

Article

Land Cover Classification for Polarimetric SAR Images Based on Mixture Models

Wei Gao *, Jian Yang and Wenting Ma

Department of Electronic Engineering, Tsinghua University, Beijing 100084, China;

E-Mails: yangjian.ee@gmail.com (J.Y.); mwt09@mails.tsinghua.edu.cn (W.M.)

* Author to whom correspondence should be addressed; E-Mail: gaow05@mails.tsinghua.edu.cn; Tel.: +86-10-6278-2397; Fax: +86-10-6277-0317.

Received: 7 February 2014; in revised form: 26 March 2014 / Accepted: 1 April 2014 /

Published: 28 April 2014

Abstract: In this paper, two mixture models are proposed for modeling heterogeneous regions in single-look and multi-look polarimetric SAR images, along with their corresponding maximum likelihood classifiers for land cover classification. The classical Gaussian and Wishart models are suitable for modeling scattering vectors and covariance matrices from homogeneous regions, while their performance deteriorates for regions that are heterogeneous. By comparison, the proposed mixture models reduce the modeling error by expressing the data distribution as a weighted sum of multiple component distributions. For single-look and multi-look polarimetric SAR data, complex Gaussian and complex Wishart components are adopted, respectively. Model parameters are determined by employing the expectation-maximization (EM) algorithm. Two maximum likelihood classifiers are then constructed based on the proposed mixture models. These classifiers are assessed using polarimetric SAR images from the RADARSAT-2 sensor of the Canadian Space Agency (CSA), the AIRSAR sensor of the Jet Propulsion Laboratory (JPL) and the EMISAR sensor of the Technical University of Denmark (DTU). Experiment results demonstrate that the new models fit heterogeneous regions preferably to the classical models and are especially appropriate for extremely heterogeneous regions, such as urban areas. The overall accuracy of land cover classification is also improved due to the more refined modeling.

Keywords: land cover classification; mixture model; polarimetric SAR

1. Introduction

In the past two decades, land cover classification has been one of the central topics in the radar remote sensing community. Many classification schemes for polarimetric SAR images have been proposed. The first method that utilizes the complete polarimetric information is the Bayes classifier based on the complex Gaussian distribution for single-look polarimetric SAR data, presented by Kong [1] and Lim [2]. Lee [3] extended their method and developed an optimal classifier for multi-look polarimetric SAR data based on the complex Wishart distribution, namely the Wishart classifier, which is a milestone in the research of polarimetric SAR data classification. These methods are widely applied, since they describe each land cover type by comparatively simple probability density functions and do not require much computation. However, these methods make the assumption that the land cover is homogeneous, which is improper for complicated SAR scenes. Other statistical models for land cover classification include the K-distribution [4,5] and the G-distribution [6,7], which belong to the category of product models [8]. These models employ a positive random variable for characterizing heterogeneity of the scattering medium and have higher accuracy than the Wishart model in heterogeneous regions.

The above-mentioned classifiers are pixel-based, *i.e.*, the class label to a pixel is completely determined by the pixel itself. An alternative to a pixel-based classification scheme would be that based on regions. Typically, a region-based classification scheme explores spatial correlations between adjacent pixels [9]. Images are first segmented into non-overlapping small regions, each of which contains pixels with similar properties. The class label to a region is determined collectively by all the pixels in that region. As an example, Wu [10] proposed such a method that combines the Wishart distribution with the Markov random field (MRF). Although additional information of spatial correlations may improve classification accuracy, it also introduces extra complexity in modeling and computation. Pixel-based classification schemes that can achieve comparably satisfactory accuracy are still appealing and of equal importance. Therefore, discussions will be focused on pixel-based classification schemes in this paper.

Accurate classification of polarimetric SAR data relies on appropriate statistical modeling of each land cover type. As pointed out by Lee [3], scattering vectors from a homogeneous region follow a complex Gaussian distribution, and the corresponding covariance matrices have a complex Wishart distribution. These distributions give the fundamental descriptions of single-look and multi-look polarimetric SAR data. Classifiers based on these models achieve the optimal classification accuracy in homogeneous regions. While in heterogeneous regions, their performance deteriorates and more refined statistical models are needed. In research fields of image processing and computer vision, irregular distributions are frequently approximated by the Gaussian mixture model [11,12]. Inspired by these applications, we propose to describe heterogeneous regions in polarimetric SAR images using mixture models. For single-look data, the true distribution of the scattering vector is approximated by a weighted sum of complex Gaussian distributions. For multi-look data, the true distribution of the covariance matrix is approximated by a weighted sum of complex Wishart distributions. The principle of this modeling can be accounted for from two perspectives. From the physical point of view, heterogeneity of a certain land cover type can be partly attributed to the blending of pixels reflecting different scattering mechanisms [13] and partly to the finer spatial resolution [14]. The proposed model is based on the

assumption that samples of each land cover type are a mixture of multiple subclasses, each of which has a complex Gaussian distribution in the single-look case, or a complex Wishart distribution in the multi-look case. From the mathematical point of view, the mixture component distributions serve as the kernels in estimating an irregular probability density function of the data. Based on the proposed mixture models, two maximum likelihood classifiers can be constructed. The expectation-maximization (EM) algorithm [15] is employed to estimate the model parameters using preselected training samples. If independence is further assumed among data from multiple radar frequencies, the classifiers can be extended to apply to multi-frequency polarimetric SAR images with no effort.

Some researchers have applied the mixture model to the problem of unsupervised clustering of polarimetric SAR data. Kersten [16] proposed an EM clustering scheme based on the Wishart mixture model and compared it with fuzzy clustering algorithms. Doulgeris [17] developed an automated clustering procedure using a mixture of K-distributions to accommodate non-Gaussian polarimetric SAR data. Both clustering methods model the whole image by using a mixture model, with each mixture component corresponding to a cluster. Therefore, each class is essentially described by a single-component distribution. While the primary goal of this research is to investigate whether a mixture of multiple distributions within a single class would better fit the real polarimetric SAR data and, thus, improve the classification performance in a supervised framework. Additionally, in the above clustering methods, it is the EM algorithm and the merging and splitting strategies that actually perform the classification. In our method, classification is based on the maximum likelihood criterion. The EM algorithm is employed only to determine mixture model parameters for each class.

The remainder of this paper is organized as follows. Some statistical properties of polarimetric SAR data are summarized in Section 2, along with the principle of fundamental classifiers. In Section 3, classification schemes based on the proposed mixture models are presented in detail. Experiment results are given in Section 4 with discussions. Finally, the research is concluded in Section 5.

2. Background

Generally, polarimetric SAR data are provided in either of two formats: the single-look format and the multi-look format. In this section, some statistical properties of polarimetric SAR data will be briefly reviewed, along with the principle of fundamental classifiers for each data format.

2.1. Polarimetric SAR Data

A polarimetric SAR measures the backscattering coefficient of the scene under illumination using multiple combinations of transmitting and receiving antenna polarizations. The most frequently used antenna polarizations are the horizontal polarization (h) and the vertical polarization (v), which are orthogonal. In the case of a fully polarimetric SAR, each resolution unit is described by a 2×2 complex scattering matrix \mathbf{S} , or the Sinclair matrix, as shown in Equation (1).

$$\mathbf{S} = \begin{bmatrix} S_{hh} & S_{hv} \\ S_{vh} & S_{vv} \end{bmatrix} \quad (1)$$

The elements S_{pq} in the scattering matrix represent the complex scattering coefficient, with q denoting the transmitting polarization and p denoting the receiving polarization. If reciprocity holds, the cross-polarizations are considered as equal, *i.e.*,

$$S_{hv} = S_{vh}. \quad (2)$$

In such a case, the scattering matrix becomes symmetric and may be reduced to a three-dimensional single-look scattering vector :

$$\mathbf{k} = \begin{bmatrix} S_{hh} & \sqrt{2}S_{hv} & S_{vv} \end{bmatrix}^T \quad (3)$$

where T denotes the transpose of a matrix. For the purpose of speckle reduction, scattering vectors of adjacent resolution units are incoherently averaged, or multi-look processed, according to Equation (4).

$$\mathbf{Z} = \frac{1}{n} \sum_{i=1}^n \mathbf{k}_i \mathbf{k}_i^H \quad (4)$$

In Equation (4), H denotes the conjugate transpose of a matrix, \mathbf{k}_i are the scattering vectors to be averaged, and n is defined as the number of looks. The resulting matrix \mathbf{Z} is called the n -look covariance matrix. Other forms of polarimetric SAR data include the coherency matrix and the Kennaugh matrix. In this paper, covariance matrices are mainly dealt with.

2.2. Statistical Models of Polarimetric SAR Data

The backscatter from a distributed target is random in nature. A great many statistical models have been proposed for single-look and multi-look polarimetric SAR data, among which, the Gaussian model and the Wishart model are fundamental. For homogeneous regions, the scattering vector \mathbf{k} can be modeled as having a complex Gaussian distribution with zero mean, the probability density function of which is:

$$p(\mathbf{k}|\mathbf{C}) = \frac{\exp \{-\mathbf{k}^H \mathbf{C}^{-1} \mathbf{k}\}}{\pi^d |\mathbf{C}|}. \quad (5)$$

In Equation (5), \mathbf{C} denotes the covariance of the distribution. As demonstrated by Goodman [18], if the scattering vectors in Equation (4) are independent, then \mathbf{Z} has a complex Wishart distribution, the probability density function of which is:

$$p(\mathbf{Z}|\mathbf{C}) = \frac{n^{nd} |\mathbf{Z}|^{n-d} \exp \{-n \text{Tr}(\mathbf{C}^{-1} \mathbf{Z})\}}{R(n, d) |\mathbf{C}|^n}. \quad (6)$$

In Equation (6), \mathbf{C} denotes the centroid of the distribution and is identical to the covariance \mathbf{C} in Equation (5). Tr denotes the trace of a matrix. The normalization term $R(n, d)$ in the denominator can be evaluated according to:

$$R(n, d) = \pi^{\frac{1}{2}d(d-1)} \Gamma(n) \cdots \Gamma(n - d + 1). \quad (7)$$

In Equation (7), $\Gamma(\cdot)$ represents the gamma function. The parameter d in Equations (5), (6) and (7) is the dimension of the scattering vector. For a fully polarimetric SAR under reciprocity, $d = 3$.

The Gaussian model and the Wishart model are appropriate for modeling homogeneous regions, while for heterogeneous regions, more advanced product models are frequently used. A product model

employs a positive random variable with unit mean to describe the texture of the scattering medium. The covariance matrix \mathbf{Z} is modeled as the product of the texture variable z and a Wishart distributed matrix \mathbf{X} , *i.e.*,

$$\mathbf{Z} = z\mathbf{X}. \quad (8)$$

The probability density function of \mathbf{Z} , or $p_{\mathbf{Z}}(\mathbf{Z})$, can be determined by the following integration in Equation (9), if a distribution $p_z(z)$ is assumed for the texture variable z .

$$p_{\mathbf{Z}}(\mathbf{Z}) = \int p_z(z) \cdot p_{\mathbf{Z}|z}(\mathbf{Z}|z) dz \quad (9)$$

When the texture has a gamma distribution, \mathbf{Z} has a K-distribution, the probability density function of which is:

$$p(\mathbf{Z}|\alpha, \mathbf{C}) = \frac{2|\mathbf{Z}|^{n-d}(n\alpha)^{\frac{1}{2}(\alpha+nd)} K_{\alpha+nd} \left(2\sqrt{n\alpha \text{Tr}(\mathbf{C}^{-1}\mathbf{Z})} \right)}{R(n, d)|\mathbf{C}|^n \Gamma(\alpha) \text{Tr}(\mathbf{C}^{-1}\mathbf{Z})^{-\frac{1}{2}(\alpha+nd)}}. \quad (10)$$

In Equation (10), $K(\cdot)$ represents the modified Bessel function of the second kind. α and \mathbf{C} correspond to the shape parameter of z and the centroid of \mathbf{X} , respectively. When α tends to infinity, z becomes a constant, and the K-distribution degrades to the Wishart distribution, as a special case. Numerous instances fall into the category of product models [14], since different distributions of the texture variable lead to different distributions of the covariance matrix. In this paper, discussions and comparisons are made mainly with respect to the three models mentioned above.

2.3. The Wishart Classifier

Lee [3] has proposed an optimal Bayes classifier for multi-look polarimetric SAR data based on the Wishart model. In his method, each land cover type is assumed to be homogeneous and is modeled by a Wishart distribution, with the centroid characterizing the scattering property of the associated medium. To estimate the centroid, averaging the training samples of each land cover type is simply needed. Denote the M classes of land covers by ω_m , $m = 1, 2, \dots, M$, and the corresponding class priors by $P(\omega_m)$. The Wishart centroid and conditional distribution of each class are represented by \mathbf{C}_m and $p(\mathbf{Z}|\mathbf{C}_m)$, or simply $p_m(\mathbf{Z})$. The Wishart classifier labels a test sample \mathbf{Z}_n with the class ω_i , if:

$$p(\mathbf{Z}_n|\mathbf{C}_i)P(\omega_i) > p(\mathbf{Z}_n|\mathbf{C}_j)P(\omega_j) \quad (11)$$

holds for all $j \neq i$. For computational simplicity, the log-likelihood is actually computed instead of the true likelihood. Furthermore, if all classes have equal prior probabilities, the Bayes classifier degrades to a maximum likelihood classifier, *i.e.*, a sample \mathbf{Z}_n is labeled with the class ω_i , if:

$$p(\mathbf{Z}_n|\mathbf{C}_i) > p(\mathbf{Z}_n|\mathbf{C}_j) \quad (12)$$

holds for all $j \neq i$. A distance measure is also derived by Lee [3], representing the distance between a test sample \mathbf{Z}_n and a class ω_m , *i.e.*,

$$d_{\text{Wishart}}(\mathbf{Z}_n, \omega_m) = \log |\mathbf{C}_m| + \text{Tr}(\mathbf{C}_m^{-1}\mathbf{Z}_n) \quad (13)$$

which is minimized among all classes when performing classification to \mathbf{Z}_n . The Wishart classifier has great significance in that it is the first classification scheme for multi-look polarimetric SAR data

that utilizes the full polarimetric information, and it does not require much computation. The classifier proposed by Kong [1] is similar in rationale, except that it is based on the complex Gaussian distribution and applies to single-look data. For homogeneous regions, these classifiers are optimal in the sense of minimizing the classification error rate. In Section 4, classification results will be presented.

3. The Proposed Method

In this section, the proposed mixture models for polarimetric SAR data in heterogeneous regions will be presented in detail, along with the maximum likelihood classifiers based on these models. The parameter estimation procedure will also be discussed.

3.1. Mixture Modeling of Polarimetric SAR Data

The Wishart classifier for multi-look polarimetric SAR data is of great significance. It has been extensively used as basic components in many polarimetric SAR applications [19,20], although the homogeneity assumption usually makes it suboptimal in practice. As has been demonstrated in [5], some homogeneous regions do not obey the Wishart distribution, due to the finer spatial resolution; other regions to be modeled are even not homogeneous, especially the urban areas. Applying the Wishart model to these areas would cause considerable modeling errors. The same will happen to the Gaussian model for single-look polarimetric SAR data. To adapt to heterogeneity, we propose to model polarimetric SAR data using mixture models. The mixture models are powerful in approximating irregular distributions and have gained much success in various research fields, such as image processing and computer vision [11,12]. For single-look scattering vectors and multi-look covariance matrices, the true probability density function is estimated by a weighted sum of complex Gaussian distributions and complex Wishart distributions, respectively.

Take the multi-look case for example. Suppose that a region is modeled by a mixture of K Wishart components. The centroids of these Wishart components are denoted by \mathbf{C}_k , $k = 1, 2, \dots, K$. Each Wishart component is associated with a non-negative weight π_k . The weights should satisfy:

$$\sum_{k=1}^K \pi_k = 1. \quad (14)$$

From a statistical point of view, each sample from the Wishart mixture distribution is considered to be generated by one of the K Wishart components, with π_k representing the probability that the sample is drawn from the k -th component. Based on this model, the probability density function of the Wishart mixture distribution can be expressed by:

$$\begin{aligned} p(\mathbf{Z}|K, \mathbf{C}_1, \dots, \mathbf{C}_K, \pi_1, \dots, \pi_K) &= \sum_{k=1}^K \pi_k q(\mathbf{Z}|\mathbf{C}_k) \\ &= \sum_{k=1}^K \pi_k \frac{n^{nd} |\mathbf{Z}|^{n-d} \exp \{-n \text{Tr}(\mathbf{C}_k^{-1} \mathbf{Z})\}}{R(n, d) |\mathbf{C}_k|^n}. \end{aligned} \quad (15)$$

The mixture modeling of polarimetric SAR data relies on the assumption that heterogeneity of land covers results from the blending of pixels with diverse scattering properties. This is true for polarimetric

SAR images with moderate or coarse spatial resolution. For those images with very fine spatial resolution, the non-Gaussian and non-Wishart phenomena are partly due to the lack of enough random scatterers within each resolution unit [14]. In such a case, the mixture models are still applicable, as the component distributions can be viewed as kernels for approximating an unknown probability density function, from a mathematical perspective. Since the proposed mixture models take into account the heterogeneity issue of practical radar scenes, they should fit real polarimetric SAR data more precisely. Section 4 will validate the effectiveness of the mixture models by experiments.

3.2. Parameter Estimation Using the EM Algorithm

The EM algorithm is an iterative algorithm that solves the maximum likelihood estimation of model parameters when the observed data are incomplete. In the case of a mixture model, each sample is assumed to be drawn from one of the mixture components, but the definite component to which a sample belongs to is unknown. Therefore, the EM algorithm is quite suitable for the parameter estimation problem of mixture models, as indicated in [15]. Since the solving procedures are quite similar for the Gaussian mixture model and the Wishart mixture model, the latter is taken for illustration. Suppose that the set of training samples for class ω_m is $\{\mathbf{Z}_{m,1}, \mathbf{Z}_{m,2}, \dots, \mathbf{Z}_{m,N_m}\}$, and the class distribution is the weighted sum of K_m Wishart components with centroids $\{\mathbf{C}_{m,1}, \mathbf{C}_{m,2}, \dots, \mathbf{C}_{m,K_m}\}$ and weights $\{\pi_{m,1}, \pi_{m,2}, \dots, \pi_{m,K_m}\}$. The k -th mixture component is denoted by $q(\mathbf{Z}|\mathbf{C}_{m,k})$, or simply $q_{m,k}(\mathbf{Z})$. For succinctness, the subscript m will be left out in the following derivation. Additionally assuming that the training samples are drawn independently, the EM algorithm will maximize the likelihood function of Equation (16) by iteratively updating \mathbf{C}_k and π_k .

$$L(K, \mathbf{C}_1, \dots, \mathbf{C}_K, \pi_1, \dots, \pi_K) = \prod_{n=1}^N \sum_{k=1}^K \pi_k q(\mathbf{Z}_n | \mathbf{C}_k) \quad (16)$$

Following a similar derivation to [15], the model parameters can be solved by the following iterative procedure.

Step 1: Specify the initial number of Wishart components K . A merging strategy will be described later to control the number of mixture components in each iteration.

Step 2: Initialize \mathbf{C}_k and π_k , $k = 1, 2, \dots, K$. \mathbf{C}_k are initialized by randomly selected \mathbf{Z}_{n_k} from the training samples, and π_k are initialized to be equal to each other.

$$\mathbf{C}_k^{(0)} = \mathbf{Z}_{n_k} \quad (17)$$

$$\pi_k^{(0)} = \frac{1}{K} \quad (18)$$

Step 3: Compute $\gamma_{kn}^{(i)}$, $k = 1, 2, \dots, K$, $n = 1, 2, \dots, N$, according to Equation (19). These are intermediate variables in the i -th iteration of the EM algorithm. The $\mathbf{C}_k^{(i-1)}$ and $\pi_k^{(i-1)}$ used in Equation (19) are the current values of Wishart centroids and weights.

$$\gamma_{kn}^{(i)} = \frac{\pi_k^{(i-1)} q(\mathbf{Z}_n | \mathbf{C}_k^{(i-1)})}{\sum_{k=1}^K \pi_k^{(i-1)} q(\mathbf{Z}_n | \mathbf{C}_k^{(i-1)})} \quad (19)$$

Step 4: Update Wishart centroids $\mathbf{C}_k^{(i)}$ and weights $\pi_k^{(i)}$ according to Equations (20) and (21).

$$\mathbf{C}_k^{(i)} = \frac{\sum_{n=1}^N \gamma_{kn}^{(i)} \mathbf{Z}_n}{\sum_{n=1}^N \gamma_{kn}^{(i)}} \quad (20)$$

$$\pi_k^{(i)} = \frac{1}{N} \sum_{n=1}^N \gamma_{kn}^{(i)} \quad (21)$$

Step 5: Repeat Step 3 and Step 4 until convergence. The iteration converges if $\mathbf{C}_k^{(i)}$ and $\pi_k^{(i)}$ do not change much compared with $\mathbf{C}_k^{(i-1)}$ and $\pi_k^{(i-1)}$, according to Equations (22) and (23). The left side of Equation (22) is a symmetrized LogDet divergence between Hermitian positive definite matrices [21]. δ_C and δ_π are predefined thresholds.

$$\frac{1}{2} \text{Tr} \left(\mathbf{C}_k^{(i)} \mathbf{C}_k^{(i-1)-1} + \mathbf{C}_k^{(i-1)} \mathbf{C}_k^{(i)-1} \right) - d < \delta_C \quad (22)$$

$$|\pi_k^{(i)} - \pi_k^{(i-1)}| < \delta_\pi \quad (23)$$

3.3. Determine the Number of Mixture Components

In Section 3.2, an iterative procedure is developed that alternately updates the parameters and weights of all component distributions until convergence is reached, while how to determine the number of mixture components K is left undiscussed. Typically, there are two approaches to determining the number of mixture components. One can either start the iteration with only one component, splitting and merging current components every few iterations, or initialize the iteration by many components, merging those that are sufficiently close in further iterations. Doulgeris [17] adopted the former approach and proposed a goodness-of-fit test based on the matrix log-cumulant to control splitting and merging, which is theoretically sound and can be incorporated in the EM iteration. The following describes the method used in our experiments, which falls into the latter category. At the end of every five iterations, a merge-and-remove step (Step 6) is inserted in the iterative procedure, merging close components and removing those with negligible weights.

Step 6: First compute the LogDet distance between each pair of Wishart centroids $\mathbf{C}_{k_1}^{(i)}$ and $\mathbf{C}_{k_2}^{(i)}$. Merge them into a new component if the distance is less than the predefined threshold δ_C , i.e.,

$$\frac{1}{2} \text{Tr} \left(\mathbf{C}_{k_1}^{(i)} \mathbf{C}_{k_2}^{(i)-1} + \mathbf{C}_{k_2}^{(i)} \mathbf{C}_{k_1}^{(i)-1} \right) - d < \delta_C. \quad (24)$$

The Wishart centroid and weight of the new component are computed according to Equations (25) and (26).

$$\mathbf{C}_{k_1 \cup k_2}^{(i)} = \frac{\pi_{k_1}^{(i)} \mathbf{C}_{k_1}^{(i)} + \pi_{k_2}^{(i)} \mathbf{C}_{k_2}^{(i)}}{\pi_{k_1}^{(i)} + \pi_{k_2}^{(i)}} \quad (25)$$

$$\pi_{k_1 \cup k_2}^{(i)} = \pi_{k_1}^{(i)} + \pi_{k_2}^{(i)} \quad (26)$$

Then, remove the components with weights less than the predefined threshold δ_π , i.e.,

$$\pi_k^{(i)} < \delta_\pi. \quad (27)$$

It has been observed that the overall accuracy of land cover classification generally improves when the number of mixture components K is increased up to 10, while there is only slight improvement if K is further increased, at the cost of more computation. Thus, for efficiency, K is initialized to 6, δ_C and δ_π are set to 10^{-3} , and the procedure usually converges within 50 iterations.

3.4. The Complete Classification Scheme

Two supervised maximum likelihood classifiers are constructed based on the proposed mixture models for single-look and multi-look polarimetric SAR images. Training samples for each of the M land cover types are selected manually. For each class, parameters of a Gaussian mixture model or a Wishart mixture model are estimated using the EM algorithm, including the number of components K_m , the Gaussian covariances or Wishart centroids $\{C_{m,1}, C_{m,2}, \dots, C_{m,K_m}\}$, and the associated weights $\{\pi_{m,1}, \pi_{m,2}, \dots, \pi_{m,K_m}\}$. Then, the distribution of each land cover type can be expressed by the weighted sum of all components, as in Equation (15). The mixture classifiers label a test sample with the class that has the largest likelihood value. Furthermore, the proposed classification scheme can be extended to apply to multi-frequency polarimetric SAR images if it is assumed that the images from all radar frequencies are independent. In Section 4, classification results produced by the mixture classifiers are compared with those produced by classical ones.

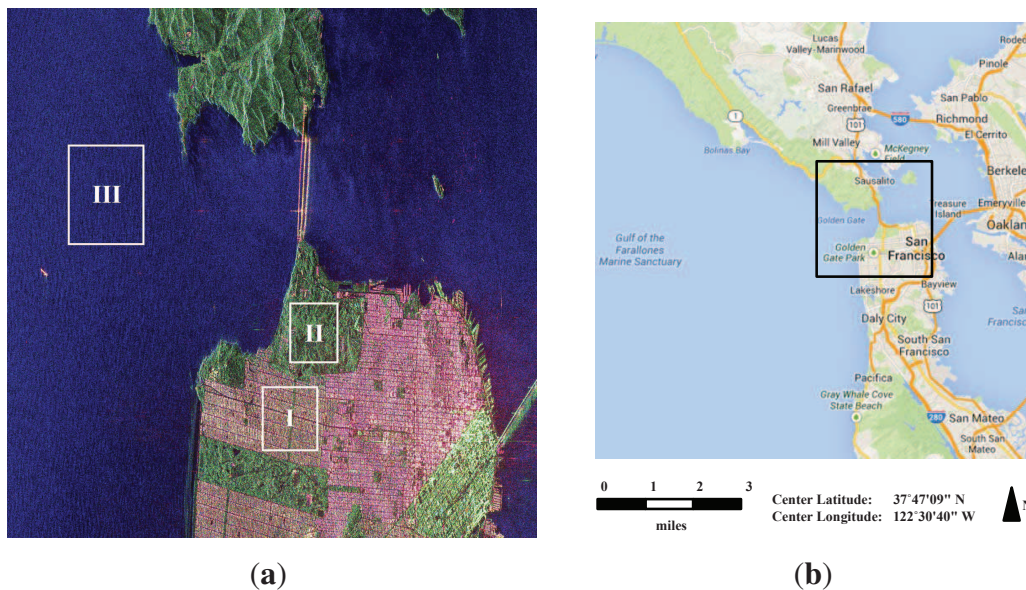
4. Results and Discussion

To demonstrate the effectiveness of the proposed mixture models, experiment results will be presented with discussions. These models are compared with existing models in the aspects of distribution fitting and land cover classification using real polarimetric SAR images.

4.1. Fitting Distributions of Heterogeneous Regions

In the first experiment, distributions of heterogeneous regions are fitted for both single-look and multi-look polarimetric SAR images using mixture models and classical models. The data set used in this experiment is a C-band fully polarimetric SAR image acquired by the RADARSAT-2 sensor over the San Francisco bay area. Figure 1 shows the pseudocolor image of San Francisco and its geographic information. The original data is in single-look format, and the multi-look data is obtained by incoherently averaging 2×2 (range \times azimuth) neighboring pixels. The scene mainly contains three typical land cover types, corresponding to urban areas (I), forest areas (II) and ocean areas (III), as shown in Figure 1a. The selected urban and forest regions in Figure 1a are both heterogeneous by observation. Applying the standard Gaussian and Wishart models to these regions will cause large modeling errors, as will be seen later.

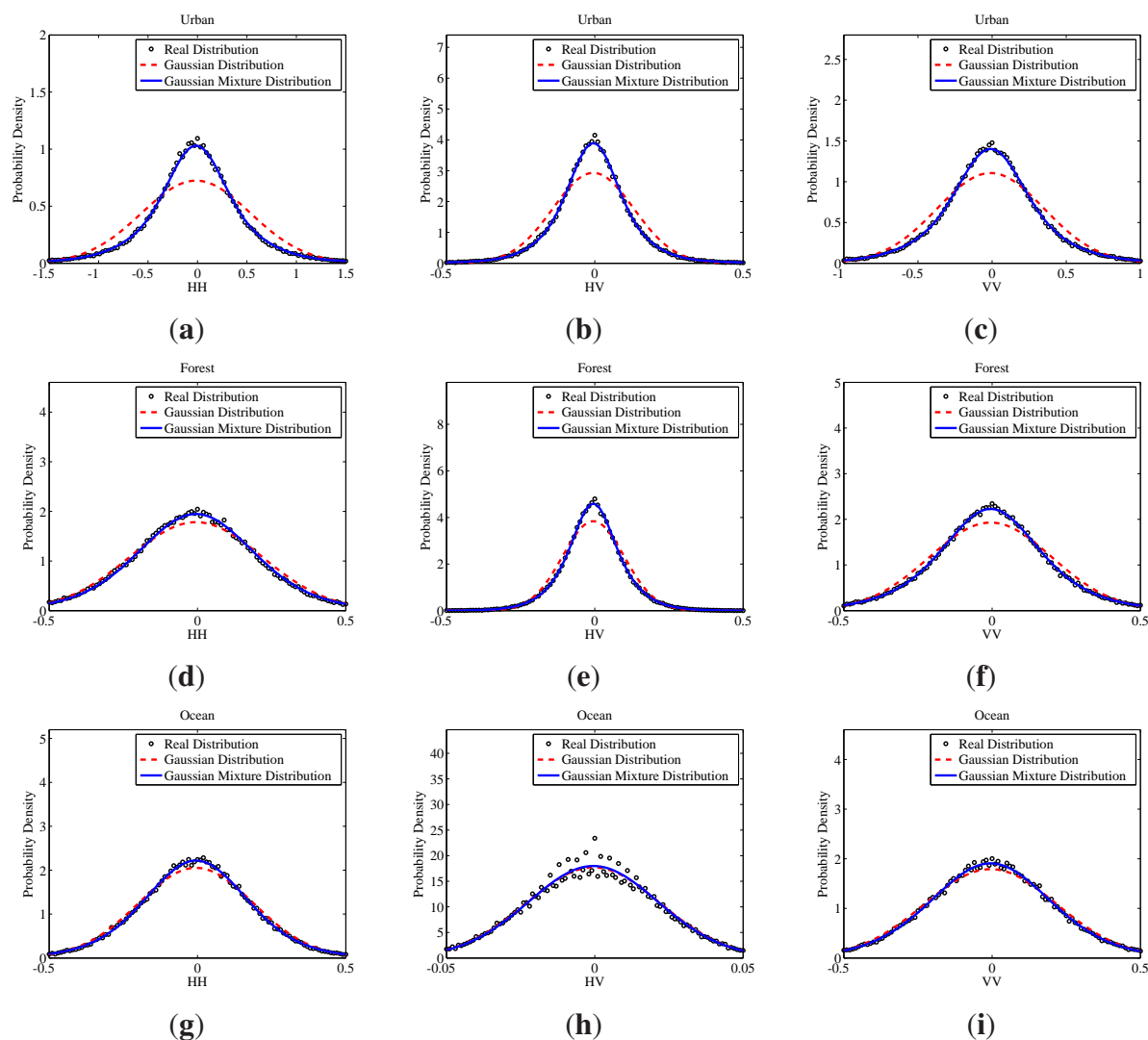
Figure 1. (a) Pseudocolor image of San Francisco with three land cover types, red = $|S_{hh} - S_{vv}|$, green = $|S_{hv}|$, blue = $|S_{hh} + S_{vv}|$; (b) geographic information.



In the single-look case, the mixture Gaussian model is compared with the standard Gaussian model. For each land cover type, the distribution of real and imaginary parts of the scattering vectors is fitted. The fitting results are shown in Figure 2, in which each row corresponds to a land cover type and each column corresponds to a polarization channel. In the multi-look case, the mixture Wishart model is compared with the standard Wishart model and the product model with a K-distribution. A multi-look covariance matrix can be recognized as a nine-dimensional random vector, and visualization will be tedious if all dimensions are concerned. Therefore, only the fitting results of the diagonal elements are presented in Figure 3, which correspond to the scattering coefficients of HH, HV and VV channels. The subfigures in Figure 3 are similarly arranged.

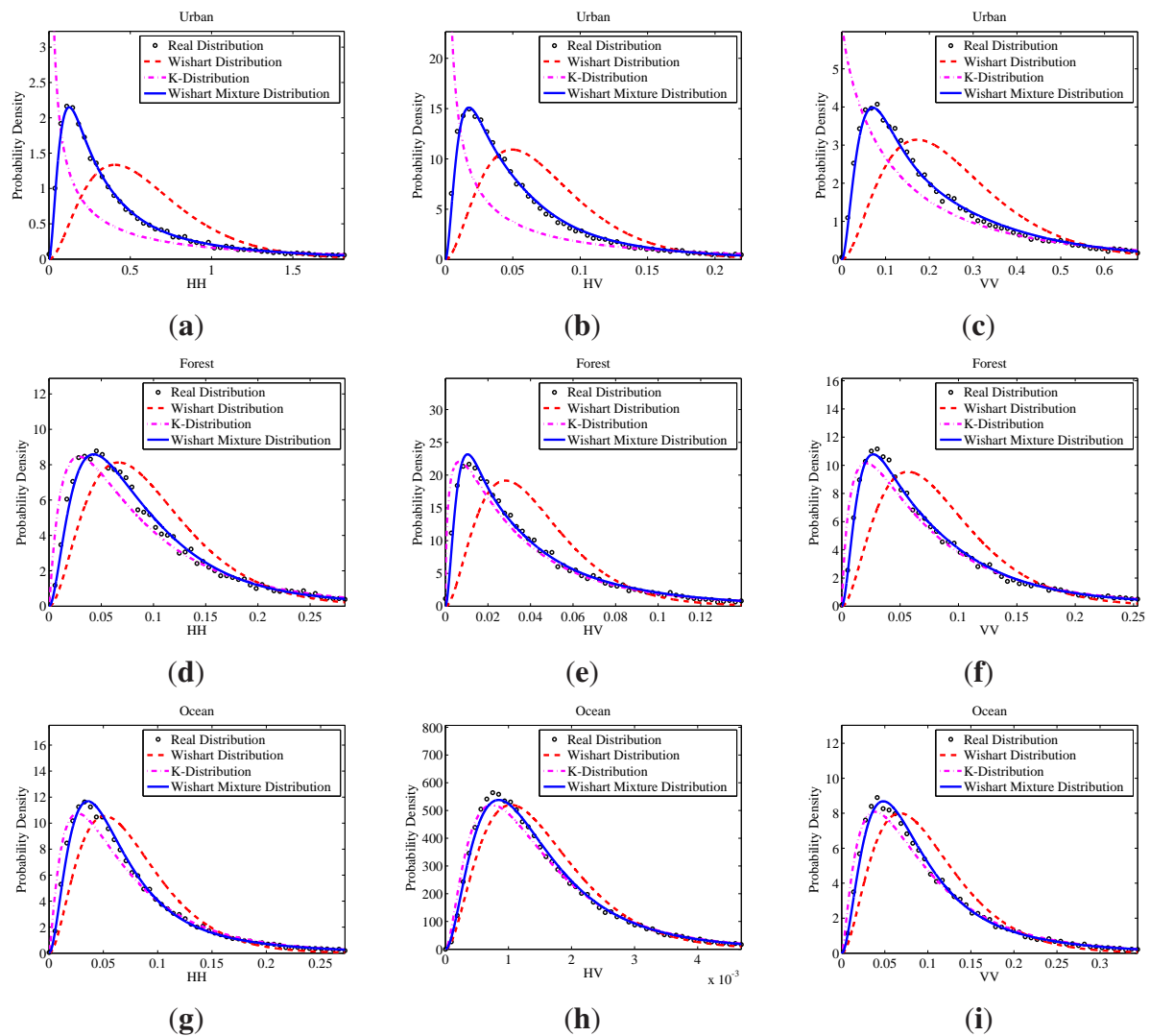
As can be observed in Figure 2a–c, in urban areas, the fitted Gaussian distribution (red dashed line) deviates largely from the real distribution (black circle) for each polarization channel. This is an indication that the urban area is highly heterogeneous. Applying the Gaussian model to these regions leads to considerable modeling errors. By contrast, the distribution from the Gaussian mixture model (blue line) coincides tightly with the real distribution for all channels, demonstrating that the Gaussian mixture model is effective in describing heterogeneous regions. The forest areas are less heterogeneous. In the selected forest regions, the Gaussian model could fit the HH channel with tolerable error, but is still not so accurate as the Gaussian mixture model for the HV and VV channels, as shown in Figure 2d–f. The ocean areas are nearly homogeneous. Both models fit the data well in all channels with slight errors, as indicated by Figure 2g–i.

Figure 2. Fitting results for single-look data. (a–c) HH, HV and VV of urban areas; (d–f) HH, HV and VV of forest areas; (g–i) HH, HV and VV of ocean areas.



The fitting results for multi-look data presented in Figure 3 also support the superiority of mixture models. It can be noticed in Figure 3a–c that the Wishart model is not satisfactory for modeling the urban areas, since significant differences can be observed between the real distributions (black circle) and those from the Wishart model (red dashed line). The product model with K-distribution (magenta dotted dashed line) can handle moderate heterogeneity rather well [5], but has difficulty in describing extremely heterogeneous regions, such as urban areas, since the gamma texture assumption does not hold anymore. Compared with these two models, the Wishart mixture model (blue line) fits all the distributions with less errors, not only in the forest and ocean areas, but also in the extremely heterogeneous urban areas.

Figure 3. Fitting results for multi-look data. (a–c) HH, HV and VV of urban areas; (d–f) HH, HV and VV of forest areas; (g–i) HH, HV and VV of ocean areas.



4.2. Classification of Polarimetric SAR Images

In the second experiment, the classifiers based on the mixture models are evaluated using both single-look and multi-look polarimetric SAR images. A coastal region in the San Francisco image is of interest, as shown in Figure 4. To classify the pixels in this region into three classes that correspond to urban, forest and ocean areas, the Gaussian classifier and its mixture version are applied to the single-look image, while the Wishart classifier and its mixture version are applied to the multi-look image. Figure 5 shows the classification maps produced by the four classifiers, with red, green and blue pixels representing urban, forest and ocean areas, respectively. The confusion matrix and the overall accuracy of each classifier are listed in Tables 1 and 2. Classification rates and overall accuracies are boldfaced.

Figure 4. (a) Pseudocolor image of San Francisco, red = $|S_{hh} - S_{vv}|$, green = $|S_{hv}|$, blue = $|S_{hh} + S_{vv}|$; (b) geographic information.

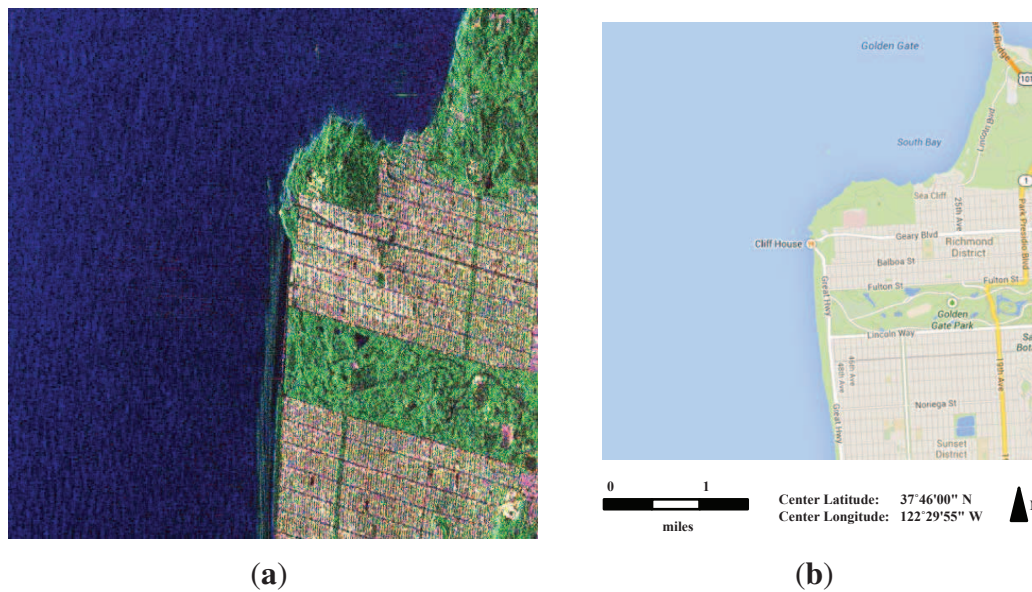


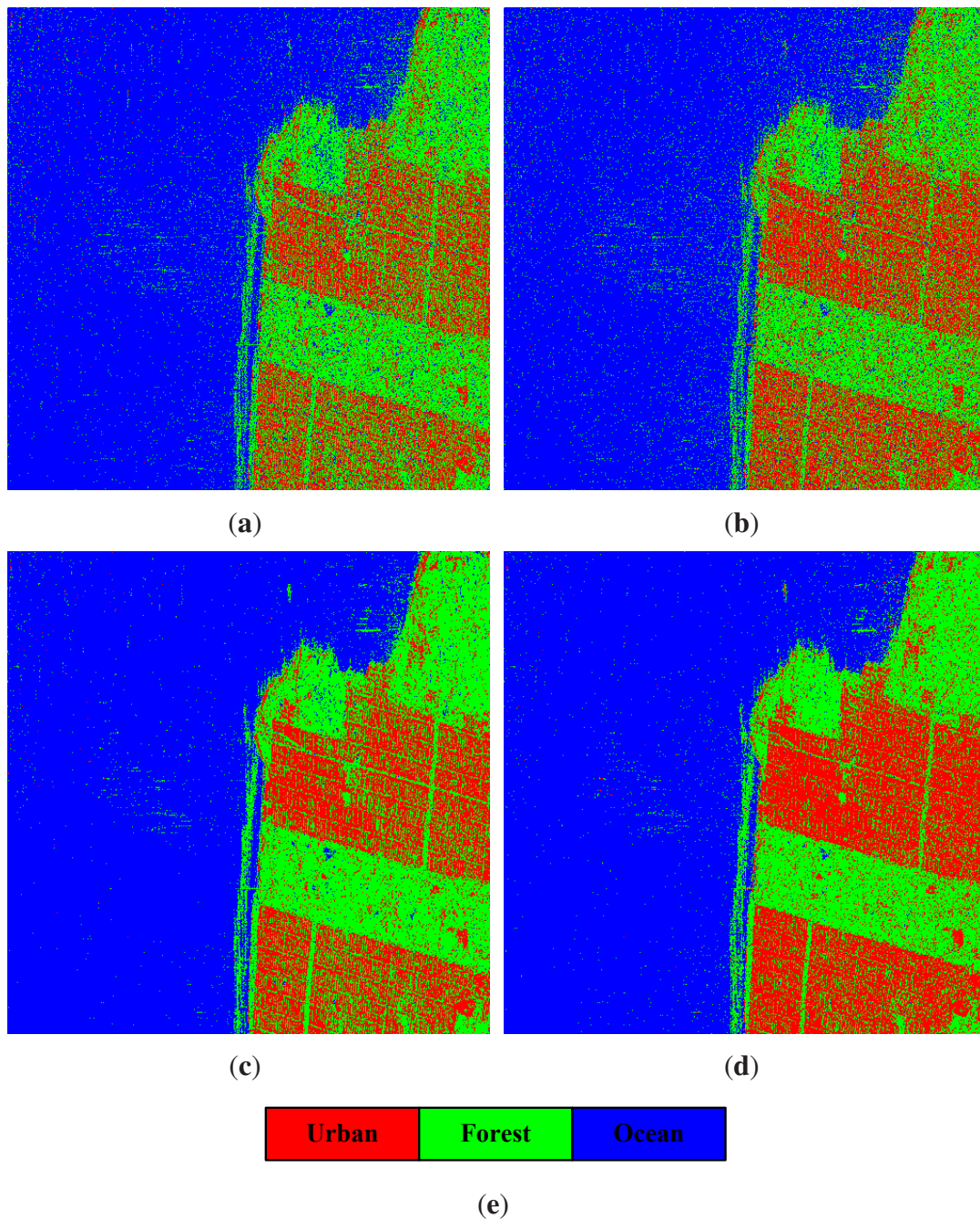
Table 1. Classification results for the single-look San Francisco image; an average of 100 realizations.

Confusion Matrix (%)	Gaussian			Gaussian Mixture		
	Urban	Forest	Ocean	Urban	Forest	Ocean
Urban	56.07	40.77	3.16	65.35	31.83	2.82
Forest	11.02	79.30	9.68	17.13	75.43	7.44
Ocean	0.25	1.93	97.82	0.18	3.17	96.65
Overall Accuracy (%)	77.73			79.14		

Table 2. Classification results for the multi-look San Francisco image; an average of 100 realizations.

Confusion Matrix (%)	Wishart			Wishart Mixture		
	Urban	Forest	Ocean	Urban	Forest	Ocean
Urban	68.88	31.01	0.11	80.37	19.55	0.08
Forest	8.08	89.44	2.48	11.62	87.41	0.97
Ocean	0.07	0.41	99.52	0.04	0.59	99.37
Overall Accuracy (%)	85.95			89.05		

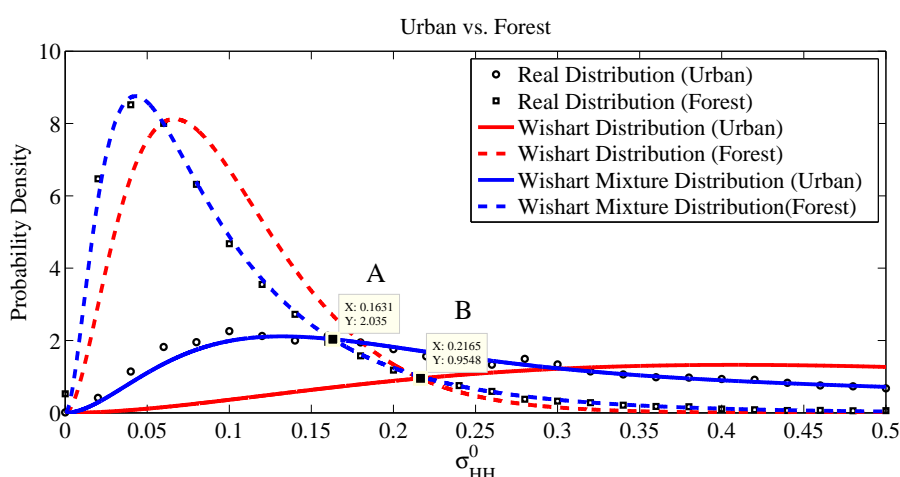
Figure 5. Classification maps of San Francisco. (a) Gaussian classifier; (b) Gaussian mixture classifier; (c) Wishart classifier; (d) Wishart mixture classifier; (e) legend.



Two major points are illustrated by the classification maps in Figure 5. Firstly, multi-look processing of polarimetric SAR images improves the classification results by reducing the speckle noise. The classification maps produced by the multi-look image are less noisy than those produced by the single-look image. Secondly, the mixture models enhance the classification rate of urban areas by reducing modeling errors. More urban pixels appear in the classification maps of the mixture classifiers. Improvement is observed more clearly from the figures in Tables 1 and 2. The classification rate of urban areas is significantly increased by the mixture models from 56.07% to 65.35% for the single-look image, and from 68.88% to 80.37% for the multi-look image.

The drop in the classification rate of forest areas can be interpreted by the shift of the decision boundary. As a one-dimensional simplification, the scattering coefficient of the HH polarization is analyzed. As shown in Figure 6, Point B is the maximum likelihood decision boundary determined by the Wishart model, which is apparently suboptimal. When the Wishart mixture model is applied, the decision boundary shifts from Point B to Point A, causing more samples to be classified as urban areas. This will increase the classification rate of urban areas, while decreasing that of forest areas. However, since Point A is nearly optimal, improvement in the overall accuracy is guaranteed.

Figure 6. Illustration of the decision boundary shift.



4.3. Classification of Multi-Frequency Polarimetric SAR Images

The mixture-model-based classifiers derived in Section 3 can be generalized to apply to multi-frequency polarimetric SAR images. As mentioned by Lee [3], polarimetric SAR images of different bands can be considered as statistically independent if the radar frequencies are sufficiently separated. In this case, the joint log-likelihood function is the sum of the log-likelihood functions from all bands.

Classification experiments are run on a set of multi-look, multi-frequency polarimetric SAR images acquired by the AIRSAR sensor with P, L and C bands over a cultivated area at Flevoland in the Netherlands. Pseudocolor images of the three bands are presented in Figure 7, along with the associated geographic information. According to the ground-truth data provided by Hoekman [22], there are mainly 14 types of crops in the scene. The ground-truth data used in this experiment is visualized in Figure 8. The Wishart classifier and the Wishart mixture classifier are employed for comparison using various combinations of the P-, L- and C-band images. Table 3 lists the classification rates under each experiment configuration. In Table 3, each row corresponds to a classification experiment with a certain configuration. The first column indicates which bands are employed in the experiment. The following 14 columns list the classification rate of each type of crop. The last column shows the overall accuracy. For easy comparison, classification results produced by the Wishart mixture classifier are boldfaced, and the best classification rates and overall accuracy are underlined.

From Table 3, it can be noticed that the Wishart mixture classifier performs better than the Wishart classifier in the overall classification accuracy under all configurations. The overall accuracy is improved by at least 4.5%, and the largest improvement of 11.9% occurs at the P band. This once again

demonstrates that the mixture model is generally a better choice for heterogeneous regions. It is obvious that combining multiple bands for classification enhances the overall accuracy, since the information provided by different bands are complementary. The best overall accuracy for single-band classification using the Wishart mixture classifier is given by the L-band image as 87.4%, this is greatly increased to 97.9% when images of all three bands are fused. Another point that worth noting is the superiority of the L-band image over the others in crop classification, which can be explained by the property of electromagnetic waves. P-band waves have good penetration into vegetation, though their wavelengths are too long to discriminate between similar crops. C-band waves have shorter wavelengths, but limited penetrating ability, so the volume scattering mechanism is not fully exploited. L-band waves are a compromise, possessing both abilities of penetration and discrimination.

Figure 7. (a–c) P-, L- and C-band pseudocolor images of Flevoland, red = $|S_{hh} - S_{vv}|$, green = $|S_{hv}|$, blue = $|S_{hh} + S_{vv}|$; (d) geographic information.

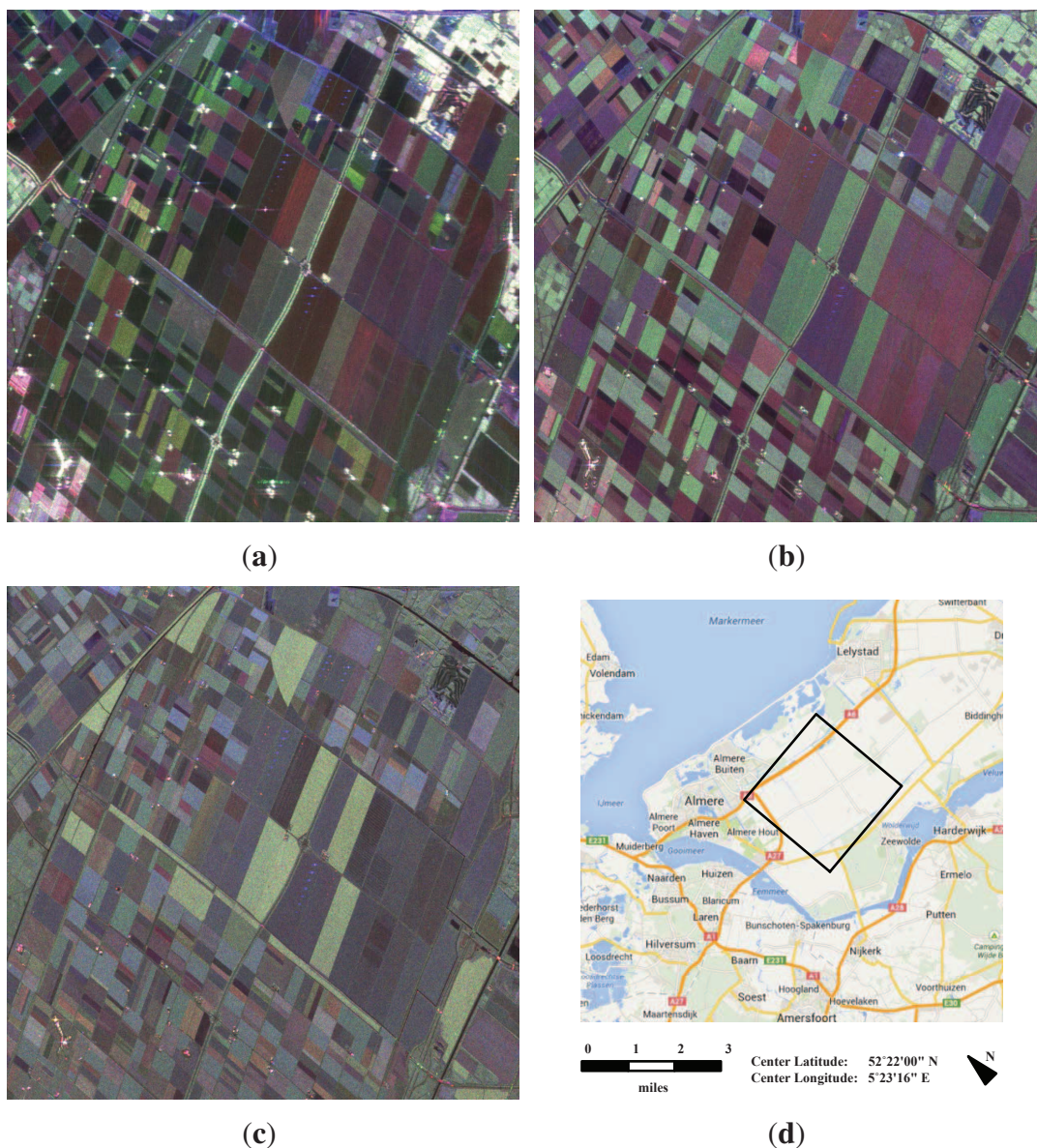
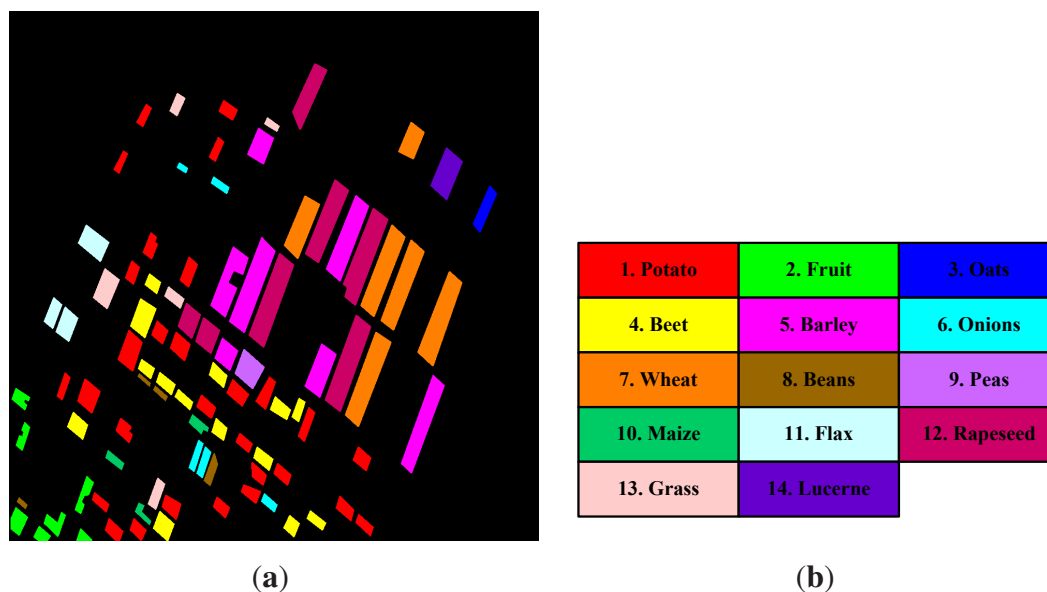


Figure 8. (a) Ground-truth data of Flevoland; (b) Legend.**Table 3.** Classification results for the Flevoland image; an average of 100 realizations. BD, band ; OA, overall accuracy.

BD	Classification Rate (%)														OA (%)
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
P	44.5	98.6	83.6	65.3	55.6	22.0	90.6	18.6	95.4	50.7	23.7	78.6	43.8	95.1	61.9
	77.7	99.3	83.8	43.7	71.8	44.7	93.0	40.6	95.3	61.1	69.2	84.9	72.8	94.9	73.8
L	89.5	88.0	94.9	61.0	76.2	41.0	89.9	88.5	96.1	57.2	95.4	80.8	71.6	95.8	80.4
	89.4	91.7	92.0	82.1	84.2	65.0	91.5	86.8	95.6	77.2	96.9	90.1	86.7	94.8	87.4
C	56.1	29.1	55.6	23.7	52.0	58.2	75.9	94.5	42.1	42.7	76.5	96.4	56.4	38.7	57.0
	53.4	29.5	56.8	26.6	52.3	67.0	69.9	96.1	43.7	42.2	81.0	96.2	64.7	39.5	58.5
PL	96.2	99.5	97.1	85.6	84.1	39.4	98.8	89.8	99.7	65.9	94.5	92.2	75.5	99.5	87.0
	98.0	100.0	96.2	87.7	92.3	78.8	99.5	89.4	99.8	84.9	96.2	98.4	96.2	99.4	94.1
PC	85.0	99.5	92.4	72.4	83.2	64.3	96.1	96.7	97.7	73.9	74.1	99.0	72.9	98.1	86.1
	94.7	99.8	93.2	72.0	88.3	86.7	97.2	99.1	97.9	80.0	88.9	99.1	87.7	97.5	91.6
LC	95.3	93.6	97.6	74.3	90.7	73.1	97.7	98.0	98.2	84.7	96.6	98.9	73.7	98.6	90.8
	96.0	95.2	96.0	90.5	94.0	90.3	96.8	99.4	97.8	89.5	98.4	99.4	92.5	98.1	95.3
PLC	99.0	99.7	<u>98.8</u>	87.3	94.1	69.0	99.5	97.4	99.8	88.7	95.8	99.8	78.4	<u>99.9</u>	93.4
	99.7	100.0	98.2	92.5	97.3	94.8	99.9	99.2	99.9	93.3	98.1	99.9	97.7	99.8	97.9

4.4. Further Comparison with Existing Classifiers

To fully assess the performance of the mixture classifiers, comparisons based on a number of recently proposed statistical models are employed in the following classification experiment, including the K-Wishart model derived from the scale mixture of Gaussian (SMoG) distribution of scattering vectors [8,17], the G_p^0 model from the inverse gamma texture [7], the KummerU model from the Fisher texture [23,24] and the Wishart-Kotz model from the Kotz-type distribution of scattering vectors [25].

Though the underlying assumptions of these models are different, their common destination is addressing the heterogeneity issue by incorporating non-Gaussianity. Classification is performed on a polarimetric SAR image acquired by the EMISAR sensor over the Foulum area in Denmark, which mainly contains seven types of land covers. The pseudocolor image and ground-truth data [26] are depicted in Figures 9 and 10, respectively. Classification results are listed in Table 4 for each classifier, with underlined figures indicating the best classification rates and overall accuracy. The Wishart classifier can be regarded as the baseline classifier, since it ignores heterogeneity in the scattering medium, while the others do not.

Figure 9. (a) Pseudocolor image of Foulum, red = $|S_{hh} - S_{vv}|$, green = $|S_{hv}|$, blue = $|S_{hh} + S_{vv}|$; (b) geographic information.

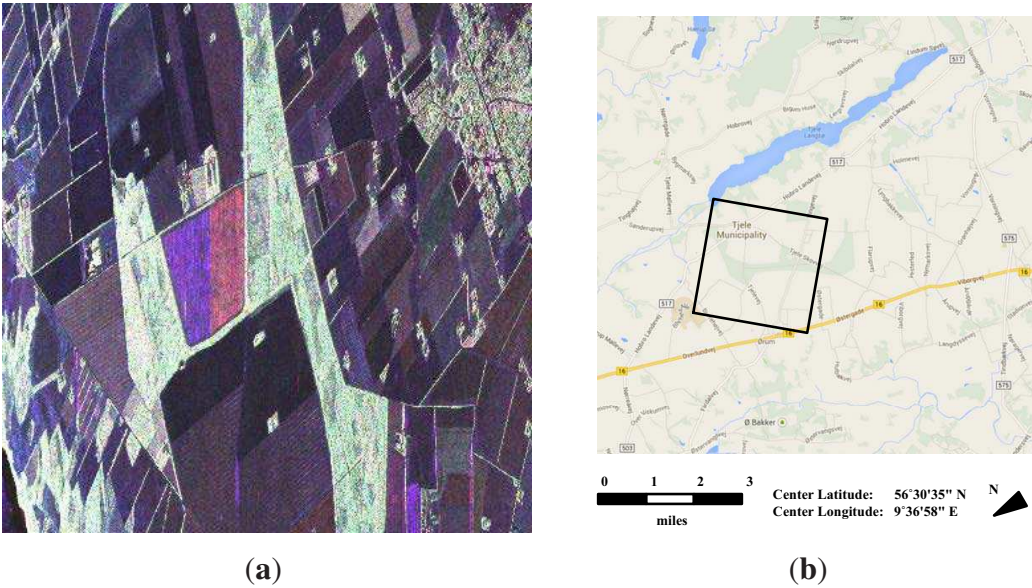
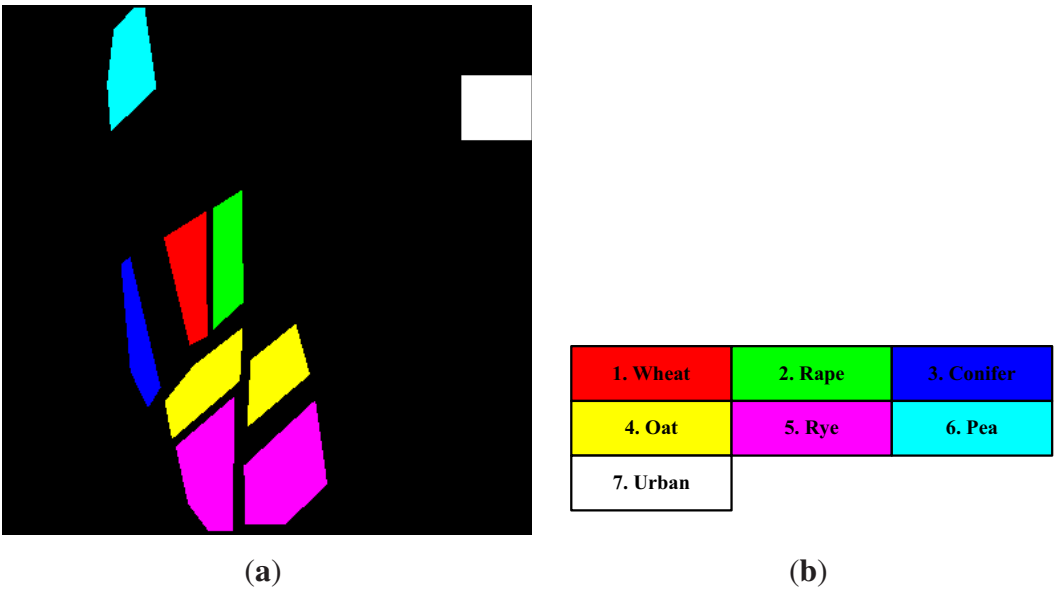


Figure 10. (a) Ground-truth data of Foulum; (b) legend.



As illustrated in Table 4, the classification rates of highly textured conifer (Class 3) and urban (Class 7) areas are remarkably increased when heterogeneity is taken into account. The overall classification

accuracy is improved in different degrees by the non-Gaussian classifiers, among which, the Wishart mixture classifier gives the highest of 81.4%. Actually, the product model can be viewed as a continuous mixture model. The weighted sum in Equation (15) is replaced by the integration in Equation (9), with each conditional distribution $p_{\mathbf{Z}|z}(\mathbf{Z}|z)$ corresponding to a component distribution $q(\mathbf{Z}|\mathbf{C}_k)$, and each prior probability $p_z(z)$ to a component weight π_k . For the product model, parameters of all conditional distributions (e.g., the conditional Wishart centroids in the multi-look case) are identical, except for a scalar multiplication, whereas parameters of all component distributions can take on arbitrary reasonable values for the mixture model. Such flexibility makes the mixture models applicable to a wide range of heterogeneous land covers in polarimetric SAR images.

Table 4. Comparison of classifiers for the Foulum image; an average of 100 realizations. OA, overall accuracy.

Classifier	Classification Rate (%)							OA (%)
	1	2	3	4	5	6	7	
Wishart	<u>93.7</u>	91.8	77.2	<u>96.5</u>	69.4	68.9	46.2	77.7
K-Wishart	92.7	<u>92.4</u>	<u>92.4</u>	91.1	68.7	70.4	52.6	80.0
G _p ⁰	89.9	87.0	92.2	92.2	69.5	69.1	48.3	78.3
KummerU	93.0	91.3	84.9	93.2	69.2	70.1	54.4	79.5
Wishart-Kotz	92.8	89.9	92.2	95.2	70.5	70.2	56.5	81.0
Wishart Mixture	92.8	92.1	88.5	94.2	<u>72.7</u>	<u>71.8</u>	<u>57.5</u>	81.4

5. Conclusions

In this paper, two mixture models have been proposed for modeling heterogeneous regions in polarimetric synthetic aperture radar (SAR) images. For single-look and multi-look data, the mixture of complex Gaussian and complex Wishart distributions were employed, respectively. The expectation-maximization (EM) algorithm was applied for solving the model parameters. Effectiveness of the mixture models, especially in characterizing extremely heterogeneous regions, such as urban areas, has been validated by fitting the models to real polarimetric SAR data from the RADARSAT-2 sensor. Two maximum likelihood classifiers based upon the mixture models have been constructed and then generalized to apply to multi-frequency polarimetric SAR images, assuming independence among images from separate radar frequencies. These classifiers were evaluated using polarimetric SAR images from the RADARSAT-2 sensor, the AIRSAR sensor and the EMISAR sensor. Experiment results have demonstrated that, due to the more precise modeling of heterogeneous regions, the mixture-model-based classifiers clearly improve the overall accuracy of land cover classification, and an average promotion of 5% in the overall accuracy has been noticed compared with the classical Gaussian and Wishart classifiers. By further comparison with classifiers based on recently proposed non-Gaussian statistical models, the mixture classifier was found to give the highest overall accuracy. Beyond classification, statistical modeling of polarimetric SAR data is equally essential in image segmentation and denoising, where the proposed mixture models may also find their applications.

Acknowledgments

This work was in part supported by the National Natural Science Foundation of China (NSFC, No. 41171317), in part supported by the key project of the NSFC (No. 61132008) and in part supported by the Research Foundation of Tsinghua University. The authors would like to thank Gabriel Vasile of the Grenoble Images Parole Signal Automatique Laboratory (GIPSA-Lab, Grenoble, France) for the constructive suggestion on implementing the KummerU model. The authors would also like to thank the anonymous reviewers for the pertinent comments and corrections that helped to improve this article.

Author Contributions

The general idea of mixture modeling for polarimetric SAR data was proposed by Wei Gao. Experiments were carried out by Wei Gao and Wenting Ma, who also prepared the figures. Jian Yang provided much in-depth analysis of the results. This manuscript was written by Wei Gao.

Conflicts of Interest

The authors declare no conflicts of interest.

References

1. Kong, J.A.; Swartz, A.A.; Yueh, H.A.; Novak, L.M.; Shin, R.T. Identification of terrain cover using the optimum polarimetric classifier. *J. Electromagn. Waves Appl.* **1987**, *2*, 171–194.
2. Lim, H.H.; Swartz, A.A.; Yueh, H.A.; Kong, J.A.; Shin, R.T. Classification of earth terrain using polarimetric synthetic aperture radar images. *J. Geophys. Res.* **1989**, *94*, 7049–7057.
3. Lee, J.S.; Grunes, M.R.; Kwok, R. Classification of multi-look polarimetric SAR imagery based on complex Wishart distribution. *Int. J. Remote Sens.* **1994**, *15*, 2299–2311.
4. Yueh, S.H.; Kong, J.A.; Jao, J.K.; Shin, R.T.; Novak, L.M. K-distribution and polarimetric terrain radar clutter. *J. Electromagn. Waves Appl.* **1989**, *3*, 747–768.
5. Lee, J.S.; Schuler, D.L.; Lang, R.H.; Ranson, K.J. K-Distribution for Multi-Look Processed Polarimetric SAR Imagery. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Pasadena, PA, USA, 8–12 August 1994; pp. 2179–2181.
6. Frery, A.C.; Müller, H.J.; Freitas, C.C.; Sant’Anna, S.J.S. A model for extremely heterogeneous clutter. *IEEE Trans. Geosci. Remote Sens.* **1997**, *35*, 648–659.
7. Freitas, C.C.; Frery, A.C.; Correia, A.H. The polarimetric G distribution for SAR data analysis. *Environmetrics* **2005**, *16*, 13–31.
8. Doulgeris, A.P.; Anfinson, S.N.; Eltoft, T. Classification with a non-Gaussian model for PolSAR data. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 2999–3009.
9. Dong, Y.; Milne, A.K.; Forster, B.C. Segmentation and classification of vegetated areas using polarimetric SAR image data. *IEEE Trans. Geosci. Remote Sens.* **2001**, *39*, 321–329.
10. Wu, Y.; Ji, K.; Yu, W.; Su, Y. Region-based classification of polarimetric SAR images using Wishart MRF. *IEEE Geosci. Remote Sens. Lett.* **2008**, *5*, 668–672.

11. Murtagh, F.; Contreras, P.; Starck, J.L. Scale-based Gaussian coverings: Combining intra and inter mixture models in image segmentation. *Entropy* **2009**, *11*, 513–528.
12. Chen, J.; Wang, Y.; Wu, H. A coded aperture compressive imaging array and its visual detection and tracking algorithms for surveillance systems. *Sensors* **2012**, *12*, 14397–14415.
13. Peng, Y.; Chen, J.; Xu, X.; Pu, F. SAR images statistical modeling and classification based on the mixture of alpha-stable distributions. *Remote Sens.* **2013**, *5*, 2145–2163.
14. Gao, G. Statistical modeling of SAR images: A survey. *Sensors* **2010**, *10*, 775–795.
15. Dempster, A.P.; Laird, N.M.; Rubin, D.B. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. Ser. B* **1977**, *39*, 1–38.
16. Kersten, P.R.; Lee, J.S.; Ainsworth, T.L. Unsupervised classification of polarimetric synthetic aperture radar images using fuzzy clustering and EM clustering. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 519–527.
17. Doulgeris, A.P.; Anfinson, S.N.; Eltoft, T. Automated non-Gaussian clustering of polarimetric synthetic aperture radar images. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 3665–3676.
18. Goodman, N.R. Statistical analysis based on a certain multivariate complex Gaussian distribution (An introduction). *Ann. Math. Stat.* **1963**, *34*, 152–177.
19. Chen, J.; Chen, Y.; An, W.; Cui, Y.; Yang, J. Nonlocal filtering for polarimetric SAR data: A pretest approach. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 1744–1754.
20. Yu, P.; Qin, A.K.; Clausi, D.A. Unsupervised polarimetric SAR image segmentation and classification using region growing with edge penalty. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 1302–1317.
21. Kulis, B.; Sustik, M.A.; Dhillon, I.S. Low-rank kernel learning with Bregman matrix divergences. *J. Mach. Learn. Res.* **2009**, *10*, 341–376.
22. Hoekman, D.H.; Vissers, M.A.M. A new polarimetric classification approach evaluated for agricultural crops. *IEEE Trans. Geosci. Remote Sens.* **2003**, *41*, 2881–2889.
23. Bombrun, L.; Beaulieu, J.M. Fisher distribution for texture modeling of polarimetric SAR data. *IEEE Geosci. Remote Sens. Lett.* **2008**, *5*, 512–516.
24. Bombrun, L.; Vasile, G.; Gay, M.; Totir, F. Hierarchical segmentation of polarimetric SAR images using heterogeneous clutter models. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 726–737.
25. Kersten, P.R.; Anfinson, S.N.; Doulgeris, A.P. The Wishart-Kotz Classifier for Multilook Polarimetric SAR Data. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Munich, Germany, 22–27 July 2012; pp. 3146–3149.
26. Skriver, H.; Dall, J.; Le Toan, T.; Quegan, S.; Ferro-Famil, L.; Pottier, E.; Lumsdon, P.; Moshammer, R. Agriculture Classification Using PolSAR Data. In Proceedings of the 2nd International Workshop on Science and Applications of SAR Polarimetry and Polarimetric Interferometry, Frascati, Italy, 17–21 January 2005; pp. 213–218.