cover class, there is no straight-forward 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. This meta-Gaussian approach is studied here for Gamma marginals, and the results of tests on a multi-temporal series of ERS-1 multi-look images are presented.

1 Introduction

With a growing number of satellite sensors, the coverage of the earth in space, time and 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 multi-sensor, multi-scale and multi-temporal images [1], which is currently being validated. In this paper we consider modeling of a time series of radar images based on a transformation method that permits the computation of joint distributions for a series of intensity radar images.

In multi-temporal classification, we can model temporal context both in terms of inter-pixel class dependency (variations in the class label) and inter-pixel feature correlation (variations in the class properties). In [2] it was assumed that multi-temporal feature vectors were conditionally independent and only inter-pixel class dependency was modeled. In this paper we derive new distributions for multivariate radar images and use this model for multi-temporal SAR images. Hence we assume that the underlying class label image does not change in the time period, but that the class properties change and that there may be temporal feature correlation.

More generally, let us assume that we have a set of radar images that have been acquired over a given area, with approximately the same acquisition geometry. The images will generally appear somewhat different, e.g. because they were:

- not acquired simultaneously (multi-temporal)
- acquired by sensors with different wavelengths (multi-frequency)
- acquired with different polarization combinations (polarimetric)

For corresponding pixels belonging to a given ground cover class (e.g. a certain kind of agricultural field or forest), there will in many cases be inter-image correlation. In the case of multi-temporal images, this may e.g. be related to the phenological evolution of the vegetation. Such correlation can easily be taken into account for single-look complex (SLC) radar images, where a multivariate complex circular Gaussian distribution is well suited. However, for amplitude or intensity images (single- or multi-look) there is no straight-forward way of expressing the joint distribution.

Assuming fully developed speckle [3] and ignoring spatial correlations, the intensity $\bar{I}$ of a pixel in a multi-look radar image is Gamma distributed

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

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 [4].

Some multivariate Gamma distributions are presented in [5]. 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 [5]. 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.

The outline of the paper is as follows: Meta-Gaussian distributions are presented in section 2. Sections 3 and 4 describe classification based on meta-Gaussian distributions and estimation procedures for meta-Gaussian distributions, respectively. In section 5 we apply the meta-Gaussian distributions for classification of a multi-temporal series of real radar images and a simulated multivariate radar data set.

2 Meta-Gaussian distribution

The basic idea of meta-Gaussian distributions [6] is to transform the marginal values so that they become Gaussian, model the correlation on the Gaussian scale, and invert the transformation.

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

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

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

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

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

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

where $\mathbf{y}(\mathbf{x}) = (y_1(x_1), \ldots, y_N(x_N))^T$ and $y_j(x_j) = G_j^{-1}(\Phi(x_j))$.

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

3 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

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

Contextual classification methods can also be applied in the ordinary way, using e.g. 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

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

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) [7] are usually applied. An efficient algorithm for obtaining global maxima has been presented in [8]. Alternative posterior estimates, including uncertainty measures, can be obtained by employing Markov chain Monte Carlo (MCMC) techniques [9].

4 Estimation

The multivariate density (4) involves parameters which needs to be specified. In addition to the covariance matrix $\mathbf{\Sigma}$, each marginal density g_j contain additional parameters $\boldsymbol{\gamma}_j$. For marginal Gamma densities, $\boldsymbol{\gamma}_j = (L_j, R_j)$, see (1).

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 $\boldsymbol{\gamma}_j$, $j = 1, \dots, N$, 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 $\boldsymbol{\gamma}_j$, $j = 1, \dots, N$.

Based on theory on estimation functions [10], it can be shown that the estimates obtained are asymptotically consistent and normally distributed. The asymptotic variances for estimation function (EF) estimates will differ from the ML estimates, but in our experience the efficiency loss is small.

5 Results

The pixelwise Bayes classification rule (5) has been used to examine whether the use of meta-Gaussian distributions 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 [11]. 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 1 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 probability of correct classification (PCC) on the basis of the classification rule.

We compare several approaches. One assumes that all components are independent with Gamma marginals, and ML is used to estimate the parameters involved. We denote this method by independent maximum likelihood (IML). The second approach is the meta-Gaussian with Gamma marginals. For this model, both ML estimation and a simpler approach based on estimation functions are considered. These methods are denoted by MGML and MGEF, respectively. For both Gamma models, the marginal distributions (1) of a class are described by the parameters $\boldsymbol{\gamma}_j = (L_j, R_j)$, $j = 1, \dots, N$. For the meta-Gaussian approach, the inter-image dependencies are described by the covariance matrix $\mathbf{\Sigma}$ on the Gaussian scale.

Table 1: Result of classification of multi-temporal series of ERS-1 images into 15 classes with the IML, MGML and MGEF methods.

Class	Size training	Size test	IML PCC	IML Sens.	MGML PCC	MGML Sens.	MGEF PCC	MGEF Sens.
forest	2559	11985	0.472	0.579	0.481	0.576	0.462	0.594
orchard	48	66	0.394	0.005	0.424	0.005	0.424	0.006
hard wheat	2985	8195	0.431	0.619	0.462	0.602	0.453	0.616
soft wheat*	2264	5782	0.384	0.427	0.352	0.422	0.337	0.426
maize*	2876	10598	0.184	0.661	0.228	0.700	0.253	0.709
sunflower	2384	5479	0.334	0.457	0.346	0.480	0.351	0.473
barley*	141	161	0.447	0.023	0.422	0.026	0.460	0.025
oilseed rape	2749	7012	0.428	0.514	0.451	0.517	0.441	0.522
peas	623	1573	0.528	0.239	0.521	0.239	0.531	0.237
clover	488	793	0.295	0.052	0.314	0.057	0.319	0.055
prairie	722	1899	0.351	0.170	0.354	0.174	0.336	0.170
bare soil	1162	2993	0.241	0.218	0.282	0.238	0.276	0.240
road*	404	923	0.556	0.259	0.484	0.283	0.603	0.263
water	537	1990	0.828	0.883	0.837	0.857	0.853	0.856
urban area*	1581	4008	0.415	0.562	0.358	0.683	0.452	0.736

The overall portion of correctly classified pixels in the test set for the three methods were 0.387 (IML), 0.398 (MGML) and 0.404 (MGEF). In Table 1, the probability of correct classification (PCC) for each class is given together with a *sensitivity* measure (Sens.). This sensitivity measure is defined as the number of correctly classified pixels divided by the total number of pixels classified to that class. The sensitivity measure is included because only considering probability rates inside a class do not fully describe the properties of that class. A large PCC can be combined with a large number of pixels from other classes being classified to this class. The sensitivity measure quantifies this latter aspect and should be close to one. We would expect that strong correlations within a class would give less confusion with other classes when taking the covariances into account (MGML and MGEF) than when assuming independence (IML). This is mostly the case (classes with strong correlations are marked with an asterisk in Table 1), but there are exceptions. The modest improvements obtained with the meta-Gaussian approach are probably due to the relatively weak inter-image correlations in this data set. The maximum correlations range from 0.2 to 0.75 for the different classes.

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. These classes are given in Table 2. In this case, the overall portion of correctly classified pixels for the three classification rules were 0.620 (IML),

Table 2: Result of classification of multi-temporal series of ERS-1 images into a reduced set of 5 classes with the IML, MGML and MGEF methods.

Class	Size training	Size test	IML PCC	IML Sens.	MGML PCC	MGML Sens.	MGEF PCC	MGEF Sens.
soft wheat*	2264	5782	0.721	0.847	0.762	0.839	0.712	0.858
maize*	2876	10598	0.600	0.823	0.632	0.834	0.629	0.895
barley*	141	161	0.727	0.028	0.708	0.032	0.696	0.037
road*	404	923	0.611	0.377	0.548	0.252	0.706	0.234
urban area*	1581	4008	0.523	0.676	0.493	0.752	0.648	0.764

0.638 (MGML) and 0.659 (MGEF), i.e., a non-negligible improvement is obtained by incorporating covariance through meta-Gaussian distributions, especially with the MGEF method.

We also investigated the impact of the models on simulated data. The simulated data were obtained by simulating from a meta-Gaussian distribution (Gamma marginals) for each class. The class parameters were given from the MGEF estimates from the real data, where the parameter estimates from the marginal distribution are kept untouched. To obtain stronger correlations between images, the off-diagonal elements of the Cholesky decomposed covariance matrix were multiplied by a factor 5. This procedure gives stronger correlations in the covariance matrix and the same relative change for each class. Note that the marginal distributions of each class are Gamma with approximately the same parameters as for the real data. The only change is stronger correlation. The number of training and tests samples are also the same. The overall portion of correctly classified pixels for the three classification rules were here 0.422 (IML), 0.482 (MGML) and 0.648 (MGEF), i.e., a significant improvement is obtained by incorporating covariance through meta-Gaussian distributions, especially in the case of EF estimation. Information on the individual classes is given in Table 3. The large difference between the results of MGEF and MGML is probably due to a more robust estimation procedure for MGEF. The MGML method seems to encounter numerical problems in maximizing a very complex likelihood function.

6 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

Table 3: Result of classification of a simulated data set with strong inter-image covariance into 15 classes with the IML, MGML and MGEF methods.

Class	Size	Size	IML		MGML		MGEF	
	training	test	PCC	Sens.	PCC	Sens.	PCC	Sens.
forest	2559	11985	0.477	0.624	0.534	0.533	0.553	0.811
orchard	48	66	0.485	0.006	0.303	0.009	0.758	0.026
hard wheat	2985	8195	0.422	0.645	0.537	0.683	0.593	0.808
soft wheat	2264	5782	0.349	0.368	0.586	0.506	0.720	0.617
maize	2876	10598	0.284	0.779	0.214	0.565	0.650	0.803
sunflower	2384	5479	0.381	0.509	0.590	0.477	0.607	0.674
barley	141	161	0.565	0.045	0.627	0.071	0.807	0.090
oilseed rape	2749	7012	0.409	0.616	0.578	0.771	0.579	0.816
peas	623	1573	0.594	0.261	0.690	0.389	0.713	0.418
clover	488	793	0.398	0.075	0.521	0.127	0.554	0.174
prairie	722	1899	0.312	0.173	0.465	0.235	0.491	0.313
bare soil	1162	2993	0.365	0.394	0.659	0.394	0.718	0.562
road	404	923	0.469	0.122	0.092	0.085	0.931	0.405
water	537	1990	0.902	0.823	0.876	0.935	0.985	0.925
urban area	1581	4008	0.582	0.562	0.100	0.450	0.881	0.795

with stronger inter-image covariance. Partially polarimetric radar images are of particular interest. The choice of parameter estimation method is decisive. The use of estimation functions seems more robust than maximization of the likelihood, and it is also much faster.

Acknowledgment

This study was supported by the Norwegian Research Council (NFR) in the framework of the EOtools project.

References

[1] G. Storvik, R. Fjørtoft, and A. H. S. Solberg, "Integration of multi-source 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] A. H. S. Solberg, T. Taxt, and A. K. Jain, "A Markov random field model for classification of multisource satellite imagery," *IEEE Trans. Geoscience and Remote Sensing*, vol. 34, no. 1, pp. 100–113, 1996.

[3] 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.

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

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

[6] K. S. Kelly and R. S. Krzysztofowicz, "A bivariate meta-Gaussian density for use in hydrology," *Stochastic hydrology and hydraulics*, vol. 11, no. 1, pp. 17–31, 1997.

[7] 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.

[8] 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.

[9] W. Gilks, S. Richardson, and D. J. Spiegelhalter, *Markov chain Monte Carlo in practice*. London: Chapman & Hall, 1996.

[10] 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.

[11] A. Lopès, R. Fjørtoft, D. Ducrot, P. Marthon, and C. Lemaréchal, "Segmentation of SAR images in homogeneous regions," in *Information Processing for Remote Sensing*, C. H. Chen, Ed. Singapore: World Scientific Publishing Co., 1999.