Goodness-of-Fit Tests for Multilook Polarimetric Radar Data Based on the Mellin Transform

Stian Normann Anfinsen, *Student Member, IEEE*, Anthony Paul Doulgeris, *Student Member, IEEE* and Torbjørn Eltoft, *Member, IEEE*

***Abstract*—The advent of polarimetric synthetic aperture radar has spurred a growing interest in statistical models for complex-valued covariance matrices, which is the common representation of multilook polarimetric radar images. In this paper, we respond to an emergent need by proposing statistical tests for the simple and composite goodness-of-fit problem for a class of compound matrix distributions. The tests are based on Mellin kind matrix cumulants. These are derived from a novel characteristic function for positive definite Hermitian random matrices, defined in terms of a matrix-variate Mellin transform instead of the conventional Fouriér transform, and belong to a new framework for statistical analysis of multilook polarimetric radar data recently introduced by the authors. The cumulant-based tests are easy to compute and the asymptotic sampling distribution of the test statistic is chi-square distributed in the simple hypothesis case. Under the composite hypothesis, the sampling distribution is obtained by Monte Carlo simulations. We evaluate the power of the proposed goodness-of-fit tests with simulated data. We also use them to assess the fit of several matrix distributions to real data acquired by Radarsat-2 in fine quad polarisation mode.**

***Index Terms*—Radar polarimetry, synthetic aperture radar, probability density functions, goodness-of-fit, parameter estimation, log-statistics, Mellin transform, matrix-variate statistics**

I. Introduction

STATISTICAL modelling of radar data in terms of probability density functions (PDFs) is an important exercise which forms the basis of many radar image analysis techniques. Experience with single polarisation radar data has shown that they are well suited for parametric modelling, and a number of distribution families have been proposed for the purpose. Some are based on the simplistic assumption that the scattering coefficient is complex Gaussian random variable (RV), such as the Rayleigh distribution for single-look amplitude data, the exponential distribution for single-look intensity data, and the gamma distribution for multilook intensity data. Other models, such as the Weibull distribution and the log-normal distribution, provide added flexibility and the ability to model data with non-Gaussian characteristics (See [1] for a review of all the mentioned models). One of the most successful and accurate distribution models for radar data is arguably the $\mathcal{K}$ distribution family [1], [2], derived from the doubly stochastic product model, and also described as a compound distribution. The more recent $\mathcal{G}^0$ distribution family [3] is another versatile model derived with the same approach.

The complex Wishart distribution [4] was the first model proposed for multilook polarimetric radar data, and is still the most common, largely due to its mathematical tractability. It allows for a simplified analysis based on the assumption of Gaussian statistics for the complex scattering coefficients, which translates to complex Wishart statistics in the domain of the polarimetric covariance matrix. Its shortcomings has been amended with alternative compound densities, such as the polarimetric $\mathcal{K}$ distribution [5], $\mathcal{G}^0$ distribution [6], and $\mathcal{U}$ distribution [7]. These distributions account for non-Gaussianity and thus provide a more realistic model for high resolution radar images, whose scale increases the presence of heterogeneous targets and partially developed speckle.

In most publications where new PDFs have been proposed for the polarimetric covariance matrix, they have been justified only by visual comparison of fitted model densities against histograms of data for a single polarimetric channel at the time. One exception is [8], where the likelihood function was used to assess model fit for a set of different compound models. However, the use of the likelihood function as a goodness-of-fit measure is generally discouraged, since it does not carry

The authors are with the Department of Physics and Technology, University of Tromsø, NO-9037 Tromsø, Norway (e-mail: {stian.normann.anfinsen;torbjorn.eltoft}@uit.no).

much information as a test statistic [9]. None of the standard goodness-of-fit tests (such as Pearson's χ^2 test, the Kolmogorov-Smirnov test, the Anderson-Darling test or the Cramér-von Mises test) have seemed applicable to the matrix distributions under study. They require binning and ordering of data points, or an expression for the cumulative distribution function (CDF), which will in this case be defined on the cone of positive definite Hermitian random matrices. Although such a matrix CDF has been defined in the complex Wishart case [10], it is difficult to evaluate for a compound distribution. No adequate alternative has, to the best of our knowledge, been suggested in the literature, which is what we try to remedy.

In [11], Li and Papodopoulos provide a general framework for the design of moment-based goodness-of-fit tests. Their simple idea is to compare sample moments with population moments, and to combine these in a test statistic which is asymptotically normal or χ^2 distributed, which makes it easy to perform hypothesis testing or to obtain a p-value for a given data sample. We have applied their theory to a set of compound distribution models for the polarimetric covariance matrix, based on the matrix log-cumulants (MLCs) introduced in [12], [13].

MLCs are matrix-variate generalisations of the log-cumulants derived and successfully applied to the analysis of single polarisation synthetic aperture radar (SAR) data in a series of publications by Nicolas et al. (See e.g. [14], [15], [16]). We give a geometrical interpretation of the approach by using a diagram of the space spanned by the MLCs to illustrate the statistical distance between given models and empirical data. This is a polarimetric extension of the diagram introduced by Nicolas for the univariate case [14], [15]. It provides intuition about the capabilities of the different models, and how they adapt to the non-Gaussian data found in textured and heterogeneous areas.

The appropriateness of non-Gaussian models increase as the spatial resolution improves. Thus, our developments are highly relevant to high resolution SAR instruments. However, the methods are in principle applicable to data from any imaging radar, and are therefore presented as general analysis tools for radar image data.

Section II gives the necessary theoretical background. It presents the data format, the distribution models and the Mellin kind statistics (MKS) used in our derivations. Section III presents the proposed goodness-of-fit tests, after an introductory definition of the problem and review of the literature. In Section IV we present the results of simulations with random generated data, in also test the model fit to some real data samples. We give our conclusions in Section V.

Our convention for notation is that scalar values are denoted as lower or upper case standard weight characters, vectors are lower case boldface characters, and matrices are upper case boldface characters. Except for scalar random variables, we do not distinguish between random variables and instances of random variables, as such can be ascertained through context. A list of acronyms is provided for convenience:

NOMENCLATURE

CDF	cumulative distribution function
CF	characteristic function
EDF	empirical distribution function
GoF	goodness-of-fit
MAL	maximum asymptotic likelihood
MLC	matrix log-cumulant
MLM	matrix log-moment
MoLC	method of log-cumulants
MoMLC	method of matrix log-cumulants
MT	Mellin transform
MKS	Mellin kind statistics
PDF	probability density function
RV	random variable
SAR	synthetic aperture radar

II. POLARIMETRIC RADAR DATA

A. Data Format

The measurable of a polarimetric radar is the Sinclair scattering matrix

$$\mathbf{S} = \begin{bmatrix} S_{xx} & S_{xy} \\ S_{yx} & S_{yy} \end{bmatrix}, \tag{1}$$

or equivalently, the scattering vector $\mathbf{s}$, which is simply the vectorised version of $\mathbf{S}$:

$$\mathbf{s} = \mathrm{vec}(\mathbf{S}^T) = \begin{bmatrix} S_{xx} \\ S_{xy} \\ S_{yx} \\ S_{yy} \end{bmatrix}, \tag{2}$$

where $(\cdot)^T$ and $\mathrm{vec}(\cdot)$ denote the transposition and vectorisation operator, respectively. The entries of $\mathbf{S}$ and $\mathbf{s}$ are the scattering coefficients of the d polarimetric channels. These are complex-valued, dimensionless numbers that describe the transformation of incident to backscattered electromagnetic field for all combinations of two orthogonal transmit and receive polarisations, denoted by x and y.

In the following, we shall only be concerned with multilook complex data. Multilooking is an averaging process, applied either during or after focusing of the radar image, which suppresses the noise-like effect of speckle at the cost of reduced spatial resolution. Assume that a set $\{\mathbf{s}_\ell\}_{\ell=1}^{L}$ of scattering vectors are averaged. We refer to $\mathbf{s}_\ell$ as a look and L as the number of looks. Then, multilook polarimetric radar data is represented in the intensity domain by:

$$\mathbf{C} = \frac{1}{L}\sum_{\ell=1}^{L} \mathbf{s}_\ell \mathbf{s}_\ell^H \,, \tag{3}$$

or a linearly transformed version of $\mathbf{C}$, where $(\cdot)^H$ is the Hermitian (conjugate transposition) operator. We refer to $\mathbf{C} \in \Omega_+ \subset \mathbb{C}^{d\times d}$ as the multilook polarimetric covariance matrix, and note that $\mathbf{C}$ is a random matrix defined on the cone Ω_+ of positive definite complex Hermitian matrices.

B. Distribution Models

We base our work upon the multilook polarimetric product model [6], which decomposes $\mathbf{C}$ as

$$\mathbf{C} = T\,\widetilde{\mathbf{W}}\,. \tag{4}$$

The strictly positive and unit mean scalar random variable T models texture, which is here defined as spatial variation in the mean backscatter due to target variability. It represents natural variations in the radar return for pixels that could be labelled as one class, as opposed to variation attributed to the inherent interference produced by coherent imaging. The latter contribution, known as speckle or clutter, is modelled by $\widetilde{\mathbf{W}} \sim s\mathcal{W}_d^{\mathbb{C}}(L,\boldsymbol{\Sigma})$, a scaled complex Wishart matrix[1], which follows the distribution

$$f_{\widetilde{\mathbf{W}}}(\widetilde{\mathbf{W}};L,\boldsymbol{\Sigma}) = \frac{L^{Ld}}{\Gamma_d(L)}\frac{|\widetilde{\mathbf{W}}|^{L-d}}{|\boldsymbol{\Sigma}|^{L}}\,\mathrm{etr}(-L\boldsymbol{\Sigma}^{-1}\widetilde{\mathbf{W}}) \tag{5}$$

[1]The matrix $\mathbf{W} = L\widetilde{\mathbf{W}}$ follows a true complex Wishart distribution, denoted $\mathbf{W} \sim \mathcal{W}_d^{\mathbb{C}}(L,\boldsymbol{\Sigma})$ [4].

where $\boldsymbol{\Sigma} = \mathrm{E}\{\widetilde{\mathbf{W}}\}$ is the scale matrix, $|\cdot|$ is the determinant, $\mathrm{etr}(\cdot) = \exp(\mathrm{tr}(\cdot))$ where $\mathrm{tr}(\cdot)$ is the trace operator, $\Gamma_d(L)$ is the multivariate gamma function of the complex kind [6], and $L \geq d$ assures that $\mathbf{C}$ is nonsingular.

The simplest model for the PDF of $\mathbf{C}$ assumes that the scattering coefficients are jointly circular complex Gaussian. This is strictly justified only for homogeneous regions of the image characterised by fully developed speckle and no texture, which may be expressed as the probability $P(T=1)=1$, or $f_T(t) = \delta(t-1)$, where $\delta(\cdot)$ is the Dirac delta function. This results in $\mathbf{C} \sim s\mathcal{W}_d^{\mathbb{C}}(L,\boldsymbol{\Sigma})$.

When the PDF of T is not degenerate, we obtain a more complicated distribution for $\mathbf{C}$, which depends on the distribution of T through

$$f_{\mathbf{C}}(\mathbf{C}) = \int_0^\infty f_{\mathbf{C}|T}(\mathbf{C}|t) f_T(t)\,dt \tag{6}$$

where $\mathbf{C}|T \sim s\mathcal{W}_d^{\mathbb{C}}(L,\boldsymbol{\Sigma})$. For instance, we obtain the matrix-variate $\mathcal{K}$ distribution [6] for gamma distributed texture (denoted $T\sim\bar{\gamma}(\alpha)$), the matrix-variate $\mathcal{G}^0$ distribution [6] for inverse gamma distributed texture ($T\sim\bar{\gamma}^{-1}(\lambda)$), and the $\mathcal{U}$ distribution [7] for texture that follows a Fisher-Snedecor distribution ($T\sim\bar{\mathcal{F}}(\alpha,\lambda)$).

The distributions are shown in Table I. All have been normalised to unit mean, indicated by the bar over the distribution symbol. The normalisation explains why the number of parameters is one less than for the nominal distribution. The $\mathcal{K}$ distribution and the $\mathcal{U}$ distribution have got their name from special functions that occur within their PDF: respectively $K_\nu(\cdot)$, the second kind modified Bessel function of order ν, and $U(\cdot,\cdot,\cdot)$, the second kind confluent hypergeometric Kummer function.

C. Mellin Kind Statistics

Mellin kind statistics for complex random matrices were defined in [12], [13] and evaluated for the distributions in Table I. We here repeat the expressions needed in the rest of the paper.

Let $\mathbf{C} \in \Omega_+$ be a $d\times d$ complex covariance matrix whose PDF is $f_{\mathbf{C}}(\mathbf{C})$. The complex matrix-variate Mellin transform (MT) of a general real-valued function $g(\mathbf{C}) : \Omega_+ \to \mathbb{R}$ is

$$\begin{aligned} G(s) &= \mathcal{M}\{g(\mathbf{C})\}(s) \\ &= \int_{\Omega_+} |\mathbf{C}|^{s-d} g(\mathbf{C})\,d\mathbf{C} \end{aligned} \tag{7}$$

TABLE I
TEXTURE AND COVARIANCE MATRIX DISTRIBUTIONS UNDER THE DOUBLY STOCHASTIC PRODUCT MODEL

$f_T(t)$ of texture variable T		$f_{\mathbf{C}}(\mathbf{C})$ of covariance matrix $\mathbf{C}$		Ref.
Constant	$\delta(t-1)$	$s\mathcal{W}_d^{\mathbb{C}}(\mathbf{\Sigma}, L)$	$\frac{L^{Ld}}{\Gamma_d(L)} \frac{\lvert\mathbf{C}\rvert^{L-d}}{\lvert\mathbf{\Sigma}\rvert^{L}} \operatorname{etr}(-L\mathbf{\Sigma}^{-1}\mathbf{C})$	[4]
$\bar{\gamma}(\alpha)$	$\frac{\alpha^\alpha}{\Gamma(\alpha)} t^{\alpha-1} \exp(-\alpha t)$	$\mathcal{K}_d(\mathbf{\Sigma}, L, \alpha)$	$\frac{2\lvert\mathbf{C}\rvert^{L-d}(L\alpha)^{\frac{\alpha+Ld}{2}}}{\lvert\mathbf{\Sigma}\rvert^{L}\Gamma_d(L)\Gamma(\alpha)} \big(\operatorname{tr}(\mathbf{\Sigma}^{-1}\mathbf{C})\big)^{\frac{\alpha-Ld}{2}} K_{\alpha-Ld}\big(2\sqrt{L\alpha \operatorname{tr}(\mathbf{\Sigma}^{-1}\mathbf{C})}\big)$	[5]
$\bar{\gamma}^{-1}(\lambda)$	$\frac{(\lambda-1)^\lambda}{\Gamma(\lambda)} \frac{1}{t^{\lambda+1}} \exp\big(-\frac{\lambda-1}{t}\big)$	$\mathcal{G}_d^0(\mathbf{\Sigma}, L, \lambda)$	$\frac{L^{Ld}\lvert\mathbf{C}\rvert^{L-d}}{\Gamma_d(L)\lvert\mathbf{\Sigma}\rvert^{L}} \frac{\Gamma(Ld+\lambda)(\lambda-1)^\lambda}{\Gamma(\lambda)} \big(L\operatorname{tr}(\mathbf{\Sigma}^{-1}\mathbf{C}) + \lambda - 1\big)^{-\lambda-Ld}$	[6]
$\bar{\mathcal{F}}(\alpha, \lambda)$	$\frac{\Gamma(\alpha+\lambda)}{\Gamma(\alpha)\Gamma(\lambda)} \frac{\alpha}{\lambda-1} \frac{\big(\frac{\alpha}{\lambda-1}t\big)^{\alpha-1}}{\big(\frac{\alpha}{\lambda-1}t+1\big)^{\alpha+\lambda}}$	$\mathcal{U}_d(\mathbf{\Sigma}, L, \alpha, \lambda)$	$\frac{L^{Ld}\lvert\mathbf{C}\rvert^{L-d}}{\Gamma_d(L)\lvert\mathbf{\Sigma}\rvert^{L}} \frac{\Gamma(\alpha+\lambda)}{\Gamma(\alpha)\Gamma(\lambda)} \Big(\frac{\alpha}{\lambda-1}\Big) \Gamma(Ld+\lambda)$ $\times U\big(Ld+\lambda, Ld-\alpha+1, L\operatorname{tr}(\mathbf{\Sigma}^{-1}\mathbf{C})\alpha/(\lambda-1)\big)$	[7]

with transform variable $s \in \mathbb{C}$, whenever the integral exists. It generally does for the matrix distributions we study. The MT of $f_{\mathbf{C}}(\mathbf{C})$ is defined as the Mellin kind characteristic function (CF) of the random matrix $\mathbf{C}$:

$$\phi_{\mathbf{C}}(s) = \mathrm{E}\{|\mathbf{C}|^{s-d}\} = \mathcal{M}\{f_{\mathbf{C}}(\mathbf{C})\}(s)\,. \tag{8}$$

When it exists, the νth-order matrix log-moment (MLM) is derived from

$$\mu_\nu\{\mathbf{C}\} = \mathrm{E}\{(\ln|\mathbf{C}|)^\nu\} = \frac{d^\nu}{ds^\nu}\phi_{\mathbf{C}}(s)\bigg|_{s=d}\,. \tag{9}$$

The Mellin kind cumulant generating function is defined as

$$\varphi_{\mathbf{C}}(s) = \ln\phi_{\mathbf{C}}(s) \tag{10}$$

and the νth-order matrix log-cumulant (MLC) as

$$\kappa_\nu\{\mathbf{C}\} = \frac{d^\nu}{ds^\nu}\varphi_{\mathbf{C}}(s)\bigg|_{s=d}\,. \tag{11}$$

MLMs and MLCs are related by

$$\kappa_\nu\{\mathbf{C}\} = \mu_\nu\{\mathbf{C}\} - \sum_{i=1}^{\nu-1}\binom{\nu-1}{i-1}\kappa_i\{\mathbf{C}\}\mu_{\nu-i}\{\mathbf{C}\}\,. \tag{12}$$

For instance, the first three MLCs are

$$\kappa_1 = \mu_1 \tag{13}$$
$$\kappa_2 = \mu_2 - \mu_1^2 \tag{14}$$
$$\kappa_3 = \mu_3 - 3\mu_1\mu_2 + 2\mu_1^3 \tag{15}$$

where the argument of the MLMs and MLCs has been suppressed for brevity. We continue this practice in the following, whenever there is no confusion about what stochastic entity the statistic is computed from. More relations between moments and cumulants are given in Appendix A.[2]

The νth-order sample MLM of $\mathbf{C}$, denoted $\langle\mu_\nu\{\mathbf{C}\}\rangle$, can be computed from a set of n independent and identically distributed covariance matrices, $\mathscr{C} = \{\mathbf{C}_i\}_{i=1}^n$, using the sample mean estimator:

$$\langle\mu_\nu\{\mathbf{C}\}\rangle = \frac{1}{n}\sum_{i=1}^{n}(\log|\mathbf{C}_i|)^\nu\,. \tag{16}$$

The sample MLCs $\langle\kappa_\nu\{\mathbf{C}\}\rangle$ are computed from (12) with the population MLMs and MLCs replaced by $\langle\mu_\nu\{\mathbf{C}\}\rangle$ and $\langle\kappa_\nu\{\mathbf{C}\}\rangle$.

For the multilook polarimetric product model, the Mellin kind CF is expressed as [13]

$$\phi_{\mathbf{C}}(s) = \phi_T(d(s-d)+1)\,\phi_{\widetilde{\mathbf{W}}} \tag{17}$$

where $\phi_T(s)$ is the univariate Mellin kind CF of a general texture RV T. The Mellin kind CF of a scaled complex Wishart matrix is [13]

$$\phi_{\widetilde{\mathbf{W}}}(s) = L^{ds}\frac{\Gamma_d(L+s+d)}{\Gamma_d(L)}|\mathbf{\Sigma}|^{(s-d)}\,. \tag{18}$$

This yields the population MLCs

$$\kappa_\nu\{\mathbf{C}\} = \kappa_\nu\{\widetilde{\mathbf{W}}\} + d^\nu\kappa_\nu\{T\}\,. \tag{19}$$

[2]Remark that (12) is valid for moments and cumulants of all kinds (i.e., moments of scalars, vectors or matrices – logarithmic or not), since the formula relies on the definition of the cumulant generating function as the natural logarithm of the CF. It is easily derived using Leibniz' rule for differentiation of a product [17].

TABLE II
MELLIN KIND STATISTICS OF UNIVARIATE DISTRIBUTIONS FOR REAL POSITIVE TEXTURE VARIABLES

$f_T(t)$	Characteristic function $\phi_T(s)$	Log-cumulants $\kappa_\nu(T)$
$\bar{\gamma}(\alpha)$	$\alpha^{1-s}\frac{\Gamma(\alpha+s-1)}{\Gamma(\alpha)}$	$\kappa_1 = \psi^{(0)}(\alpha) - \ln(\alpha)$ $\kappa_{\nu>1} = \psi^{(\nu-1)}(\alpha)$
$\bar{\gamma}^{-1}(\lambda)$	$(\lambda-1)^{s-1}\frac{\Gamma(\lambda+1-s)}{\Gamma(\lambda)}$	$\kappa_1 = -\psi^{(0)}(\lambda) + \ln(\lambda-1)$ $\kappa_{\nu>1} = (-1)^\nu\psi^{(\nu-1)}(\lambda)$
$\bar{\mathcal{F}}(\alpha,\lambda)$	$\left(\frac{\lambda-1}{\alpha}\right)^{s-1}\frac{\Gamma(\alpha+s-1)}{\Gamma(\alpha)}\frac{\Gamma(\lambda+1-s)}{\Gamma(\lambda)}$	$\kappa_1 = \psi^{(0)}(\alpha) - \psi^{(0)}(\lambda) + \ln\left(\frac{\lambda-1}{\alpha}\right)$ $\kappa_{\nu>1} = \psi^{(\nu-1)}(\alpha) + (-1)^\nu\psi^{(\nu-1)}(\lambda)$

We note that the speckle contribution

$$\kappa_1\{\widetilde{\mathbf{W}}\} = \psi_d^{(0)}(L) + \ln|\mathbf{\Sigma}| - d\ln L \quad (20a)$$

$$\kappa_{\nu>1}\{\widetilde{\mathbf{W}}\} = \psi_d^{(\nu-1)}(L) \quad (20b)$$

is separated from the texture contribution (i.e., the second term of (19)). The texture part is determined by the distribution of the univariate RV T. Univariate Mellin kind characteristic functions, $\phi_T(s)$, and univariate log-cumulants, $\kappa_\nu\{T\}$, are listed in Table II for the texture distributions presented in Table I.

Finally note the following key property of the MLCs: They depend only on the texture parameters and L for $\nu > 1$, while the first-order MLC also depends on $\mathbf{\Sigma}$.

III. GOODNESS-OF-FIT TESTS

A. Theory and Literature Review

1) Definitions: A formal GoF test[3] is a procedure for testing the null hypothesis H_0 that a set of random variates follow a given PDF. The procedure measures the conformity or the discrepancy of the data sample with respect to the distribution model. It provides a test statistic, which is used to decide whether H_0 should be accepted or rejected [18], [19]. From the test statistic and its sampling distribution, we can also compute the p-value, defined as the probability of obtaining a realisation of the test statistic at least as extreme as the one observed.

When the model is fully specified, i.e., all parameters of the hypothesised PDF are known, we say that H_0 is a simple hypothesis. If some or all of the parameters are unknown and must be estimated, then H_0 is a composite hypothesis and defines a *composite GoF problem* [20], [21], [22]. In the latter case, we measure the fit to a distribution family, rather than a specific distribution. The alternative hypothesis H_1 is in the context of our work a composite one. It simply states that H_0 is wrong and contains no other information about the data distribution.[4]

We will discuss both simple and composite GoF tests. The composite GoF problem describes most practical applications, but represents a far more difficult setting, since the sampling distribution of the test statistic becomes much harder to derive. It will generally depend on the functional form of the distribution family under H_0, the true values of the unknown parameters and the estimator used to determine them, as well as the available number of samples [22]. For some distributions, it has been possible to find modifications of known GoF tests for the simple hypothesis, that allow us to use the same formalism in the composite case. This becomes increasingly difficult as the complexity of the distribution and the dimension of the data increase. In many cases, the sampling distribution must be simulated by computer intensive methods, such as Monte Carlo methods and bootstrap sampling.

2) Types of GoF tests: There are many different approaches to formal GoF testing [18]. Some of the most important categories are:

- Pearson's χ^2 test and variations thereof
- Tests based on the empirical distribution function (EDF)

[3]We distinguish formal methods from informal methods by their use of probabilistic decision theory, as opposed to e.g. graphical methods that prepare for visual inspection and decisions that are more or less subjective.

[4]What we describe here are one sample GoF tests. The two sample test (or generally the k-sample test) assesses whether or not two (or k) samples come from the same distribution.

- Tests based on the empirical characteristic function (ECF)
- Tests based on entropy
- Tests based on moments
- Tests based on regression and correlation

The χ^2 tests [18], [23] require binning of the data domain, which is not suitable for matrix distributions defined on Ω_+. The EDF and the ECF are estimates of the CDF and the CF, respectively. Among the EDF-based tests, we find the popular Kolmogorov-Smirnov, Anderson-Darling, and Cramér-von Mises tests [18]. These have been applied to distributions of radar intensity [24], but as far as we know, not to matrix distributions. This is possibly because matrix-variate CDFs are little known, difficult to derive, and also since the relevant CDFs contain a hypergeometric function of matrix argument (or another special function), which is difficult to implement and costly to compute.

Models can be represented equivalently by the PDF, the CDF and the CF, as these functions can be retrieved from each other. It is therefore logical that tests based on the ECF [25], [26] or sample entropy [27] (computed by integration over PDF estimates) yield results comparable to those based on the EDF, which is indeed the case. Scalar moments, on the other hand, only capture certain aspects of a distribution, and do not assemble the same amount of statistical information. Still, moment-based approaches to GoF testing have been proposed [11], [28], and these have inspired the tests that we present in the next section.

Before algorithms are presented, we outline three different problem settings that affect the design of the GoF test, and discuss their relation to practical applications:

Case I (All parameters specified): This is the simple hypothesis case, where we want to test data against a fully specified distribution. A potential application could be within a model-based image analysis algorithm, e.g. segmentation, classification or clustering. The GoF test could be used to decide whether separate segments, classes or clusters should be splitted or merged, for cluster validation, or to estimate the number of classes. The assumption that the distribution of the segments/classes/clusters are fully specified, is of course a simplification, which may be accepted when the number of samples is high, or in order to obtain an efficient algorithm. Practical applications where the parameters are truly known, are hard to exemplify.

Case II (Texture parameters unknown): The composite hypothesis case with known scale matrix $\mathbf{\Sigma}$ and unknown texture parameters is also not very realistic in practice. However, in a setting where we test a number of competing distribution models derived from the multilook polarimetric product model, it is possible to disregard the scale matrix. Note that the maximum likelihood estimate of the scale matrix is the same for all models. Further recall from (19) and (20) that the MLCs of order $\nu > 1$ are independent of $\mathbf{\Sigma}$, and depend only on the texture parameters and L, which is assumed a known constant. By using an MLC-based GoF test, we only need to estimate the texture parameters, and can thereby avoid the nuisance parameter $\mathbf{\Sigma}$.

Case III (All parameters unknown): Finally assume that all parameters are unknown and must be estimated. This is the most difficult, and also the most realistic setting, which is faced in the composite hypothesis case by all other GoF tests than those founded on MLCs, where the scale matrix cannot be decoupled from the texture parameters.

B. Simple Hypothesis Tests

In this section we derive GoF tests for the simple hypothesis case. The tests are based on sample MLCs. We start by deriving the asymptotic distribution of sample MLMs and sample MLCs. Afterwards, we propose test statistics whose true sampling distribution is approximated by the asymptotic sampling distribution.

1) Asymptotic Distribution of the Sample MLMs: Let $\mathscr{C} = \{\mathbf{C}_1, \cdots, \mathbf{C}_n\}$ be a size n sample of independent and identically distributed covariance matrices drawn from the PDF $f_{\mathbf{C}}(\mathbf{C})$. The νth-order sample matrix log-moment

$$\langle \mu_\nu \rangle = \frac{1}{n} \sum_{i=1}^{n} (\log |\mathbf{C}_i|)^\nu \tag{21}$$

is a random variable with mean $\mathrm{E}\{\langle \mu_\nu \rangle\} = \mu_\nu$ and variance $\mathrm{Var}\{\langle \mu_\nu \rangle\} = \sigma^2_{\mu_\nu}/n$, where $\sigma^2_{\mu_\nu} = \mathrm{Var}\{(\log |\mathbf{C}|)^\nu\}$. Note that the explicit reference to the random matrix variable will be suppressed hereafter in the notation of both MLMs and MLCs, in writing μ_ν instead of $\mu_\nu\{\mathbf{C}\}$. By the central limit

theorem,

$$\sqrt{n}(\langle\mu_\nu\rangle - \mu_\nu) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2_{\mu_\nu}) \tag{22}$$

denoting convergence in distribution to a univariate Gaussian RV with zero mean and variance $\sigma^2_{\mu_\nu}$.

If the MLMs of $\mathbf{C}$ exist up to order ν, then all sample MLMs up to this order have expectation. If $\mathbf{C}$ has MLMs up to order 2ν, then all sample MLMs have finite variance. More specifically, we have according to [29]:

$$\mathrm{E}\{\langle\mu_\nu\rangle\} = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left\{(\log|\mathbf{C}_i|)^\nu\right\} = \mu_\nu \tag{23}$$

and

$$\begin{aligned}
&\mathrm{Cov}\{\langle\mu_\upsilon\rangle, \langle\mu_\nu\rangle\} \\
&= \mathrm{E}\left\{\left(\frac{1}{n}\sum_{i=1}^{N}(\log|\mathbf{C}_i|)^\upsilon - \mu_\upsilon\right)\right. \\
&\qquad \left.\times\left(\frac{1}{n}\sum_{j=1}^{N}(\log|\mathbf{C}_j|)^\nu - \mu_\nu\right)\right\} \\
&= \frac{1}{n}\big(\mu_{\upsilon+\nu} - \mu_\upsilon\mu_\nu\big)
\end{aligned} \tag{24}$$

with

$$\mathrm{Var}\{\langle\mu_\nu\rangle\} = \frac{1}{n}\big(\mu_{2\nu} - \mu_\nu^2\big) \tag{25}$$

as a special case of (24).

A multivariate version of the asymptotic distribution for the vector of joint sample log-moments can be formulated. Let

$$\langle\boldsymbol{\mu}_\nu\rangle = [\langle\mu_1\rangle, \langle\mu_2\rangle, \cdots, \langle\mu_\nu\rangle]^T \tag{26}$$

and

$$\boldsymbol{\mu}_\nu = [\mu_1, \mu_2, \cdots, \mu_\nu]^T, \tag{27}$$

such that $\mathrm{E}\{\langle\boldsymbol{\mu}_\nu\rangle\} = \boldsymbol{\mu}_\nu$. The central limit theorem, in conjunction with the Cramér-Wold theorem [30], proves that

$$\sqrt{n}(\langle\boldsymbol{\mu}_\nu\rangle - \boldsymbol{\mu}_\nu) \xrightarrow{\mathcal{D}} \mathcal{N}_\nu(\mathbf{0}, \mathbf{M}_\nu) \tag{28}$$

where $\mathcal{N}_\nu(\cdot,\cdot)$ denotes a ν-variate normal distribution, the mean vector $\mathbf{0}$ is a length ν column of zeros, and the $\nu \times \nu$ covariance matrix

$$\mathbf{M}_\nu = n\,\mathrm{E}\{(\langle\boldsymbol{\mu}_\nu\rangle - \boldsymbol{\mu}_\nu)(\langle\boldsymbol{\mu}_\nu\rangle - \boldsymbol{\mu}_\nu)^T\} \tag{29}$$

has entries $[\mathbf{M}_\nu]_{ij} = n\,\mathrm{Cov}\{\langle\mu_i\rangle, \langle\mu_j\rangle\} = \mu_{i+j} - \mu_i\mu_j$, which can be verified from (24).

2) Asymptotic Distribution of the Sample MLCs: From the asymptotic distribution of the sample MLMs, we now derive the asymptotic distribution of the sample MLCs. The MLCs can be written as a combination of the MLMs up to the same order, following (12), with the first six moment-to-cumulant relations listed in Appendix A. In general, we may write

$$\kappa_\nu = g_\nu(\mu_1, \mu_2, \cdots, \mu_\nu) = g_\nu(\boldsymbol{\mu}_\nu) \tag{30}$$

with the family of moment-to-cumulant transformation functions, $g_\nu : \mathbb{R}^\nu \to \mathbb{R}$, determined by (12). The same relations are valid when population moments are replaced with sample moments.

We assume in the following that all MLMs of $\mathbf{C}$ exist up to order 2ν. Hence, so do the MLCs up to order ν, as defined by (30). Furthermore, we know that $g_\nu(\boldsymbol{\mu}_\nu)$ is a polynomial in the MLMs, and therefore continuously differentiable. The multivariate delta method proposition [30] states that, given the result in (28), then

$$\sqrt{n}(\langle\kappa_\nu\rangle - \kappa_\nu) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2_{\kappa_\nu}) \tag{31}$$

where we define

$$\begin{aligned}
\sigma^2_{\kappa_\nu} &= \mathrm{Var}\{\kappa_\nu\} \\
&= n\,\mathrm{Var}\{\langle\kappa_\nu\rangle\} = \boldsymbol{\nabla}g_\nu^T\,\mathbf{M}_\nu\boldsymbol{\nabla}g_\nu
\end{aligned} \tag{32}$$

using

$$\boldsymbol{\nabla}g_\nu = \left[\frac{\partial g_\nu(\boldsymbol{\mu}_\nu)}{\partial\mu_1}, \cdots, \frac{\partial g_\nu(\boldsymbol{\mu}_\nu)}{\partial\mu_\nu}\right]^T. \tag{33}$$

This is the asymptotic distribution of the sample MLCs. In order to put this result into practical use, we derive the variances

$$\sigma^2_{\kappa_1} = \kappa_2 \tag{34a}$$
$$\sigma^2_{\kappa_2} = \kappa_4 + 2\kappa_2^2 \tag{34b}$$
$$\sigma^2_{\kappa_3} = \kappa_6 + 9\kappa_4\kappa_2 + 9\kappa_3^2 + 6\kappa_2^3 \tag{34c}$$

by evaluating (32). We note that specification of the asymptotic distribution for the νth-order sample MLC requires knowledge of the population MLCs up to order 2ν.

As for the sample MLMs, we can extend (31) to the multivariate case to obtain the asymptotic distribution of the vector of joint sample MLCs. We introduce the notation

$$\langle\boldsymbol{\kappa}_\nu\rangle = [\langle\kappa_1\rangle, \langle\kappa_2\rangle, \cdots, \langle\kappa_\nu\rangle]^T \tag{35}$$

and

$$\boldsymbol{\kappa}_\nu = [\kappa_1, \kappa_2, \cdots, \kappa_\nu]^T, \tag{36}$$

such that $\mathrm{E}\{\langle\boldsymbol{\kappa}_\nu\rangle\} = \boldsymbol{\kappa}_\nu$. Then, (31) together with the Cramér-Wold theorem again asserts that

$$\sqrt{n}(\langle\boldsymbol{\kappa}_\nu\rangle - \boldsymbol{\kappa}_\nu) \xrightarrow{\mathcal{D}} \mathcal{N}_\nu(\mathbf{0}, \mathbf{K}_\nu)\,, \tag{37}$$

where the scaled covariance matrix

$$\mathbf{K}_\nu = n\,\mathrm{E}\{(\langle\boldsymbol{\kappa}_\nu\rangle - \boldsymbol{\kappa}_\nu)(\langle\boldsymbol{\kappa}_\nu\rangle - \boldsymbol{\kappa}_\nu)^T\} \tag{38}$$

has entries $[\mathbf{K}_\nu]_{ij} = n\,\mathrm{Cov}\{\langle\kappa_i\rangle, \langle\kappa_j\rangle\}$. Let $\mathbf{J}_\nu$ denote the Jacobian matrix of the moment-to-cumulant transformations up to order ν. Thus $\mathbf{J}_\nu$ has entries $[J_\nu]_{ij} = \partial g_i(\boldsymbol{\mu}_\nu)/\partial\mu_j$ and $\mathbf{J}_\nu = [\boldsymbol{\nabla}g_1, \cdots, \boldsymbol{\nabla}g_\nu]^T$. The asymptotic covariance matrix of the sample MLCs can then be written as

$$\mathbf{K}_\nu = \mathbf{J}_\nu \mathbf{M}_\nu \mathbf{J}_\nu^T\,, \tag{39}$$

where $\mathbf{M}_\nu$ is the asymptotic covariance matrix of the sample MLMs, as defined in (29). The matrices $\mathbf{M}_\nu$, $\mathbf{J}_\nu$ and $\mathbf{K}_\nu$ are given in Appendix A for $\nu = 4$.

3) Normal Approximation: We are now ready to test the simple null hypothesis H_0 stating that the sample $\mathscr{C} = \{\mathbf{C}_i\}_{i=1}^n$ is drawn from a PDF $f_{\mathbf{C}}(\mathbf{C}; \boldsymbol{\Sigma}_0, \boldsymbol{\theta}_0)$ with specified parameters, where $\boldsymbol{\Sigma}_0 \in \Omega_+$ is the scale matrix, $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$ is a vector of q texture parameters, and the parameter space $\boldsymbol{\Theta}$ is generally a subset of $\mathbb{R}^q$. We generalise the approach by using p sample MLCs of selected orders $\{\nu_1, \nu_2, \cdots, \nu_p\}$ in the test. The test is thus based on the vector

$$\langle\boldsymbol{\kappa}\rangle = [\langle\kappa_{\nu_1}\rangle, \langle\kappa_{\nu_2}\rangle, \cdots, \langle\kappa_{\nu_p}\rangle]^T\,, \tag{40}$$

with mean vector

$$\mathrm{E}\{\langle\boldsymbol{\kappa}\rangle\} = \boldsymbol{\kappa} = [\kappa_{\nu_1}, \kappa_{\nu_2}, \cdots, \kappa_{\nu_p}]^T \tag{41}$$

and scaled covariance matrix

$$\mathbf{K} = n\,\mathrm{E}\left\{\big(\langle\boldsymbol{\kappa}\rangle - \boldsymbol{\kappa}\big)\big(\langle\boldsymbol{\kappa}\rangle - \boldsymbol{\kappa}\big)^T\right\}. \tag{42}$$

The sampling distribution of $\langle\boldsymbol{\kappa}\rangle$ depends on $\boldsymbol{\Sigma}_0$ and $\boldsymbol{\theta}_0$ through MLCs up to order

$$2\,\nu_{\max} = 2 \cdot \max\{\nu_1, \cdots, \nu_p\}\,, \tag{43}$$

and we assume that they all exist. Under the multinormal assumption on $\langle\boldsymbol{\kappa}\rangle$, these MLCs determine the mean vector $\boldsymbol{\kappa}$ and scaled covariance matrix $\mathbf{K}$.

If we use only one MLC in the test ($p = 1$), then $\langle\boldsymbol{\kappa}\rangle = \langle\kappa_\nu\rangle$, where ν is the selected order, and we can define the test statistic

$$T_\nu = \frac{\langle\kappa_\nu\rangle - \kappa_\nu}{\sqrt{\sigma^2_{\kappa_\nu}/n}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1) \tag{44}$$

whose asymptotic sampling distribution under H_0 is standard normal, as indicated. A test with size (significance level) α_c is given by

$$|T_\nu| \underset{\omega_0}{\overset{\omega_1}{\gtrless}} z_{\alpha_c/2}\,. \tag{45}$$

Here ω_0 and ω_1 denote acceptance and rejection of H_0, respectively. The threshold $z_{\alpha_c/2}$ is the upper $\alpha_c/2$ percentile of a standardised normal distribution, which must be inverted from

$$P(|T_\nu| > z_{\alpha_c/2}) = 1 - \mathrm{erf}\left(\frac{z_{\alpha_c/2}}{\sqrt{2}}\right) = \alpha_c \tag{46}$$

with the Gauss error function defined as

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\,dt\,. \tag{47}$$

We remark that the test statistic T_ν is indexed by the order of the MLC it is based on, and refer to ν as the order of the normal approximation test.

4) Chi-square Approximation: The normal approximation test utilises only a single MLC at the time, thus the GoF is measured with respect to a limited aspect of the model distribution. We now construct a test that utilises multiple sample MLCs, and thereby captures more statistical information about the data.

Consider the test statistic

$$Q_p = n(\langle\boldsymbol{\kappa}\rangle - \boldsymbol{\kappa})^T \mathbf{K}^{-1} (\langle\boldsymbol{\kappa}\rangle - \boldsymbol{\kappa}) \tag{48}$$

which uses information from the p MLCs in $\boldsymbol{\kappa}$. The asymptotic distribution of Q_p follows readily from the assumption of $\sqrt{n}(\langle\boldsymbol{\kappa}\rangle - \boldsymbol{\kappa}) \xrightarrow{\mathcal{D}} \mathcal{N}_p(\mathbf{0}, \mathbf{K})$ as

$$Q_p \xrightarrow{\mathcal{D}} \chi^2(p) \tag{49}$$

where $\chi^2(p)$ denotes a central χ^2 distribution with p degrees of freedom. Most importantly, note that the sampling distribution is independent of $\boldsymbol{\Sigma}_0$ and $\boldsymbol{\theta}_0$. A test with significance level α_c is given by

$$Q_p \underset{\omega_0}{\overset{\omega_1}{\gtrless}} z_{\alpha_c}\,. \tag{50}$$

Algorithm 1 MLC-based GoF test of the simple hypothesis using a χ^2 approximation of the test statistic sampling distribution

1) Determine the significance level α_c.
2) Determine the orders $\{\nu_1, \cdots, \nu_p\}$ of the MLCs used in the test.
3) From the dataset $\mathscr{C}$ of n covariance matrices, compute the sample MLMs of the data up to the maximum MLC order: $\max\{\nu_1, \cdots, \nu_q\}$ with the sample mean estimator.
4) Use the obtained sample MLMs to compute the required sample MLCs by means of the moment-to-cumulant transformations.
5) Compute the population MLCs of the hypothesised model up to twice the maximum MLC order: $2 \cdot \max\{\nu_1, \cdots, \nu_q\}$, by plugging the known parameters into the parametric expressions.
6) Use the population MLCs to form the mean vector $\boldsymbol{\kappa}$ and the scaled covariance matrix $\mathbf{K}$.
7) Compute the test statistic Q_p, the threshold z_α, and perform the hypothesis test.

The threshold z_{α_c} is the upper α_c percentile of the $\chi^2(p)$ distribution, found by inversion of

$$P(Q_p > z_{\alpha_c}) = \int_{z_{\alpha_c}}^{\infty} \frac{(1/2)^{p/2}}{\Gamma(p/2)} t^{\frac{p}{2}-1} e^{-\frac{p}{2}} dt = \alpha_c \,. \quad (51)$$

A stepwise description of the GoF test based on a $\chi^2(p)$ approximation of the test statistic Q_p is given in Algorithm 1.

C. Composite Hypothesis Test

We next consider a test of a the composite null hypothesis H_0, declaring that the sample $\mathscr{C}$ is drawn from a parametric distribution family $f_{\mathbf{C}}(\mathbf{C}; \boldsymbol{\Sigma}, \boldsymbol{\theta})$, where the true parameters $\boldsymbol{\Sigma}_0$ and $\boldsymbol{\theta}_0$ are unknown and must be replaced by the estimates $\hat{\boldsymbol{\Sigma}}$ and $\hat{\boldsymbol{\theta}}$. We omit the normal approximation, since more powerful tests using multiple MLCs will be preferred, and go straight to tests built upon $\langle \boldsymbol{\kappa} \rangle$ for $p > 1$.

1) Quadratic Test Statistic: We can safely assume that the MLCs in $\boldsymbol{\kappa}$ are continuous function of $\boldsymbol{\Sigma}$ and $\boldsymbol{\theta}$. Further assume that we have estimators $\hat{\boldsymbol{\Sigma}}(\mathscr{C})$ and $\hat{\boldsymbol{\theta}}(\mathscr{C})$ that produce consistent estimates of $\boldsymbol{\Sigma}_0$ and $\boldsymbol{\theta}_0$. It follows that $\boldsymbol{\kappa}(\hat{\boldsymbol{\Sigma}}, \hat{\boldsymbol{\theta}})$ and $\mathbf{K}(\hat{\boldsymbol{\theta}})$ will be consistent estimates of $\boldsymbol{\kappa}(\boldsymbol{\Sigma}_0, \boldsymbol{\theta}_0)$ and $\mathbf{K}(\boldsymbol{\theta}_0)$, where we have written $\boldsymbol{\kappa}$ and $\mathbf{K}$ with the (estimated) parameters as arguments in order to highlight the dependencies. We remark that $\mathbf{K}$ is a function of MLCs of order $\nu \geq 2$, and therefore depends only on $\boldsymbol{\theta}$, while $\boldsymbol{\kappa}$ is generally a function of both $\boldsymbol{\Sigma}$ (when $\langle \boldsymbol{\kappa} \rangle$ contains $\langle \kappa_1 \rangle$) and $\boldsymbol{\theta}$.

Define the test statistic

$$\begin{aligned} Q'_p = n\big(\langle \boldsymbol{\kappa} \rangle - \boldsymbol{\kappa}(\hat{\boldsymbol{\Sigma}}, \hat{\boldsymbol{\theta}})\big)^T \mathbf{K}(\hat{\boldsymbol{\theta}})^{-1} \\ \times \big(\langle \boldsymbol{\kappa} \rangle - \boldsymbol{\kappa}(\hat{\boldsymbol{\Sigma}}, \hat{\boldsymbol{\theta}})\big) \end{aligned} \quad (52)$$

where the mean vector and the covariance matrix from (48) have been replaced by estimates. The exact sampling distribution of Q'_p generally depends on the sample size n, the true parameters $\boldsymbol{\Sigma}_0$ and $\boldsymbol{\theta}_0$, the estimators producing $\hat{\boldsymbol{\Sigma}}$ and $\hat{\boldsymbol{\theta}}$, and the functional form of $f_{\mathbf{C}}(\mathbf{C}; \boldsymbol{\Sigma}, \boldsymbol{\theta})$ [22]. We have not been able to find any approximation for the sampling distribution that works over a sufficiently wide range of n and $\hat{\boldsymbol{\theta}}$ values, and have therefore resorted to Monte Carlo simulation.

2) Monte Carlo Simulation: The advantage of Monte Carlo simulation is that we obtain a sampling distribution which is accurate for low values of n, and not only as we approach the asymptotic limit. The downside is obviously the computational cost. Since we are interested in tail probabilities of the sampling distribution, we must generate a large number m of realisations of Q'_p: $\{Q'_p(i)\}_{i=1}^m$. This involves random generation of determinants $|\mathbf{C}_i|$ under H_0, which is much easier than generating the full matrix data. We then compute sample MLCs from the simulated data, and calculate the estimates $\hat{\boldsymbol{\Sigma}}$ and $\hat{\boldsymbol{\theta}}$ from them. An alternative would be to random generate the sample MLCs directly from their asymptotic multinormal sampling distribution. We discard this option, because it would abandon the accuracy obtained for finite n.

The biggest practical problem is that the true parameters $\boldsymbol{\Sigma}_0$ and $\boldsymbol{\theta}_0$ are unknown, and we do not have a specified distribution to random generate from. It would be possible to choose a prior distribution for the true parameters, conditioned on the estimated parameters, which could take into account the covariance matrix of the estimated parameters. Our solution is simpler, and keeps the computational requirements at an acceptable level: We use the parameter estimates directly as a best guess of the true parameters. The departure of the estimated parameter values from the true ones will inevitably affect the Monte Carlo simulated sample of test statistics. We assume, nevertheless, that the

resulting p-value will on average equal the one we would have obtained from the true sampling distribution. The validity of this assumption is tested by simulations.

3) Estimation of model parameters: The parameter estimators are vital elements in the composite GoF test. The maximum likelihood (i.e., sample mean) estimator can be used for the scale matrix $\mathbf{\Sigma}$ for all distributions derived from the product model. However, our experience is that $\mathbf{\Sigma}$ can be excluded from the computations with good results. The reason is: Unless L is very large, the variance of $\hat{\mathbf{\Sigma}}$ is so high that there is little information to be gained from the first-order sample MLC about GoF. Instead, we use higher-order sample MLCs (i.e., $\langle \kappa_{\nu>1} \rangle$) that are independent of $\mathbf{\Sigma}$, and need only be concerned with estimators of $\boldsymbol{\theta}$.

The estimators that we shall use are based on minimisation of Q_p. Numerical results recently obtained by the authors [13] show that these estimators are superior to all known alternatives, both in terms of bias and variance. We follow in the footsteps of Parr and Schucany [31] and Boos [32], among others, who discuss the coupling of the GoF problem and the estimation problem. Both references propose minimum distance estimators that produce their estimates by minimising popular GoF test statistics interpreted as distances between data and model. In cases when the test statistic has an asymptotic χ^2 distribution, the approach has been termed minimum chi-square estimation.

The method can also be classified as maximum asymptotic likelihood (MAL) estimation, since the asymptotic log-likelihood function under the multinormal assumption for $\langle \boldsymbol{\kappa} \rangle$ is

$$\ell(\mathbf{\Sigma}, \boldsymbol{\theta} | \mathscr{C}) = -\frac{1}{2} \ln |\mathbf{K}(\boldsymbol{\theta})| - \frac{1}{2n} Q_p(\mathbf{\Sigma}, \boldsymbol{\theta}) + C \quad (53)$$

with C is a constant. Minimisation of (50) and maximisation of (53) yield asymptotically equivalent estimates.

We formally write our texture parameter estimator as

$$\hat{\boldsymbol{\theta}} = \arg \left\{ \min_{\boldsymbol{\theta}} \{Q_p\} \right\} \quad (54)$$

and refer to it as the MAL estimator.

4) Implementation: The complete MLC-based test for the composite problem is described in Algorithm 2. As described in the previous section, the computation of the test statistic Q_p' is performed jointly with the estimation of the texture parameters $\boldsymbol{\theta}$ in an iterative search procedure. We have implemented this using Brent's optimisation algorithm [33], which combines quadratic interpolation with the golden section algorithm to achieve a good compromise of speed and robustness.

Algorithm 2 MLC-based GoF test of the composite hypothesis with Monte Carlo simulation of the test statistic sampling distribution

1) Determine the significance level α_c.
2) Determine the orders $\{\nu_1, \cdots, \nu_p\}$ of the MLCs used in the test.
3) From the dataset $\mathscr{C}$ of n covariance matrices (or the dataset $\mathscr{D} = \{|\mathbf{C}_i|\}_{i=1}^n$ of corresponding matrix determinants), compute the sample MLMs of the data up to the maximum MLC order: $\max\{\nu_1, \cdots, \nu_q\}$ with the sample mean estimators.
4) Use the obtained sample MLMs to compute the required sample MLCs by means of the moment-to-cumulant transformations.
5) Estimate the texture parameters $\boldsymbol{\theta}$ of the hypothesised distribution model from the sample MLCs with the MAL estimator and, if necessary, the scale matrix $\mathbf{\Sigma}$ from $\mathscr{C}$ with the maximum likelihood estimator.
6) Store the value of Q_p' obtained in the joint estimation of $\boldsymbol{\theta}$ and minimisation of Q_p.
7) Random generate m matrix determinant samples of size n under the hypothesised model. For each sample, repeat step 3-5 and store the simulated test statistics as $\{Q_p^*(i)\}_{i=1}^m$.
8) Count the number of simulated test statistics that are larger than the test statistic Q_p' obtained in step 6 and compute the fraction with the respect to the number of Monte Carlo simulations. This yields the p-value.
9) Perform the hypothesis test by comparing the p-value to the significance level.

After obtaining Q_p' and $\hat{\boldsymbol{\theta}}$, we Monte Carlo simulate the sampling distribution of Q_p'. With $\hat{\boldsymbol{\theta}}$ in place of the unknown $\boldsymbol{\theta}_0$, we random generate a set $\mathscr{D} = \{|\mathbf{C}_i|\}_{i=1}^m$ of size n matrix determinant samples under H_0. This can be done efficiently with a standard random generator of gamma variates. Using these samples, we repeat the minimisation procedure to produce a set $\{Q_p^*(i)\}_{i=1}^m$ of m Monte

Carlo simulated test statistics. These are used to determine the p-value empirically as

$$p_{\text{mc}} = \frac{1}{m}\sum_{i=1}^{m} \mathbb{I}_{(Q_p^*(i) > Q'_p)} \,, \tag{55}$$

where $\mathbb{I}_{(.)}$ is the indicator function subject to the superscripted condition. That is, we compute the fraction of simulated $Q_p^*(i)$ that are larger than Q'_p. The Monte Carlo simulated p-value, p_{mc}, is then evaluated against the chosen significance level, α_c, in the test:

$$p_{\text{mc}} \underset{\omega_1}{\overset{\omega_0}{\gtrless}} \alpha_c \,. \tag{56}$$

D. Geometrical Interpretation

We now give a geometrical interpretation of the MLC-based GoF tests. Figure 1 introduces the MLC diagram, where we plot the third-order MLC against the second-order MLC to show simultaneously: (i) the manifolds spanned by the theoretical population MLCs that can be attained under given distribution models, and (ii) points that represent the empirical sample MLCs computed from data samples. A more general definition is given in [13], together with a discussion of MLC space manifolds and their dimension, corresponding to the number of texture parameters associated with the model. The $s\mathcal{W}_d^{\text{C}}$ distribution, with no texture parameters, is represented by a point (black circle in the MLC diagram); the $\mathcal{K}_d$ and $\mathcal{G}_d^0$ distributions, with one texture parameter, are represented by curves (red and blue, respectively); and the $\mathcal{U}_d$ distribution, with its two texture parameters, is represented by a surface (yellow). This is an extension of the univariate log-cumulant diagram, introduced by Nicolas in [14], [15].

Equations (19) and (20) show that MLCs with order higher than two are independent of the scale matrix $\boldsymbol{\Sigma}$ under the polarimetric product model. Assuming that L is a global constant for the dataset, a diagram with the selected MLC orders shows the solitary impact of the texture parameters upon the models. It provides insight about how the texture parameters are estimated from MLCs (see [13] for details), and also how the proposed GoF test procedure is executed.

A sample MLC vector $\langle \boldsymbol{\kappa} \rangle = [\langle \kappa_3 \rangle, \langle \kappa_2 \rangle]^T$ computed from data is shown in the figure as the black '$\times$' symbol. In the simple GoF problem, we measure the distance between the sample MLC vector and a given distribution model in terms of Q_p with $p=2$, where the test statistic is interpreted as a distance measure. For instance, we measure the distance between $\langle \boldsymbol{\kappa} \rangle$ and the $\mathcal{K}_d$ distribution with specified texture parameter α, represented in MLC space by $\boldsymbol{\kappa} = [\kappa_2(\alpha), \kappa_3(\alpha)]$. This distance, Q_p, is pictured as the upper leftmost arrow in Figure 1. Then consider the composite problem. For a fixed $\langle \boldsymbol{\kappa} \rangle$, the distance Q'_p is a function of the texture parameters of a given model, and can be minimised with respect to these. For a $\mathcal{K}_d$ distribution hypothesis, we minimise the distance with respect to α. The parameter value providing a minimum of Q'_p is the MAL estimate of α, as defined in (54). The resulting distance, Q'_p, is shown as the lower rightmost arrow in the figure. Note that Q'_p is always the shortest statistical distance to the manifold of population MLCs, and therefore consistently underestimates Q_p.

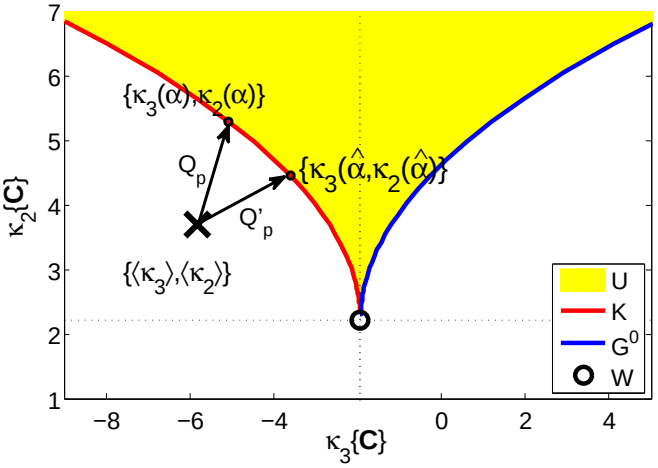

Fig. 1. MLC space geometrical interpretation of the goodness-of-fit test proposed for simple and composite hypotheses. The test statistics Q_p and Q'_p are interpreted as distance between data (sample MLC vector) and model (point on the manifold of population MLC vectors).

Parameter estimation is visualised as a projection of a sample MLC vector onto the manifold representing the model. To test the GoF, we measure the distance between data and model, and then assess the probability of obtaining the resulting distance using the sampling distribution of Q_p or Q'_p. We note that the number of MLCs required by the described GoF test procedure is one more than the number of texture parameters. For instance, the distance to the surface representing the $\mathcal{U}_d$ distribution must be measured with respect to a point in three-dimensional MLC space, thus requiring three sample MLCs.

IV. Results

In this section we use experiments with simulated and real data to demonstrate the capabilities of the proposed procedures for GoF testing.

A. Assessment of the χ^2 Assumption

We first test the departure of the test statistic Q_p from the χ^2 distribution for finite n. We have simulated data from the scaled Wishart distribution with $L=4$ number of looks, polarimetric dimension $d=3$ and scale matrix

$$\boldsymbol{\Sigma}_0 = 10^{-4}\times\begin{bmatrix} 35.8 & & \\ -8.40-6.31\jmath & 16.9 & \\ -0.45+0.45\jmath & -0.17-0.13\jmath & 4.17 \end{bmatrix},$$

where the upper triangle is the complex conjugate of the given lower triangle. The same scale matrix, and hence the same polarimetric dimension, is used in all the simulations of Section IV. The data are tested against the simple hypothesis

$$H_0 : \mathbf{C} \sim s\mathcal{W}_d^{\mathbb{C}}(L=4, \boldsymbol{\Sigma}=\boldsymbol{\Sigma}_0).$$

The test statistic, Q_2, is based on the second and third-order MLC, and is evaluated for the sample sizes $n=\{8, 64, 512, 4096\}$.

The results in the top and middle panel of Figure 2 compares the asymptotic PDF, which is a χ^2 distribution with two degrees of freedom, with empirical PDFs computed by the kernel density estimator [34] from $m = 10,000$ Monte Carlo simulations of Q_2. The figure shows that there is a large discrepancy for small sample sizes, but the true sampling distribution converges quickly towards the χ^2 approximation for moderate sample sizes of $n > 100$.

The top and middle panel show the PDFs on linear and logarithmic scale, respectively. The logarithmic scale emphasises the behaviour at the tail of the distribution, which is critical in GoF testing, and confirms that the approximation is good here also. The fluctuations of the empirical PDFs around the $\chi^2(2)$ approximation for $n=512$ and $n=4096$ can be explained by estimator variance.

In the bottom panel, we compare the specified size of GoF test with the measured size. The size of a statistical hypothesis test is the same as the significance level α_c, defined as the probability of incorrectly rejecting H_0. Again, we see that the specified and the measured size differ greatly for

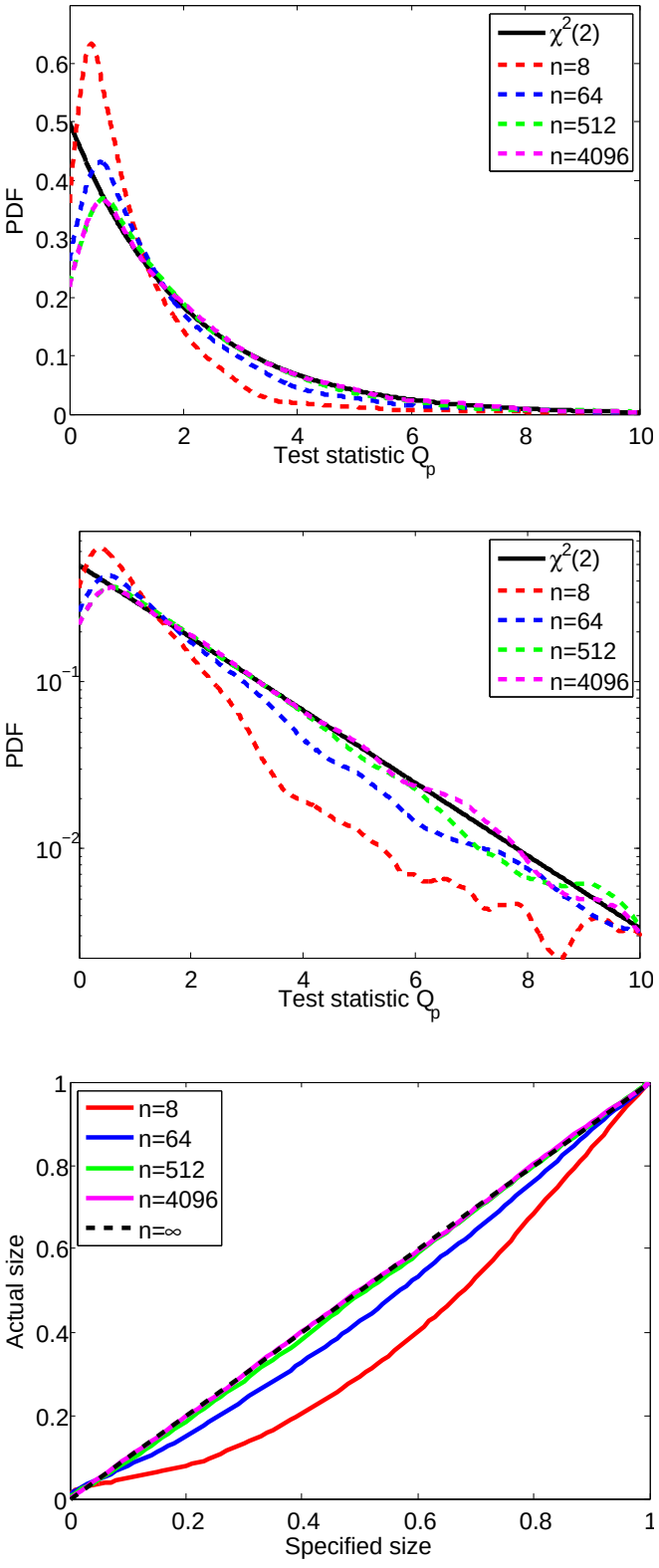

Fig. 2. Assessing validity of $\chi^2(2)$ distribution model for Q_2 with finite sample sizes. The simulated data follow the distribution specified under $H_0 : \mathbf{C} \sim s\mathcal{W}_d^{\mathbb{C}}(L=4, \boldsymbol{\Sigma}=\boldsymbol{\Sigma}_0)$. The plots show estimated sampling distributions of Q_2 on linear scale (upper panel) and logarithmic scale (middle), and a diagram of specified test size against measured test size (bottom) for various sample sizes n.

small sample sizes, but converge quickly as n grows past some hundred samples.

B. Simple Test Applied to Simulated Data

1) Test datasets for the simple hypotheses: Figure 3 shows the data used to assess the MLC-based GoF test of the simple hypothesis. The MLC diagrams display the second and third-order population MLC for the null hypothesis, marked by the symbol '$\times$', and for the simulated datasets used as input to the tests, marked by symbol '$+$'. The respective panels represent tests of the hypotheses:

$$H_0 : \mathbf{C} \sim s\mathcal{W}_d^{\mathbf{C}}(L=4, \boldsymbol{\Sigma}=\boldsymbol{\Sigma}_0) \quad \text{(top)}$$
$$H_0 : \mathbf{C} \sim \mathcal{K}_d(L=4, \boldsymbol{\Sigma}=\boldsymbol{\Sigma}_0, \alpha=8) \text{ (middle)}$$
$$H_0 : \mathbf{C} \sim \mathcal{G}_d^0(L=4, \boldsymbol{\Sigma}=\boldsymbol{\Sigma}_0, \lambda=8) \text{ (bottom).}$$

The input data applied to the tests are matrix-variate $\mathcal{K}$, $\mathcal{G}^0$, and $\mathcal{U}$ distributed with different choices of texture parameters, as specified in the figure. The same datasets are shown in Figures 4 and 5 as marginal PDFs of a single polarimetric channel on linear and logarithmic scale, respectively. All datasets have the same scale matrix, and the marginal PDF is displayed with unit mean intensity for illustration purposes.

When compared to Figure 3, we clearly see the strength of the MLC diagram as a visualisation tool in its ability to discriminate between datasets with equal mean intensity, but different texture. The datasets applied to the test of the Wishart hypothesis have very similar statistical properties, and it difficult to visually separate the PDFs on linear scale (Figure 4, upper panel). The statistical distances between the datasets in the $\mathcal{K}_d$ and $\mathcal{G}_d^0$ hypothesis tests are larger, but it is still difficult to distinguish well between many of the datasets, especially on linear scale. On logarithmic, the distinct behaviour at the tail becomes more visible. Nevertheless, the MLC diagram is far superior to the marginal PDFs in terms of ability to discriminate between datasets.

2) Performance results: The results of the simple hypothesis GoF tests are shown in Figure 6, which displays the power of the tests as function of the sample size. The power of a statistical test is defined as the probability of correctly rejecting H_0. The test of the $s\mathcal{W}_d^{\mathbf{C}}$ hypothesis (shown in the upper panel) performs best for the $\mathcal{K}_d$ distributed dataset with texture parameter $\alpha=16$ (hereafter denoted $\mathcal{K}_d(16)$) and worst for the $\mathcal{K}_d(64)$ dataset. The ranking of the datasets in terms of detectability, i.e. test power, corresponds well at first eye-cast with their statistical

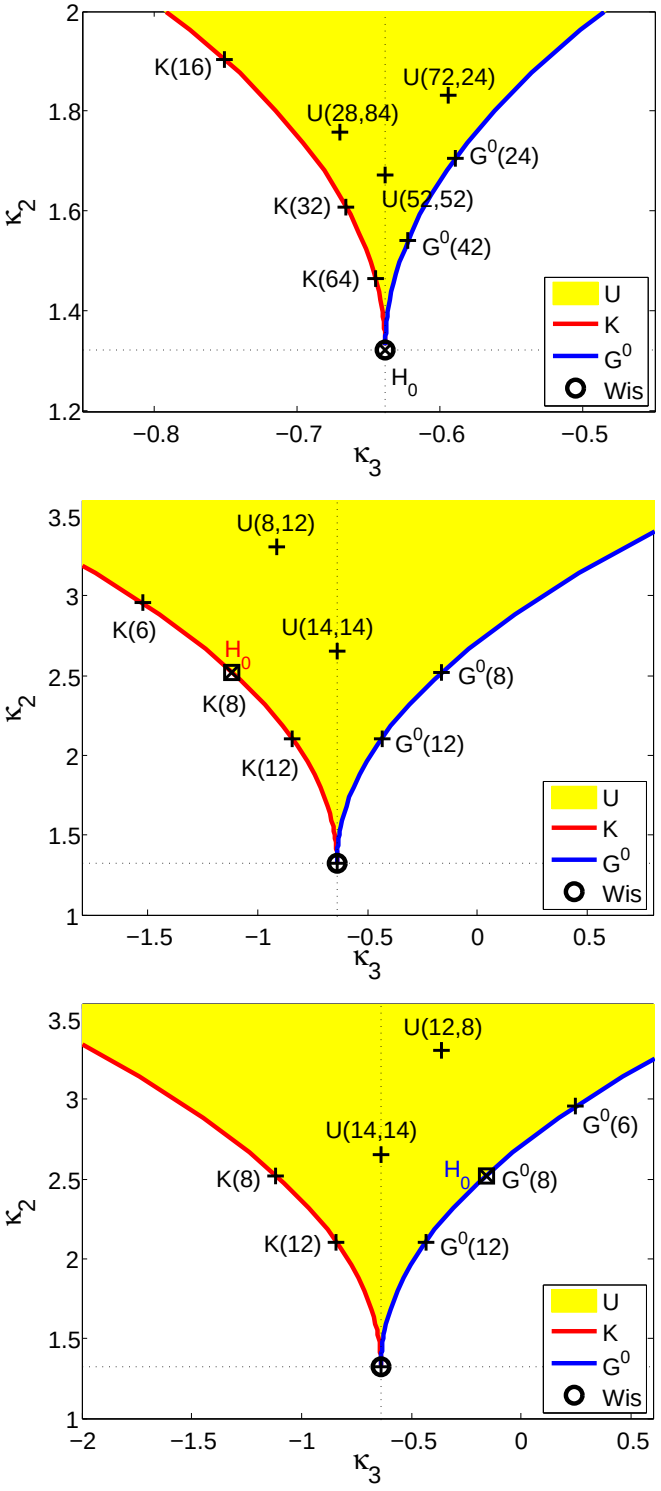

Fig. 3. Data applied to the $\chi^2(2)$ tests of $H_0 : \mathbf{C} \sim \mathcal{W}_d^{\mathbf{C}}(L=4, \boldsymbol{\Sigma}=\boldsymbol{\Sigma}_0)$ (top), $H_0 : \mathbf{C} \sim \mathcal{K}(L=4, \boldsymbol{\Sigma}=\boldsymbol{\Sigma}_0, \alpha=8)$ (middle), and $H_0 : \mathbf{C} \sim \mathcal{G}^0(L=4, \boldsymbol{\Sigma}=\boldsymbol{\Sigma}_0, \lambda=8)$ (bottom), with population MLCs of the datasets represented by symbols '+'.

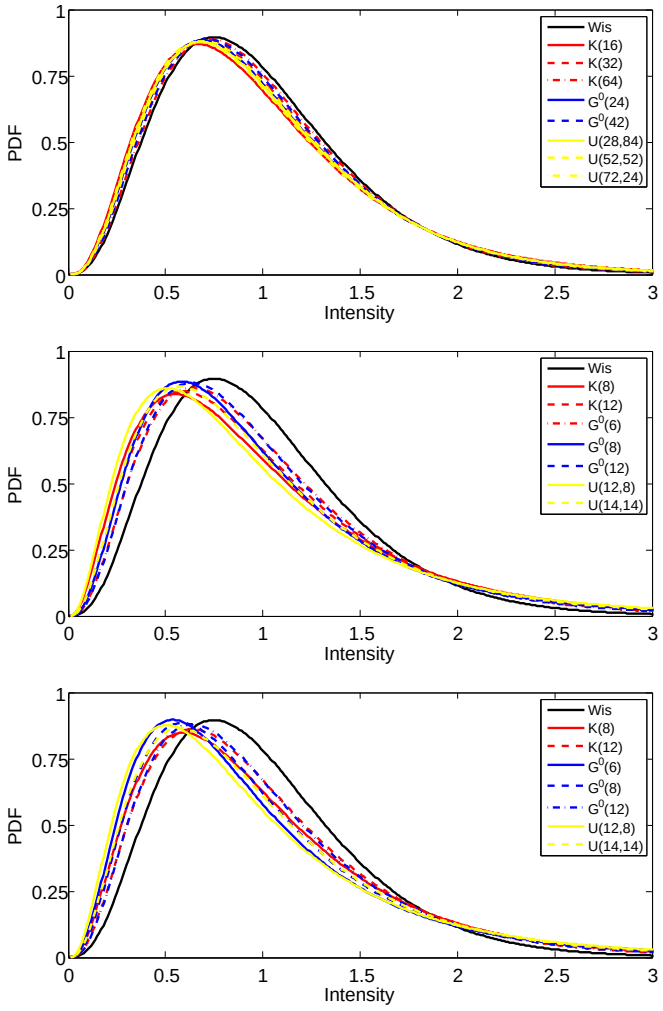

Fig. 4. Data applied to the $\chi^2(2)$ tests of the simple hypotheses H_0 : $\mathbf{C} \sim \mathcal{W}_d^{\mathbf{C}}(L=4, \mathbf{\Sigma}=\mathbf{\Sigma}_0)$ (top), $H_0 : \mathbf{C} \sim \mathcal{K}(L=4, \mathbf{\Sigma}=\mathbf{\Sigma}_0, \alpha=8)$ (middle), and $H_0 : \mathbf{C} \sim \mathcal{G}^0(L=4, \mathbf{\Sigma}=\mathbf{\Sigma}_0, \lambda=8)$ (bottom), shown as marginal PDFs with unit mean intensity on linear scale.

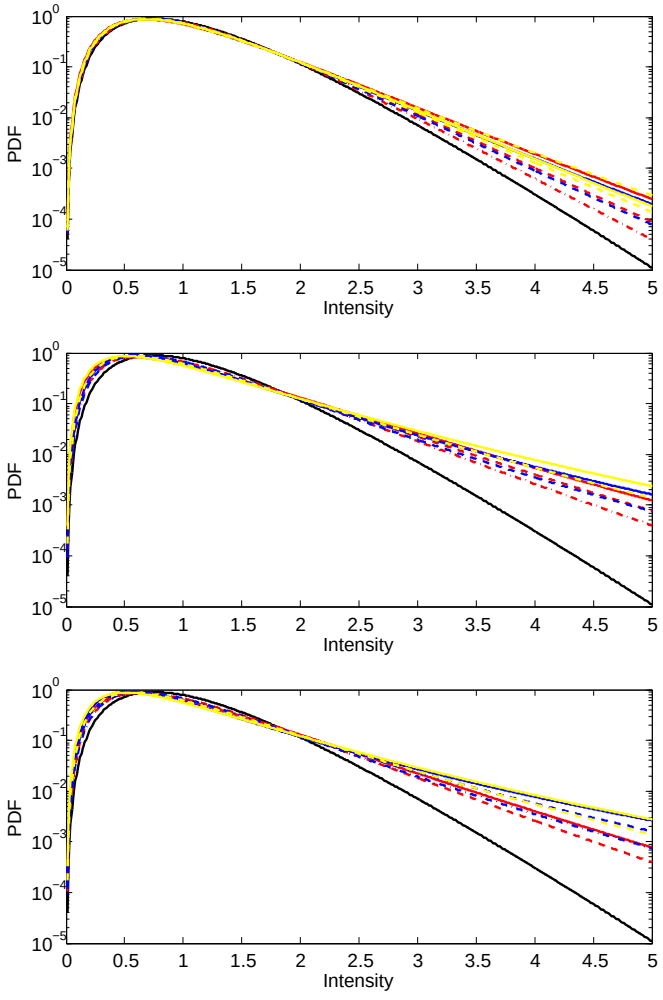

Fig. 5. Data applied to the $\chi^2(2)$ tests of the simple hypotheses H_0 : $\mathbf{C} \sim \mathcal{W}_d^{\mathbf{C}}(L=4, \mathbf{\Sigma}=\mathbf{\Sigma}_0)$ (top), $H_0 : \mathbf{C} \sim \mathcal{K}(L=4, \mathbf{\Sigma}=\mathbf{\Sigma}_0, \alpha=8)$ (middle), and $H_0 : \mathbf{C} \sim \mathcal{G}^0(L=4, \mathbf{\Sigma}=\mathbf{\Sigma}_0, \lambda=8)$ (bottom), shown as marginal PDFs with unit mean intensity on logarithmic scale.

distance from H_0, as perceived in the MLC diagram of Figure 3; The further away from H_0 in the MLC diagram, the more easily a dataset is rejected by the test. The same observation is made for the test of the $\mathcal{K}_d$ hypothesis (middle panel) and the $\mathcal{G}_d^0$ hypothesis (bottom panel).

Upon closer examination of the $\mathcal{K}_d(8)$ and $\mathcal{G}_d^0(8)$ hypothesis tests, we note that its seems more difficult to distinguish datasets that are separated along the κ_3 axis than along the κ_2 axis. For instance, the test of $H_0 : \mathbf{C} \sim \mathcal{K}_d(8)$ has problems with the $\mathcal{G}_d^0(8)$ distributed dataset, just like the test of $H_0 : \mathbf{C} \sim \mathcal{G}_d^0(8)$ has with $\mathcal{K}_d(8)$ data. Both tests struggle most of all with the dataset $\mathcal{U}_d(14, 14)$, whose separation from the null hypotheses is very small along κ_2. This is logical, as the detectability of the first two cases will rely completely on $\langle\kappa_3\rangle$, and the latter mainly on $\langle\kappa_3\rangle$. Furthermore, the higher order of $\langle\kappa_3\rangle$ with respect to $\langle\kappa_2\rangle$ implies that it has larger variance, and therefore less discriminative power. The difference in estimability of κ_2 and κ_3 also explains why several curves in Figure 6 cross each other, indicating that the internal ranking of detectability changes with sample size.

3) Interpretation in terms of equiprobability curves: The reasoning above is supported by the numerical results of Figure 7, which shows curves of equiprobable sample MLCs under different models. In the upper panel, we see equiprobability curves of a $s\mathcal{W}_d^{\mathbf{C}}$ model, a $\mathcal{K}_d(16)$ model and a $\mathcal{G}_d^0(16)$ model for a sample size of $n=1024$. The multiple ellipses represent the set of significance levels:

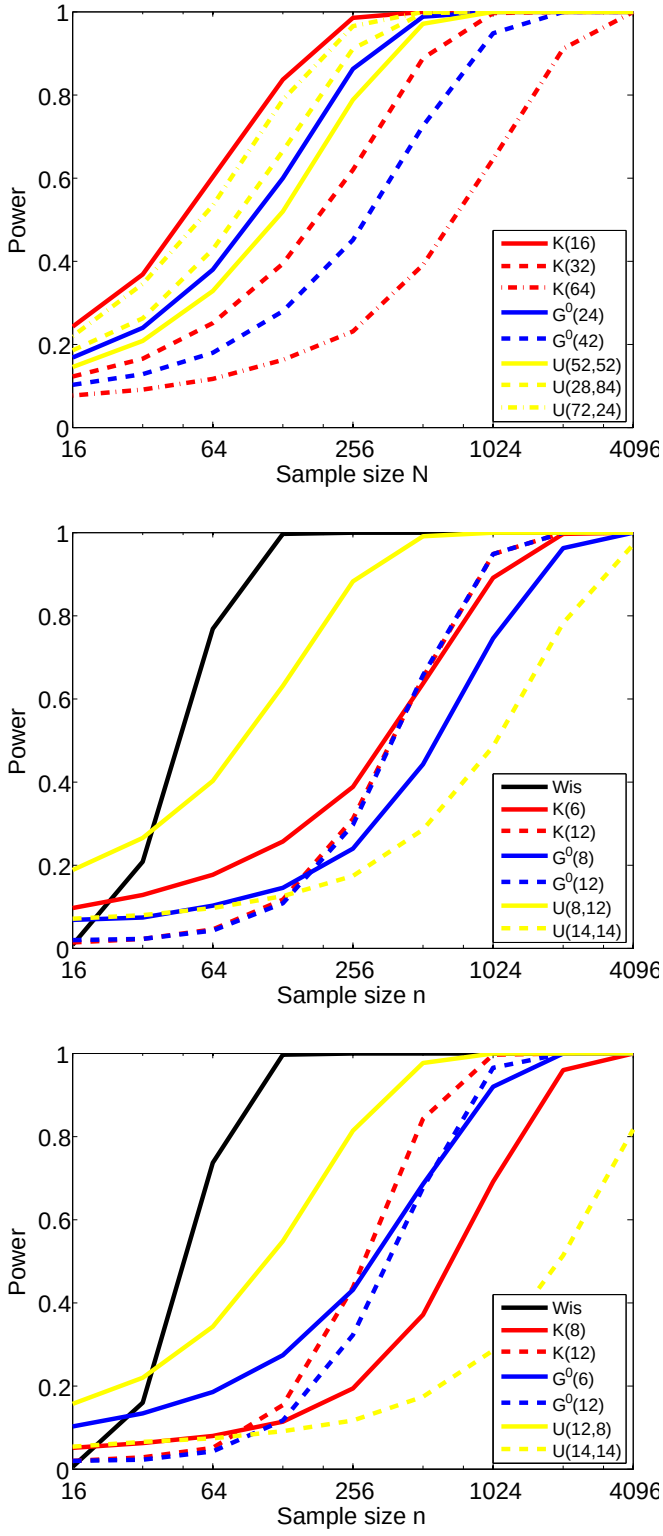

Fig. 6. Power of the Q_p test of the simple hypotheses $H_0 : \mathbf{C} \sim \mathcal{W}_d^{\mathbb{C}}(L = 4, \mathbf{\Sigma} = \mathbf{\Sigma}_0)$ (top), $H_0 : \mathbf{C} \sim \mathcal{K}(L = 4, \mathbf{\Sigma} = \mathbf{\Sigma}_0, \alpha = 8)$ (middle), and $H_0 : \mathbf{C} \sim \mathcal{G}^0(L = 4, \mathbf{\Sigma} = \mathbf{\Sigma}_0, \lambda = 8)$ (bottom) at the $\alpha_c = 5\%$ significance level with $\nu = \{2, 3\}$ for various data distributions, as function of sample size n.

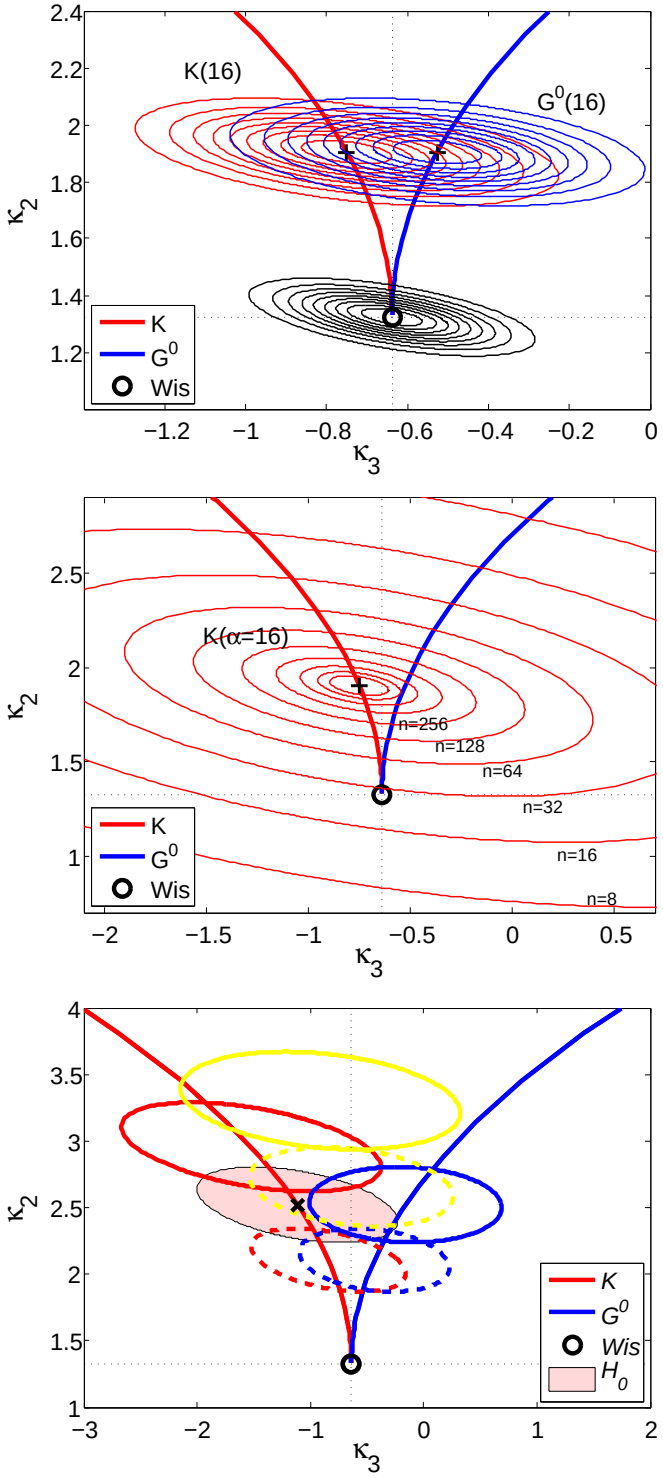

Fig. 7. Top panel: Equiprobability curves defining the $\{10, 20, 30, \cdots, 90\}$th percentiles of the models: $\mathbf{C} \sim \mathcal{W}_d^{\mathbb{C}}(L = 4, \mathbf{\Sigma} = \mathbf{\Sigma}_0)$ (black ellipses), $\mathbf{C} \sim \mathcal{W}_d^{\mathbb{C}}(L = 4, \mathbf{\Sigma} = \mathbf{\Sigma}_0, \alpha = 16)$ (red) and $\mathbf{C} \sim \mathcal{W}_d^{\mathbb{C}}(L = 4, \mathbf{\Sigma} = \mathbf{\Sigma}_0, \lambda = 16)$ (blue) for $N = 1024$ samples based on $Q_2 \sim \chi^2(2)$ with $\nu = \{2, 3\}$. Middle panel: Equiprobability curves defining the 50th percentile of the model $\mathcal{K}(L = 4, \mathbf{\Sigma} = \mathbf{\Sigma}_0, \alpha = 16)$ based on $Q \sim \chi^2(2)$ with $\nu = \{2, 3\}$ as a function of sample size N. Bottom panel: 95th percentile equiprobability curves for sample size $N = 1024$ of the datasets used to test the χ^2 test of $H_0 : \mathbf{C} \sim \mathcal{K}(L = 4, \mathbf{\Sigma} = \mathbf{\Sigma}_0, \alpha = 8)$.

$\alpha_c = \{10, 20, \cdots, 90\}\%$. We see that the equiprobability ellipses extend much further along the κ_3 axis than along the κ_2, as explained in the previous. The orientation of the ellipses depends on the covariance between $\langle\kappa_2\rangle$ and $\langle\kappa_3\rangle$, which is a function of L and the texture parameters, as seen from the expression given in Appendix A. However, the covariance value is dominated by L, and the orientation of the equiprobability ellipses therefore appears nearly constant. We also observe a large overlap between the ellipses of the $\mathcal{K}_d(16)$ model and the $\mathcal{G}_d^0(16)$ model, which means that there will be much confusion when testing either model using input data drawn from the other. This is exactly what we have experienced.

The middle panel of Figure 7 shows the evolution of the equiprobability ellipses for the $\mathcal{K}_d(16)$ model at the $\alpha_c = 50\%$ significance level as a function of n. It explicates the difficulty of GoF testing for small sample sizes due to the large statistical variation of the sample MLCs. The bottom panel again displays the datasets applied to the test of the $\mathcal{K}_d(8)$ hypotheses, but this time as equiprobability ellipses in the MLC diagram for $n = 1024$. The null hypothesis is shown as a '$\times$' symbol, surrounded by an ellipse (filled with pink), which delimits its acceptance region at the $\alpha_c = 0.05$ level. The other ellipses are the 95th percentile equiprobability curves of all datasets applied to the test of the $\mathcal{K}_d(8)$ hypothesis, with colours and line styles corresponding to those used and defined in Figure 4 and 6. The intersection areas between the H_0 ellipse and the dataset ellipses can be approximately related to the probability of falsely accepting H_0 with the respective dataset as input, which is equal to one minus the test power. This interpretation corresponds with the test powers measured in the middle panel of Figure 6 for $n = 1024$.

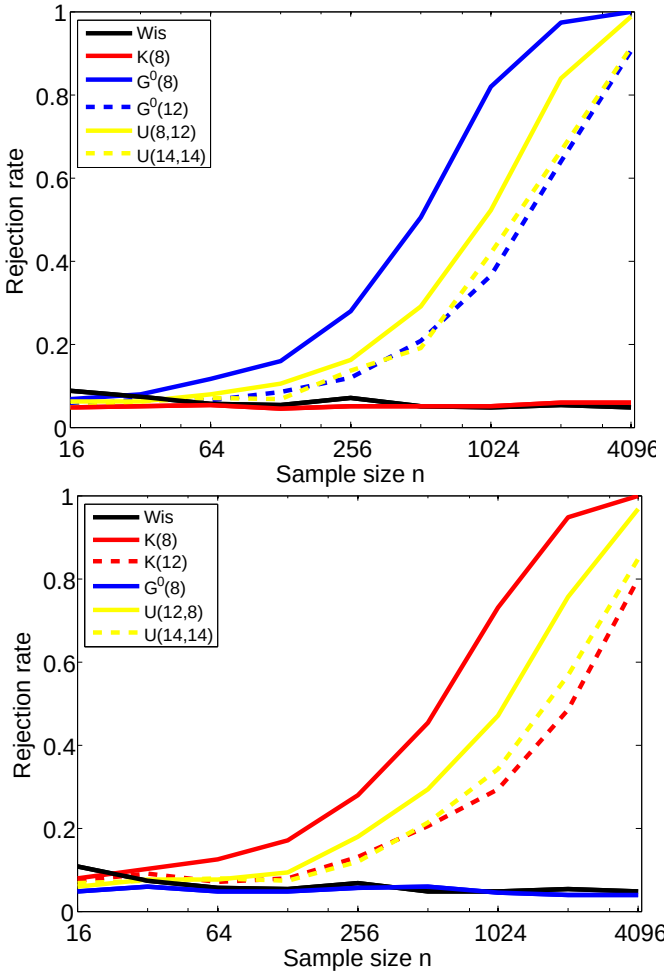

Fig. 8. Rejection rate of the Q'_p test of the composite hypotheses: $H_0 : \mathbf{C} \sim \mathcal{K}(L, \mathbf{\Sigma}, \alpha)$ (top) and $H_0 : \mathbf{C} \sim \mathcal{G}^0(L, \mathbf{\Sigma}, \lambda)$ (bottom) at the $\alpha_c = 5\%$ significance level with $\nu = \{2, 3\}$ for various data distributions, as function of sample size n.

C. Composite Test Applied to Simulated Data

1) Test datasets for the composite hypotheses: For the tests of the composite GoF hypothesis, we assume that the number of looks, L, is known or can be estimated for the dataset as a whole (See [35] for a review of estimation procedures). The scale matrix and the texture parameters are unspecified. We avoid the problem of estimating $\mathbf{\Sigma}$ by using tests based on the second and third order MLC only. The texture parameters are estimated with the MAL estimator described in Section III-C3 and evaluated in [13]. The test statistic Q'_p is applied to two composite hypotheses:

$$H_0 : \mathbf{C} \sim \mathcal{K}_d(L = 4, \mathbf{\Sigma}, \alpha)$$
$$H_0 : \mathbf{C} \sim \mathcal{G}_d^0(L = 4, \mathbf{\Sigma}, \lambda)$$

No dedicated test for the $s\mathcal{W}_d^{\mathbb{C}}$ distribution is performed in the composite hypothesis case. Because the $s\mathcal{W}_d^{\mathbb{C}}$ distribution has no texture parameters that need to be estimated, the composite test of a $s\mathcal{W}_d^{\mathbb{C}}$ hypothesis reduces to the simple test when it is based on MLCs of order $\nu = 2$ and higher.

The datasets applied to the composite tests are the same that we applied to the simple tests. These are presented in Figure 3-5. For the test of the $\mathcal{K}_d$

hypothesis, we limit the testing to one $\mathcal{K}_d$ distributed dataset, in addition to the $s\mathcal{W}_d^{\mathbb{C}}$ distributed dataset, noting that the $s\mathcal{W}_d^{\mathbb{C}}$ distribution is a special case of the $\mathcal{K}_d$ distribution obtained as a limiting case when $\alpha \to \infty$. These datasets are used to check whether we meet the specified test size with the algorithm based on Monte Carlo simulation of the sampling distribution for the test statistic Q'_p. The same approach is taken for the test of the $\mathcal{G}_d^0$ hypothesis, and we equivalently note that the $\mathcal{G}_d^0$ converges in distribution to the $s\mathcal{W}_d^{\mathbb{C}}$ distribution as $\lambda \to \infty$.

2) Performance results: Figure 8 shows the results of the composite GoF test of the $\mathcal{K}_d$ hypothesis (upper panel) and the $\mathcal{G}_d^0$ hypothesis (bottom panel). Note that the figures present rejection rate, instead of test power. The explanation is: When the input dataset belongs to the distribution family under H_0, the rejection rate is the probability of falsely rejecting H_0, which is the same as the test size. When the input dataset belongs to a different family, the rejection rate is the probability of correctly rejecting H_0, previously defined as the test power. The tests are performed at the $\alpha_c = 0.05$ significance level.

Because we have used the same input data and the same test sizes, the performance of the composite tests can be directly compared to the simple tests in Figure 6. For the test datasets that belong to another distribution family than H_0, the test power increases with n as expected, but at a slower rate than for the simple tests. The ranking of the datasets in terms of detectability has changed, and we do not observe any crossing of the rejection rate curves, as we did in Figure 6. This may reflect that the sampling distributions of Q'_p and Q_p are different, but also that the Monte Carlo simulation method yields the true sampling distribution, while the χ^2 approximation used for the simple tests is an approximation, whose validity increases with n. The approximately flat curves at the 0.05 rejection rate level depict the measured test size when the input dataset satisfy H_0. If we disregard fluctuations that can be attributed to expected statistical variations of the Monte Carlo simulations, the measured test size seems to meet the specified test size. The exception is for $s\mathcal{W}_d^{\mathbb{C}}$ distributed input data with small sample sizes, where the measured size exceeds the specified size.

D. Composite Test Applied to Real Data

It remains to test the GoF tests against real data. We have selected three datasets acquired by the Radarsat-2 C-band SAR instrument in fine quad polarisation mode. The scenes are from: 1) Flevoland, The Netherlands, 2) San Francisco, USA and 3) Oberpfaffenhofen, Germany. From the full scenes, we have extracted the subsets shown in the upper row of Figures 9-11. From each subset, we have cropped four image samples, selected to be as homogeneous as possible. The size of each image sample is $n = 16 \times 16 = 256$ pixels. We make the simplifying assumption that the pixel represent independent measurements, even though they are in reality correlated.

The image samples are outlined by the small coloured squares in the upper row images, and enlarged versions are shown in the middle row of the figures. The bottom row of each figure shows an MLC diagram, where sample MLCs of each homogeneous image sample has been plotted on top of the population MLC manifolds of the $s\mathcal{W}_d^{\mathbb{C}}$, $\mathcal{K}_d$ and $\mathcal{G}_d^0$ distribution. Multiple sample MLCs are obtained from each image sample by collecting 64 bootstrap samples of size $n_{bs} = 128$ from the total $n = 256$ covariance matrix samples. This way, we can visualise the statistical spread of the sample MLC distribution, which differs a lot between the image samples. The equivalent number of looks was estimated to $L = 8.0$.

From the Flevoland subset in Figure 9, we have extracted a water sample (magenta coloured square), an urban sample (cyan square), and two vegetation samples, labelled A and B (orange and indigo squares). The false colour RGB images are composites made from intensity channels of the Pauli decomposition [36]. The well-known colour interpretation of so-called Pauli images in terms of scattering mechanisms tells us that the blue water sample is dominated by surface scattering, the pinkish urban sample by double bounce scattering, and the green vegetation A sample by volume scattering. The turquoise appearance of the vegetation B sample reveals a mixture of volume and surface scattering.

In the MLC diagram, the collection of sample MLCs for the water, vegetation A and the vegetation B sample (shown as magenta, orange and indigo dots, respectively) are all well clustered. The clus-

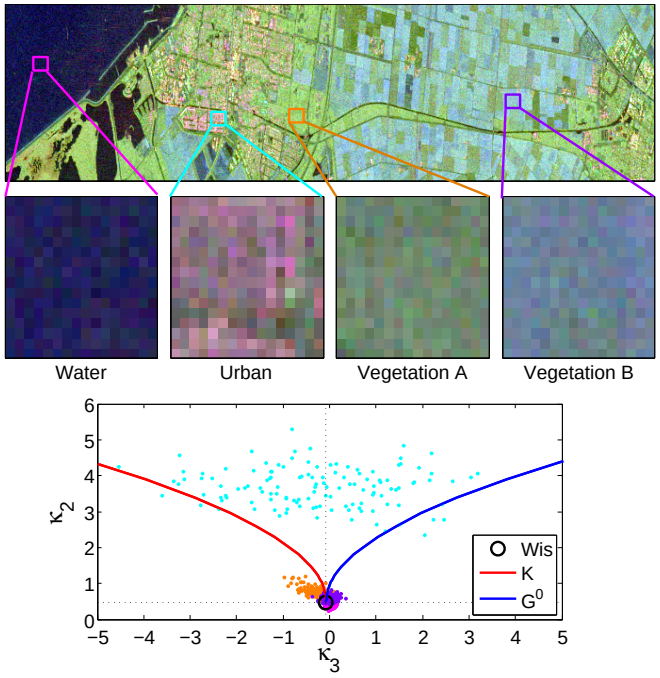

Fig. 9. **Top:** Subset of Radarsat-2 fine quad polarisation mode image of Flevoland, The Netherlands, acquired on 2 March 2008. **Middle:** Homogeneous samples of a water body, urban area, and two vegetated areas (labelled A and B). **Bottom:** MLC diagram with sample MLCs computed from the image samples.

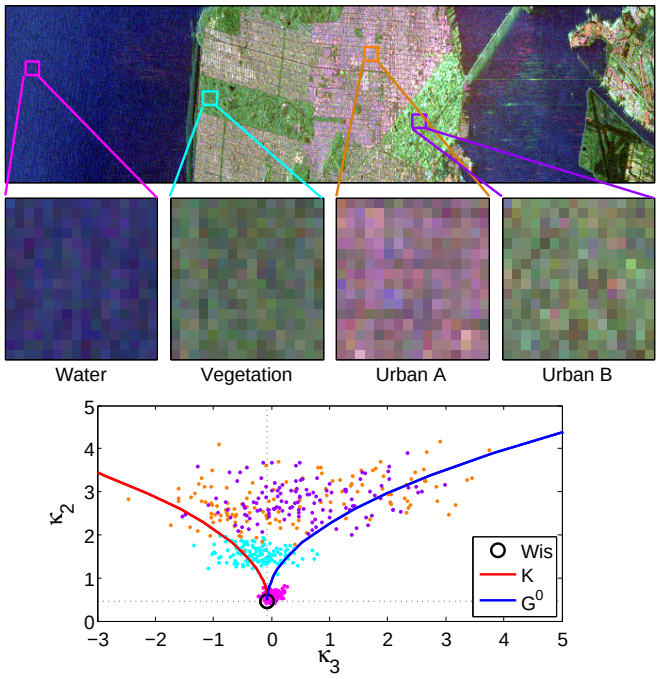

Fig. 10. **Top:** Subset of Radarsat-2 fine quad polarisation mode image of San Francisco, United States, acquired on 9 March 2008. **Middle:** Homogeneous samples of a water body, a vegetated area, and two urban areas (labelled A and B). **Bottom:** MLC diagram with sample MLCs computed from the image samples.

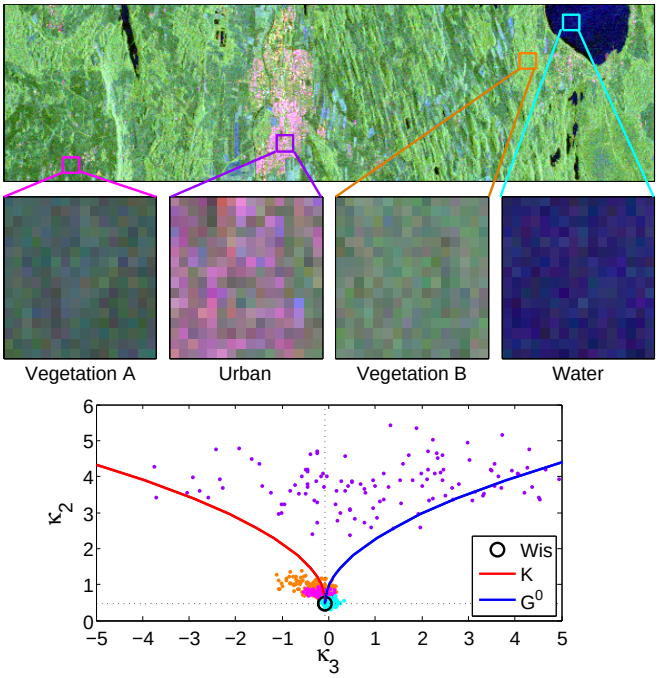

Fig. 11. **Top:** Subset of Radarsat-2 fine quad polarisation mode image of Oberpfaffenhofen, Germany, acquired on 6 March 2008. **Middle:** Homogeneous samples of a vegetated area (labelled A), an urban area, a vegetated area (labelled B), and a water body. **Bottom:** MLC diagram with sample MLCs computed from the image samples.

TABLE III
p-VALUES OF GOF TESTS FOR FLEVOLAND DATASET

	Water	Urban	Veget. A	Veget. B
$s\mathcal{W}_d^{\mathbf{C}}(L,\boldsymbol{\Sigma})$	2.7%	0%	0%	0.1%
$\mathcal{K}_d(L,\boldsymbol{\Sigma},\alpha)$	7.1%	10.8%	10.0%	5.7%
$\mathcal{G}_d^0(L,\boldsymbol{\Sigma},\lambda)$	7.2%	11.4%	12.3%	5.1%

TABLE IV
p-VALUES OF GOF TESTS FOR SAN FRANCISCO DATASET

	Water	Veget.	Urban A	Urban B
$s\mathcal{W}_d^{\mathbf{C}}(L,\boldsymbol{\Sigma})$	0%	0%	0%	0%
$\mathcal{K}_d(L,\boldsymbol{\Sigma},\alpha)$	6.5%	3.5%	13.0%	12.1%
$\mathcal{G}_d^0(L,\boldsymbol{\Sigma},\lambda)$	5.5%	8.4%	7.6%	7.3%

TABLE V
p-VALUES OF GOF TESTS FOR OBERPFAFFENHOFEN DATASET

	Veget. A	Urban	Veget. B	Water
$s\mathcal{W}_d^{\mathbf{C}}(L,\boldsymbol{\Sigma})$	0%	0%	0%	6.3%
$\mathcal{K}_d(L,\boldsymbol{\Sigma},\alpha)$	2.3%	13.9%	5.9%	7.6%
$\mathcal{G}_d^0(L,\boldsymbol{\Sigma},\lambda)$	3.8%	7.7%	11.9%	7.7%

ters may seem to fall close the black square, representing the $s\mathcal{W}_d^{\mathbb{C}}$ distribution, and also the red and blue line, corresponding to the $\mathcal{K}_d$ and $\mathcal{G}_d^0$ distribution, respectively. The sample MLCs for the urban sample have a wide spread, and fall into the region between the $\mathcal{K}_d$ and $\mathcal{G}_d^0$ distribution curves, occupied by the $\mathcal{U}_d$ distribution. The p-values of the composite GoF tests, computed from the complete image samples and presented in Figure III, show that neither model is a very good fit to the image samples. All image samples fail the $s\mathcal{W}_d^{\mathbb{C}}$ distribution test on the 5% level. They all pass the $\mathcal{K}_d$ distribution test and the $\mathcal{G}_d^0$ distribution test on the 5% level, the urban sample and the vegetation A sample also pass on the 10% level by a small margin, but the highest p-value recorded is a modest 12.3%.

From the San Francisco subset in Figure 10, we have extracted a water sample (magenta square), a vegetation sample (cyan square), and two urban samples, labelled A and B (orange and indigo squares). The urban samples are distinguished by their respective pink and green tinged tone. The green appearance of the urban B sample occurs because the city blocks are aligned at an angle to the radar, inducing a strong cross-polarised return [37], which may be mistakenly interpreted as volume scattering. The MLC diagram reveals that both urban samples have the same statistical texture properties. Their sample MLCs have a large variance, and are located in the $\mathcal{U}_d$ distribution region, characteristic of scattering from a mixture of urban objects. The vegetation sample has moderate, but pronounced texture, while the water sample is closer to the $s\mathcal{W}_d^{\mathbb{C}}$ distribution. The p-values in Table IV gives the judgement of the GoF tests: The $s\mathcal{W}_d^{\mathbb{C}}$ distribution in a bad fit. The $\mathcal{K}_d$ distribution hypothesis is passed by the water sample on the 5% level, and by the urban samples on the 10% level. The $\mathcal{G}_d^0$ distribution hypothesis is passed by all samples, but only at the 5% level.

The image samples selected from the Oberpfaffenhofen subset in Figure 11 are two vegetation samples, labelled A and B (magenta and orange squares), an urban sample (indigo square) and a water sample (cyan square). In the Pauli images, the vegetation samples seem to be distinguished mainly by their intensity. The MLC diagram shows that the vegetation B sample has more texture than the vegetation A sample, and that both are located relatively close to the $\mathcal{K}_d$ distribution curve. The water sample appears to be close to the $s\mathcal{W}_d^{\mathbb{C}}$ distribution, while the urban sample MLCs lies in the $\mathcal{U}_d$ distribution region. The p-values in Table V show that only the water sample passes the $s\mathcal{W}_d^{\mathbb{C}}$ distribution test at the 5% level. The vegetation B sample and the water sample pass the $\mathcal{K}_d$ distribution test at the 5% level, and the urban sample at the 10% level. The urban sample and the water sample pass the $\mathcal{G}_d^0$ distribution test at the 5% level, and the vegetation B sample pass at the 10% level.

V. Conclusions

We have proposed goodness-of-fit tests for composite matrix distributions derived under the multilook polarimetric product model. These are based on a newly developed framework for statistical analysis of polarimetric radar data, called matrix-variate Mellin kind statistics. The test procedure can be applied to both simple and composite hypotheses. We have tested them on simulated data for the scaled Wishart distribution, the $\mathcal{K}_d$ distribution and the $\mathcal{G}_d^0$ distribution. The simulations prove that the sampling distribution of the test statistic in the simple hypothesis case is well approximated by the χ^2 distribution for moderate sample sizes and upwards. In the composite hypothesis case, we must resort to Monte Carlo simulations to find the sampling distribution for the test statistic. This approach has a higher computational cost, but produces the true sampling distribution regardless of the sample size. Assessment of the test power proves that the tests are useful contributions that provide a hitherto missing formal procedure for model selection. Experiments with real data from the Radarsat-2 C-band instrument demonstrate the utility of the tests.

Appendix A
Moment and Cumulant Relations

This appendix provides explicit expressions for conversion between moments and cumulants, which is needed in the computations of the MLC-based GoF tests. It also presents covariance matrices of the sample moments and sample cumulants, and relations between them. The transformation are valid for all kinds of moments and cumulants, only requiring that the cumulant generating function is the logarithm of the moment generating function.

The first six moment to cumulant transformations are:

$$\kappa_1 = \mu_1\,, \tag{57}$$
$$\kappa_2 = \mu_2 - \mu_1^2\,, \tag{58}$$
$$\kappa_3 = \mu_3 - 3\mu_1\mu_2 + 2\mu_1^3\,, \tag{59}$$
$$\kappa_4 = \mu_4 - 4\mu_1\mu_3 - 3\mu_2^2 + 12\mu_1^2\mu_2 - 6\mu_1^4\,, \tag{60}$$
$$\kappa_5 = \mu_5 - 5\mu_1\mu_4 - 10\mu_2\mu_3 + 20\mu_1^2\mu_3 + 30\mu_1\mu_2^2 - 60\mu_1^3\mu_2 + 24\mu_1^5\,, \tag{61}$$
$$\kappa_6 = \mu_6 - 6\mu_1\mu_5 - 15\mu_2\mu_4 + 30\mu_1^2\mu_4 - 10\mu_3^2 + 120\mu_1\mu_2\mu_3 - 120\mu_1^3\mu_3 + 30\mu_2^3 - 270\mu_1^2\mu_2^2 + 360\mu_1^4\mu_2 - 120\mu_1^6\,. \tag{62}$$

The first eight cumulant to moment transformations are:

$$\mu_1 = \kappa_1\,, \tag{63}$$
$$\mu_2 = \kappa_2 + \kappa_1^2\,, \tag{64}$$
$$\mu_3 = \kappa_3 + 3\kappa_2\kappa_1 + \kappa_1^3\,, \tag{65}$$
$$\mu_4 = \kappa_4 + 4\kappa_3\kappa_1 + 3\kappa_2^2 + 6\kappa_2\kappa_1^2 + \kappa_1^4\,, \tag{66}$$
$$\mu_5 = \kappa_5 + 5\kappa_4\kappa_1 + 10\kappa_3\kappa_2 + 10\kappa_3\kappa_1^2 + 15\kappa_2^2\kappa_1 + 10\kappa_2\kappa_1^3 + \kappa_1^5\,, \tag{67}$$
$$\mu_6 = \kappa_6 + 6\kappa_5\kappa_1 + 15\kappa_4\kappa_2 + 15\kappa_4\kappa_1^2 + 10\kappa_3^2 + 60\kappa_3\kappa_2\kappa_1 + 20\kappa_3\kappa_1^3 + 15\kappa_2^3 + 45\kappa_2^2\kappa_1^2 + 15\kappa_2\kappa_1^4 + \kappa_1^6\,, \tag{68}$$
$$\mu_7 = \kappa_7 + 7\kappa_6\kappa_1 + 21\kappa_5\kappa_2 + 21\kappa_5\kappa_1^2 + 35\kappa_4\kappa_3 + 105\kappa_4\kappa_2\kappa_1 + 35\kappa_4\kappa_1^3 + 70\kappa_3^2\kappa_1 + 105\kappa_3\kappa_2^2 + 210\kappa_3\kappa_2\kappa_1^2 + 35\kappa_3\kappa_1^4 + 105\kappa_2^3\kappa_1 + 105\kappa_2^2\kappa_1^3 + 21\kappa_2\kappa_1^5 + \kappa_1^7\,, \tag{69}$$
$$\mu_8 = \kappa_8 + 8\kappa_7\kappa_1 + 28\kappa_6\kappa_2 + 28\kappa_6\kappa_1^2 + 56\kappa_5\kappa_3 + 168\kappa_5\kappa_2\kappa_1 + 56\kappa_5\kappa_1^3 + 35\kappa_4^2 + 280\kappa_4\kappa_3\kappa_1 + 210\kappa_4\kappa_2^2 + 420\kappa_4\kappa_2\kappa_1^2 + 70\kappa_4\kappa_1^4 + 280\kappa_3^2\kappa_2 + 280\kappa_3^2\kappa_1^2 + 840\kappa_3\kappa_2^2\kappa_1 + 560\kappa_3\kappa_2\kappa_1^3 + 56\kappa_3\kappa_1^5 + 105\kappa_2^4 + 420\kappa_2^3\kappa_1^2 + 210\kappa_2^2\kappa_1^4 + 28\kappa_2\kappa_1^6 + \kappa_1^8\,. \tag{70}$$

Recall that $\mathbf{M}_4$ and $\mathbf{K}_4$ was defined as the covariance matrices of the sample moment vector $\langle\boldsymbol{\mu}_4\rangle = [\langle\mu_1\rangle, \langle\mu_2\rangle, \langle\mu_3\rangle, \langle\mu_4\rangle]^T$ and the sample cumulant vector $\langle\boldsymbol{\kappa}_4\rangle = [\langle\kappa_1\rangle, \langle\kappa_2\rangle, \langle\kappa_3\rangle, \langle\kappa_4\rangle]^T$, respectively. These are related by

$$\mathbf{K}_4 = \mathbf{J}_4\mathbf{M}_4\mathbf{J}_4^T \tag{71}$$

with the fourth-order Jacobian matrix of the moment to cumulant transformations given by

$$\mathbf{J}_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -2\mu_1 & 1 & 0 & 0 \\ -3(\mu_2 - 2\mu_1^2) & -3\mu_1 & 1 & 0 \\ J_{41} & J_{42} & -4\mu_1 & 1 \end{bmatrix} \tag{72}$$

where

$$J_{41} = -4(\mu_3 - 6\mu_1\mu_2 + 6\mu_1^3)\,, \tag{73}$$
$$J_{42} = -6(\mu_2 - 2\mu_1^2)\,. \tag{74}$$

Explicit expressions for the elements of the sample moment covariance matrix are given as

$$[\mathbf{M}]_{11} = \kappa_2 \tag{75}$$
$$[\mathbf{M}]_{12} = \kappa_3 + 2\kappa_1\kappa_2 \tag{76}$$
$$[\mathbf{M}]_{13} = \kappa_4 + 3\kappa_1\kappa_3 + 3\kappa_2^2 + 3\kappa_1^2\kappa_2 \tag{77}$$
$$[\mathbf{M}]_{14} = \kappa_5 + 4\kappa_1\kappa 4 + 10\kappa_2\kappa_3 + 6\kappa_1^2\kappa_3 + 12\kappa_1\kappa_2^2 + 4\kappa_1^3\kappa_2 \tag{78}$$
$$[\mathbf{M}]_{22} = \kappa_4 + 4\kappa_1\kappa_3 + 2\kappa_2^2 + 4\kappa_1^2\kappa_2 \tag{79}$$
$$[\mathbf{M}]_{23} = \kappa_5 + 5\kappa_1\kappa_4 + 9\kappa_2\kappa_3 + 9\kappa_1^2\kappa_3 + 12\kappa_1\kappa_2^2 + 6\kappa_1^3\kappa_2 \tag{80}$$
$$[\mathbf{M}]_{24} = \kappa_6 + 6\kappa_1\kappa_5 + 14\kappa_2\kappa_4 + 14\kappa_1^2\kappa_4 + 10\kappa_3^2 + 56\kappa_1\kappa_2\kappa_3 + 16\kappa_1^3\kappa_3 + 12\kappa_2^3 + 36\kappa_1^2\kappa_2^2 + 8\kappa_1^4\kappa_2 \tag{81}$$
$$[\mathbf{M}]_{33} = \kappa_6 + 6\kappa_1\kappa_5 + 15\kappa_2\kappa_4 + 15\kappa_1^2\kappa_4 + 9\kappa_3^2 + 54\kappa_1\kappa_2\kappa_3 + 18\kappa_1^3\kappa_3 + 15\kappa_2^3 + 36\kappa_1^2\kappa_2^2 + 9\kappa_1^4\kappa_2 \tag{82}$$
$$[\mathbf{M}]_{34} = \kappa_7 + 7\kappa_1\kappa_6 + 21\kappa_2\kappa_5 + 21\kappa_1^2\kappa_6 + 34\kappa_3\kappa_4 + 102\kappa_1\kappa_2\kappa_4 + 34\kappa_1^3\kappa_4 + 66\kappa_1\kappa_3^2 + 102\kappa_2^2\kappa_3 + 192\kappa_1^2\kappa_2\kappa_3 + 30\kappa_1^4\kappa_3 + 96\kappa_1\kappa_2^3 + 84\kappa_1^3\kappa_2^2 + 12\kappa_1^5\kappa_2 + \kappa_1^7 \tag{83}$$
$$[\mathbf{M}]_{44} = \kappa_8 + 8\kappa_1\kappa_7 + 28\kappa_2\kappa_6 + 28\kappa_1^2\kappa_6 + 56\kappa_3\kappa_5 + 168\kappa_1\kappa_2\kappa_5 + 56\kappa_1^3\kappa_5 + 34\kappa_4^2 + 272\kappa_1\kappa_3\kappa_4 + 204\kappa_2^2\kappa_4 + 408\kappa_1^2\kappa_2\kappa_4 + 68\kappa_1^4\kappa_4 + 280\kappa_2\kappa_3^2 + 264\kappa_1^2\kappa_3^2 + 816\kappa_1\kappa_2^2\kappa_3 + 512\kappa_1^3\kappa_2\kappa_3 + 48\kappa_1^5\kappa_3 + 96\kappa_2^4 + 384\kappa_1^2\kappa_2^3 + 168\kappa_1^4\kappa_2^2 + 16\kappa_1^6\kappa_2 + \kappa_1^8 \tag{84}$$

The sample cumulant covariance matrix becomes

$$\mathbf{K}_4 = \begin{bmatrix} \kappa_2 & \kappa_3 & \kappa_4 & \kappa_5 \\ \kappa_3 & \kappa_4 + 2\kappa_2^2 & \kappa_5 + 6\kappa_2\kappa_3 & K_{24} \\ \kappa_4 & \kappa_5 + 6\kappa_2\kappa_3 & K_{33} & K_{34} \\ \kappa_5 & K_{42} & K_{43} & K_{44} \end{bmatrix} \quad (85)$$

where

$$K_{24} = K_{42} = \kappa_6 + 8\kappa_2\kappa_4 + 6\kappa_3^2\,, \quad (86)$$

$$K_{33} = \kappa_6 + 9\kappa_2\kappa_4 + 9\kappa_3^2 + 6\kappa_2^3\,, \quad (87)$$

$$K_{34} = K_{43} = \kappa_7 + 12\kappa_2\kappa_5 + 30\kappa_3\kappa_4 + 36\kappa_2^2\kappa_3\,, \quad (88)$$

$$K_{44} = \kappa_8 + 16\kappa_2\kappa_6 + 48\kappa_3\kappa_5 + 34\kappa_4^2 + 72\kappa_2^2\kappa_4 + 144\kappa_2\kappa_3^2 + 24\kappa_2^4\,. \quad (89)$$

ACKNOWLEDGEMENT

The authors wish to thank Gabriele Moser at the University of Genoa for his valuable comments on the manuscript. We would also like to thank MacDonald, Dettwiler and Associates (MDA) for making available the Radarsat-2 datasets used in the paper.

REFERENCES

[1] C. Oliver and S. Quegan, *Understanding Synthetic Aperture Radar Images*, 2nd ed. Raleigh, USA: SciTech Publishing, 2004.

[2] E. Jakeman and P. N. Pusey, "A model for non-Rayleigh sea echo," *IEEE Trans. Antennas Propag.*, vol. 24, no. 6, pp. 806–814, Nov. 1976.

[3] A. C. Frery, H.-J. Müller, 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.

[4] N. R. Goodman, "Statistical analysis based on a certain multivariate complex Gaussian distribution (an introduction)," *Ann. Math. Statist.*, vol. 34, no. 1, pp. 152–177, Mar. 1963.

[5] 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 Int. Geosc. Remote Sens. Symp., IGARSS'94*, vol. 4, Pasadena, USA, Aug. 1994, pp. 2179–2181.

[6] 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.

[7] 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.

[8] A. P. Doulgeris and T. Eltoft, "Scale mixture of Gaussian modelling of polarimetic SAR data," *Eurasip J. Adv. Sig. Proc.*, vol. 2010, 12 pp., Jan. 2010.

[9] J. Heinrich, "Pitfalls of goodness-of-fit from likelihood," in *Proc. PHYSTAT2003 Conf. on Statist. Problems in Particle Phys., Astrophys., and Cosmology*, Stanford, USA, Oct. 2003, pp. 52–55.

[10] A. Grant, "Rayleigh fading multi-antenna channels," *EURASIP J. Appl. Sig. Proc.*, vol. 3, no. 1, pp. 316–329, jan 2002.

[11] G. Li and A. Papadopoulos, "A note on goodness of fit test using moments," *Statistica*, vol. 62, no. 1, pp. 71–86, 2002.

[12] S. N. Anfinsen, "Statistical analysis of multilook polarimetric radar images with the Mellin transform," Ph.D. dissertation, University of Tromsø, Tromsø, Norway, in prep, 2010.

[13] S. N. Anfinsen and T. Eltoft, "Application of the matrix-variate Mellin transform to analysis of polarimetric radar images," *IEEE Trans. Geosci. Remote Sens.*, 2010, submitted, Available: http://www.phys.uit.no/˜stiann/sna-mts-submitted.pdf.

[14] J.-M. Nicolas, "Introduction aux statistique de deuxième espèce: Application des logs-moments et des logs-cumulants à l'analyse des lois d'images radar," *Traitement du Signal*, vol. 19, no. 3, pp. 139–167, 2002, in French.

[15] ——, "Application de la transformée de Mellin: Étude des lois statistiques de l'imagerie cohérente," Ecole Nationale Supérieure des Télécommunications, Paris, France, Tech. Rep. 2006D010, 2006, in French.

[16] C. Tison, J.-M. Nicolas, F. Tupin, and H. Maître, "A new statistical model for Markovian classification of urban areas in high-resolution SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 10, pp. 2046–2057, Oct. 2004.

[17] L. Mattner, "What are cumulants?" *Documenta Mathematica*, vol. 4, pp. 601–622, 1999.

[18] R. B. D'Agostino and M. A. Stephens, *Goodness-of-Fit Techniques*, ser. Statistics: textbooks and monographs. New York, USA: Marcel Dekker, 1986, vol. 68.

[19] J. C. W. Rayner, O. Thas, and D. J. Best, *Smooth Tests of Goodness of Fit*, 2nd ed. New York, USA: John Wiley & Sons, 2009.

[20] A. van der Vaart, *Asymptotic Statistics*. Cambridge, UK: Cambridge University Press, 1998.

[21] E. del Barrio, P. Deheuvels, and S. van de Geer, *Lectures on Empirical Processes: Theory and Statistical Applications*. Zurich, Switzerland: European Mathematical Society, 2007.

[22] A. DasGupta, "Goodness of fit with estimated parameters," in *Asymptotic Theory of Statistics and Probability*. New York, USA: Springer, 2008, ch. 28.

[23] P. E. Greenwood and M. S. Nikulin, *A Guide to Chi-Squared Testing*. New York, USA: John Wiley & Sons, 1996, ch. 2.

[24] E. Conte, A. De Maio, and A. Farina, "Statistical tests for higher order analysis of radar clutter," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 41, no. 1, pp. 205–218, Jan. 2005.

[25] T. W. Epps, "Tests for location-scale families based on the empirical characteristic function," *Metrika*, vol. 62, no. 1, pp. 99–114, Sep. 2005.

[26] S. G. Meintanis, "Goodness-of-fit tests and minimum distance estimation via optimal transformation to uniformity," *J. Statist. Planning and Inference*, vol. 139, no. 2, pp. 100–108, Feb. 2009.

[27] C. E. Marchetti and G. S. Mudholkar, "Characterization theorems and goodness-of-fit tests," in *Goodness-of-Fit Tests and Model Validity*, C. Huber-Carol, N. Balakrishnan, M. S. Nikulin, and M. Mesbah, Eds. New York, USA: Birkhaüser Boston, 2002, ch. 10.

[28] R. Natarajan and G. S. Mudholkar, "Moment-based goodness-of-fit tests for the inverse Gaussian distribution," *Technometrics*, vol. 46, no. 3, pp. 339–347, Aug. 2004.

[29] M. Kendall, *Kendall's Advanced Theory of Statistics: Distribution Theory*, 5th ed. London, UK: Charles Griffin, 1987, vol. 1, ch. 10.

[30] M. Bilodeau and D. Brenner, *Theory of Multivariate Statistics*, ser. Springer texts in statistics. New York, USA: Springer, 1999.

[31] W. C. Parr and W. R. Schucany, "Minimum distance and robust

estimation," *J. Am. Statist. Assoc.*, vol. 75, no. 371, pp. 616–624, Sep. 1980.

[32] D. D. Boos, "Minimum distance estimators for location and goodness of fit," *J. Am. Statist. Assoc.*, vol. 76, no. 375, pp. 663–670, Sep. 1981.

[33] R. P. Brent, *Algorithms for Minimization Without Derivatives*. Englewood Cliffs, USA: Prentice Hall, 1973.

[34] M. P. Wand and M. C. Jones, *Kernel Smoothing*. London, UK: Chapman & Hall, 1995.

[35] S. N. Anfinsen, A. P. Doulgeris, 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.

[36] J.-S. Lee and E. Pottier, *Polarimetric Radar Imaging: From Basics to Applications*, ser. Optical Science and Engineering. Boca Raton, USA: CRC Press, 2009, no. 143.

[37] J.-S. Lee, M. R. Grunes, E. Pottier, and L. Ferro-Famil, "Unsupervised terrain classification preserving polarimetric scattering characteristics," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 4, pp. 722–731, Apr. 2004.