

A Physical Analysis of Polarimetric SAR Data Statistical Models

Xinping Deng, Carlos López-Martínez, *Senior Member, IEEE*, and Eduardo Makhoul Varona, *Member, IEEE*

Abstract—The random walk model is studied with the objective to obtain a physical explanation for the texture of polarimetric synthetic aperture radar (SAR) data. A simulator is designed to imitate the scattering process under different circumstances, taking into account different distributions for the scatterer number and the scatterer response, as well as mixtures of scatterers. Statistical analysis of the simulated data shows that the distribution of the scatterer response has an effect only when the number of scatterers in a resolution cell is very small, which appears in very high resolution data. Moreover, it is found that both the fluctuation of the scatterer number and the mixture of different targets can give non-Gaussian-distributed data. The mixture of point targets and distributed targets will lead to an extremely heterogeneous appearance, which may be a clue to analyze the urban areas in polarimetric SAR data.

Index Terms—Polarimetric synthetic aperture radar (PolSAR), random walk model, simulation, statistical model, texture analysis.

I. INTRODUCTION

KNOWLEDGE of the exact statistical properties of the signal plays an important role in the applications of polarimetric synthetic aperture radar (PolSAR) data. Under the assumption of Born approximation [1], the data received from a resolution cell can be regarded as a sum of a number of complex phasors, each resulting from an individual scatterer or a separate subregion of a continuous scattering surface. The addition of the independent and randomly located complex contributions gives rise to speckle. The classical model for homogeneous data assumes that the scatterer number in each resolution cell is big enough to give Gaussian-distributed speckle [2]. However, the \mathcal{K} -distribution obtained by letting the scatterer number follow a negative binomial distribution is proved to be a more accurate representation for the speckle in areas such as sea surface and forest [3]–[5].

Manuscript received May 22, 2015; revised October 22, 2015; accepted December 14, 2015. This work was supported by the projects TEC2011-28201-C02-01 and TIN2014-55413-C2-1-P funded by the Spanish MICINN and FEDER funds.

X. Deng and C. López-Martínez are with the Remote Sensing Laboratory, Signal Theory and Communications Department, Universitat Politècnica de Catalunya, 08034 Barcelona, Spain (e-mail: xinping.deng@upc.edu; carlos.lopez@tsc.upc.edu).

E. M. Varona was with the Department of Signal Theory and Communications, Universitat Politècnica de Catalunya, 08034 Barcelona, Spain. He is now with isardSAT, 08042 Barcelona, Spain (e-mail: eduard.makhoul@isardsat.cat).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TGRS.2015.2510399

The \mathcal{K} -distribution can be formulated into a product of two independent random variables: one is gamma distributed with a long spatial correlation length, and the other follows a Gaussian distribution with a much shorter spatial correlation length [6]. The former is referred to as a texture parameter, and the latter is referred to as a speckle [7], [8]. Based on this product scheme, a number of models have been proposed by altering the distribution of the texture parameter. For instance, the \mathcal{G}^0 -distribution, suggested to model urban areas, is obtained by introducing an inverse gamma-distributed texture parameter [9]. The Kummer- \mathcal{U} distribution has been recently proposed, which models the texture component with a Fisher distribution [10], [11], providing a wide variety of textures. Other statistical models include the generalized \mathcal{K} -distribution [12], the generalized Gaussian distribution [13], the \mathcal{W} -distribution, and the \mathcal{M} -distribution [14].

Furthermore, assuming that the electromagnetic wave sees different geometrical or dielectric properties of the target according to the wave polarization, several models known as multitexture models allow different texture parameters for polarimetric channels. For example, Lombardo *et al.* assume that the texture parameters follow a partially correlated gamma distribution [15], [16], Khan *et al.* suppose that the co-pol and cross-pol channels have independent textures that follow generalized inverse Gaussian distributions [17], and in the work of Eltoft and Doulgeris [18]–[20], the texture parameters of co-pol and cross-pol channels can be either independent or identical.

However, unlike the Gaussian distribution and the \mathcal{K} -distribution, most of the aforementioned models lack a physical explanation of the underlying scattering problem. They give no clues about why the data following a specific distribution are obtained. In this paper, a simulator based on the random walk model [2], [3] is developed. With it, we can simulate data under certain scattering scenarios by controlling the factors that may appear in the PolSAR data formation process. Then, a suitable texture model, \mathcal{K} -distribution or \mathcal{G}^0 -distribution for example, can be found for the simulated data via statistical analysis. Thus, the scattering process and the statistical data models are bridged, and possible physical explanations can be given to some models.

The remainder of this paper is organized as follows. Section II presents a brief review of the statistical models in PolSAR data, along with the tools employed to analyze those distributions. Section III details the simulator. Algorithms on how to simulate different scatterer types and scatterer numbers are given. Experimental results are shown in Sections IV and V, and the conclusion is in Section VI.

TABLE I
COMPARISON OF POLSAR DATA MODELS

Models	Texture Parameter	Texture Distribution $p(\tau)$	Sample Covariance Matrix Distribution $p(\mathbf{C}_L)$
Wishart	None	$p(\tau) = \delta(\tau - 1)$	$\frac{L^d}{I(L,d)} \times \frac{ \mathbf{C}_L ^{L-d}}{ \boldsymbol{\Sigma} ^L} \exp(-L \text{Tr}(\boldsymbol{\Sigma}^{-1} \mathbf{C}_L))$
\mathcal{K}	α	$\frac{(\alpha\tau)^\alpha}{\tau\Gamma(\alpha)} \exp(-\alpha\tau)$	$\frac{2}{I(L,d)\Gamma(\alpha)} \times (L\alpha)^{\frac{\alpha+Ld}{2}} \times \frac{ \mathbf{C}_L ^{L-d}}{ \boldsymbol{\Sigma} ^L} (\text{Tr}(\boldsymbol{\Sigma}^{-1} \mathbf{C}_L))^{\frac{\alpha-Ld}{2}} \times K_{\alpha-Ld} \left(2\sqrt{L\alpha \text{Tr}(\boldsymbol{\Sigma}^{-1} \mathbf{C}_L)} \right)$
\mathcal{G}^0	λ	$\frac{(\lambda-1)^\lambda}{\Gamma(\lambda)\tau^{\lambda+1}} \exp\left(-\frac{\lambda-1}{\tau}\right)$	$\frac{\Gamma(Ld+\lambda)}{I(L,d)\Gamma(\lambda)} \times L^{Ld} (\lambda-1)^\lambda \times \frac{ \mathbf{C}_L ^{L-d}}{ \boldsymbol{\Sigma} ^L} (L \text{Tr}(\boldsymbol{\Sigma}^{-1} \mathbf{C}_L) + \lambda - 1)^{-\lambda-Ld}$
Kummer- \mathcal{U}	ξ, ζ	$\frac{\xi\Gamma(\xi+\zeta)}{(\zeta-1)\Gamma(\xi)\Gamma(\zeta)} \left(\frac{\xi}{\zeta-1}\tau\right)^{\xi-1} \left(\frac{\xi}{\zeta-1}\tau+1\right)^{\xi+\zeta}$	$\frac{\Gamma(\xi+\zeta)\Gamma(Ld+\zeta)}{I(L,d)\Gamma(\xi)\Gamma(\zeta)} \times \left(\frac{L\xi}{\zeta-1}\right)^{Ld} \times \frac{ \mathbf{C}_L ^{L-d}}{ \boldsymbol{\Sigma} ^L} \times U\left(Ld+\zeta, Ld-\xi+1, \frac{L \text{Tr}(\boldsymbol{\Sigma}^{-1} \mathbf{C}_L)\xi}{\zeta-1}\right)$
\mathcal{W}	ξ, ζ	$\frac{\xi\Gamma(\zeta)}{\zeta\Gamma(\xi)\Gamma(\zeta-\xi)} \left(\frac{\xi\tau}{\zeta}\right)^{\xi-1} \times \left(1 - \frac{\xi\tau}{\zeta}\right)^{\zeta-\xi-1}$	$\frac{\Gamma(\zeta)}{I(L,d)\Gamma(\xi)} \times \left(\frac{L\xi}{\zeta}\right)^{\frac{\xi+Ld-1}{2}} \times \frac{ \mathbf{C}_L ^{L-d}}{ \boldsymbol{\Sigma} ^L} (\text{Tr}(\boldsymbol{\Sigma}^{-1} \mathbf{C}_L))^{\frac{\xi-Ld-1}{2}} \exp\left(-\frac{L\xi}{\zeta} \text{Tr}(\boldsymbol{\Sigma}^{-1} \mathbf{C}_L)\right) \times W_{Ld+\xi+1-2\zeta, \frac{Ld-\xi}{2}} \left(\frac{L\xi}{\zeta} \text{Tr}(\boldsymbol{\Sigma}^{-1} \mathbf{C}_L)\right)$
\mathcal{M}	ξ, ζ	$\frac{(\zeta-1)\Gamma(\zeta)}{(\xi-1)\Gamma(\xi)\Gamma(\zeta-\xi)} \left(\frac{\zeta-1}{\xi-1}\tau\right)^{-\zeta} \times \left(\frac{\zeta-1}{\xi-1}\tau-1\right)^{\zeta-\xi-1}$	$\frac{\Gamma(\zeta)\Gamma(Ld+\xi)}{I(L,d)\Gamma(\xi)\Gamma(Ld+\zeta)} \times \left(\frac{L(\zeta-1)}{\xi-1}\right)^{\frac{Ld-\zeta}{2}} \times \frac{ \mathbf{C}_L ^{L-d}}{ \boldsymbol{\Sigma} ^L} (\text{Tr}(\boldsymbol{\Sigma}^{-1} \mathbf{C}_L))^{\frac{-Ld-\zeta}{2}} \exp\left(-\frac{L(\zeta-1)}{2(\xi-1)} \text{Tr}(\boldsymbol{\Sigma}^{-1} \mathbf{C}_L)\right) \times M_{\zeta-Ld-2\xi, \frac{Ld+\zeta-1}{2}} \left(-\frac{L(\zeta-1)}{\xi-1} \text{Tr}(\boldsymbol{\Sigma}^{-1} \mathbf{C}_L)\right)$

Note: K_v is the modified Bessel function of the second kind of order v , U is the hyper-geometric function of the second kind, W is the Whittaker W function, and M is the hypergeometric function of the first kind (Kummer M function).

II. POLARIMETRIC SAR DATA MODELS

A. Product Model

The observed data of a PolSAR system can be formulated as scattering vectors, which can be written as follows under the assumption of the product model [6]:

$$\mathbf{k} = \sqrt{\tau} \mathbf{z}. \quad (1)$$

Here, τ is the texture parameter, a positive random variable with a mean value equal to 1, and \mathbf{z} is the speckle vector, following a multivariate complex Gaussian distribution with covariance matrix given by $\boldsymbol{\Sigma} = E\{\mathbf{z}\mathbf{z}^\dagger\}$, where $E\{\cdot\}$ is the expectation operator and $(\cdot)^\dagger$ is the transpose conjugate operation. The matrix $\boldsymbol{\Sigma}$ contains the polarimetric information.

For speckle reduction, the scattering vectors are frequently multilook processed and represented by sample covariance matrices. Assume that the texture has a higher spatial correlation than the speckle [21], which means that the texture parameter is constant over the multilook processing window, and the sample covariance matrix can be expressed as

$$\mathbf{C}_L = \frac{1}{L} \sum_{i=1}^L \mathbf{k}_i \mathbf{k}_i^\dagger = \frac{\tau}{L} \sum_{i=1}^L \mathbf{z}_i \mathbf{z}_i^\dagger \quad (2)$$

where L is the number of looks. It is known that the conditional distribution of \mathbf{C}_L given on τ is a Wishart distribution [21]. Let d denote the number of dimensions and $|\cdot|$ denote the matrix

determinant; then, the marginal probability density function (pdf) is obtained by averaging all possible τ

$$p(\mathbf{C}_L; L, \boldsymbol{\Sigma}) = \int_0^\infty \frac{L^d |\mathbf{C}_L|^{L-d}}{I(L,d)\tau^{Ld} |\boldsymbol{\Sigma}|^L} \times \exp\left(-\frac{L}{\tau} \text{Tr}(\boldsymbol{\Sigma}^{-1} \mathbf{C}_L)\right) p(\tau) d\tau \quad (3)$$

where $\text{Tr}(\cdot)$ denotes the matrix trace operation and $I(L, d)$ is a normalization factor given by

$$I(L, d) = \pi^{\frac{d(d-1)}{2}} \prod_{i=1}^d \Gamma(L - i + 1) \quad (4)$$

with Γ representing the gamma function.

A number of models have been proposed in the literature by assuming different distributions for the texture parameter τ , among which the Wishart distribution [22], the \mathcal{K} -distribution [5], [21], [23], the \mathcal{G}^0 -distribution [9], [24], and the Kummer- \mathcal{U} distribution [10], [25] are extensively studied. The distributions of the corresponding texture parameter and the sample covariance matrix are shown in Table I.

B. Log-Cumulants

To test if a model is suitable for the data, a measurable statistic or possible method for parameter estimation is required. It

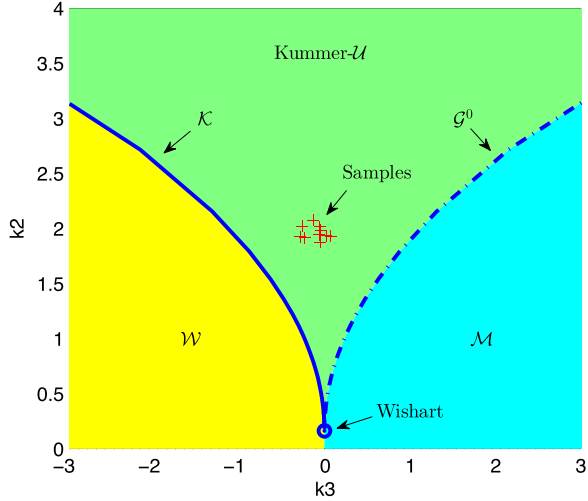


Fig. 1. Complete coverage of log-cumulant $\kappa_2 - \kappa_3$ space where different distributions are placed in different regions.

is demonstrated that Mellin kind statistics are of great value for the analysis of the sample covariance matrix and that they can be employed to derive estimators for the distribution parameters with low bias and variance compared with classical algorithms [26], [27].

Define the Mellin kind matrix-variate characteristic function as the Mellin transform of the pdf

$$\phi(s) = \int_{\Omega_+} |\mathbf{Z}|^{s-d} p(\mathbf{Z}) d\mathbf{Z} \quad (5)$$

then, the v th-order log-cumulant or Mellin kind cumulant is given by

$$\kappa_v = \left. \frac{d^v}{ds^v} \ln \phi(s) \right|_{s=d}. \quad (6)$$

Meanwhile, the sample log-cumulants can be estimated from the data by

$$\hat{\kappa}_v = \hat{\mu}_v - \sum_{i=1}^{v-1} \binom{v-1}{i-1} \hat{\kappa}_i \hat{\mu}_{v-i} \quad (7)$$

where $\hat{\mu}_v$ is the estimated log-moments

$$\hat{\mu}_v = \frac{1}{M} \sum_{i=1}^M (\ln |\mathbf{C}_i|)^v \quad (8)$$

with M denoting the number of samples and \mathbf{C}_i denoting the i th sample covariance matrix.

To see if a signal model is applicable, we can compare the log-cumulants calculated from the pdf (κ_v) and those estimated from the sample data ($\hat{\kappa}_v$). In [26], a diagram is proposed to visualize the comparison by plotting the second-order log-cumulants κ_2 against third-order log-cumulants κ_3 in a plane, where different distributions are placed in different regions. Fig. 1 shows the distributions making up a complete coverage of the diagram [14].

III. POLARIMETRIC SAR DATA SIMULATOR

A. Random Walk Model

Based on the Born approximation or simple scattering approximation [1], if there are N individual scatterers randomly distributed inside a resolution cell, the received polarimetric data can be written as the sum of their responses

$$\mathbf{k} = \frac{1}{\sqrt{N}} \sum_{i=1}^N \mathbf{s}_i \quad (9)$$

where \mathbf{s}_i denotes the response from a single scatterer. The normalizing factor $1/\sqrt{N}$ is used to assure that the average intensity of the speckle is independent of N . In this model, the correlation between resolution cells is not considered; therefore, N is a positive integer. Equation (9) provides a physical insight into the scattering problem. The dimension of \mathbf{k} is taken to be 4, corresponding to the polarimetric channels S_{hh} , S_{hv} , S_{vh} , and S_{vv} , respectively, where h denotes the horizontal polarization state and v denotes the vertical one. For the backscattering from a reciprocal medium, the cross polarization channels are equal ($S_{hv} = S_{vh}$), and then, the dimension is reduced to 3, which is the case mainly discussed here, without loss of generality. The covariance matrix is given by

$$\mathbf{C} = E\{\mathbf{k}\mathbf{k}^\dagger\} = E\{\mathbf{s}\mathbf{s}^\dagger\}. \quad (10)$$

There are several factors affecting the distribution of the vector \mathbf{k} . The first one is the scatterer type that characterizes the response of each scatterer [28]. Let a denote the amplitude and ϕ denote the phase, and a scatterer can be represented by

$$\mathbf{s} = [a_{hh}e^{j\phi_{hh}}, a_{hv}e^{j\phi_{hv}}, a_{vv}e^{j\phi_{vv}}]^T. \quad (11)$$

For each polarimetric channel, the amplitudes and the phases are statistically independent. Moreover, the phases are assumed to be uniformly distributed over $[0, 2\pi)$. This assumption arises from the fact that the path differences in one resolution cell are much greater than the wavelength so that any value of the phase is equally probable [7]. Different scatterer types can be distinguished by the distributions of their amplitudes.

The second factor is the scatterer number N . On one hand, a lot of research has shown that the fluctuation of scatterer numbers will give rise to the texture [3], [5], which conveys information that may enable a user to identify different cover types. On the other hand, for the high-resolution data, where the scatterer number is finite and small, it is known that the classical models such as Gaussian distribution usually fail to give an accurate representation [29].

At last, the heterogeneity that appeared in PolSAR data can result from the mixture of different targets. For instance, from an urban area which usually consists of discrete objects, like houses, trees, and roads, the backscattering is a combination of different scattering mechanisms. To represent this kind of data, a simple model would be inappropriate [30]. Mixture models, instead, could achieve a reasonable level of accuracy [30]–[32].

Taking into account all of the aforementioned factors, a simulator that helps us to study the statistics of PolSAR data is designed. The details of the implementation are given in the following sections, provided that the algorithms that are used

```

1: procedure POINTSCATTERER( $s'$ )
2:    $\phi \sim U(0, 2\pi)$  ▷  $U$ , uniform distribution
3:    $\mathbf{s} \leftarrow \mathbf{s}' \cdot e^{j\phi}$ 
4:   return  $\mathbf{s}$ 
5: end procedure

```

Fig. 2. Algorithm to simulate a point scatterer.

to sample the univariate uniform, Gaussian, gamma, and beta distributions exist.

B. Scatterer Type

When the radar transmits a perfectly monochromatic wave and this wave reaches a fixed or stationary target, it results into a perfectly polarized scattered wave. The scattering process can be completely represented by the scattering vector. This type of targets is referred to as point scatterers or coherent scatterers. However, most of the targets under observation are situated in a dynamically changing environment and are subjected to spatial and temporal variations. Despite the radar system transmitting a perfectly polarized wave, the wave scattered by the scatterer is partially polarized. Such scatterers are called distributed targets, and they are described by the polarimetric covariance or coherency matrices [33]. The Gaussian scatterer, K -scatterer, constant scatterer, and beta scatterer shown as follows belong to distributed targets.

1) *Point Scatterer*: Man-made targets can be treated as point-like scatterers. It is important to simulate this kind of targets when modeling urban areas. The response from a point scatterer to the incident wave should be deterministic. Consider several point scatterers of the same kind; however, the scattered waves are random due to their different positions. Let ϕ denote the phase difference due to the position and \mathbf{s}' denote the scattering vector regardless of the position, and the simulation of a point scatterer can be described as Fig. 2. The phase difference is uniformly distributed over $[0, 2\pi)$ by assuming that the scatterers are randomly located.

2) *Gaussian Scatterer*: A Gaussian scatterer can be represented by a vector that follows a multivariate complex Gaussian distribution with zero mean:

$$p(\mathbf{s}) = \frac{1}{\pi^3 |\mathbf{C}|} \exp(-\mathbf{s}^\dagger \mathbf{C}^{-1} \mathbf{s}). \quad (12)$$

For each polarimetric channel, the amplitude is Rayleigh distributed. Gaussian distributions are widely used to analyze random variables whose distributions are not known. Moreover, many complex problems can be derived analytically in explicit form when the relevant variables are assumed to follow Gaussian distributions due to their mathematical tractability. Therefore, it is interesting to test Gaussian distributions as the distributions of scatterers. The procedure of simulating a Gaussian scatterer is shown in Fig. 3.

3) *K-Scatterer*: A K -scatterer is defined by the multivariate \mathcal{K} -distribution given as follows:

$$p(\mathbf{s}; \alpha) = \frac{2\alpha^\alpha}{\pi^3 \Gamma(\alpha) |\mathbf{C}|} \left(\frac{\mathbf{s}^\dagger \mathbf{C}^{-1} \mathbf{s}}{\alpha} \right)^{\frac{2\alpha-3}{4}} \times K_{\alpha-3/2}(2\sqrt{\alpha \mathbf{s}^\dagger \mathbf{C}^{-1} \mathbf{s}}) \quad (13)$$

```

1: procedure GAUSSIANSCATTERER( $\mathbf{C}$ )
2:    $\mathbf{L} \leftarrow$  Cholesky decomposition on  $\mathbf{C}$ 
3:    $R_{hh}, I_{hh}, R_{hv}, I_{hv}, R_{vv}, I_{vv} \sim N(0, 1/2)$  ▷  $N$ , Normal distribution
4:    $\mathbf{u} \leftarrow [R_{hh} + jI_{hh}, R_{hv} + jI_{hv}, R_{vv} + jI_{vv}]^T$ 
5:    $\mathbf{s} \leftarrow \mathbf{L}\mathbf{u}$ 
6:   return  $\mathbf{s}$ 
7: end procedure

```

Fig. 3. Algorithm to simulate a Gaussian scatterer.

```

1: procedure KSCATTERER( $\mathbf{C}, \alpha$ )
2:    $\mathbf{z} \leftarrow$  GAUSSIANSCATTERER( $\mathbf{C}$ )
3:    $\tau \sim \Gamma(\alpha, 1/\alpha)$  ▷  $\Gamma$ , Gamma distribution
4:    $\mathbf{s} \leftarrow \tau \cdot \mathbf{z}$ 
5:   return  $\mathbf{s}$ 
6: end procedure

```

Fig. 4. Algorithm to simulate a K -scatterer.

where α is the shape parameter and K_ν is the modified Bessel function of the second kind of order ν . The multivariate \mathcal{K} -distribution has a wide range of shapes, which makes it well suited to describe scenes where there are scatterers with a variety of geometries and orientations [23]. The \mathcal{K} -distributed random variable can be viewed as the product of two random variables: one follows a Gaussian distribution and the other follows a gamma distribution. Therefore, to simulate a K -scatterer (Fig. 4), we can first simulate a Gaussian scatterer and then multiply it by a gamma-distributed random variable τ

$$p(\tau) = \frac{(\alpha\tau)^\alpha}{\tau\Gamma(\alpha)} \exp(-\alpha\tau). \quad (14)$$

4) *Constant Scatterer*: For each polarimetric channel, the response of a constant scatterer to the incident wave has a nonrandom amplitude [28]. In reality, it is difficult to find any constant scatterers, but as a contrast to the variable amplitudes, the study of constant ones is instructive. The constant scatterer is equivalent to the equal length component in the random walk model [34]. Consider the simplest case first, where the amplitudes of all polarimetric channels are equal to 1

$$\mathbf{u} = [e^{j\phi_{hh}}, e^{j\phi_{hv}}, e^{j\phi_{vv}}]^T. \quad (15)$$

The covariance matrix is given by $\mathbf{R} = E\{\mathbf{u}\mathbf{u}^\dagger\}$, which is also the correlation matrix. If all diagonal elements of the covariance matrix are equal to 1, like the \mathbf{R} here, the corresponding scatterer is referred to as a normalized scatterer. To the best of our knowledge, there is neither closed-form expression for the pdf nor well-known simulation algorithm of the random vector \mathbf{u} for a given covariance matrix. A simulation procedure based on a numerical analysis is designed in this paper.

Given a set of normalized Gaussian scatterers with covariance matrix \mathbf{R}_g , if we let all of the amplitudes be 1, they are transformed into a set of normalized constant scatterers with a different covariance matrix \mathbf{R} . From tests with different correlations [see Fig. 6(a)], it is found that, for the matrix

```

1: procedure CONSTANTSCATTERER(C)
2:    $\mathbf{R} \leftarrow$  Calculate (17)
3:    $\mathbf{R}_g \leftarrow$  Calculate (16)
4:    $\mathbf{z} \leftarrow$  GAUSSIANSCATTERER( $\mathbf{R}_g$ )
5:    $\mathbf{u} \leftarrow$  Set amplitudes of  $\mathbf{z}$  to 1
6:    $\mathbf{s} \leftarrow (\text{diag}(\mathbf{C}))^{-1/2} \mathbf{u}$ 
7:   return  $\mathbf{s}$ 
8: end procedure

```

Fig. 5. Algorithm to simulate a constant scatterer.

entries, $\mathbf{R}_g(i, k) = \rho_g e^{j\phi_g}$ and $\mathbf{R}(i, k) = \rho e^{j\phi}$, the relationship is given approximately by

$$\begin{cases} \rho_g = 0.3701\rho^5 - 1.1215\rho^4 + 0.9406\rho^3 \\ \quad - 0.5643\rho^2 + 1.3875\rho - 0.0078 \\ \phi_g = \phi. \end{cases} \quad (16)$$

This means that a normalized constant scatterer can be obtained by setting the amplitudes of a normalized Gaussian scatterer to 1, and the covariance matrix used to simulate the latter is calculated according to (16).

To simulate a scatterer for any covariance matrix \mathbf{C} , we first simulate a normalized constant scatterer \mathbf{u} with the covariance matrix given by

$$\mathbf{R}(i, k) = \frac{\mathbf{C}(i, k)}{\sqrt{\mathbf{C}(i, i)\mathbf{C}(k, k)}} \quad (17)$$

where i and k are the row number and column number of the matrix entries. Then, multiply a factor for each polarimetric channel $\mathbf{s} = (\text{diag}(\mathbf{C}))^{-1/2} \mathbf{u}$ with $\text{diag}(\mathbf{C})$, denoting the matrix of the diagonal elements of \mathbf{C} . The total simulation process is described in Fig. 5.

5) *Beta Scatterer*: The amplitudes of the responses from the beta scatterers are assumed to follow a beta distribution. A useful property of the beta distribution is that it has nonzero value over only a finite interval, and therefore, it does not allow amplitudes larger than a certain maximum [28]. For $0 \leq a \leq 1$ and shape parameters $\alpha, \beta > 0$, the pdf of the beta distribution is given by

$$p(a; \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} a^{\alpha-1} (1-a)^{\beta-1}. \quad (18)$$

It has been shown that the statistic of the speckle will follow a beta distribution if introducing a constraint to the total wave intensity received by the sensor [35], which is also known as the saturation effect. However, here the introduction of the beta scatterer is mainly because it allows exploration of the effects of various shapes of possible distributions, rather than a particular scattering problem. A multivariate generalization of beta distribution, known as the Dirichlet distribution, has a limited dependence structure, where the correlation coefficients are negative [36]. It is not suitable to model the amplitudes of different polarimetric channels. In order to model a more general correlation, the numerical method based on copulas theorem is employed, where correlated beta random variables are produced by transforming correlated normal random variables [37].

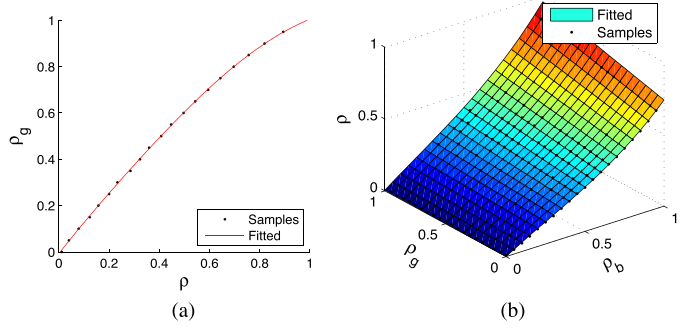


Fig. 6. Fitting Monte Carlo simulations with constant scatterer and beta scatterer. (a) Constant scatterer. (b) Beta scatterer.

```

1: procedure BETASCATTERER(C,  $\alpha$ ,  $\beta$ )
2:    $\mathbf{R} \leftarrow$  Calculate (17)
3:    $\phi_g \leftarrow$  Calculate (19)
4:    $\mathbf{R}_g(i, k) \leftarrow 1 \times e^{j\phi_g}$ 
5:    $\mathbf{z} \leftarrow$  GAUSSIANSCATTERER( $\mathbf{R}_g$ )
6:    $\rho_b \leftarrow$  Calculate (19)
7:    $\mathbf{a} \sim B(\alpha, \beta, \rho_b) \quad \triangleright B, \text{Beta distribution}$ 
8:    $\mathbf{u} \leftarrow$  Set amplitudes of  $\mathbf{z}$  as  $\mathbf{a}$ 
9:    $\mathbf{s} \leftarrow (\text{diag}(\mathbf{C}))^{-1/2} \mathbf{u}$ 
10:  Return  $\mathbf{s}$ 
11: end procedure

```

Fig. 7. Algorithm to simulate a beta scatterer.

Similar as the constant scatterers, beta scatterers are obtained by replacing the amplitudes of Gaussian scatterers with correlated beta random variables. Through the Monte Carlo simulations, it is found that the interchannel correlation of the normalized beta scatterer ($\rho e^{j\phi}$) is a function of that of the normalized Gaussian scatterer ($\rho_g e^{j\phi_g}$) and the correlation of beta-distributed amplitudes (ρ_b) [see Fig. 6(b)]. The relationship is approximated by

$$\begin{cases} \rho = 0.6721\rho_b + 0.0050\rho_g - 0.2884\rho_b^2 \\ \quad + 0.145\rho_b\rho_g + 0.3563\rho_b^3 + 0.0756\rho_b^2\rho_g \\ \phi = \phi_g. \end{cases} \quad (19)$$

By letting ρ_g be 1, the phases of the beta scatterers are obtained from Gaussian scatterers and amplitudes from correlated beta distributions. The simulation procedure is described as Fig. 7.

C. Scatterer Number

As the PolSAR sensor moves along the azimuth direction, the scatterers going in and out of the resolution cells can be modeled by a stochastic process. The number of scatterers, as a result, varies from a resolution cell to another. The fluctuation of the scatterer number plays an important role in the distribution of the scattering vector. In our simulator, only integers for scatterer number are considered, including the following cases.

1) *Constant*: For middle- or low-resolution data, it is reasonable to assume that there are a large number of scatterers inside each resolution cell, and in different resolution cells, the number is kept the same. At this point, the central limit theorem can be brought to bear. No matter what distribution the scatterers follow, the resulting scattering vector will be

TABLE II
COMPOUNDING DISTRIBUTIONS

Distribution	PDF
Gamma	$f(\lambda; \alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} \lambda^{\alpha-1} \exp(-\lambda/\beta)$
Inverse Gamma	$f(\lambda; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{-\alpha-1} \exp(-\beta/\lambda)$
Beta	$f(\lambda; \alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \lambda^{\alpha-1} (1-\lambda)^{\beta-1}, \lambda \in [0, 1]$
Beta Prime	$f(\lambda; \alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \lambda^{\alpha-1} (1+\lambda)^{-\alpha-\beta}$

- 1: **procedure** COMPOUNDNUM(α)
- 2: Get a sample λ from $f(\lambda; \alpha)$
- 3: Get a sample from Poisson distribution, $N \sim Po(\lambda)$
- 4: **end procedure**

Fig. 8. Simulation of compound scatterer number.

asymptotically Gaussian distributed [2]. In contrast, the area covered by a resolution cell becomes smaller in the high-resolution data, and the central limit theorem does not hold any more. The distribution of the received data will be affected by the scatterer type [28]. To meet all of these situations, any positive integer is allowed as the scatterer number.

2) *Negative Binomial*: In many cases of practical interest, the number of scatterers N is a random variable [3]. Then, the distribution of the resulting field is non-Gaussian. For example, if N follows a negative binomial distribution:

$$p(N; \alpha, p) = \binom{N + \alpha - 1}{N} p^\alpha (1-p)^N \quad (20)$$

and the mean value $\mu = \alpha(1-p)/p$ is large, then the scattering vector will follow a multivariate \mathcal{K} -distribution [5]. The scatterer numbers and scatterers are supposed to fluctuate at different scales, and the frequency of randomness of the scatterer numbers is lower than that of the scatterers.

3) *Compound Distribution*: In [38], the number of scatterers is modeled by a Poisson distribution with the mean value λ also randomly distributed according to some other distribution $f(\lambda; \alpha)$. Integrating the intermediate parameter λ out, the distribution of the scatterer number is then given by

$$p(N; \alpha) = \int_0^\infty \frac{\lambda^N e^{-\lambda}}{N!} f(\lambda; \alpha) d\lambda \quad (21)$$

with parameters α . There are several distributions from the Pearson system suggested for f , including gamma distribution, inverse gamma distribution, beta distribution, and beta prime distribution [38] (see Table II). They provide a wide variety of the scatterer number fluctuations. In particular, when λ is gamma distributed, the resulting distribution is equivalent to negative binomial distribution. Fig. 8 shows the simulation of a scatterer number that follows compound distributions.

Notice that, if $Y = 1/X$, where X is gamma distributed, then Y follows an inverse gamma distribution. Moreover, Y follows a beta prime distribution if $Y = X/(1-X)$ and X is beta distributed. In addition, a beta distribution can be obtained

by a Mellin product of a gamma distribution and an inverse gamma distribution. Therefore, it is possible to draw samples from inverse gamma distributions and beta prime distributions by transforming samples from gamma distributions or beta distributions.

D. Mixture of Scatterers

Normally, the scatterers in (9) are assumed to belong to the same category, one of the Gaussian scatterer, K -scatterer, constant scatterer, or beta scatterer. However, there are many cases where more than one scatterer type appears in the same resolution cell. In addition, a region under analysis can be divided into several subregions on many occasions, each with a different scatterer type. These will be regarded as mixtures of scatterers at the pixel level (in the same resolution cell) or at the spatial level (in different resolution cells).

1) *Pixel Level*: The mixture at pixel level assumes that each resolution cell contains scatterers from more than one type. Taking the urban areas for example, the backscattering can be viewed as the combination of echoes from distributed targets (trees and grass) and point targets (buildings). Furthermore, the scatterers inside a resolution cell can be classified into volume scattering, double-bounce scattering, surface scattering, and so on according to scattering mechanisms. In this case, the random walk model can be written as

$$\mathbf{k} = \sqrt{\frac{1}{N_1 + N_2 + \dots}} \left(\sum_{i=1}^{N_1} \mathbf{s}_i^{(a)} + \sum_{i=1}^{N_2} \mathbf{s}_i^{(b)} + \dots \right) \quad (22)$$

where $N = N_1 + N_2 + \dots$ is the total number of scatterers.

2) *Spatial Level*: In PolSAR data, the bright clutters and dark ones usually appear alternatively in the forest areas, corresponding to strong returns from the crowns of trees with shadows behind them. A similar phenomenon also appears in the sea surface when there exist wave crests and troughs. The received data can be treated as the mixture of two types of scatterers located alternatively. Assume that the region of interest (ROI) can be modeled by a mixture of K components, and then, the overall pdf of the scattering vector can be written as [30]

$$p(\mathbf{k}; \boldsymbol{\theta}) = \sum_{i=1}^K w_i p_i(\mathbf{k}; \boldsymbol{\theta}_i) \quad (23)$$

where $\sum_{i=1}^K w_i = 1$ is the mixing proportions and $\boldsymbol{\theta}$ is a vector collecting all of the parameters of the distribution. It has been shown that, for complicated regions with more irregular histograms (multimodal and spiky), the mixture model is more accurate than a single distribution [30], [32].

IV. SIMULATED SAR DATA

A. High-Resolution Data

In high-resolution data, where the resolution cell is only a few wavelength wide, the usual hypothesis underlying the speckle phenomena, i.e., a large number of scatterers per resolution cell, may not be verified [28]. The distribution of the data

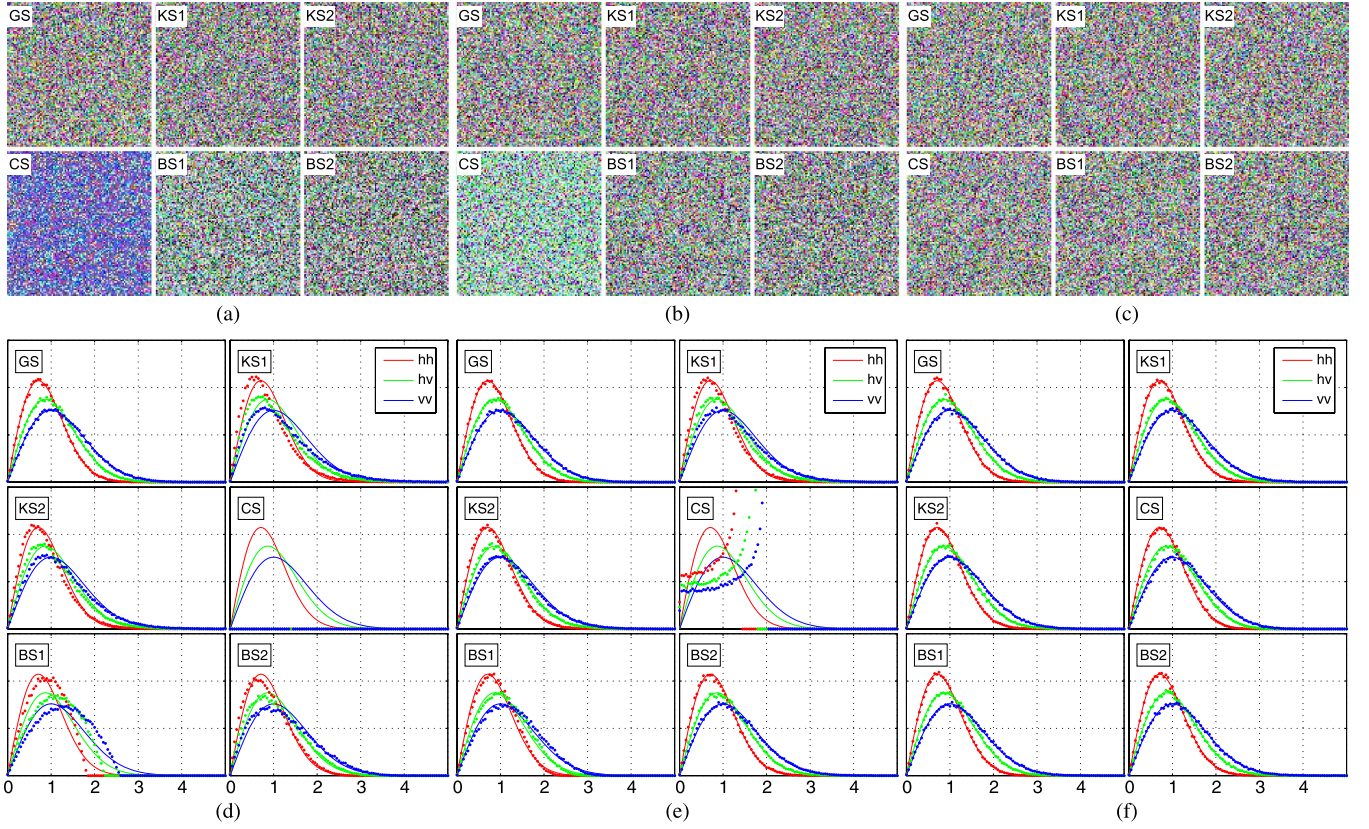


Fig. 9. Pauli decomposition and pdf of simulated high-resolution data by changing the scatterer number N . The first column (a) and (d) shows the Pauli decomposition and pdf when $N = 1$. The second column (b) and (e) shows the results when $N = 2$, and the last column (c) and (f) is when $N = 10$.

may depend on both the number of scatterers within a resolution cell and the distribution characterizing the individual scatterers. To validate this hypothesis, simulations with different scatterer numbers and scatterer types are implemented.

In each simulation, all of the resolution cells have the same number of scatterers. Integer values from 1 to 10 are tested as the scatterer number, and Gaussian scatterer (GS), K -scatterer (KS), constant scatterer (CS), and beta scatterer (BS) are tested as the scatterer type. All scatterers share the same covariance matrix

$$\mathbf{C}_1 = \begin{bmatrix} 1 & \frac{3\sqrt{3}}{10}e^{-j\frac{\pi}{2}} & \frac{7\sqrt{2}}{10}e^{j\frac{\pi}{5}} \\ \frac{3\sqrt{3}}{10}e^{j\frac{\pi}{2}} & 3 & \frac{\sqrt{6}}{5}e^{j\frac{\pi}{4}} \\ \frac{7\sqrt{2}}{10}e^{-j\frac{\pi}{5}} & \frac{\sqrt{6}}{5}e^{-j\frac{\pi}{4}} & 2 \end{bmatrix} \quad (24)$$

which is chosen for academic study only, and different values for the intensities of polarimetric channels as well as correlation coefficients between them are considered. For the K -scatterer, two values are chosen to test different variances: $\alpha = 3$ (KS1) with a large variance and $\alpha = 6$ (KS2) with a relatively small variance, where the former represents a rougher surface than the latter. For the beta scatterers, different parameters are chosen to test different skewness values: $\alpha = 2, \beta = 2$ (BS1) with skewness equal to zero and $\alpha = 2, \beta = 5$ (BS2) with a positive skewness. The skewness indicates the relation between the number of strong scatterers and that of weak scatterers. In the first case, the number of strong scatterers and the number of weak scatterers are comparable, whereas there are more weak scatterers

in the second case. For each testing case, a 300×300 pixel single-look complex image is generated.

The first row of Fig. 9 shows the Pauli decomposition of the simulated data, where the scatterer numbers within each resolution cell are 1, 2, and 10, respectively. All simulated images are homogeneous, but there are rather different colors in the Pauli decompositions, especially for the constant scatterers. The empirical pdfs of the amplitudes are compared with the Rayleigh distributions. The results show that not all homogeneous data have Gaussian statistics. The statistic information plays an important role in defining the homogeneous data besides the covariance matrix.

To quantitatively evaluate the fit of the Rayleigh distribution to the simulated data, the Kolmogorov–Smirnov distance (l) and the correlation coefficient (ρ) are employed [10], [30], as shown in Table III. The Kolmogorov–Smirnov distance quantifies a distance between the empirical distribution function of the samples and the cumulative distribution function of the reference distribution. The correlation coefficient provides a simple quantitative measure of similarity between two distributions.

As it can be seen, the statistical distribution of the sum of N Gaussian scatterers is independent of the value of N . This is because the addition of Gaussian scatterers on an amplitude basis does not change the statistics. The results calculated from Gaussian scatterers provide a reference for the comparison. The value of N from where the Rayleigh distribution can be concerned depends on the scatterer type. For KS1, KS2, CS, and BS1, when the scatterer number is equal or larger than 50,

TABLE III
CORRELATION COEFFICIENTS AND KOLMOGOROV–SMIRNOV DISTANCES

Scatterer Number	GS		KS1		KS2		CS		BS1		BS2	
	l	ρ	l	ρ	l	ρ	l	ρ	l	ρ	l	ρ
1	0.0019	99.96%	0.0713	97.56%	0.0376	99.29%	0.6088	16.65%	0.0651	96.79%	0.0348	99.21%
2	0.0034	99.95%	0.0364	99.31%	0.0179	99.82%	0.1352	61.81%	0.0264	99.58%	0.0095	99.92%
3	0.0025	99.96%	0.0262	99.65%	0.0122	99.90%	0.0580	93.52%	0.0158	99.83%	0.0046	99.96%
4	0.0022	99.96%	0.0199	99.78%	0.0114	99.91%	0.0438	98.44%	0.0129	99.87%	0.0053	99.95%
5	0.0029	99.96%	0.0164	99.85%	0.0077	99.94%	0.0250	99.57%	0.0097	99.93%	0.0037	99.96%
6	0.0031	99.96%	0.0138	99.88%	0.0068	99.94%	0.0190	99.73%	0.0080	99.94%	0.0045	99.95%
7	0.0023	99.96%	0.0118	99.91%	0.0072	99.95%	0.0184	99.80%	0.0076	99.95%	0.0042	99.95%
8	0.0027	99.96%	0.0100	99.93%	0.0069	99.94%	0.0170	99.82%	0.0057	99.95%	0.0032	99.96%
9	0.0024	99.97%	0.0095	99.92%	0.0054	99.95%	0.0138	99.88%	0.0064	99.94%	0.0034	99.97%
10	0.0034	99.95%	0.0091	99.94%	0.0052	99.96%	0.0132	99.88%	0.0054	99.96%	0.0031	99.96%
50	0.0020	99.97%	0.0032	99.97%	0.0035	99.97%	0.0026	99.97%	0.0032	99.97%	0.0031	99.96%
1000	0.0020	99.96%	0.0030	99.97%	0.0022	99.97%	0.0025	99.97%	0.0021	99.96%	0.0028	99.97%

the resulting distributions can be well approximated by Rayleigh distributions. The correlation coefficients between the estimated and empirical distributions are higher than 99.97%, the Kolmogorov–Smirnov distances are less than 0.035, and for BS2, the value is 9. It has been shown that modern spaceborne SAR (e.g., Cosmo-SkyMed and TerraSAR-X) can record data with an equivalent number of scatterers less than 10 based on the classical surface model [39]. In this case, the scatterer type should be concerned to understand the distribution of data.

Multilook PolSAR data are usually represented by sample covariance matrices, to which analyzing statistics such as matrix log-cumulants can be applied. When performing the multilook processing on the high-resolution data, special attention should be paid, as this procedure is equivalent to increasing the number of scatterers in resolution cells. In Fig. 10, we show the log-cumulants of the sample covariance matrices which are obtained after applying the multilook processing with 2×2 and 5×5 rectangular windows on the simulated data when $N = 1$.

It demonstrates that, after applying the multilooking process with a large window size, e.g., 5×5 , the resulting sample covariance matrices can be approximated by a Wishart distribution even if there is only one K -scatterer or constant scatterer in each resolution cell. The statistics of the PolSAR data approach Gaussian statistics more rapidly after the multilooking procedure. Therefore, for high-resolution data, preference should be given to the scattering vectors over the covariance matrices when coming to statistical analysis. In addition, it is interesting to note that the sample covariance matrices of BS1 after the multilook processing can be modeled by a \mathcal{K} -distribution.

B. Scatterer Number Fluctuation

In the middle- or low-spatial-resolution data, a resolution cell covers a large area, and the number of scatterers inside it is considered to be large. Due to the variations of the properties of the targets and the roughness of the observing surface, the scatterer number will fluctuate from cell to cell. This situation will give rise to the “texture,” bunching clutters in the data [6]. In the following, simulations of different distributions for

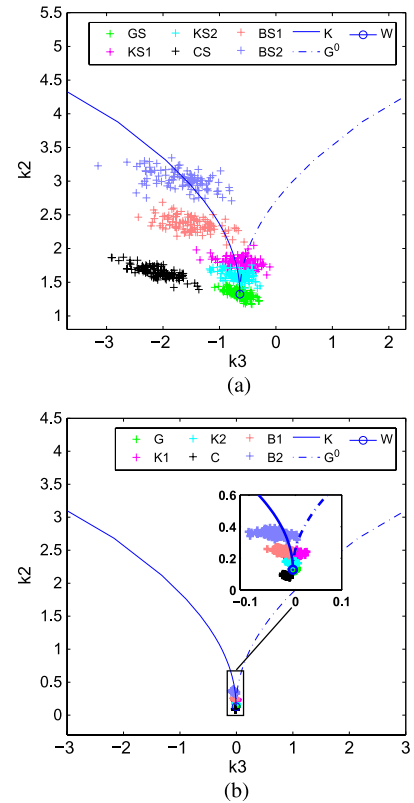


Fig. 10. Log-cumulants of the simulated high-resolution data after applying multilooking with different window sizes. (a) 2×2 . (b) 5×5 .

the scatterer number are shown, including negative binomial distribution and compound Poisson distributions.

Again, in each case, a 300×300 pixel single-look image is produced to ensure that there are enough samples. For the negative binomial distribution, two cases are tested with parameter α given by 5 and 6, denoted as NB1 and NB2, respectively. Inverse gamma, beta, and beta prime distributions are examined as the compounding distribution in (21). To make results comparable, a same shape parameter $\alpha = 5$ is tested for each of them, denoted as IG, B, and BP, respectively. In all simulations, the average value of scatterer numbers is larger than 1000 by

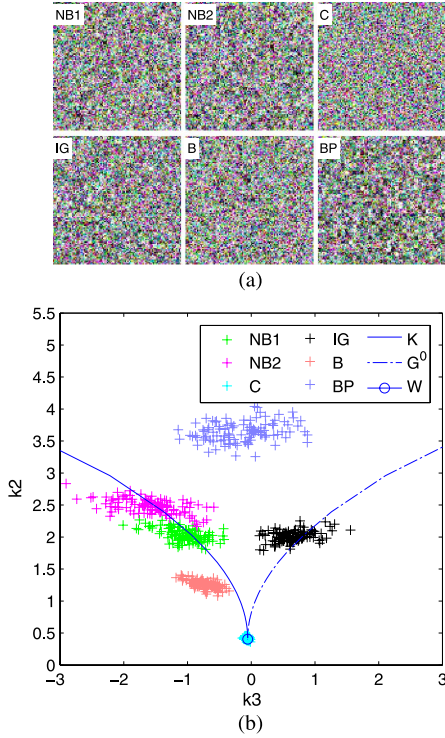


Fig. 11. Pauli decomposition and log-cumulants by changing the distribution of the scatterer number. (a) Pauli decomposition. (b) Log-cumulants.

choosing a proper scale parameter β . The sample covariance matrices are obtained by applying a multilook process with a 3×3 window. In the simulations, the randomness of the scatterer numbers is assured to have a lower frequency than that of the scatterers by letting the scatterer numbers of the pixels inside a neighboring window be the same.

Here, only the results of the Gaussian scatterers are shown since the scatterer type has a little influence on the final result when the scatterer number is large. From Fig. 11, it can be seen that we obtained \mathcal{K} , \mathcal{G}^0 , and Kummer- \mathcal{U} -distributed data by varying the distributions of the scatterer number. In the results of Pauli decomposition, there is bunching of scatterers, which is known as heterogeneity described as the “texture.” Altering the shape parameters of the distributions for the scatterer number, data with different texture parameters are obtained (see NB1 and NB2 for example).

C. Mixture

Mixtures of ground cover types within a resolution cell or a region are common in PolSAR images since the observing scene usually covers a large area. To see the influence on the statistical properties of PolSAR data, mixtures at pixel and spatial levels with different proportions are simulated.

First, the simulation of mixing two Gaussian scatterers of different classes is implemented: one is given by the covariance matrix (24), and the other is given by

$$\mathbf{C}_2 = \begin{bmatrix} 1 & 0.5e^{j\pi/3} & 0.4e^{-j\pi/4} \\ 0.5e^{-j\pi/3} & 1 & 0.5e^{j\pi/6} \\ 0.4e^{j\pi/4} & 0.5e^{-j\pi/6} & 1 \end{bmatrix}. \quad (25)$$

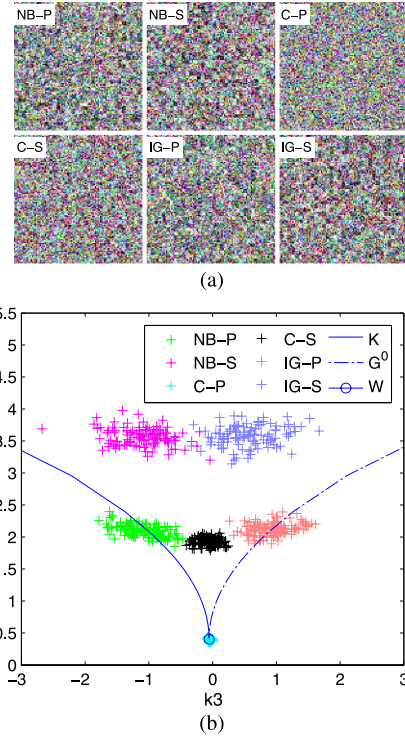


Fig. 12. Pauli decomposition and log-cumulants by mixing targets at pixel level (P) and spatial level (S). (a) Pauli decomposition. (b) Log-cumulants.

The covariance matrix has different elements as those of (24) and is chosen only for academic study. For the scatterer number, we tested the negative binomial distributed random variables (NB), Poisson distribution compounding with inverse gamma distribution (IG), and constant values (C). Mixtures at both pixel level (P) and spatial level (S) are applied, with each class taking up 50% proportion. The simulated data and the log-cumulants calculated from them are shown in Fig. 12. It is noticed that the mixture at spatial level gives heterogeneous data that follow Kummer- \mathcal{U} distributions. The mixture of distributed scatterers at pixel level changes the covariance matrix, but the texture of the obtained data is determined by the fluctuations of the scatterer numbers.

Fig. 13 shows the log-cumulants by changing the mixing proportion of the first class, given by covariance matrix (24), from 0% to 100%. It is observed that changing the proportions leads the log-cumulants of the simulated data to cover all of the $\kappa_2 - \kappa_3$ space. This is also validated by Nicolas on single channel data [27]. In another simulation, the covariance matrices (24) and (25) are changed into

$$\mathbf{C}_s = \begin{bmatrix} 1.56 & 0 & 2.12 - 0.03j \\ 0 & 0.09 & 0 \\ 2.12 + 0.03j & 0 & 2.90 \end{bmatrix} \quad (26)$$

$$\mathbf{C}_v = \frac{1}{8} \begin{bmatrix} 3 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 3 \end{bmatrix} \quad (27)$$

where the former covariance matrix represents the X-Bragg surface scattering [40] and the latter represents volume scattering

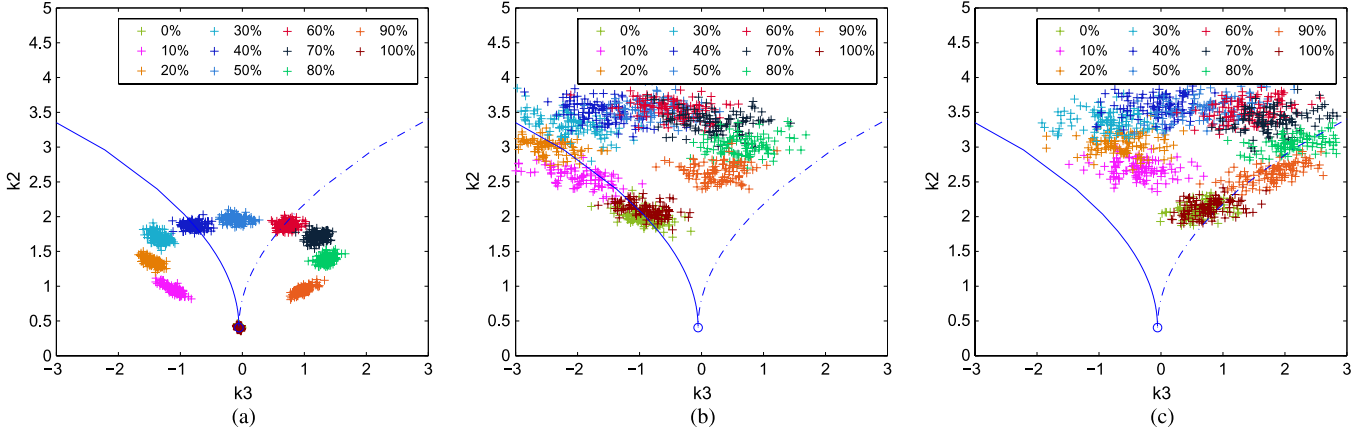


Fig. 13. Log-cumulants of the mixture of targets C_1 and C_2 at spatial level with different proportions. (a) Scatterer number is constant. (b) Scatterer number follows a negative binomial distribution. (c) Scatterer number follows a compound Poisson distribution.

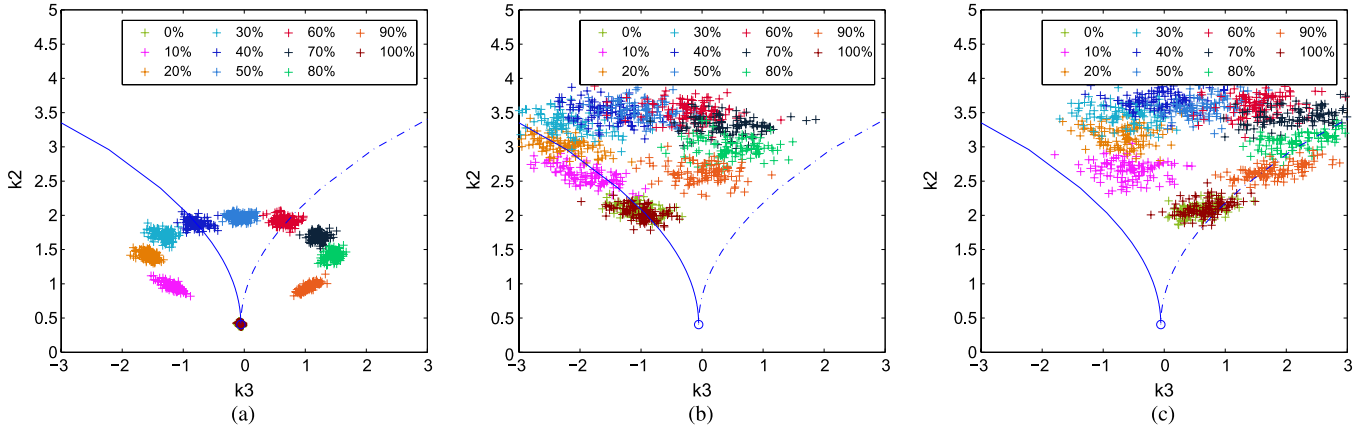


Fig. 14. Log-cumulants of the mixture of targets C_s and C_v at spatial level with different proportions. (a) Scatterer number is constant. (b) Scatterer number follows a negative binomial distribution. (c) Scatterer number follows a compound Poisson distribution.

[33]. The results are shown in Fig. 14, which is similar as those of the previous simulation.

From Figs. 11 and 13 (or 14), we notice that changing either the distribution of scatterer numbers or mixing proportion can give \mathcal{K} -, Kummer- \mathcal{U} -, and \mathcal{G}^0 -distributed data according to the log-cumulants. However, they have rather different polarimetry meanings. In the first case, the data represent only one type of scattering mechanism, while the second case represents a complex combination of different polarimetric scattering mechanisms.

In urban areas, man-made targets and natural targets are usually arranged alongside each other. The received data of a PolSAR system will be a mixture of responses from both distributed scatterers and point scatterers. To show the statistics of these areas, distributed scatterers with the covariance matrix given by (24) and point scatterers given by (28) are generated and mixed

$$\begin{aligned} \mathbf{S}_1 &= [5, 0, 0]^T \\ \mathbf{S}_2 &= [0, 5, 0]^T \\ \mathbf{S}_3 &= [10, 0, 0]^T. \end{aligned} \quad (28)$$

Different percentages of point scatterers are tested (from 0.1% to 90%). Fig. 15(a)–(c) shows the log-cumulants of the simulated data where the distributed scatterers are mixed with one point scatterer \mathbf{S}_1 , two point scatterers \mathbf{S}_1 and \mathbf{S}_2 , and one point scatterer \mathbf{S}_3 , respectively. As it can be seen, adding up point scatterers of different types increases the heterogeneity of the simulated data. Comparing Fig. 15(a) and (c), it is shown that increasing the amplitudes of the point scatterers results into more heterogeneity.

V. REAL SAR DATA

Apart from the simulated data, real PolSAR images, including the RADARSAT-2 fine quad-pol data (RST2), the ALOS-2 level 1.1 data in high-sensitive full-pol mode (ALOS2), and the TerraSAR-X Stripmap quad-pol data (TSX), are analyzed. The first two images were acquired over Barcelona (Spain), and the third one was over Vancouver (Canada). Original data are in the single-look complex format, from which the sample covariance matrices are obtained by applying the multilook processing with a sliding rectangular window. Table IV shows the size of the sliding window, and as the pixels are correlated, the equivalent number of looks (ENL) is estimated using the

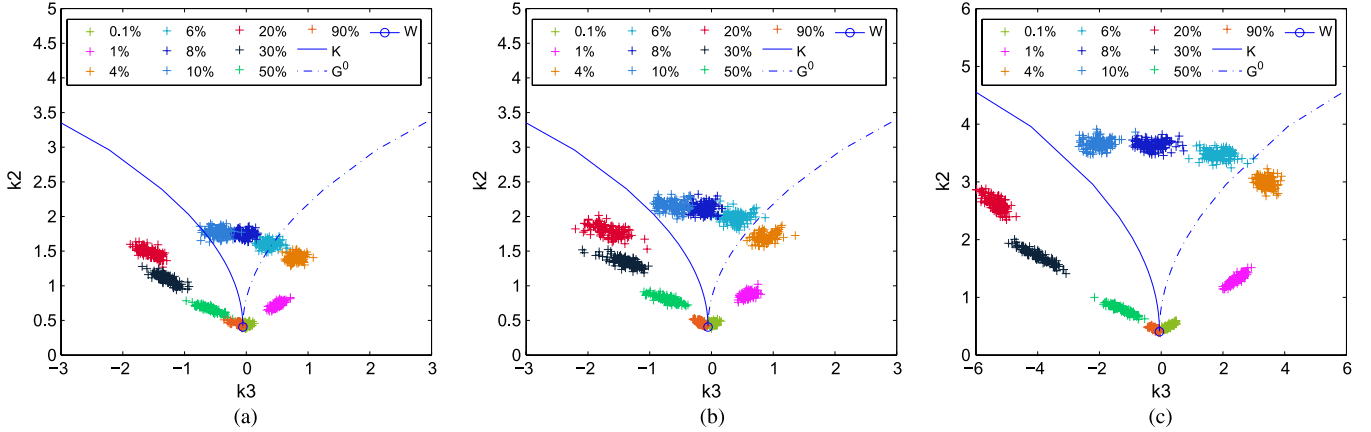


Fig. 15. Log-cumulants of the mixture of distributed targets C_1 and point targets. (a) Mixture with point targets S_1 . (b) Mixture with point targets S_1 and S_2 . (c) Mixture with point targets S_3 .

TABLE IV
REAL POLSAR DATA PARAMETERS

ID	Freq.	Resolution Rg \times Az (m)	Incidence Angle (degs)	Multilook Processing Window	ENL
RST2	C Band	11.1 \times 7.6	28.9	3 \times 3	5.61
ALOS2	L Band	3.49 \times 3.84	33.9	5 \times 5	15.15
TSX	X Band	1.18 \times 6.60	32.6/24.6	5 \times 5	7.77

log-determinant moment-based estimator [41]. Additional parameters such as the wave frequency, the incidence angle, and the spatial resolution are also listed as a comparison. We selected ROIs in the urban area, the agriculture area, the ocean, and the forest area. To avoid the SNR problem of TerraSAR-X data [42], two images with different incidence angles are included: the one with a larger incidence angle covers urban, forest, and agriculture areas, and the other is over wide-open ocean. The Pauli decomposition and ROIs are shown in Fig. 16.

The second row of Fig. 16 shows the second-order and third-order log-cumulants of the ROIs. As we can see, the urban areas (red and green rectangles) can be modeled by the \mathcal{G}^0 or the Kummer- \mathcal{U} distributions, which have the capability to model heterogeneous areas [9], [10]. The two ROIs in the urban area of the RST2 data and the ALOS2 data represent two different urban structures: one is of tall and densely distributed apartments, and the other is of short and sparse houses. This may be an explanation why different statistics, the \mathcal{G}^0 versus the Kummer- \mathcal{U} , are obtained. In agriculture areas (cyan and yellow rectangles), \mathcal{K} -distribution is shown to be the most suitable model due to the mixtures of different crop types. For a single type, the Wishart distribution is validated (see the yellow ROI of the TSX data). The forest area (black rectangle) shows a weak texture in the RST2 and TSX data. However, in the ALOS2 data, there is a strong fluctuation in the backscattering due to the radar foreshortening. To eliminate the effect of radar image distortions, another forest region (purple rectangle) is analyzed, which is found to follow a \mathcal{K} -distribution. In most cases, texture is not observed in the sea areas, with the exception of wide-open oceans where strong waves may exist (see the purple ROI in the TSX data for example). We must remind that the data of

the TerraSAR-X Stripmap quad-pol mode are an experimental product, and before applying statistical analysis to the TSX data, a simple calibration is accomplished, where pixels with intensities below the noise level are discarded. This may be the reason why the results on TSX data indicate a different behavior as those on RST2 and ALOS2 data.

Besides the spaceborne sensors, data from a ground-based SAR sensor, the UPC RiskSAR sensor [43], are also tested. The area under analysis covers different regions near Barcelona (Spain; see Fig. 17). The forest areas show a weak texture, while the urban areas are extremely heterogeneous, where the second-order log-cumulant can reach as large as 100.

VI. CONCLUSION

The random walk model has been studied, with the objective to give a physical insight of statistical PolSAR data texture models. Variable aspects of this model, including scatterer types, distributions of the scatterer number, and mixtures at pixel and spatial levels, are considered, and corresponding simulation algorithms are provided. When simulating the scattering process, the scalar product model has been considered by assuming the same texture for all of the polarimetric channels due to the complexity to properly simulate this type of variability. Statistical analysis is applied on the simulated data obtained under different assumptions, as well as real SAR data acquired with different SAR sensors, wave frequencies, and incidence angles.

From the simulated data, it is demonstrated that, according to the log-cumulants, i.e., the $\kappa_2 - \kappa_3$ diagram, the same PolSAR data distribution could come from different scattering scenarios.

In case of homogeneous data, it is demonstrated that the type of data distribution depends on both the scatterer number and the type of scatterer. Nevertheless, it is necessary to distinguish between low- and high-spatial-resolution data. In the former case, the PolSAR data distribution is essentially determined by the distribution of the scatterer number, whereas in the later, the data distribution is affected by the distribution of the scatter number as well as by the type of scatterer. The difference between these two cases is due to the number of scatterers inside the resolution cell. When this number is large,

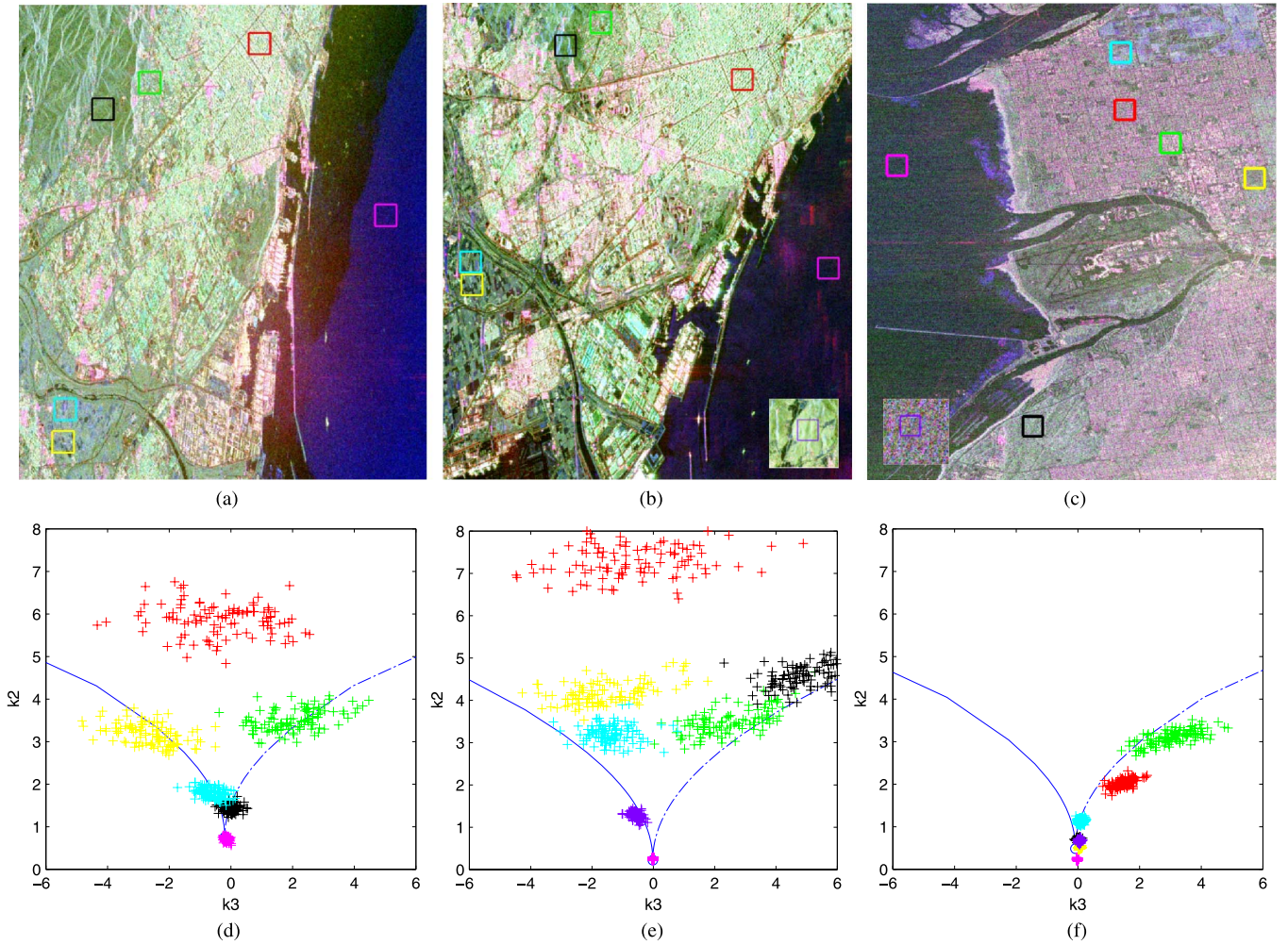


Fig. 16. Pauli decomposition and log-cumulants of real SAR data. The first column (a) and (d) shows ROIs and log-cumulants on the RST2 data, the second column (b) and (e) shows results on the ALOS2 data, and the last column (c) and (f) is on the TSX data.

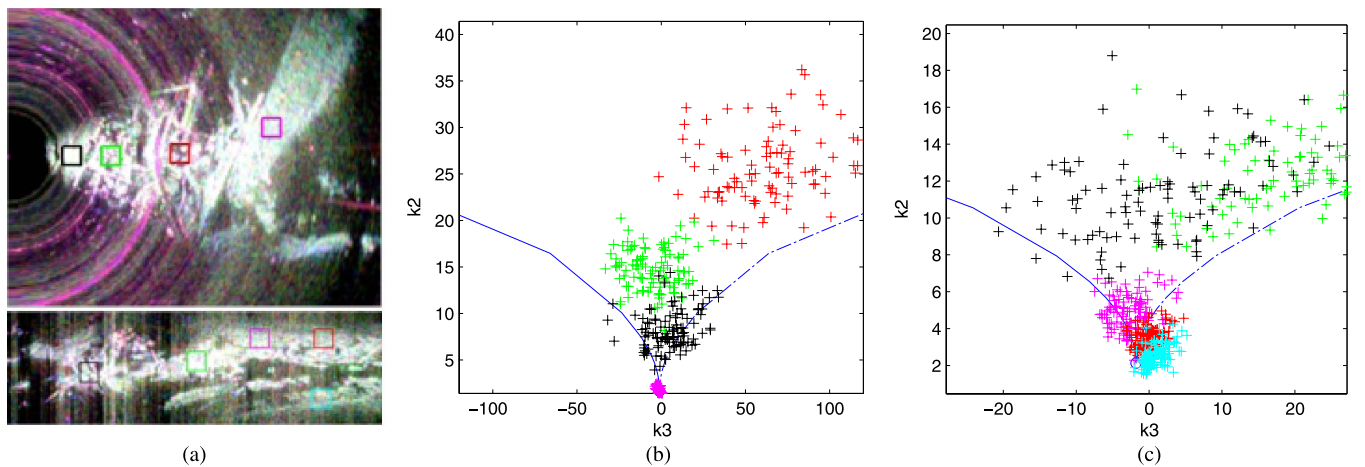


Fig. 17. Pauli decomposition and log-cumulants of ground-based SAR data. (a) ROIs of two ground-based SAR images. (b) Log-cumulant of the top image. (c) Log-cumulant of the bottom image.

larger than N_{\max} for example, the obtained data will fully develop speckle. It has also been demonstrated that the value of N_{\max} depends on the scatterer type. In addition, the effect of the scatterer type can be easily removed by the multilook processing. When a relatively large window is employed during

the multilook processing, the non-Gaussianity mainly comes from the fluctuation of the scatterer numbers.

In case of heterogeneous data, the PolSAR data distribution is determined by the mixture of different scatterers. This mixture can be considered at pixel as well as at spatial level.

The mixture of distributed scatterers at pixel level changes the covariance matrix, but the texture of the obtained data is determined by the fluctuations of the scatterer numbers. In the case of mixture at spatial level, the data distribution is determined not only by the distribution of the scatterer number but also by the proportion of the different scatter types in the mixture. Finally, the mixture of different targets will lead to extremely heterogeneous data, which provides a clue to analyze heterogeneous areas in PolSAR data, instead of introducing distributions with many parameters. A question that remains open in the case of mixture of scatterer types at pixel and spatial levels is the correct interpretation of the global polarimetric information in terms of the covariance or coherency matrices.

The configuration of a SAR sensor, e.g., frequency, incidence angle, and spatial resolution, could affect the distribution of some targets, as the information obtained by a different configurations is not the same. Normally, with higher frequency and higher resolution, more details could be observed, which will give a stronger texture. A further thorough study on this topic is required.

As it can be observed, the same PolSAR data distribution may be obtained by different scattering scenarios. One possible physical explanation for the different statistical PolSAR data models is the fluctuations of the scatterer number of each resolution cell, especially negative binomial distribution for the \mathcal{K} -distribution and compound Poisson distribution for Kummer- \mathcal{U} and \mathcal{G}^0 -distribution. Another possible explanation is the mixture of scatterers, where different data statistical models represent different mixture proportions. In other words, there is an ambiguity between the concept of texture and the concept of mixture in terms of the $\kappa_2 - \kappa_3$ log-cumulants. How to distinguish these two explanations or to eliminate this ambiguity, however, is not clear at this moment yet. The introduction of higher order statistics or the spatial correlation needs to be further explored.

ACKNOWLEDGMENT

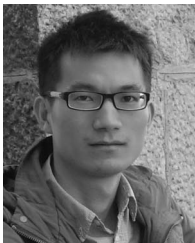
The authors would like to thank the German Aerospace Center (DLR) for the TerraSAR-X data provided in the framework of DRA proposal MTH0791, the European Space Agency (SAR) for the RADARSAT-2 data provided in the framework of the AgriSAR 2009 campaign, the Japan Aerospace Exploration Agency (JAXA) for the ALOS-2 data provided in the framework of the 4th ALOS Research Announcement for ALOS-2, and the reviewers for their valuable comments to improve this paper.

REFERENCES

- [1] J. R. Taylor, *Scattering Theory: The Quantum Theory of Nonrelativistic Collisions*. New York, NY, USA: Dover, 2012.
- [2] J. W. Goodman, "Some fundamental properties of speckle," *J. Opt. Soc. Amer.*, vol. 66, no. 11, pp. 1145–1150, Nov. 1976.
- [3] E. Jakeman and P. Pusey, "Significance of K distributions in scattering experiments," *Phys. Rev. Lett.*, vol. 40, no. 9, pp. 546–550, Feb. 1978.
- [4] E. Jakeman, "On the statistics of K -distributed noise," *J. Phys. A, Math. Gen.*, vol. 13, no. 1, pp. 31–48, Jan. 1980.
- [5] S. H. Yueh, J. A. Kong, J. K. Jao, R. T. Shin, and L. M. Novak, " K -distribution and polarimetric terrain radar clutter," *J. Electromagn. Waves Appl.*, vol. 3, no. 8, pp. 747–768, Jan. 1989.

- [6] K. Ward, "Compound representation of high resolution sea clutter," *Electron. Lett.*, vol. 17, no. 16, pp. 561–563, Aug. 1981.
- [7] C. Oliver, "The interpretation and simulation of clutter textures in coherent images," *Inv. Probl.*, vol. 2, no. 4, pp. 481–518, Nov. 1986.
- [8] F. T. Ulaby, F. Kouyate, B. Brisco, and T. L. Williams, "Textural information in SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. GE-24, no. 2, pp. 235–245, Mar. 1986.
- [9] C. C. Freitas, A. C. Frery, and A. H. Correia, "The polarimetric G distribution for SAR data analysis," *Environmetrics*, vol. 16, no. 1, pp. 13–31, Feb. 2005.
- [10] L. Bombrun and J.-M. Beaulieu, "Fisher distribution for texture modeling of polarimetric SAR data," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 3, pp. 512–516, Jul. 2008.
- [11] L. Bombrun, G. Vasile, M. Gay, and F. Totir, "Hierarchical segmentation of polarimetric SAR images using heterogeneous clutter models," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 2, pp. 726–737, Feb. 2011.
- [12] Y. Bian and B. Mercer, "Multilook polarimetric SAR data probability density function estimation using a generalized form of multivariate K -distribution," *Remote Sens. Lett.*, vol. 5, no. 7, pp. 682–691, Jul. 2014.
- [13] G. Moser, J. Zerubia, and S. B. Serpico, "SAR amplitude probability density function estimation based on a generalized Gaussian model," *IEEE Trans. Image Process.*, vol. 15, no. 6, pp. 1429–1442, Jun. 2006.
- [14] L. Bombrun, S. N. Anfinsen, and O. Harant, "A complete coverage of log-cumulant space in terms of distributions for polarimetric SAR data," in *Proc. PolInSAR*, Frascati, Italy, 2011, pp. 1–8.
- [15] P. Lombardo and A. Farina, "Coherent radar detection against K -distributed clutter with partially correlated texture," *Signal Process.*, vol. 48, no. 1, pp. 1–15, Jan. 1996.
- [16] Y. Yu, "Textural-partially correlated polarimetric K -distribution," in *Proc. IEEE IGARSS*, 1998, vol. 1, pp. 60–62.
- [17] S. Khan and R. Guida, "The new dual-texture G distribution for single-look PolSAR data," in *Proc. IEEE IGARSS*, Munich, Germany, 2012, pp. 1469–1472.
- [18] A. P. Doulgeris, S. N. Anfinsen, and T. Eltoft, "Segmentation of polarimetric SAR data with a multi-texture product model," in *Proc. IEEE IGARSS*, Munich, Germany, 2012, pp. 1437–1440.
- [19] T. Eltoft, S. N. Anfinsen, and A. P. Doulgeris, "A multitexture model for multilook polarimetric radar data," in *Proc. IEEE IGARSS*, Vancouver, BC, Canada, 2011, pp. 1048–1051.
- [20] Q. Zou, Y. Pi, and G. Liu, "Polarimetric speckle reduction by using multi-texture maximum likelihood method," in *Proc. IEEE IGARSS*, 2000, vol. 5, pp. 2287–2289.
- [21] J. S. Lee, D. L. Schuler, R. H. Lang, and K. J. Ranson, " K -distribution for multi-look processed polarimetric SAR imagery," in *Proc. IEEE IGARSS*, Pasadena, CA, USA, 1994, vol. 4, pp. 2179–2181.
- [22] J. S. Lee, R. G. Mitchell, and R. Kwok, "Classification of multi-look polarimetric SAR imagery based on complex Wishart distribution," *Int. J. Remote Sens.*, vol. 15, no. 11, pp. 2299–2311, Jul. 1994.
- [23] A. P. Doulgeris, S. N. Anfinsen, and T. Eltoft, "Classification with a non-Gaussian model for PolSAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 10, pp. 2999–3009, Oct. 2008.
- [24] A. C. Frery, H. J. Muller, C. C. F. Yanasse, and S. J. S. Sant'Anna, "A model for extremely heterogeneous clutter," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 3, pp. 648–659, May 1997.
- [25] A. P. Doulgeris, V. Akbari, and T. Eltoft, "Automatic PolSAR segmentation with the U -distribution and Markov random fields," in *Proc. 9th EUSAR Conf.*, 2012, pp. 183–186.
- [26] S. N. Anfinsen and T. Eltoft, "Application of the matrix-variate Mellin transform to analysis of polarimetric radar images," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2281–2295, Jun. 2011.
- [27] J.-M. Nicolas, "Introduction aux statistiques de deuxième espèce: Applications des logs-moments et des logs-cumulants à l'analyse des lois d'images radar," *Traitement Signal*, vol. 19, no. 3, pp. 139–167, 2002.
- [28] J. W. Goodman, "Speckle with a finite number of steps," *Appl. Opt.*, vol. 47, no. 4, pp. A111–A118, Feb. 2008.
- [29] E. Jakeman, "Speckle statistics with a small number of scatterers," *Opt. Eng.*, vol. 23, no. 4, Jul. 1984, Art. ID 234 453.
- [30] G. Moser, J. Zerubia, and S. B. Serpico, "Dictionary-based stochastic expectation-maximization for SAR amplitude probability density function estimation," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 1, pp. 188–200, Jan. 2006.
- [31] S. N. Anfinsen, "Statistical unmixing of SAR images," in *Proc. 10th EUSAR Conf.*, 2014, pp. 1–4.
- [32] V. Krylov *et al.*, "Modeling the statistics of high resolution SAR images," INRIA, Rocquencourt, France, Tech. Rep., 2008.
- [33] J.-S. Lee and E. Pottier, *Polarimetric Radar Imaging: From Basics to Applications*. Boca Raton, FL, USA: CRC Press, 2009.

- [34] J. W. Goodman, *Speckle Phenomena in Optics: Theory and Applications*. Englewood, CO, USA: Roberts & Company, 2007.
- [35] J. W. Goodman and E. G. Rawson, "Statistics of modal noise in fibers: A case of constrained speckle," *Opt. Lett.*, vol. 6, no. 7, pp. 324–326, Jul. 1981.
- [36] S. Magnussen, "An algorithm for generating positively correlated beta-distributed random variables with known marginal distributions and a specified correlation," *Comput. Stat. Data Anal.*, vol. 46, no. 2, pp. 397–406, Jun. 2004.
- [37] C. T. dos Santos Dias, A. Samaranyaka, and B. Manly, "On the use of correlated beta random variables with animal population modelling," *Ecol. Model.*, vol. 215, no. 4, pp. 293–300, Jul. 2008.
- [38] Y. Delignon and W. Pieczynski, "Modeling non-Rayleigh speckle distribution in SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 6, pp. 1430–1435, Jun. 2002.
- [39] G. Di Martino, A. Iodice, D. Riccio, and G. Ruello, "Equivalent number of scatterers for SAR speckle modeling," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2555–2564, May 2014.
- [40] I. Hajnsek, E. Pottier, and S. R. Cloude, "Inversion of surface parameters from polarimetric SAR," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 4, pp. 727–744, Apr. 2003.
- [41] S. N. Anfinsen, A. P. Douleris, and T. Eltoft, "Estimation of the equivalent number of looks in polarimetric synthetic aperture radar imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 11, pp. 3795–3809, Nov. 2009.
- [42] E. Makhoul, C. López-Martínez, and A. Broquetas, "Exploiting polarimetric TerraSAR-X data for sea clutter characterization," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 358–372, Jan. 2016.
- [43] R. Iglesias *et al.*, "Ground-based polarimetric SAR interferometry for the monitoring of terrain displacement phenomena—Part I: Theoretical description," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 3, pp. 980–993, Mar. 2015.



Xinpeng Deng received the M.Sc. degree in signal and information processing from Wuhan University, Wuhan, China, in 2011. He is currently working toward the Ph.D. degree at the Universitat Politècnica de Catalunya, Barcelona, Spain.

From September 2009 to June 2011, he was with the Signal Processing Laboratory, Wuhan University, working on target extraction, especially linear target extraction, of SAR and PolSAR data. He joined the Remote Sensing Laboratory, Universitat Politècnica de Catalunya, in April 2012, where his main work

focuses on statistical analysis of PolSAR data. His research interests include texture analysis of polarimetric SAR data and its application, digital image processing techniques and feature extraction, and machine learning.



Carlos López-Martínez (S'97–AM'03–M'04–SM'11) received the M.Sc. degree in electrical engineering and the Ph.D. degree from the Universitat Politècnica de Catalunya, Barcelona, Spain, in 1999 and 2003, respectively.

From October 2000 to March 2002, he was with the Frequency and Radar Systems Department (HR), German Aerospace Center (DLR), Oberpfaffenhofen, Germany. From June 2003 to December 2005, he was with the Image and Remote Sensing Group—SAPHIR Team, Institute of Electronics and Telecommunications of Rennes (IETR—CNRS UMR 6164), Rennes, France. In January 2006, he joined the Universitat Politècnica de Catalunya as a Ramón-y-Cajal researcher, where he is currently an Associate Professor in the area of remote sensing and microwave technology. He is the author or coauthor of more than 100 articles in journals, books, and conference proceedings in radar remote sensing and image analysis literature. His research interests include SAR and multidimensional SAR, radar polarimetry, physical parameter inversion, digital signal processing, estimation theory, and harmonic analysis.

Dr. López-Martínez is an Associate Editor of the *IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING*, and he served as a Guest Editor of the *EURASIP Journal on Advances in Signal Processing*. He has organized different invited sessions in international conferences on radar and SAR polarimetry. He has presented advanced courses and seminars on radar polarimetry to a wide range of organizations and events. He received the Student Prize Paper Award at the EUSAR 2002 Conference, and he is the coauthor of the paper awarded with the First Place Student Paper Award at the EUSAR 2012 Conference. He also received the IEEE-GRSS 2013 GOLD Early Career Award.



Eduardo Makhoul Varona (S'09–M'15) was born in Homs, Syria, in 1985. He received the M.Sc. degree in electrical engineering and the Ph.D. degree (*cum laude*) from the Universitat Politècnica de Catalunya (UPC), Barcelona, Spain, in 2009 and 2015, respectively.

In 2008, he worked on his Master's thesis in the framework of high-resolution wide-swath synthetic aperture radar (SAR) systems at the German Aerospace Center (DLR). In 2009, he joined the Remote Sensing Laboratory, UPC, researching on aspects of ground moving target indication (GMTI) for future spaceborne SAR systems. From 2009 to 2011, he worked on the radiometric budget analysis of the Spanish mission PAZ. From 2010–2013, he worked in the analysis and study of SAR-GMTI systems and techniques for future European missions in the framework of NEWA and SIMTISYS projects. In 2013 (April to August), he was a Visiting Researcher with the DLR, working on SAR-GMTI processing of airborne and spaceborne data over maritime scenarios. In August 2015, he joined isardSAT, an experienced Earth Observation research company in Barcelona, involved in the design, calibration, and maintenance of civil remote sensing instruments. His main research interests include SAR, GMTI, radar system design, mission performance, calibration, array signal processing, sea/ocean surveillance, and polarimetry.